

RESEARCH

Open Access



Molecular mimicry between tumor associated antigens and microbiota-derived epitopes

Concetta Ragone^{1†}, Carmen Manolio^{1†}, Angela Mauriello¹, Beatrice Cavalluzzo¹, Franco M. Buonaguro², Maria Lina Tornesello², Maria Tagliamonte^{1*†} and Luigi Buonaguro^{1*†} 

Abstract

Background: The gut microbiota profile is unique for each individual and are composed by different bacteria species according to individual birth-to-infant transitions. In the last years, the local and systemic effects of microbiota on cancer onset, progression and response to treatments, such as immunotherapies, has been extensively described. Here we offer a new perspective, proposing a role for the microbiota based on the molecular mimicry of tumor associated antigens by microbiome-associated antigens.

Methods: In the present study we looked for homology between published TAAs and non-self microbiota-derived epitopes. Blast search for sequence homology was combined with extensive bioinformatics analyses.

Results: Several evidences for homology between TAAs and microbiota-derived antigens have been found. Strikingly, three cases of 100% homology between the paired sequences has been identified. The predicted average affinity to HLA molecules of microbiota-derived antigens is very high (< 100 nM). The structural conformation of the microbiota-derived epitopes is, in general, highly similar to the corresponding TAA. In some cases, it is identical and contact areas with both HLA and TCR chains are indistinguishable. Moreover, the spatial conformation of TCR-facing residues can be identical in paired TAA and microbiota-derived epitopes, with exactly the same values of planar as well as dihedral angles.

Conclusions: The data reported in the present study show for the first time the high homology in the linear sequence as well as in structure and conformation between TAAs and peptides derived from microbiota species of the Firmicutes and the Bacteroidetes phyla, which together account for 90% of gut microbiota. Cross-reacting CD8⁺ T cell responses are very likely induced. Therefore, the anti-microbiota T cell memory may turn out to be an anti-cancer T cell memory, able to control the growth of a cancer developed during the lifetime if the expressed TAA is similar to the microbiota epitope. This may ultimately represent a relevant selective advantage for cancer patients and may lead to a novel preventive anti-cancer vaccine strategy.

[†]Concetta Ragone and Carmen Manolio are equally contributed to this work. Maria Tagliamonte and Luigi Buonaguro equally contributed as Senior Authors to this work.

*Correspondence: m.tagliamonte@istitutotumori.na.it; l.buonaguro@istitutotumori.na.it

¹ Lab of Innovative Immunological Models, Istituto Nazionale per lo Studio e la Cura dei Tumori - IRCCS, "Fondazione Pascale", Via Mariano Semmola, 52, 80131 Naples, Italy

Full list of author information is available at the end of the article



Keywords: Tumor associated antigens, Microbiota-derived antigens, Molecular mimicry, Cross-reacting T cells

Introduction

The human gastrointestinal tract is colonized by approximately 10^{14} microbes equating to roughly 1000 times the number of cells and 10,000 times the DNA content of the human body [1].

Gut microbiota include several species of microorganisms but the *Firmicutes* and *Bacteroidetes* phyla account for 90% of the total [2].

In addition, type of delivery [3], methods of milk feeding [4] and weaning period [5] have a major impact on the microbiota composition in the first year of life.

The child's gut microbiota composition and diversity reaches the adulthood characteristics at approximately three years of age as consequence of genetics, environment, diet, lifestyle, and gut physiology [6]. Several factors may induce variations in the gut microbiota during the life, including lifestyle, dietary and cultural habits [7].

Gut bacteria regulate food digestion along the gastrointestinal tract and prevent bacteria invasion by maintaining the intestinal epithelium integrity [8]. In addition, they have been shown to regulate the development, homeostasis, and function of innate and adaptive immunity [9].

Besides the physiological role, the microbiota have a role in human diseases [10, 11] and influence both tumor development and treatment response [12–19]. In particular, specific microbiota populations have been shown to affect cancer genesis and development via the production of selected metabolites [20]. Indeed, while components such as LPS or MPL activate T cell-mediated anti-cancer responses [21], the bacterial virulence factor Fap2 can inhibit Natural Killers [22]. Moreover, the polysaccharide A and TLR signaling to dendritic cells may drive the Tregs development [23, 24] as well as segmented filamentous bacteria (SFB) may induce Th17 cells [25].

In addition to such “generic” mechanisms of modulation, the impact of the gut microbiota on the anti-cancer immune response can be driven by a “molecular mimicry” between bacteria and tumor associated antigens.

The gut microbiome encodes over 3 million genes as whole, whereas the entire human genome consists of approximately 23,000 genes [7]. Therefore, the probability of homology between microbiota and human antigens is high, resulting in an overlapping peptidome representation. Highly similar epitopes can be targeted by the same CD8⁺ T cell receptor (TCR), given that a single TCR is cross-reactive recognizing at least 10^6 different MHC-bound peptides [26, 27]. Indeed, the epitope binds to the

HLA molecule with specific residues in fixed positions along the sequence (anchor residues) and only the central residues are exposed for recognition by the TCR (<http://www.cbs.dtu.dk/services/NetMHC/logos.php>) [28, 29]. Therefore, two unrelated antigens sharing the same TCR-facing central residues, or showing conservative variations at those positions, are very likely recognized by the same TCRs if the structural conformation of the entire epitope is saved.

Based on this assumption, we have recently shown that TAAs (<https://caped.icp.ucl.ac.be/Peptide/list>) share sequence homology to viral sequences [30]. This would suggest that viral antigens might elicit memory CD8⁺ T cells cross-reacting with tumor antigens, able to control the growth of a cancer lesion, if expressing a TAA similar to the viral epitope. This may ultimately represent a relevant selective advantage for cancer patients and may lead to a novel preventive anti-cancer vaccine strategy [31–33].

Here we report that significant sequence and conformational homology do exist also between TAAs and peptides derived from microbiota species of the *Firmicutes* and *Bacteroidetes* phyla. This suggests a much broader molecular mimicry between cancer and pathogens antigens with high potential cross-reactive T cell response and impact on tumor progression.

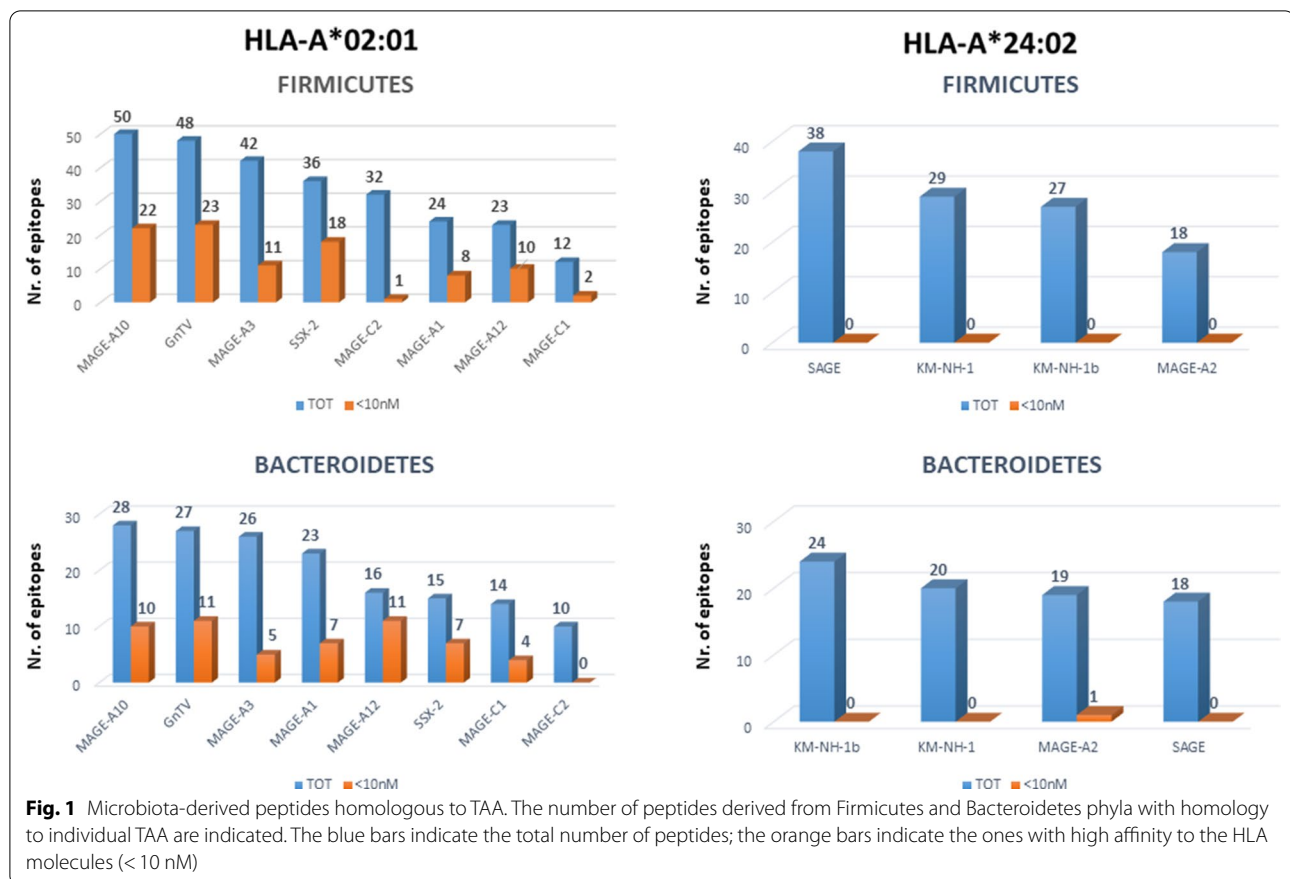
Materials and methods

Epitope prediction analysis.

All the peptides selected in the study were predicted with the NetMHCpan 4.1 and the NetMHCstabpan 1.0 predictive algorithms (<https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1>; <https://services.healthtech.dtu.dk/service.php?NetMHCstabpan-1.0>).

The peptides deposited at the cancer antigenic peptide database (<https://caped.icp.ucl.ac.be/Peptide/list>) were used to interrogate NetMHCpan 4.1 tool. Nanomer peptides for the MHC class I HLA-A*0201 and 2402 alleles (<http://www.allelefrequencies.net>) have been selected with a predicted affinity value < 100 nM (Strong Binders, SB).

Likewise, microbiota nanomer peptides identified by the BLAST homology search were used to interrogate NetMHCstabpan 1.0 tool for the four most prevalent MHC class I HLA-A*0201 and 2402 alleles. Strong binder peptides were selected with a predicted affinity value < 100 nM and stability > 1 h.



BLAST homology search

The TAAs selected as SB according to NetMHCpan 4.1 prediction tool have been submitted to BLAST for a protein homology search against Firmicutes (taxid:1239) and Bacteroidetes (taxid:976) sequences within the non-redundant protein sequences database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Homologous microbiota protein sequences have been extracted from the protein database of the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>) and epitope prediction has been performed with the NetMHCstabpan 1.0 tool.

Epitope modelling and molecular docking

Specific reference structures were selected from the protein data bank PDB (<https://www.rcsb.org>) for the HLA-A*02:01 and HLA-A*24:02 molecules. The 1AO7 complex was selected which includes the structure of the HTLV-I LLFGYPVYV epitope crystallized with the HLA-A*02:01 molecule, the β 2 microglobulin, the α and β chains of the TCR (PDB <https://www.rcsb.org/structure/1AO7>). The 7JYV complex was selected which includes the structure of the influenza YFSPiRVTF epitope crystallized with the HLA-A*24:02 molecule (<https://www.rcsb.org/structure/7JYV>). The PyMol software (PyMol

Molecular graphics system, version 1.8.6.2) was used to modify the epitope sequences into the peptides analyzed in the present study. The Molsoft Mol Browser (version 3.8-7d) was used to generate the epitope modelling and molecular docking as well as to calculate the inter-atom planar and dihedral angles.

Results

Blast search for homology between Tumor and microbiota antigens

Peptides from the cancer antigenic peptide database, predicted as strong binders (SB) to the most frequent MHC class I alleles were selected, as previously described [30].

Subsequently they were subjected to global protein BLAST against the bacteria sequences of the *Firmicutes* and *Bacteroidetes* phyla within the GeneBank non-redundant protein database. The search returned a large number of HLA-A*0201 and 2402 restricted bacterial sequences sharing homology with selected TAAs and the majority of such homology was found for the HLA-A*0201. Indeed, 267 and 159 homologous peptides were identified for Firmicutes and Bacteroidetes, respectively, for HLA-A*02:01. Of these, 95 and 55 were identified with a high affinity value < 10 nM for

Table 1 Alignment of microbiota-derived peptides and corresponding TAA

MAGEA10 - firmicutes

Peptide	Aff (nM)	Seq ID
G L Y D G M E H L	6,96	
. Y I	3,48	WP_012101781.1
. . . . S . . Y V	2,89	MBQ5676691.1
. L G V	2,94	MBC8599015.1
Y . . . D	3,33	MBQ7053643.1
A E V	2,81	HBI63224.1
. N T V	3,61	MBQ9361370.1
Y . . N . I . . I	3,71	WP_118553162.1
V M Y .	3,77	MBQ2982712.1
. V . . V	3,79	WP_185142326.1
K L . S .	4,76	HIU50197.1
K I	5,01	WP_117516162.1
M R . .	5,29	EDS01593.1
S L . Q . .	5,58	MBQ7065458.1
R L	5,83	WP_132012393.1
. . . E . . N . .	7,01	CAC8447084.1
A Q . .	7,16	WP_008590666.1
. . F . S	8,99	MBQ2769940.1
. M . . M	9,72	SCY45127.1
. I Y I	9,74	GFI38123.1
Q . . N V	9,89	MBQ7360485.1
. . . E A . . E A	10,19	EDO61928.1
. . . . V . . . I	10,66	HHT96763.1
. . . F N A	11,23	MBO5371573.1
L . . . A	13,06	MBQ5522151.1
K D . .	13,93	MBR0350109.1
. . . E T	14,31	MBQ3010667.1
A . . . T	14,52	WP_173629024.1
. N R A	17,9	MBQ3790769.1
. . . S I	19,91	WP_206458099.1
. F T	20,69	WP_063441043.1
N . . . N	21,02	WP_095131333.1
. . . E . . K . .	23,37	WP_098095089.1
T N D .	26,72	MBS5732462.1
V . N	27,42	MBF1006940.1
. V F S	29,84	MBQ1394714.1
. I T A	54,65	NBH26289.1
T Q Q C	60,42	HIR09928.1
K V C	66,68	MBQ7706286.1
N I L . . .	67,82	MBS5512791.1
. V F E	72,18	EOS74277.1
. V T	86,95	WP_212975324.1
V I . . A	136,6	MBS4967628.1
. I K I	136,88	MBQ6259525.1
H . T	136,93	MBQ3104720.1
P T . .	194,53	MBQ6814910.1
S M K	216,7	MBO5284181.1
T R Q T	289,26	HIS24277.1

Table 1 (continued)

MAGEA10 - bacteroidetes

Peptide	Aff (nM)	Seq ID
G L Y D G M E H L	6,96	
Y L D . I	2,41	KAB6529547.1
F . . . S V A . .	3,11	UKI38930.1
V I . . V	3,48	WP_134985449.1
. M L . . .	5,53	RYZ62962.1
. L . N .	6,22	WP_046368370.1
T . A Y .	6,78	KAA6301718.1
Y . . N . L K . .	7,22	HIU39037.1
R D E .	8,1	WP_004328988.1
R D A .	8,45	OKY94440.1
. L A	9,65	MBN1398196.1
V . . . N	12,56	NCA71030.1
. . . . E . . Q .	14,32	WP_089782056.1
. K A	15,73	WP_135531157.1
A D .	17,29	MBQ2520985.1
. . F . . . K . .	18,76	WP_194101443.1
A . . N . . D . .	20,93	WP_071843189.1
. R . .	21,51	WP_132555660.1
. . . H . . N L .	23,83	WP_229034998.1
F V . . . T . . A	27,69	WP_188752688.1
E L . . .	50,17	MBF8294546.1
Y Q . Q F	52,87	WP_117678479.1
A I A	60,36	WP_027076713.1
L . . . P K . . .	73,05	HIR83038.1
V T V	78,18	WP_188752688.1
F V . . . L . . T	79,73	WP_188752688.1
. K T	82,74	MBC8082572.1
K T . . . L N . .	157,46	HIT15347.1
R T . E E . . Q .	641,66	WP_044301905.1

MAGEA3 - firmicutes

Peptide	Aff (nM)	Seq ID
K V A E L V H F L	11,32	
Y L T . . L . . .	2,28	MBE6834566.1
. L . . M	2,57	TGY95779.1
I L N . V	2,98	HIV17749.1
F . S I	4,79	MBS6922347.1
F Q N Y .	5,75	HIV18837.1
V L H	5,84	WP_097838615.1
V L H	5,84	MBP3617361.1
R L D . .	5,92	HIQ58336.1
. I	7,1	MBP8691684.1
. I . . M	7,41	MBP7221824.1
. L R . .	8,41	MBR0093352.1
A I	10,43	WP_054028593.1
. L Y .	10,7	MBQ6875242.1
M Q Y .	14,09	HIU31414.1
R I . . .	14,41	MBQ3503822.1
. . . . F I N . .	14,52	MBS7060157.1
Y A . . . I . . .	14,67	MBD5491386.1

Table 1 (continued)

. I T	15,99	WP_035147891.1
H I	16,89	WP_143846559.1
. . S D . L	21,03	MBQ9865489.1
N I	24,53	MBC8577983.1
T . . . D Y .	24,73	HIT57844.1
S A	29,13	HBA64155.1
. . . . G . . . L . . .	30,9	HIV20050.1
Q Q . M	33,27	MBR5337548.1
Y A X	36,76	MBD5517959.1
M A	48,55	SEA27257.1
R Q S A	54,27	HIU69400.1
M I .	56,03	MBR6359707.1
A T . . . M L	57,39	HIS65127.1
Q	63,77	WP_188696003.1
E L A .	76,88	HIR10503.1
S D . .	92,31	HBR05003.1
A . T R V	112,64	HIV10253.1
. . T D . . . Q	136,1	MBP3701611.1
I S A	144,25	RGT99395.1
. N M .	150,7	MBC8578718.1
. D H .	246,7	EDO63073.1
. V A	451,92	MBR3171252.1
. Q . T A	976,6	HIU41270.1
. Q . T A	976,6	HIR41673.1

MAGEA3 - bacteroidetes

Peptide	Aff (nM)	Seq ID
K V A E L V H F L	11,32	
L L . . N I	3,1	WP_004328936.1
V L	3,31	MBM3426064.1
. F . A L V	4,84	MBS6652293.1
. F I	5,34	MBS1647326.1
. F I	5,34	WP_207362944.1
. H V	11	WP_078062927.1
Y I E . A	15,13	BAG84094.1
. D . E . . V	16,05	WP_004328005.1
. D . E . . V	16,05	MCD7972735.1
. . T	30,35	WP_169607473.1
L I L Q	31,53	HIU39050.1
. A	31,63	NBR83089.1
N P Y	32,26	HIU38292.1
. S	34,3	WP_126321206.1
. N	41,04	WP_095918246.1
Q I T	43,36	KAB6698732.1
S I E	44,33	WP_032598421.1
Q I	48,21	NQX86506.1
G S V I	48,34	HIR37285.1
A T	62,15	MBQ3673518.1
. D A	68,76	NLC86325.1
. D A	68,76	HBT85187.1
. . S N	76,48	MBC7417057.1
. . S N	76,48	WP_202104029.1
S . S I N	119,16	KAB3841810.1
. I V I K	139,87	MBS6650716.1

Table 1 (continued)

MAGEA3-12 - firmicutes

Peptide	Aff (nM)	Seq ID
F L W G P R A L V	7,65	
. V M	3,15	QNM09157.1
. A . I	3,36	MBQ1820230.1
. A P V . I	3,4	HIR20419.1
. F S	3,71	MBQ6708923.1
Y L	4,72	NLY91350.1
. G L	5,24	HHY09163.1
. D . . S G	6,17	HIR84741.1
. F . . . I	6,28	MBC8587361.1
. A . . . I	7,06	WP_070611851.1
Y K L	9,77	EHO34756.1
. N L	10,4	MBQ6563626.1
S H	12,74	MBR4261781.1
T . . . D S	13,15	MBQ2288613.1
. T V D A	14,38	HIT19511.1
. S . G	18,08	WP_161846669.1
. K . . . Q A	19,65	MBO5094480.1
. A . . . K I	32,66	WP_209615324.1
. K I	39,11	WP_113724721.1
. Q . A	49,91	WP_203627266.1
F L H G P K A L	49,92	MBO4679606.1
S S A	65,81	QNM07838.1
I . F L	110,56	HIR46659.1

MAGEA3-12 - bacteroidetes

Peptide	Aff (nM)	Seq ID
F L W G P R A L V	7,65	
. S I	2,41	MBQ8420774.1
. E S	3,02	MBF1567568.1
. H D	4,66	TAH08235.1
. N K	4,88	WP_204474554.1
. N K	4,88	MBN2774199.1
. I G	5,24	MBE7199410.1
R I	5,81	WP_068823941.1
. M E	6,08	MBL6865741.1
. K	6,64	MBL0883181.1
. L I	8,29	WP_137344335.1
. . Y . V G . . . L	8,9	HIU38645.1
T A P	12,91	HIQ85790.1
. R . . . V	14,24	MBC8007764.1
L K	25,41	NWK64964.1
Q A	34,92	RQP14817.1
A M . S Q	47,54	HIW87166.1

Table 1 (continued)

GnTV - firmicutes

Peptide	Aff (nM)	Seq ID
V L P D V F I R V	7,59	
Y I	3,29	MBD5555280.1
L L F . .	3,49	RGQ44771.1
F . Q E I	3,99	MBS4920904.1
R I	5,55	MBP1547557.1
. . . . M . . S .	5,88	MBD5531855.1
K . V . . L . . I	6,78	EDS02017.1
K V S I	7,15	CDZ24285.1
A . . E I	7,25	HIR19623.1
.	7,59	MBP0955392.1
. I . . K .	7,67	NLO48847.1
. L . . .	7,77	MBQ2781239.1
. . . . E . . V . .	7,91	WP_191543324.1
S . . E . . S . .	7,98	HIT18232.1
. . . E I . . K .	8,15	MBE6007693.1
. . . E . . . K .	8,61	MBQ5332073.1
. . . E . . L K .	8,68	WP_154251972.1
. . . E . . L K .	8,68	CDZ23284.1
. E K .	12,88	HIR60123.1
I L	13,37	MBE5970738.1
E C	13,48	WP_073279788.1
G . . E . Y . . .	13,84	MBD5533823.1
. . D I	16,57	MBE5893125.1
A . T I	18,69	WP_055186996.1
L T . I	19,45	MBQ6833545.1
I R K .	25,79	HIS76208.1
A . . E . . R . .	31,24	EDO61119.1
. . . . T . . L	37,23	UEA73896.1
. G A	39,15	MBR2409146.1
. . . E . . R K .	39,94	HIR41446.1
N I	44,62	WP_002573979.1
Q V . I	61,93	MBS5309671.1
S C A	64,67	MBR6147590.1
F C	99,54	HEK51933.1
E . T . I	168,64	HIQ80887.1
E . T . I	168,64	HIQ80887.1

GnTV - bacteroidetes

Peptide	Aff (nM)	Seq ID
V L P D V F I R V	7,59	
I M V P .	4,53	HIR37852.1
L . S	4,93	WP_118465647.1
F I . . . I	5,07	MBR6655791.1
F I	5,54	MBR1962375.1

Table 1 (continued)

I . Q	6,72	PGH37641.1
I . Q E	7,35	KAB3859120.1
I V P I	8,09	HIR83626.1
G L . .	9,62	MBM3443229.1
G L . .	9,62	WP_136842267.1
G L . .	9,62	WP_090992641.1
. M . . A . . G .	9,8	MCD7971285.1
. . Q E	10,08	CUN36400.1
. . . . T	10,59	GFD90141.1
I M . . I . . P L	10,6	MCD7973287.1
I M V P L	12,18	HIU38465.1
. G L	17,68	MBO7397429.1
. A L	19,98	RCU43886.1
. . Q . . . V K L	24,12	KAA6301070.1
I I . . A I . . .	32,64	MCD7972313.1
I . . E . . A . M	41,77	MCD7972526.1
. V . A	43,85	WP_204246478.1
. . . . A . . . A	90,14	MBP7641748.1
. . . . I G . K I	163,27	WP_055432820.1
K A	275,19	WP_229095086.1
. . H . . . V K T	336,21	HIR36848.1
E . . E . A . . .	567,32	HIU38848.1
. . R . I . . N M	637,19	HIT16500.1

MAGEA1 - firmicutes

Peptide	Aff (nM)	Seq ID
K V L E Y V I K V	5,83	
R L . D . L . . .	2,56	HIU39589.1
. . M . . L . S I	2,99	WP_158387495.1
R L . . . I V . I	3,98	MCD7973973.1
. R .	5,57	NLJ26328.1
. A . .	6,6	WP_077853732.1
. D .	7,39	WP_206808822.1
. A . M .	7,57	SEP60213.1
. T . .	7,83	WP_186300960.1
. A . R .	10,34	MBE5895429.1
. A . . .	11,22	RKW64949.1
M . . D . M . . I	12,84	WP_156076623.1
. I . . I	13,23	MBC7087093.1
L L E . .	15,6	WP_223384067.1
T I . . F A . . .	23,38	HIT15351.1
. T I . . . N . .	24,62	HIT15315.1
N . . G	28,39	MBR3806748.1
. I . R A	34,23	MCD8014764.1
S G . . .	43,69	MBO5354362.1
. D . I	86,52	MCC2168016.1
. . . . T . R . .	212,37	MCD7974005.1
. . . . I V . T	223,62	MBQ9562458.1
. A R N .	232,38	HIR82602.1
. . K . . . E . .	575,97	WP_181573024.1
E I . . N	725,03	HIT15621.1

Table 1 (continued)

MAGEA1 - bacteroidetes

Peptide	Aff (nM)	Seq ID
K V L E Y V I K V	5,83	
A L	3,04	WP_215235361.1
H L	3,6	WP_115829840.1
. L P . .	5,07	HIT79747.1
R I	5,29	WP_130140291.1
. A . .	6,6	MBK6621957.1
S M K . .	8,29	MBE6834075.1
A V . .	8,61	WP_173416227.1
. . . I . . . L . E A	10,65	WP_227088557.1
. E . .	10,97	WP_109650090.1
. C . . .	13,52	HHE65310.1
A L L . . . T	14,29	WP_007569258.1
. Q G I	14,92	MBS5732492.1
. . . N N . . . T . .	18,55	MBE6832653.1
N V . . .	20,59	MBP5995608.1
M S L . . .	29,93	WP_195610109.1
I . G V . . .	30,03	HIR84910.1
. I H . A	31,49	WP_068855645.1
L L	33,91	MBK5280326.1
. K	53,12	MBO7562654.1
A T A T . . .	58,93	WP_215721925.1
. N E E L	69,07	WP_117829775.1
Y F E . . .	125,24	HIU50608.1
. . . Q . . . K R . .	531,53	HIT58468.1

MAGEC1 - firmicutes

Peptide	Aff (nM)	Seq ID
K V V E F L A M L	25,53	
R L	6,41	TLS53416.1
. I L	8,49	MBR5011159.1
. L . A	10,29	MCC2147852.1
. T V . . .	11,37	MBE5922580.1
A L F T	15,41	HIT18255.1
. . A D	17,69	WP_105977819.1
. . I S . . .	20,76	HBP64347.1
. I . D Q A	22,32	HIR03824.1
.	25,53	MBO5479965.1
G . I	27,04	WP_092482186.1
. I I S . .	32,53	NBI58092.1
A I	84,4	WP_223385377.1

Table 1 (continued)

MAGEC1 - bacteroidetes

Peptide	Aff (nM)	Seq ID
K V V E F L A M L	25,53	
. . A . . . V . L V	4,84	WP_195272405.1
. . I Q D V	8,68	GAP71625.1
. . I L . .	9,11	WP_111611765.1
C L N P . . .	9,39	HIR82174.1
F . . D Q . .	10,12	WP_005802251.1
. . . D L . .	19,23	WP_210489586.1
R I . D L . .	19,25	WP_008640349.1
N L I E . . .	21,85	WP_117720041.1
R I . D Q . .	22,59	WP_004320436.1
I . L	26,51	QCX40738.1
T . L	50,96	MBL7846776.1
S . A E . . .	63,35	HIT15828.1
Q . A	85,24	QTN38443.1
E L N	141,69	HIQ84870.1

MAGEC2 - firmicutes

Peptide	Aff (nM)	Seq ID
A L K D V E E R V	239,61	
Y . . E P . .	8,63	HIR60313.1
Y Y . . .	10,09	EDS00205.1
Y Y . . .	10,09	MBS5730812.1
L . S	14,52	HIS24073.1
. . S G . .	17,22	HIS65612.1
Y I D	20,38	MBE5851475.1
Y I D	20,38	MBR3769223.1
. . S	23,82	WP_125553189.1
. . N L . .	51,02	NDO19858.1
M I	51,51	MBQ7934778.1
. P	52,68	MBQ7202456.1
. P	52,68	MBQ6558180.1
Y . . N M . .	56,46	HIS24052.1
M . . E	67,68	WP_053430754.1
. . T A L . .	68,45	HIR09109.1
L Q	93,41	HBA96956.1
C . D K . .	110,75	HIT80107.1
R . . E I Q . .	119,53	HIV18755.1
. . A . I A . .	164,41	HIQ59651.1
G I . . E L .	182,42	HIQ57945.1
. D A K . . .	186,53	HIR10457.1
T . . E I Q . .	200,53	HIQ59304.1
E . M	211,39	NSI91570.1
N . Q	231,29	MCD8338125.1
E . M	231,39	WP_118486153.1
. A L	274,81	MBQ7506690.1
F I	279,03	MBD5494815.1
. . N K . . .	325,28	HIQ69970.1
. . R . L K . .	508,44	HIR83953.1
G L L .	661,49	EDS00137.1
. . . A . Q	790,13	MBR4606065.1
. . . S L . . .	1456,57	EEU95410.1

Table 1 (continued)

MAGEC2 - bacteroidetes

Peptide	Aff (nM)	Seq ID
A L K D V E E R V	239,61	
Y	28,67	WP_076782989.1
F . . S Y . . H L	39,41	WP_117928663.1
. . D	43,24	MBQ1773637.1
. . S N . V . . L	53,11	HIT47270.1
S D . S .	74,1	WP_195272339.1
I I	99,47	WP_188370161.1
. . . E . A M . .	116,23	MBS6650723.1
R . . E L . . A .	176,64	WP_117928663.1
V I N . . N S . .	437,28	WP_117928663.1
. . . S . . . K .	921,33	RIB34267.1

SSX2 - firmicutes

Peptide	Aff (nM)	Seq ID
K A S E K I F Y V	15,3	
R M I . . M . . .	1,6	HIV19461.1
F M R	2,49	MBQ8814608.1
. L L	2,59	MBO4462507.1
A L G	2,76	MBP3875601.1
A L L	3,05	NBJ01656.1
G L L	3,21	MBR2530847.1
. M . . S V . T I	3,49	HIR52329.1
G L T L	3,66	HIT09333.1
Y M . . .	4,14	WP_193137167.1
S V N	5,39	MBQ7724778.1
V T A . T	5,87	HIU41236.1
G V	6,18	OJU08501.1
A N	7,84	HHY18381.1
Q L K . . L . . .	8,93	CBL34879.1
Q L K . . L . . .	8,93	CUQ83949.1
. . A D	9,18	MBQ6843512.1
D . T	9,55	WP_039861470.1
E L . D	9,59	MBE5889045.1
. . A . . F	10,1	MBR2037171.1
S T L I . .	11,81	HIR10297.1
. . . D	16,93	MBQ8518875.1
. T A R .	20,34	EEG32253.1
T T L I . .	22,74	HIR50649.1
V Q P I	25,31	EDO62734.1
V Q P I	25,31	RGQ44838.1
I . G	26,23	WP_057897585.1
. . N Y . .	28,53	WP_084053470.1
N . A . R	30,07	MCC2230127.1
S Y . . .	37,86	WP_091273761.1
. . . A . . . L	40,14	HIT94498.1
G S	59,42	WP_098398638.1
T T A Q	59,63	HIR49715.1
. . G D R . Y . .	66,14	HIU49597.1
G Q S A	119,79	HIS76607.1
. T K . . V . . I	623,57	WP_005923638.1

Table 1 (continued)

SSX2 - bacteroidetes

Peptide	Aff (nM)	Seq ID
K A S E K I F Y V	15,3	
Y T	2,68	RYE15078.1
. L N . . .	2,75	WP_117830273.1
. I . . D V	3,01	MCD7971606.1
V V . . D	4,52	MCD7972676.1
. . Y	4,93	MBQ6038980.1
. V T	5,36	MBK9284067.1
. T . . . V	9,03	HCQ15094.1
. . A . . V	10,07	MBQ9583024.1
. T T	10,46	NBP69626.1
. V L	10,49	WP_196828382.1
G . . . R	37,69	MBP1646831.1
V F .	46,62	WP_132989163.1
. I R . . .	72,59	HIS22046.1
R L Y F .	75,44	UKI40112.1
V T	75,82	WP_027124754.1

KM-NH-1 - firmicutes

Peptide	Aff (nM)	Seq ID
N Y N N F Y R F L	157,45	
. . T . . C . . F	34,47	EEU98063.1
G . Y N F	41,77	CUQ81199.1
R . . G	44,03	WP_155705275.1
. . I R	50,33	MBO4988623.1
. . . . L Q K I	60,81	PZM89890.1
K . . . Y	65,88	WP_154750440.1
. . . S . . S . .	75,01	WP_191684926.1
. . . . V	88,64	WP_088223547.1
K . . E . . . L V	90,51	WP_097770998.1
E . S A F	106,98	HIR10485.1
K . . D . . . Y .	110,61	MBS7105243.1
D . A A F	113,9	RGC41348.1
D . A A F	113,9	HIT09977.1
F N .	121,17	MBS7060054.1
S . G M	129,9	MBE5948340.1
S . . . L . . Y .	131,59	MBR1633626.1
H . T D I	142,97	HIT54277.1
D . . A . . . S F	160,2	WP_154265874.1
D . S A F	207,62	WP_143406190.1
D . S A F	207,62	HIT79450.1
. . . . I . Y . .	243,95	MBO4484513.1
. . . . V . Y . .	299,5	MBR5667449.1
. . D . . F . M .	715,92	MBQ2275888.1
L E	770	SDI89196.1
A . D E	1278,64	MBQ6105135.1
D . S . . N	1304,83	HIT53773.1
T . . E Y	2494,72	MCD8119786.1
E . D D S	3723,4	WP_097792912.1
D . D N .	3769,86	HIU69252.1

Table 1 (continued)

KM-NH-1 - bacteroidetes

Peptide	Aff (nM)	Seq ID
N Y N N F Y R F L	157,45	
M . A	18,89	MBR8416245.1
T F	20,38	WP 160633064.1
R . G	24,16	WP 134271503.1
. . . S F	25,03	MBS6268346.1
H . Q E . .	26,99	WP 195272477.1
K . D . Y . . L F	27,99	HIR83490.1
V L . .	32,39	WP 094414045.1
. S I E F	34,24	KAA6301807.1
. S I E F	34,24	MCD7973646.1
. . . D . . D K F	46,12	HIS22097.1
K . D . Y . . L I	71,66	HIS21838.1
. . S . I	80,98	HIN99713.1
. S . .	82,37	WP 109853465.1
. . . R . . D S I	87,09	EDS04678.1
. . . . G . . T I	148,56	HIT15675.1
D . . G . . A N I	323,08	UKI39263.1
H . D . L . S . .	674,97	HIR38703.1
L . D . Y . . I . .	717,7	WP 005679986.1
Y . D V . . N D .	792,26	HIS22865.1
L . . T . . H . Y	934,97	GAP72185.1

KM-NH-1bis - Firmicutes

Peptide	Aff (nM)	Seq ID
E Y L S L S D K I	249,87	
M . M D	18,82	RGU04328.1
L . . F . . N . .	31,27	HIR51220.1
Y . . . I . . . L	38,38	HIR03062.1
C I . . .	46,39	HIU72613.1
. . . . T . Q F	58,87	RGB96811.1
K N . . .	63,31	MBQ1487485.1
Q . . T . T . . .	88,74	EDO62611.1
. . . . M . T . .	102,61	MBQ5559680.1
. E . .	113,2	WP 077534359.1
K F . . . F . . .	140,25	EEG29117.1
. . . . F	153,16	HHZ06911.1
K D . . .	156,27	MBQ6638262.1
K D . . .	156,27	MBR3638856.1
. . . K . T . . .	160,24	MBQ9503861.1
. S . . .	168	WP 090091604.1
. V . . .	187,93	WP 059108549.1
. . I	234,91	MBS6171231.1
. . I	234,91	MCB5920702.1
. S . .	243,41	MBR7044961.1
.	249,87	WP 120409454.1
G . . K	284,35	MBQ9991463.1
V . . . N . . . V	398,44	HIU50603.1
. . . N I N . . .	440,45	RGQ37783.1
. . . . F . R M	538,83	HIQ58800.1
. . . T . G . . .	680,99	MBQ7765997.1
. . . . K . . L	1895,38	SCY81903.1

Table 1 (continued)

KM-NH-1bis - bacteroidetes

Peptide	Aff (nM)	Seq ID
E Y L S L S D K I	249,87	
V I . E . .	11,19	MBR3831564.1
Y . . E . . E . .	13,15	HIW86655.1
Y . . A	17,52	MBP3767014.1
S A N N .	18,35	WP 215721770.1
R	24,29	MBP3566537.1
I A . . .	25,89	MBN2349226.1
R . . A	26,76	WP 117507904.1
K T . E . .	29,18	HIS23517.1
K A . . .	35,18	HBG53681.1
N E . . .	38,78	MBT5932717.1
Q N . . .	45,22	MBN1133131.1
F N . . .	73,36	WP 077374390.1
. Q . . .	73,82	OPZ03153.1
. . . . I . N . .	96,92	MBK9212880.1
H	102,11	MBS1488621.1
F . . D . . . I .	109,3	MCD7974127.1
. . . Q . T T . .	132,29	HIT15593.1
. . I . . I . E . .	153,4	UKI39005.1
. . . W . A . . .	175,98	WP 234222759.1
D I F	183,23	WP 057279431.1
. . . . I . R I .	207,14	WP 141405227.1
. . . R I . . E . .	213,28	HIQ84544.1
. K N I .	481,6	HIW86671.1
. V V E V	1394,46	MBS6651286.1

MAGEA2 - Firmicutes

Peptide	Aff (nM)	Seq ID
E Y L Q L V F G I	135,27	
A S . . .	28,34	MBO6286936.1
V . . . V	31,9	WP 133444878.1
S . . . K	27,53	MBD5494776.1
I . . V . . Y . .	39,65	WP 173015421.1
. . M . . . A E .	41,91	HIQ79709.1
. . . . I . P . .	44,37	WP 048924608.1
Q . . H I I . . .	46,22	WP 112148660.1
. . . . I . N . .	53,19	MBQ6625054.1
. . . . G . S . .	57,87	MBE5909468.1
. . M . R	60,4	MBQ6386375.1
. . . E . L Y . .	92,27	PNH19151.1
V R Y V	221,01	HIT53822.1
. T Y L	383,41	MBT9711503.1
A K Y L Y	645,9	MBR6326373.1
. . . . I . E N V	1138,34	HIU50301.1
. N Q V	1576,42	MBP3569136.1
. Q R Q L	2014,53	MBD9003888.1
. . . E . . . N Y	3025,55	MBO4787423.1

Table 1 (continued)

MAGEA2 - bacteroidetes

Peptide	Aff (nM)	Seq ID
E Y L Q L V F G I	135,27	
T . M . M . . . F	5,53	KAA6302458.1
R . . . H	11,45	WP 071503703.1
. F H Q .	22,61	GAP71734.1
. . A P . . . E .	34,46	HIT15790.1
Y D K L	42,03	HIR83440.1
L A F	47,39	WP 016270428.1
. Y . F	50,1	RYZ30368.1
K . . . T . . . L	59,91	BAG83909.1
. M A	84,53	MBQ6688332.1
. T T K .	132,29	HIT15593.1
. S A	149,89	KAA6301763.1
. T . . C . F	325,48	OKY93250.1
. . V . M I . S M	397,78	MBS5322006.1
A F E T .	442,61	KAA6300759.1
. Q . I D .	684,22	MCD7973596.1
. . . . S . . V E V	1394,46	MBS6651286.1
. A . R E V	1772,63	WP 004330097.1
. E L R M	2767,63	HIR83269.1
. D . . H Y	4748,83	HIR82591.1

SAGE - Firmicutes

Peptide	Aff (nM)	Seq ID
L Y A T V I H D I	36,15	
Y . . S L	10,8	MBO4309970.1
I . . A I	12,64	MBE5866694.1
. . Q S E . . H .	14,81	EDO63150.1
Y . . . L	15,14	MBO4775486.1
Y . . . L V	17,07	MBD5502315.1
. A . . . E .	18,51	WP 115453929.1
H . . N . . . S W	18,65	HIS65767.1
G F . K .	24,9	MBS7060185.1
. . . A . L . E L	25,51	MBS7082787.1
. . . S . L	27,58	MBO4863242.1
. F Q . . .	27,99	MBD5156035.1
. F Q . . .	27,99	MBD5538214.1
. . F L L	28,11	WP 215696595.1
. . . A I L	30,56	MBE6011823.1
. . L . I	32,75	WP 091272910.1
. A E . . .	34,51	MBQ6590101.1
A . G E .	37	MBE5883573.1
F . . A . L Q . . .	39,33	EDO60106.1
. . S . . F Q . . .	40,63	MBD5497489.1
Y . E E .	41,72	BCN32811.1
R . V L	44,28	MBR6850879.1
A . Q . . . F . . .	47,04	RGQ38182.1
I . G A . . N . . .	62,78	HIV17791.1
A . . S . . Q . . .	64,27	AVM69455.1
F N	67,57	GFI65663.1
N . S A	74,81	MBR1931959.1
. . V A . V Q . . .	77,67	MBS7028433.1
. . . N . . . V . . .	88,62	MCC2149538.1

Table 1 (continued)

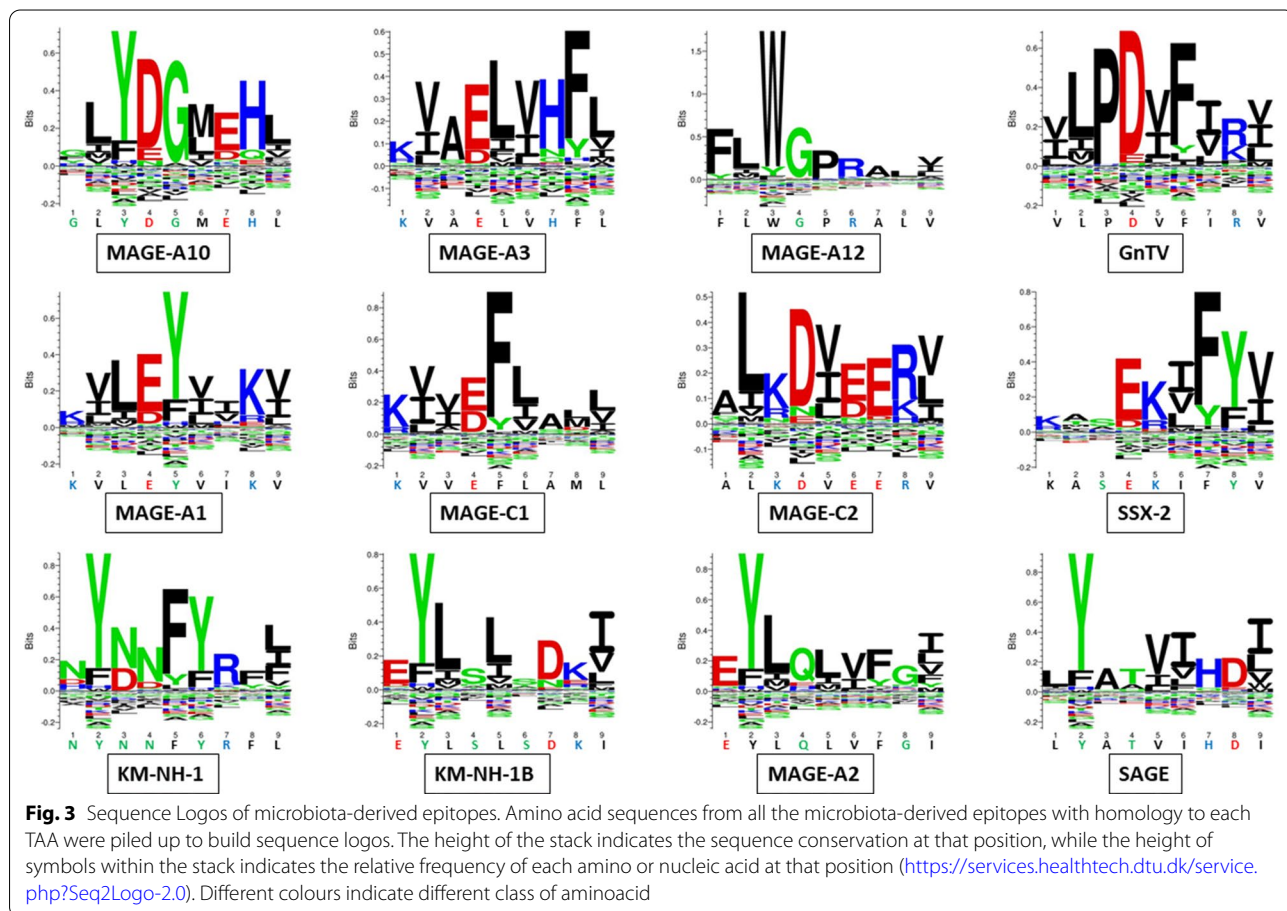
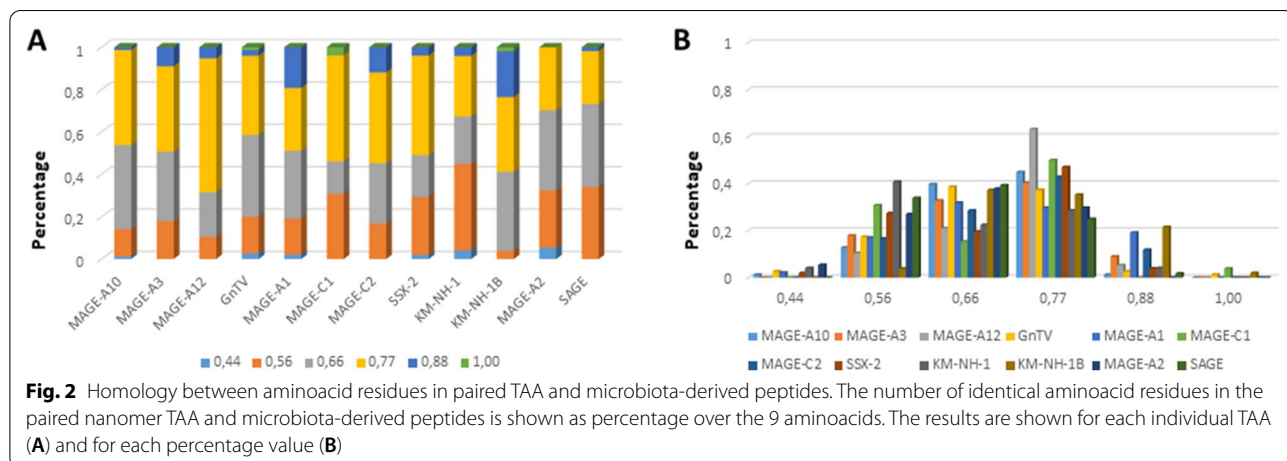
. . . S . . V . .	92,15	MBR3307355.1
. . . S . . V . .	92,15	MBQ5485557.1
. . T . . . S A .	129,48	MBS6320402.1
. T . . .	139,36	MBR2729061.1
P . V L	222,38	MBR6381150.1
P . V L	222,38	MBD5527040.1
E . . A A V . . .	282,8	HIR61685.1
. . T M	516,88	MCD7863943.1
A S Q M	556,25	HIR59982.1

SAGE - bacteroidetes

Peptide	Aff (nM)	Seq ID
L Y A T V I H D I	36,15	
K . . A . . E E .	20,8	WP 055432820.1
. . S E .	21,9	WP 204530868.1
M . . L S	33,09	KAA6303530.1
V . P L . . A . .	36,44	MCD7971608.1
. . . K . . E . W	46,82	HIQ85501.1
P . . . L . F . .	55,05	HIQ85112.1
. . . A . . W F L	77,09	HIU38446.1
R G I L	87,45	UKI39671.1
. . . S L . . . L	88,91	WP 172774650.1
. . . . L . M V L	115,97	KAA6301466.1
. . . . L N . E L	122,79	HIQ84490.1
. . . . G F G L	151,95	HIR36767.1
Y T T V I S D I	187,18	WP 117853018.1
. Y I L V	206,76	MCD7972560.1
. L V F L	232,9	HIR82641.1
. E Q N L	299,76	HIR37436.1
Y D N V I H D I	902,43	ALK82916.1
A . D N	902,43	KAA6300193.1

Dots = Identical residues, Green letters = conservative residues, Black letters = non-conservative residues. The green shades indicate indicated peptides with high affinity to HLA (<10 nM)

Firmicutes and Bacteroidetes, respectively. The MAGE-A10 was the TAA with the highest numbers of homologous peptides for both Firmicutes and Bacteroidetes (e.g. 50 and 28) while the MAGE-C1 and MAGE-C2 were the TAA with the lowest numbers of homologous peptides for Firmicutes and Bacteroidetes (e.g. 12 and 10), respectively. On the contrary, 112 and 81 homologous peptides were identified for Firmicutes and Bacteroidetes, respectively, for HLA-A*24:02. Of these, none and only 1 were identified with a high affinity value <10 nM for Firmicutes and Bacteroidetes, respectively. The SAGE and the KM-NH-1b were the TAA with the highest numbers of homologous peptides for Firmicutes and Bacteroidetes (e.g. 38 and 24), respectively; while the MAGE-A2 and SAGE were the TAA with the lowest numbers of homologous peptides for Firmicutes and Bacteroidetes (e.g. 18 and 18), respectively (Fig. 1).

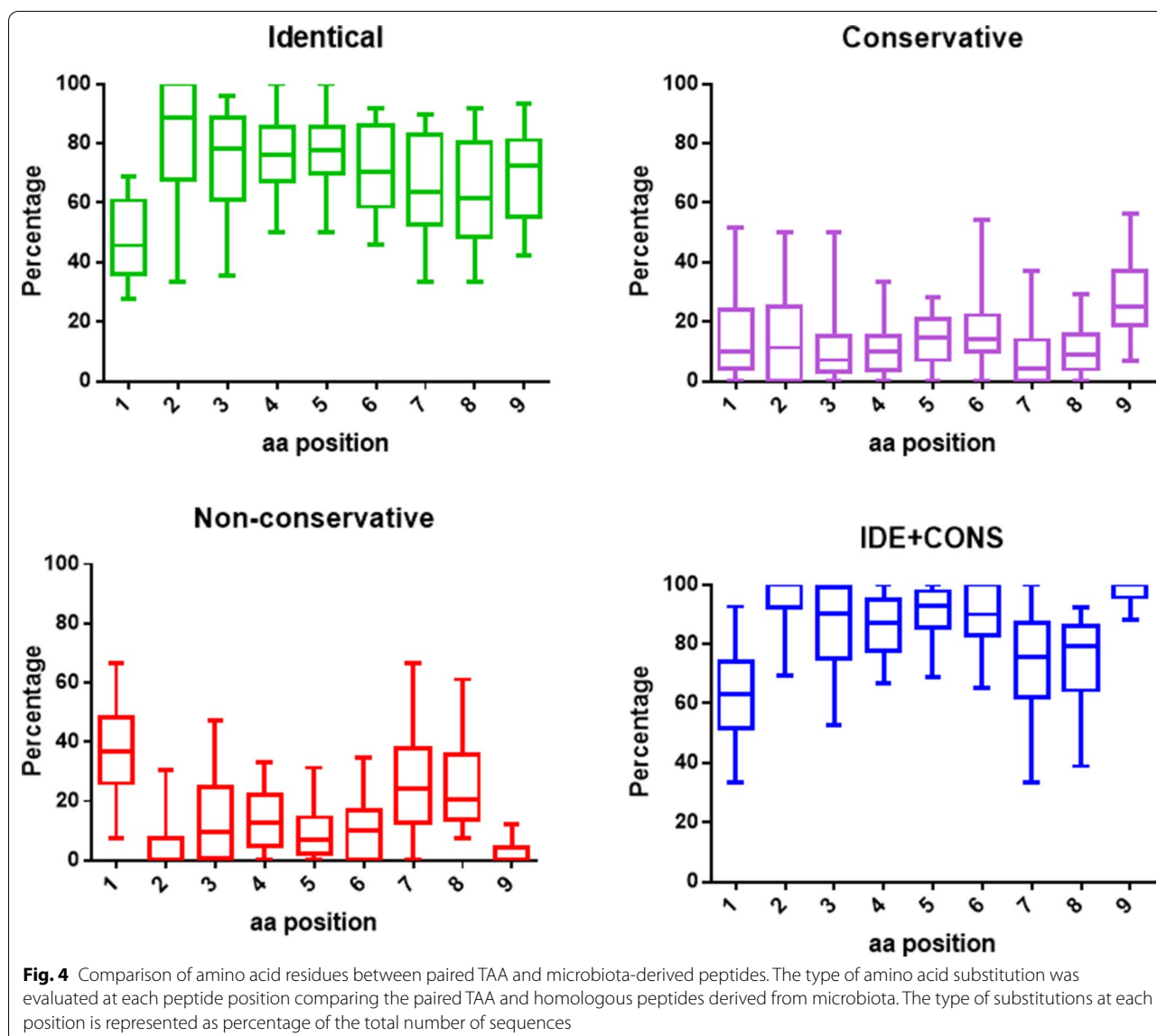


Sequence homology between tumor and microbiota antigens

The alignment of peptide sequences showed high levels of homology between TAAs and peptides from bacteria species of the two phyla. Indeed, the vast majority of microbiota-derived peptides (69.7% on average, minimum 51% for KM-NH-1—maximum 84.6% for

MAGE-A10) showed 6 or 7 residues along the sequence identical to the corresponding TAA (Table 1; Fig. 2).

Strikingly, three microbiota-derived peptides showed a sequence identical to the corresponding TAA, the GnTV (Firmicutes Oscillospiraceae bacterium Seq ID MBP0955392.1), MAGE-C1 (Firmicutes Clostridia bacterium Seq ID MBO5479965.1) and MK-NH-1 (Firmicutes



Roseburia sp. 1XD42-69 Seq ID WP_120409454.1) (Table 1). To further confirm the high homology between the paired epitopes, the pattern of the microbiota-derived peptides showed that the most frequent amino acid residues at each position perfectly matched with those in the sequence of the corresponding TAA (Fig. 3).

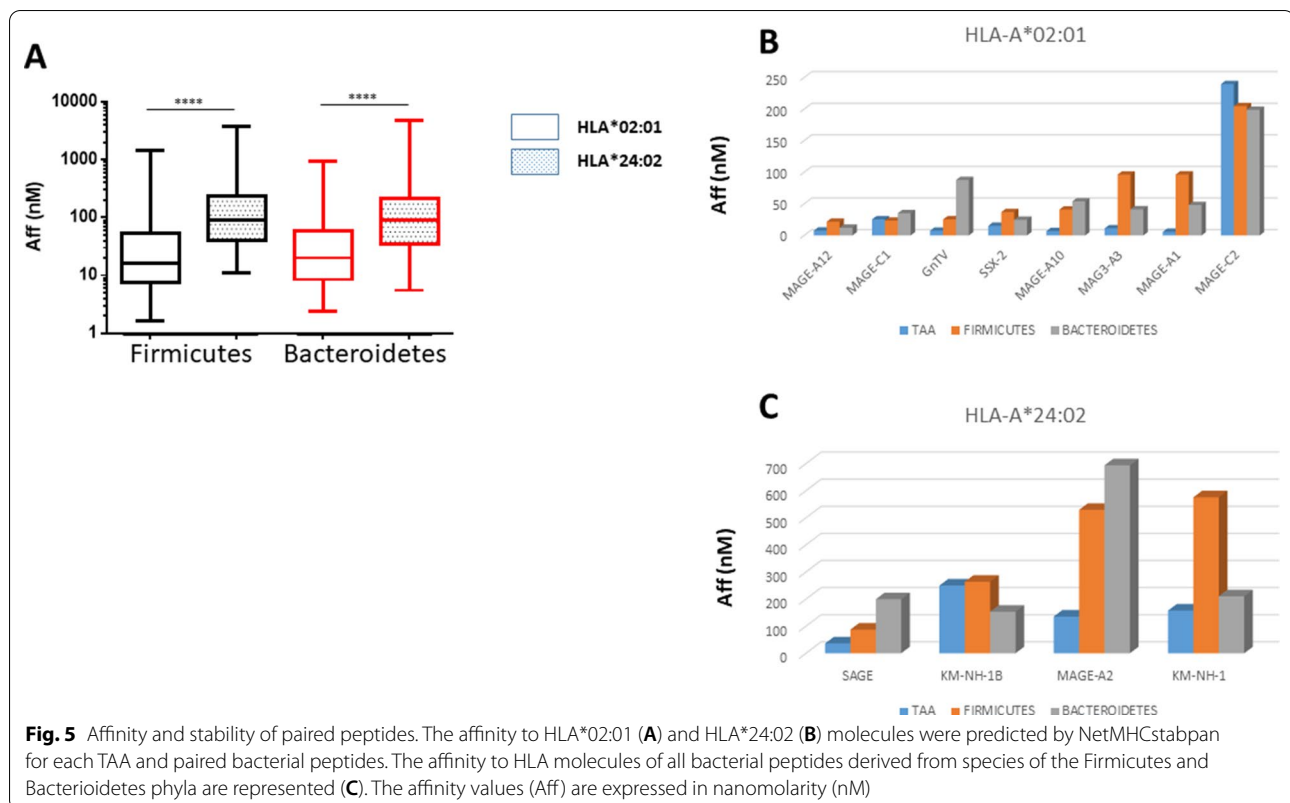
The alignment of the TAAs with homologous bacterial peptides shows that the amino acid residues at each position are mostly identical or of the same class (conservative substitution) (Table 1).

Indeed, the average percentage of identical residues at each position is 82.8%, ranging from 79.7% at position 3 to 89.5% at position 2. Summing up also the conservative substitutions, the average percentage of at each position is 93.6%, ranging from 86.7% at position 7 to 98.3% at

position 9. Consequently, the average percentage of non-conservative substitutions at each position is only 6.39%, ranging from 1.6% at position 9 to 13.2% at position 7. In such analysis, position 1 represents an outlier with the far lowest average percentage of identical (48.2%) and identical + conservative substitutions (62.4%) and the far highest average percentage of non-conservative substitutions (37.6%) (Fig. 4). The percentage of different type of substitutions at each position for the microbiota-derived peptides significantly varies among TAAs but the trend is overall confirmed (Additional file 1: Fig. S1–4).

Affinity to HLA molecules

The predicted affinity of the microbiota-derived peptides to the HLA-A molecules was, on average, very



high. In particular, the average affinity was 145.1 nM and 128.5 nM when considering all peptides derived from Firmicutes and Bacteroidetes, respectively. Interestingly, the average affinity values were significantly higher for the Firmicutes and Bacteroidetes peptides restricted for HLA*02:01 (68.6 and 56.5 nM, respectively) than for HLA*24:02 (327.5 and 335.54 nM, respectively) (Fig. 5A). Considering each HLA*02:01-restricted TAA, the homologous microbiota-derived peptides showed an affinity ranging from 21.8 (MAGE-A12) to 204.8 (MAGE-C2) (Firmicutes) and from 12.2 (MAGE-A12) to 198.9 (MAGE-C2) (Bacteroidetes). Considering each HLA*24:02-restricted TAA, the homologous microbiota-derived peptides showed an affinity ranging from 87.4 (SAGE) to 575.8 (KM-NH-1) (Firmicutes) and from 153.4 (KM-NH-1B) to 693.3 (MAGE-A2) (Bacteroidetes) (Fig. 5B and C). In all cases, more than 80% of the microbiota-derived peptides showed a binding affinity lower than the average value and 20 – 50% of them showed a binding affinity lower than the corresponding TAA (Additional file 1: Fig. S5).

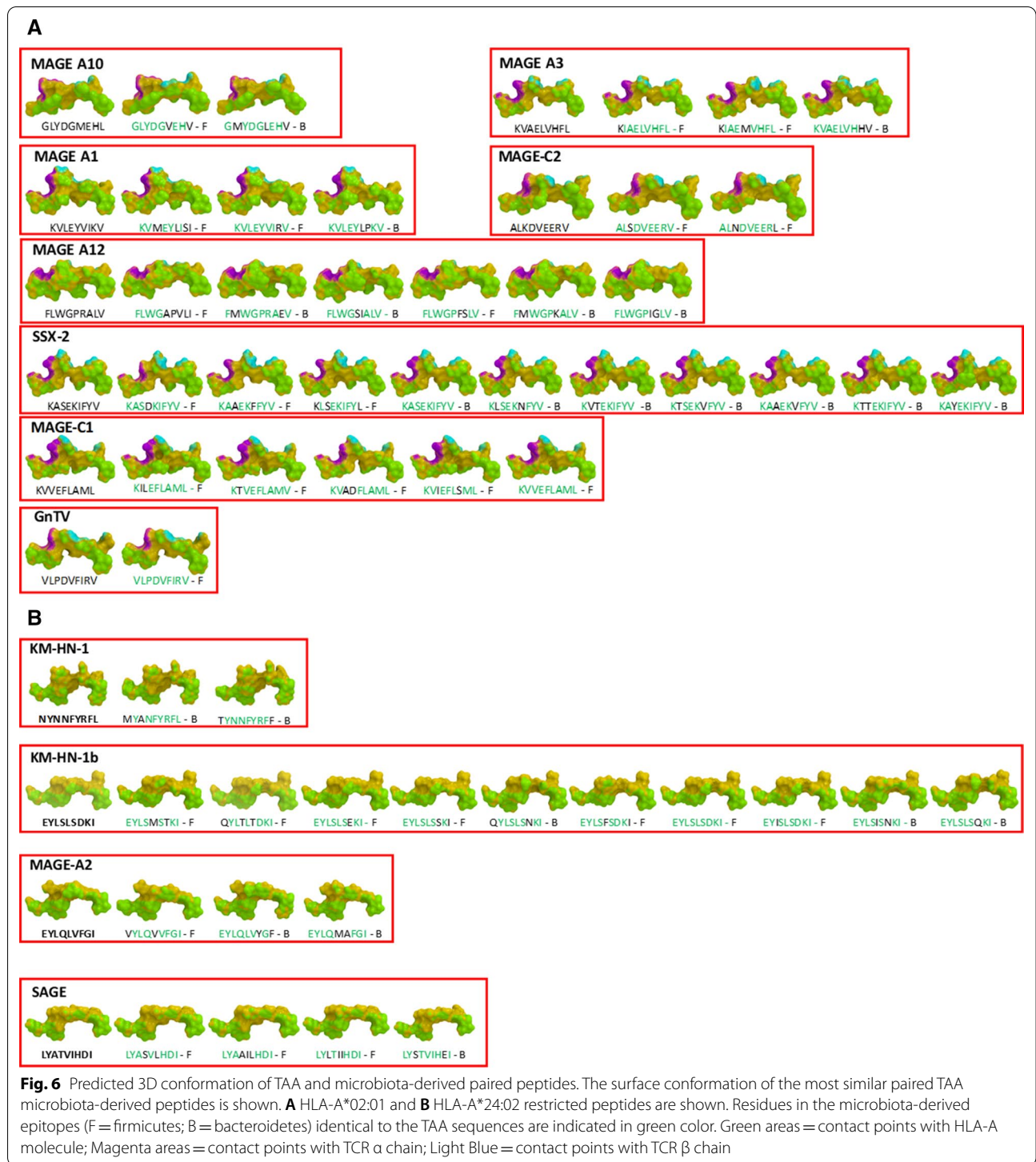
Epitope modeling and molecular docking

In order to verify whether the predicted paired TAA and bacterial epitopes share similar conformation and contact residues with the HLA molecule, epitope modeling and

molecular docking were performed taking advantage of crystallized structures deposited in PDB. In addition, the comparison of the contact residues to TCR chains was possible only for HLA-A*02:01 restricted epitopes, due to the lack of crystallized structures including both HLA and TCR for HLA-A*24:02 deposited in the PDB (Additional file 1: Fig. S6–S17).

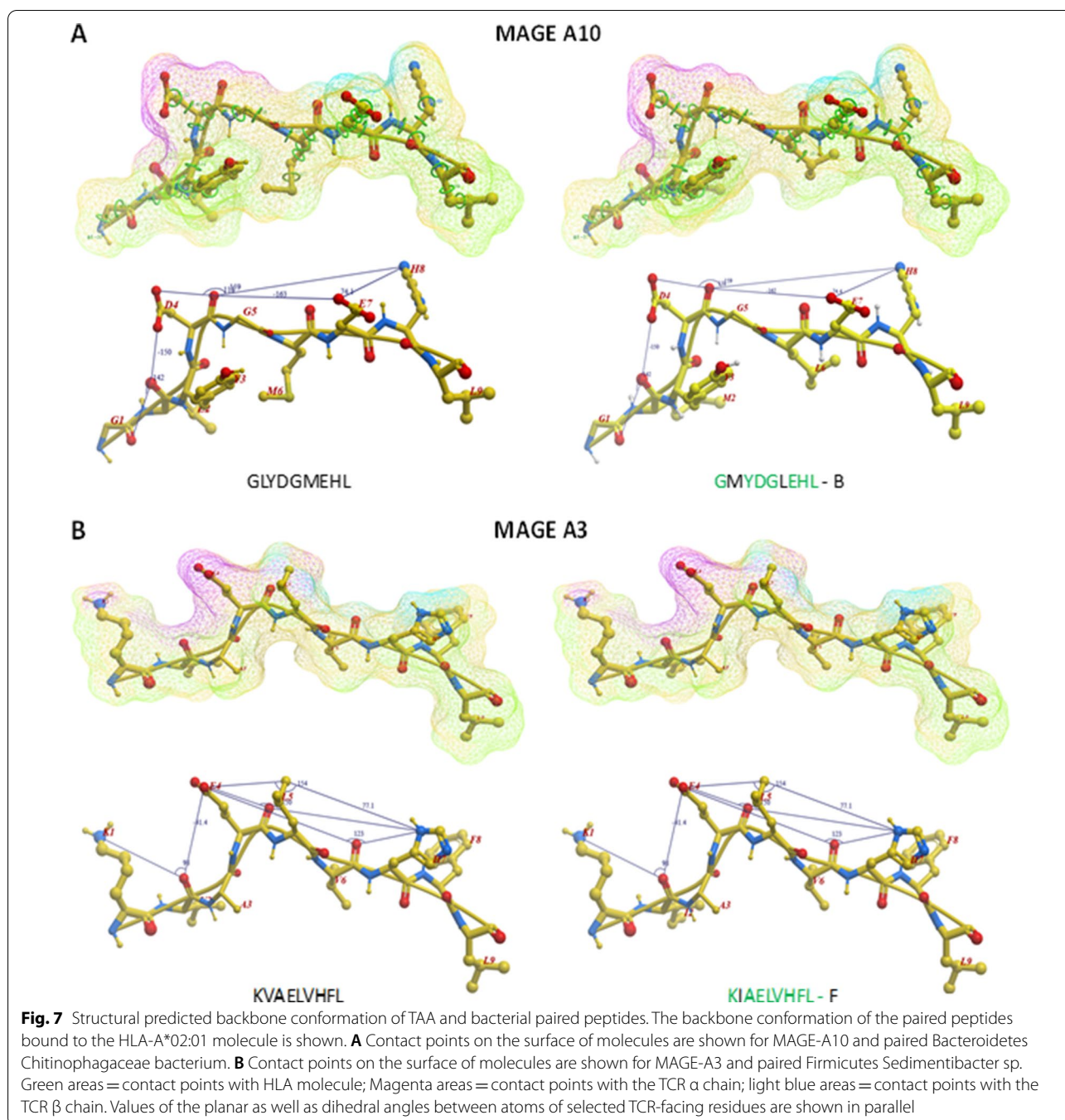
Some of the peptide structures from the microbiota-derived epitopes show a pattern of contact with the HLA molecule and the TCR chains (for the HLA-A*02:01) strikingly similar to the corresponding TAA. Indeed, in some cases, the structures of paired TAA and microbiota-derived epitopes are indistinguishable (Fig. 6A and B).

Additional information about the similarity between the paired TAA and microbiota-derived epitopes is provided by the peptide backbones. Indeed, considering the spatial position of the residues interacting with the TCR, it is possible to identify paired epitopes which share the same planar as well as dihedral angles ($\pm 10\%$) between selected atoms of the amino acid residues in the same position of the peptide. Such analysis shows that, among paired TAA and microbiota-derived epitopes with highly similar structures, it is possible to identify selected pairs with 100% identical angle values, especially in HLA-A*02:01 epitopes (Additional file 1: Fig. S18–24). As example, the GLYDGM EHL



MAGE-A10 and the GMYDGLEHV (Bacteroidetes Chitinophagaceae bacterium) peptides with M to L substitution in position 2 and an L to M substitution in position 6 shows identical spatial conformation of

TCR-facing residues (Fig. 7A). Same result is observed for the KVAELVHFL MAGE-A3 and the KIAELVHFL (Firmicutes Sedimentibacter sp.) peptides with a single I to V substitution in position 2 (Fig. 7B).



Discussion

The data reported in the present study show for the first time the high homology in the linear sequence as well as in structure and conformation between TAAs and peptides derived from microbiota species of the Firmicutes and the Bacteroidetes phyla, which together account for 90% of gut microbiota.

The number of homologous peptides with high affinity (< 10 nM) to HLA-A*02:01 molecule indicates that the molecular mimicry between TAA and microbiota-derived epitopes is not anecdotal and suggest a potential relevance in cross-reacting T cell response. In particular, about 70% of the paired epitopes show 6–7 identical residues along the 9-aa peptide sequence, 7% of them shows

8 identical residues and three of them show an identical sequence. Such long stretches of identical amino acids in a nonamer sequence suggest a great biological relevance given that a random event of an identical stretch of 6–9 amino acid in a nonamer sequence has an extremely low probability to occur (1.56×10^{-8} – 1.95×10^{-12}). To further confirm the biological relevance of such a molecular mimicry, the consensus of the microbiota-derived epitopes always perfectly matches with the sequence of the corresponding TAA frequent residues. Moreover, the amino acid substitutions at each position are mostly conservative, confirming that the replacement does not significantly impact on the charge and conformation of the peptide structure.

The average predicted affinity of the microbiota-derived epitopes to the HLA-A molecules is extremely high, especially for the HLA-A*02:01 molecule (<100 nm) and in about 50% of cases this is even higher than the corresponding TAA. According to our previous studies, such affinity values have a 100% correspondence to experimentally validated binding to HLA molecules. Therefore, this supports the concept that such microbiota-derived epitopes are indeed efficiently presented on the surface of cells to elicit a T cell response. Finally, the structural conformation of the microbiota-derived epitopes is, in general, highly similar to the corresponding TAA. In some cases, it is identical and contact areas with both HLA and TCR chains are indistinguishable. The spatial conformation of TCR-facing residues can be identical in paired TAA and microbiota-derived epitopes, with exactly the same values of planar as well as dihedral angles. This confirms that specific amino acid substitutions in the linear sequence do not affect the peptides structure and conformation which can be recognized by cross-reacting T cells. In support to this hypothesis, we have recently shown that CD8⁺ T cells cross-react with TAAs and peptides derived from viruses, sharing the same level of homology observed in the present study [24].

The biological relevance of the findings herein described is that, depending on the stage of the human life when the individuals will encounter the microbiota species, such homology may represent a favorable or unfavorable factor. Indeed, if the microbiota species have colonized the gastrointestinal tract in the first few months of life, the immune system could recognize the derived peptides as self-antigens and the specific T cell clones would be removed. In this case, a tumor lesion presenting a homologous TAA would have a selective advantage for establishing and progressing with a poor prognosis.

On the contrary, if the microbiota species have colonized the gastrointestinal tract later during life, the immune system could recognize the derived peptides

as non-self-antigens and specific memory T cell clones would be established. In this case, as for a preventive vaccine, a tumor lesion presenting a homologous TAA would promptly recall the memory T cells able to eliminate or control tumor growth with an improved prognosis.

The relevance of our findings resides in the fact that the microbiota-derived antigens show homology with TAAs shared by several cancers of different histological origin. In particular, MAGE-A3 and MAGE-A10 have been identified in non-small cell lung cancers (NSCLC), bladder cancers, esophageal and head and neck cancers, and sarcomas [34]. In particular, MAGE-A3 is over-expressed in multiple tumor types including melanoma [35] and lung cancer [36] and its presence has been associated with worse prognosis in colorectal cancer [37], cutaneous squamous cell carcinoma [38], and undifferentiated pleomorphic sarcoma/myxofibrosarcoma [39]. Consequently, structural and conformational homology of peptides derived from microbiota species with these TAAs may have a strong impact on several cancers.

These findings open a new horizon in the mutual interaction between the microbiota and immune response in humans with a potential profound impact on tumor development and progression. Moreover, they provide a completely novel class of antigens to be possibly used as anti-cancer preventive vaccination, even administered as food integrators.

If confirmed, this would completely revolutionize the cancer immunotherapy field and represent a milestone in fighting cancer.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-022-03512-6>.

Additional file 1: Figures S1–4. Percentage (average) of type of residues at each position of the microbiota-derived epitopes when aligned to the corresponding TAA. (A) Identical residues; (B) conservative residues; (C) non-conservative residues; (D) Identical + conservative residues. **Figure S5.** Affinity to HLA-A* 02:01 and 24:02 molecules of each microbiota-derived epitopes with homology to the corresponding TAA. The red dotted line indicates the affinity of the TAA. **Figure S6–17.** Predicted 3D conformation of TAA and microbiota-derived paired peptides. The surface conformation of the paired TAA microbiota-derived peptides is shown. Residues in the microbiota-derived epitopes identical to the TAA sequences are indicated in green color. Green areas = contact points with HLA-A molecule; Magenta areas = contact points with TCR α chain; Light Blue = contact points with TCR β chain. The latter are available only in **Figure S6–S13.** (HLA-A* 02:01 restricted epitopes). **Figures S18.** Percentage of identical values ($\pm 10\%$) of planare and dehidral angles between of the microbiota-derived epitopes and the corresponding TAA. **Figures S19–S24.** Structural predicted backbone conformation of paired TAA and microbiota-derived epitopes. Green areas = contact points with HLA molecule; Magenta areas = contact points with the TCR α chain; light blue areas = contact points with the TCR β chain. The latter are available only in Additional file 1: **Figure S19–22.** (HLA-A* 02:01 restricted epitopes). Values of the planar as well as dihedral angles between atoms of selected TCR-facing residues are shown in parallel.

Author contributions

CM and AM performed bioinformatics predictions of binding affinity; CR and BC performed all the epitope conformation; MLT and FMB contributed to data analysis; MT and LB designed the study, supervised the analysis and drafted the manuscript. All authors read and approved the final manuscript.

Funding

The study was funded by the Italian Ministry of Health through Institutional "Ricerca Corrente" (LB); POR FESR 2014/2020 "Campania OncoTerapie" (LB). AM is funded by "Ricerca Corrente". CM, BC; CR are funded by POR FESR 2014/2020 "NanoCAN".

Availability of data and materials

Data and material have been deposited and are publicly available at <https://doi.org/10.5281/zenodo.6759952>.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

The corresponding author has received consent for publication.

Competing interests

The authors declare no potential conflicts of interest.

Author details

¹Lab of Innovative Immunological Models, Istituto Nazionale per lo Studio e la Cura dei Tumori - IRCCS, "Fondazione Pascale", Via Mariano Semmola, 52, 80131 Naples, Italy. ²Molecular Biology and Viral Oncogenesis Unit, Istituto Nazionale per lo Studio e la Cura dei Tumori - IRCCS "Fond. G. Pascale", Via Mariano Semmola, 52, 80131 Naples, Italy.

Received: 13 June 2022 Accepted: 28 June 2022

Published online: 14 July 2022

References

- Thursby E, Juge N. Introduction to the human gut microbiota. *Biochem J*. 2017;474(11):1823–36. <https://doi.org/10.1042/BCJ20160510>.
- Arumugam M, et al. Enterotypes of the human gut microbiome. *Nature*. 2011;473(7346):174–80. <https://doi.org/10.1038/nature09944>.
- Dominguez-Bello MG, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci USA*. 2010;107(26):11971–5. <https://doi.org/10.1073/pnas.1002601107>.
- Fallani M, Young D, Scott J, Norin E, Amarri S, Adam R, Aguilera M, Khanna S, Gil A, Edwards CA, Doré J. Intestinal microbiota of 6-week-old infants across Europe: geographic influence beyond delivery mode, breast-feeding, and antibiotics. *J Pediatr Gastroenterol Nutr*. 2010;51(1):77–84. <https://doi.org/10.1097/MPG.0b013e3181d1b11e>.
- Fallani M, Amarri S, Uusjarvi A, Adam R, Khanna S, Aguilera M, Gil A, Vieites JM, Norin E, Young D, Scott JA, Doré J, Edwards CA, The Infabio Team. Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology (Reading)*. 2011;157(Pt 5):1385–92. <https://doi.org/10.1099/mic.0.042143-0>.
- Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. Human gut microbiome viewed across age and geography. *Nature*. 2012;486(7402):222–7. <https://doi.org/10.1038/nature11053>.
- Rinninella E, Raouf P, Cintoni M, Franceschi F, Miggianno GAD, Gasbarrini A, Mele MC. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms*. 2019. <https://doi.org/10.3390/microorganisms7010014>.
- Khosravi A, Mazmanian SK. Disruption of the gut microbiome as a risk factor for microbial infections. *Curr Opin Microbiol*. 2013;16(2):221–7. <https://doi.org/10.1016/j.mib.2013.03.009>.
- Brestoff JR, Artis D. Commensal bacteria at the interface of host metabolism and the immune system. *Nat Immunol*. 2013;14(7):676–84. <https://doi.org/10.1038/ni.2640>.
- Fan Y, Pedersen O. Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol*. 2021;19(1):55–71. <https://doi.org/10.1038/s41579-020-0433-9>.
- Chen Y, Zhou J, Wang L. Role and mechanism of gut microbiota in human disease. *Front Cell Infect Microbiol*. 2021;11: 625913. <https://doi.org/10.3389/fcimb.2021.625913>.
- Sepich-Poore GD, Zitvogel L, Straussman R, Hasty J, Wargo JA, Knight R. The microbiome and human cancer. *Science*. 2021. <https://doi.org/10.1126/science.abc4552>.
- McCulloch JA, et al. Intestinal microbiota signatures of clinical response and immune-related adverse events in melanoma patients treated with anti-PD-1. *Nat Med*. 2022;28:545–56.
- Sivan A, et al. Commensal *Bifidobacterium* promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science*. 2015;350:1084–9.
- Vetizou M, et al. Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science*. 2015;350:1079–84.
- Dzutsev A, Goldszmid RS, Viaud S, Zitvogel L, Trinchieri GT. role of the microbiota in inflammation, carcinogenesis, and cancer therapy. *Eur J Immunol*. 2015;45:17–31.
- Gopalakrishnan V, et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. *Science*. 2018;359:97–103.
- Matson V, et al. Te commensal microbiome is associated with anti-PD-1 efficacy in metastatic melanoma patients. *Science*. 2018;359:104–8.
- Routy B, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science*. 2018;359:91–7.
- Vivarelli S, Salemi R, Candido S, Falzone L, Santagati M, Stefani S, Torino F, Banna GL, Tonini G, Libra M. Gut microbiota and cancer: from pathogenesis to therapy. *Cancers (Basel)*. 2019. <https://doi.org/10.3390/cancers11010038>.
- Paulos CM, Wrzesinski C, Kaiser A, Hinrichs CS, Chieppa M, Cassard L, Palmer DC, Boni A, Muranski P, Yu Z, Gattinoni L, Antony PA, Rosenberg SA, Restifo NP. Microbial translocation augments the function of adoptively transferred self/tumor-specific CD8+ T cells via TLR4 signaling. *J Clin Invest*. 2007;117(8):2197–204. <https://doi.org/10.1172/JCI32205>.
- Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, Enk J, Bar-On Y, Stanitsky-Kaynan N, Copenhagen-Glazer S, Shussman N, Almogly G, Cuapio A, Hofer E, Mevorach D, Tabib A, Ortenberg R, Markel G, Miklic K, Jonjic S, Brennan CA, Garrett WS, Bachrach G, Mandelboim O. Binding of the Fap2 protein of *Fusobacterium nucleatum* to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity*. 2015;42(2):344–55. <https://doi.org/10.1016/j.immuni.2015.01.010>.
- Round JL, Mazmanian SK. Inducible Foxp3+ regulatory T-cell development by a commensal bacterium of the intestinal microbiota. *Proc Natl Acad Sci USA*. 2010;107(27):12204–9. <https://doi.org/10.1073/pnas.0909122107>.
- Shen Y, GiardinoTorchia ML, Lawson GW, Karp CL, Ashwell JD, Mazmanian SK. Outer membrane vesicles of a human commensal mediate immune regulation and disease protection. *Cell Host Microbe*. 2012;12(4):509–20. <https://doi.org/10.1016/j.chom.2012.08.004>.
- Ivanov II, Atarashi K, Manel N, Brodie EL, Shima T, Karaoz U, Wei D, Goldfarb KC, Santee CA, Lynch SV, Tanoue T, Imaoka A, Itoh K, Takeda K, Umesaki Y, Honda K, Littman DR. Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell*. 2009;139(3):485–98. <https://doi.org/10.1016/j.cell.2009.09.033>.
- Wooldridge L, Ekeruche-Makinde J, van den Berg HA, Skowera A, Miles JJ, Tan MP, Dolton G, Clement M, Llewellyn-Lacey S, Price DA, Peakman M, Sewell AK. A single autoimmune T cell receptor recognizes more than a million different peptides. *J Biol Chem*. 2012;287(2):1168–77. <https://doi.org/10.1074/jbc.M111.289488>.
- Sewell AK. Why must T cells be cross-reactive? *Nat Rev Immunol*. 2012;12(9):669–77. <https://doi.org/10.1038/nri3279>.
- Zhang L, Udaka K, Mamitsuka H, Zhu S. Toward more accurate pan-specific MHC-peptide binding prediction: a review of current methods

- and tools. *Brief Bioinform.* 2012;13(3):350–64. <https://doi.org/10.1093/bib/bbr060>.
29. Binkowski TA, Marino SR, Joachimiak A. Predicting HLA class I non-permissive amino acid residues substitutions. *PLoS ONE.* 2012;7(8): e41710. <https://doi.org/10.1371/journal.pone.0041710>.
 30. Ragone C, Manolio C, Cavalluzzo B, Mauriello A, Tornesello ML, Buonaguro FM, Castiglione F, Vitagliano L, Iaccarino E, Ruvo M, Tagliamonte M, Buonaguro L. Identification and validation of viral antigens sharing sequence and structural homology with tumor-associated antigens (TAAs). *J Immunother Cancer.* 2021. <https://doi.org/10.1136/jitc-2021-002694>.
 31. Buonaguro L, Cerullo V. Pathogens: our allies against cancer? *Mol Ther.* 2021;29(1):10–2. <https://doi.org/10.1016/j.ymthe.2020.12.005>.
 32. Tornesello ML, Buonaguro L, Buonaguro FM. Mutations of the TP53 gene in adenocarcinoma and squamous cell carcinoma of the cervix: a systematic review. *Gynecol Oncol.* 2013;128(3):442–8. <https://doi.org/10.1016/j.ygyno.2012.11.017>.
 33. Tornesello ML, Duraturo ML, Waddell KM, Biryahwaho B, Downing R, Balinandi S, Lucas SB, Buonaguro L, Buonaguro FM. Evaluating the role of human papillomaviruses in conjunctival neoplasia. *Br J Cancer.* 2006;94(3):446–9. <https://doi.org/10.1038/sj.bjc.6602921>.
 34. Juretic A, Spagnoli GC, Schultz-Thater E, Sarcevic B. Cancer/testis tumour-associated antigens: immunohistochemical detection with monoclonal antibodies. *Lancet Oncol.* 2003;4(2):104–9. [https://doi.org/10.1016/s1470-2045\(03\)00982-3](https://doi.org/10.1016/s1470-2045(03)00982-3).
 35. Gaugler B, Van den Eynde B, van der Bruggen P, Romero P, Gaforio JJ, De Plaen E, Lethé B, Brasseur F, Boon T. Human gene MAGE-3 codes for an antigen recognized on a melanoma by autologous cytolytic T lymphocytes. *J Exp Med.* 1994;179(3):921–30. <https://doi.org/10.1084/jem.179.3.921>.
 36. Sielen W, Varwerk C, Linder A, Kaiser D, Teschner M, Delire M, Stamatis G, Passlick B. Melanoma associated antigen (MAGE)-A3 expression in stages I and II non-small cell lung cancer: results of a multi-center study. *Eur J Cardiothorac Surg.* 2004;25(1):131–4. <https://doi.org/10.1016/j.ejcts.2003.09.015>.
 37. Tarnowski M, Czerewaty M, Deskur A, Safranow K, Marlicz W, Urańska E, Ratajczak MZ, Starzyńska T. Expression of cancer testis antigens in colorectal cancer: new prognostic and therapeutic implications. *Dis Markers.* 2016;2016:1987505. <https://doi.org/10.1155/2016/1987505>.
 38. Abikhair M, Roudiani N, Mitsui H, Krueger JG, Pavlick A, Lee J, Therrien JP, Meehan SA, Felsen D, Carucci JA. MAGEA3 expression in cutaneous squamous cell carcinoma is associated with advanced tumor stage and poor prognosis. *J Invest Dermatol.* 2017;137(3):775–8. <https://doi.org/10.1016/j.jid.2016.10.036>.
 39. Conley AP, Wang WL, Livingston JA, Ravi V, Tsai JW, Ali A, Ingram DR, Lowery CD, Roland CL, Somaiah N, Hwu P, Yee C, Subbiah V, Futreal A, Lazar AJ, Patel S, Roszik J. MAGE-A3 is a clinically relevant target in undifferentiated pleomorphic sarcoma/myxofibrosarcoma. *Cancers (Basel).* 2019. <https://doi.org/10.3390/cancers11050677>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

