

The importance of *Helicobacter pylori*'s genetic variability for the construction of an efficient vaccine

Filipa Vaz, Andreia Couto, Cecilia R.C. Calado*

Engineering Faculty, Catholic University of Portugal, Rio de Mouro, Portugal (*, c.calado@fe.lisboa.ucp.pt)

Abstract—*Helicobacter pylori* is responsible for several gastric diseases. The main constraints of vaccine trials against this pathogen are mainly due to the bacterium high antigenic variability and to down-regulation of the host immune responses. To counteract these factors we propose a DNA vaccine able to induce a balanced humoral and cytotoxic specific immune responses, based on multi-antigens. The selection of the antigens NapA, HpaA, VacA and HomB were conducted based on immunoproteomic data and the protein role on infection and pathogenesis. A fragment of each target-antigen was selected by *in silico* methods based on the maximization of the gene conservation and antigenicity. The set of these small fragments will be presented as a vaccine based on several conserved epitopes of multi-antigenic targets, and consequently representative of the bacterium antigenic variability.

Index Terms— *Helicobacter pylori*; DNA vaccine; multi-antigenic; genetic variability; conserved epitopes.

I. INTRODUCTION

Helicobacter pylori is a major gram-negative pathogen that colonizes the stomach of half of the world's population. The exact mechanisms of transmission are unknown, but it is thought that the main routes are oral-oral and fecal-oral and that acquisition occurs mostly during childhood by family members and close persons [1]. The prevalence of infection is higher on undeveloped countries (80%) than on developed countries (40%), probably because of improved public health measurements and hygiene conditions. This pathogen highly contributes to acute and chronic gastritis, peptic ulcer, intestinal metaplasia and neoplasia and also MALT lymphoma, being considered a class I carcinogen by the WHO. Treatment against *H. pylori* infection consists in a combination of antimicrobials, that however fails in the majority of the cases due to antibiotic resistance increase [2]. Therefore, a prophylactic and therapeutic vaccine stands up against all these problems, consisting of the pharmacological method with the highest benefit/cost ratio. Furthermore, because an asymptomatic infection is quite common, resulting in symptoms at later stages of gastric disease, a vaccine with a prophylactic character would strongly contribute to prevent severe gastric diseases as peptic ulcers and gastric carcinoma.

This bacterium presents a high genetic variability due to high mutation and recombination rates, DNA transfer and a poor DNA repair system. It also presents rearrangements at the

chromosome level, high diversity in terms of number and content of genes and high allelic diversity. The fact that an individual can be colonized by a multitude of different *H. pylori* strains, that each strain presents great genetic variability that gives high antigenic diversity and can acquire new antigens or express/repress others according to environmental conditions, present main constraints on the development of an efficient vaccine. Hence, a vaccine against this pathogen should be composed of several antigens – a multi-antigenic vaccine. The advantage of DNA vaccines over protein vaccines, is the possibility of inducing a balanced immune response by activating humoral and cellular processes [1].

II. GOALS

The main goal was to design a multi-antigenic DNA vaccine, based on a plasmid vector, against *Helicobacter pylori*. It was aimed to define a set of antigenic targets based on data of the microorganism antigenic diversity. To each target-antigen (protein) it was intended to define a smaller region conserved among species and rich in B, Th and Tc epitopes as estimated by *in silico* methods.

III. MATERIAL AND METHODS

Analysis of the variability of each target gene

The variability analysis to each antigenic target was conducted with genes from UniProtKB database. The alignment of the nucleotide sequences was done by MAFFT v6 and visualized using Jalview [3]. Genetic variability was evaluated using DnaSP v5 [4].

Selection of the most conserved and highly antigenic fragment from each whole-antigen

Conservancy and antigenicity analysis

Protein sequences from the target genes were obtained by their nucleotide sequences, translating with ORF FINDER (NCBI). After sequence alignment the consensus sequence was obtained, and the antigenicity evaluated.

B epitopes were estimated by determination of the accessibility of the fragments at the surface of the protein by Emini Surface Accessibility Prediction (Emini algorithm [5])

and Bepipred method by Bepipred Linear Epitope Prediction [6].

To estimate T epitopes, methods that evaluate peptide processing, its transport and affinity to MHC-I and -II molecules were used.

Th epitopes were estimated by the *consensus* method [7], in which it was used the MHC-II allele more frequent in Europe according to the data base dbMHC – HLA DRB1*0401. The Tc epitopes were estimated by NetCTL method [8] [9], that takes into account the affinity of the peptides to MHC-I molecules, cleavage by the proteasome and efficiency of transport by TAP (*transporter associated with antigen processing*).

The genetic conservation and antigenic analysis were done with IEDB Analysis Resource tools available at the *Immune Epitope Database* (IEDB). The conservation of the fragments for each protein from the set of sequences used was evaluated using Epitope Conservancy Analysis by Epitope linear sequence conservancy method. The regions presented a simultaneous high genetic conservancy and immunogenicity with the previous pointed methods, were also evaluated for the affinity for MHC-I molecules.

Specificity of the epitopes and genetic analysis

The specificity of the antigens towards *H. pylori* was conducted by homology analysis: i) against proteins from other microorganisms, using the Microbial DataBase with the genomic BLAST tool from BLASTP NCBI; ii) against Human Sequences of NCBI with the BLASTP by the RefSeq protein data base.

IV. RESULTS AND DISCUSSION

The target antigens selection was performed based on published immunoproteomic data, of 2D-gel electrophoresis of total protein extracts of different *H. pylori* strains, detected by antibodies from serum of human patients, and considering the role of each common antigen find in that data and its level of expression. The target antigens identified for the design of a multi-antigenic vaccine were: NapA, HpaA, VacA and HomB (Table II). Highly antigenic and/or highly expressed proteins were not considered as target antigens due to the possibility of induction of antigenic tolerance. Low expression proteins were also not considered as targets antigens as the immune response usually depends directly on the amount of antigen expressed [10] [11]. The following proteins were identified by immunoproteomics as antigenic and very common to different strains: NapA [10] [12] [13], VacA [10] [14] [15] and HpaA [10] [15] [16]. Indeed, immunization assays were already previously conducted for NapA [17] [18], VacA [17] [18] [19] and HpaA [20] [21]. All of the selected proteins are located outside the cell – VacA is secreted and the others are also present at the cell surface (Table II).

For all the target antigens selected, it was observed a high polymorphism, as for all genes, the number of alleles is almost the same as the number of sequences, and no predominant allele exists in the population (Table III). Accordingly, Go *et al.* (1996) determined 73 allelic profiles by MLST in 74

strains [28]. Suerbaum *et al.* (1998) reported that *H. pylori* shows a clonal structure only for a short period of time after transmission, identifying a high number of alleles for genes *vacA*, *flaA* e *flab* [29], like Kansau *et al.* (1996) for gene *ureC/glmM* [9] and Pan *et al.* (1997) for gene *cagA* [30].

TABLE II. ROLE OF THE ANTIGENIC TARGETS

NapA neutrophil-activating protein
Discriminator between ulcer and cancer [22], it allows the bacteria to adhere to mucin, it is a chemical attractant of monocytes, induces the production of oxygen reactive species from neutrophils and stimulates the production of IL-12 and -13, contributing to polarization of the immune response to Th1. It is speculated that has a role on DNA protection and iron reservation. <i>In vivo</i> occurs the production of antibodies anti-NapA [11].
HpaA <i>H. pylori</i> adhesin A
Colonization factor that induces maturation and presentation of antigens by dendritic cells and is essential for colonization and establishment of infection [23] [24].
VacA vacuolating cytotoxin
Implicated in several virulence mechanisms such as: fagocytosis inhibition by macrophages, formation of pores on the membrane of epithelial cells, induction of formation of vacuoli, apoptosis and modulation of immune responses. Although all strains bear the gene <i>vacA</i> , the expression and activity of the protein is dependent on the genetic heterogeneity of the signal region (s region) that codifies for the signal peptide and the middle region (m region) that codifies for the binding domain to p58. The genotype of the s region can be classified in s1 or s2 and the m region as m1 or m2. The genotype s1/m1 is correlated to high activity of the protein, s1/m2 to an intermediate activity and in genotype s2/m2 the activity is absent. Therefore, the genotype s1/m1 is associated with peptic ulcer and gastric carcinoma, being also related to a <i>cagA</i> genotype. Protein expression is variable throughout the course of infection, being thought that the improvement and aggravation of ulcers may be related to the changes of protein expression [1].
HomB
An outer membrane protein that is a co-marker of strains associated to peptic ulcer, induces inflammatory responses and plays a role on bacteria adherence [25] [26] [27].

From the sequences analyzed, genes *napA* and *hpaA* showed a recombination rate between adjacent sites (r) higher than the mutation rate (θ), indicating that in these genes the variability is mainly due to recombination phenomena (Table III). On the contrary, the genes *vacA* and *homB* showed a higher mutation rate than the recombination rate, being the genetic variability mainly dependent on spontaneous mutations.

The Tajima statistical test D (1989), that allows to evaluate if a neutral evolution is occurring on the population (i.e. if it is random), was not statistically significant for any of the genes, nor was the Fu&Li (1993) D Test, apart for gene *homB* ($P < 0,05$) (data not shown). In spite of no statistical meaning, the Tajima D test for *vacA* was positive indicating a probable balancing selection. Genes *napA*, *hpaA* e *homB* presented negative D test, indicating a tendency for a probable stabilizing selection or directional selection that may imply a subdivision of the population. Accordingly, all genes presented a dN/dS ratio much higher than 1, indicative of a genetic diversification under a positive selection.

TABLE III. DIVERSITY OF THE TARGET GENES. Genetic analysis of the target genes in relation to: polymorphic site (S); number of alleles; haplotype diversity (Hd) determined by the Nei formula (1987) that implies the substitution of $2n$ for n ; nucleotide diversity (π) that represents the average of the number of different nucleotides per site between two sequences (Nei, 1987); Watterson Test, where $\theta = 4N\mu$, N represents the population dimension and μ the mutation rate in each nucleotide by generation (Watterson, 1975); Tajima Test (D) to evaluate the null hypothesis that mutations are neutral (* indicates that the results are not significant: $P > 0,10$) (Tajima, 1989); recombination rate (R) (Hudson, 1987); recombination rate between adjacent sites ($r = R/L$, L is the average of the distance in nucleotides between the most distant sites under recombination); ratio between the number of non-synonymous substitutions (dN) and the number of synonymous substitutions (dS).

	<i>napA</i> n=29	<i>hpaA</i> n=25	<i>vacA</i> n=56	<i>homb</i> n=120
No. of polymorphic sites, S	38	36	778	836
No. of alleles	25	22	55	118
Haplotype diversity, Hd	0,9877	0,987	0,999	0,9997
Nucleotide Diversity, π	0,0329	0,04225	0,09787	0,07834
Watterson Test, θ	0,03612	0,04414	0,09225	0,10464
Tajima Test, D	-0,33150*	-0,16203*	0,2195*	-0,84256*
Recombination, R	44	75,9	13,5	65,9
R between adjacent sites, r	0,1066	0,1188	0,0035	0,0331
Ratio dN/dS	3,77	4,03	3,36	3,42

TABLE IV. CHARACTERISTICS OF THE SELECTED FRAGMENTS. The location and sequence of the fragments are not shown due to patent issues. It is only presented the fragments size in amino acid. The column “MHC-II” presents the values relatives to the method “Consensus Percentile Rank” that estimates the affinity of binding of the epitopes to the MHC-I and -II, i.e., e.g. for all the epitopes that the fragment of HpaA can generate, it is estimated that the most antigenic has a value of 1,64 and the less antigenic is 53,9. The column MHC-I represents the supertypes to which the fragments are recognized.

Fragment Size	Conservation			MHC-II		MHC-I
	100%	90%	$\geq 80\%$	lower value	higher value	
NapA						A2
22 amino acids	79,30% (23/29)			12,74	54,84	A3 A24 B8 B39
HpaA						A3
19 amino acids	72% (18/25)			1,64	53,94	A24 B44 B62
VacA						A1
30 amino acids	3,57% (2/56)	80,36% (45/56)	98,21% (55/56)	6,66	50,21	A26 B8 B39 B62
30 amino acids	10,71% (6/56)	98,21% (55/56)	98,21% (55/56)	1,94	51,77	A1 A3 A24
26 amino acids	12,50% (7/56)	85,71% (48/56)	85,71% (48/56)	2,16	70,73	A26 B27 B58 B62
Homb						A24
21 amino acids	75% (90/120)	100%		5,42	34,99	A26 B44 B62
28 amino acids	12,50% (15/120)	72,50% (87/120)	99,17% (119/120)	0,62	11,01	A26 B39 B62

Examples of dramatic changes in the host human population, that may have consequences on redirecting the selection of specific *H. pylori* antigenic variants, are: improvement of sanitary conditions and hygiene; improvement of public health infrastructures; alteration of habits as types of food; introduction of antimicrobials compounds.

It was observed that the target genes of the vaccine have a high rate of non-synonymous substitutions on the DNA sequence, corresponding to variable amino acid sequences. The codons that present non-synonymous substitutions are also rich in synonymous. As expected these highly variable genetic regions were avoided in the vaccine design. Based on the genetic conservancy analysis and the antigenic variability along the consensus sequence, defined fragments of each antigen were selected as described in Table IV. The set of these fragments defines the DNA vaccine. All of the fragments selected are specific for the target proteins, except HomB that showed homology with others OMP from *H. pylori*, namely HomA. NapA and HpaA presented homology with Nap e Hpa from *H. acinonychis* and *H. hepaticus*.

V. CONCLUSIONS

In order to construct a multi-antigenic vaccine against *Helicobacter pylori*, the set of antigens NapA, HpaA, VacA and HomB were defined based on immunoproteomic data and the role of the common antigens on the pathogenesis mechanisms. For all the genes evaluated it was observed a high genetic and antigenic variability along the consensus sequence. Based on the genetic conservation analysis and on the antigenic analysis, small fragments of each target antigen were selected that will be included in a plasmid expression system, to produce a chimeric protein representing the bacterium specific and conserved high antigenic variability.

ACKNOWLEDGMENT

This work was financially supported by the FCT research project PTDC/Bio/69242/2006.

REFERENCES

- [1] Kusters J.G., J.G. van Vliet J.G., Kuipers E.J. (2006). Pathogenesis of *Helicobacter pylori* Infection. *Clinical Microbiology Reviews* 19 (3): 449-490.
- [2] Malfertheiner P., Megraud F., O'Morain C., Bazzoli F., El-Omar E., Graham D., Hunt R., Rokkas T., Vakil N., Kuipers E. (2007). Current concepts in the management of *Helicobacter pylori* infection – The Maastricht III Consensus Report. *Gut* doi:10.1136/gut.2006.101634.
- [3] Clamp M., Cuff, J., Searle S.M., Barton, G.J. (2004). The Jalview Java Alignment Editor. *Bioinformatics* 20:426-7.
- [4] Librado P., & Rozas J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451-1452.
- [5] Emini E.A., Hughes J.V., Perlow D.S., Boger J. (1985). Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. *J Virol* 55:836-839.

- [6] Larsen M.V., Lundegaard C., Kasper Lamberth, Buus S., Brunak S., Lund O., Nielsen M. (2005). An integrative approach to CTL epitope prediction: A combined algorithm integrating MHC-I binding, TAP transport efficiency, and proteasomal cleavage predictions. *European Journal of Immunology* 35(8):2295-303.
- [7] Wang P., Sidney J., Dow C., Mothé B., Sette A., et al. (2008) A Systematic Assessment of MHC Class II Peptide Binding Predictions and Evaluation of a Consensus Approach. *PLoS Comput Biol* 4(4): e1000048. doi:10.1371/journal.pcbi.1000048
- [8] Larsen J.E., Lund O., Nielsen M. (2006). Improved method for predicting linear B-cell epitopes. *Immunome Res* 2:2.
- [9] Larsen M.V., Lundegaard C., Lamberth K., Buus S., Lund O., and Nielsen M. (2007). Large-Scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* 8:424.
- [10] Jungblut P. R., Bumann D., Haas G., Zimny-Arndt U., Holland P., Lamer S., Siejak F., Aebischer A., Meyer T.F. (2000). Comparative proteome analysis of *Helicobacter pylori* *Molecular Microbiology* 36(3), 710-725.
- [11] Del Giudice G., Covacci A., Telford J.L., Montecucco C., Rappuoli R. (2001). The Design of Vaccines against *Helicobacter pylori* and their development. *Annu. Rev. Immunol.* 19:523–63.
- [12] McAtee C.P., Lim M.Y., Fung K., Velligan M., Fry K., Chow T., Berg D.E. (1998). Identification of Potential Diagnostic and Vaccine Candidates of *Helicobacter pylori* by Two-Dimensional Gel Electrophoresis, Sequence Analysis, and Serum Profiling. *Clinical and Diagnostic Laboratory Immunology* 5(4): 537–542.
- [13] Kimmel B., Bosserhoff A., Frank R., Gross R., Goebel W., Beier D. (2000). Identification of Immunodominant Antigens from *Helicobacter pylori* and Evaluation of Their Reactivities with Sera from Patients with Different Gastrointestinal Pathologies. *Infection and Immunity* 68(2):915–920.
- [14] Sabarth N., Lamer S., Zimny-Arndt U., Jungblut P.R., Meyer T.F., Bumann D. (2002). Identification of Surface Proteins of *Helicobacter pylori* by Selective Biotinylation, Affinity Purification, and Two-dimensional Gel Electrophoresis. *The Journal of Biological Chemistry* 277(31):27896–27902.
- [15] Lock R.A., Cordwell S.J., Coombs G.W., Walsh B.J., Forbes G.M. (2001). Proteome Analysis of *Helicobacter pylori*: major proteins of type strain NCTC 11637. *Pathology* 33:365–374.
- [16] Haas G. et al. (2002). Immunoproteomics of *Helicobacter pylori* infection and relation to gastric disease *Proteomics* 2, 313–324.
- [17] Rossi G., Ruggiero P., Peppoloni S., Pancotto L., Fortuna D., Lauretti L., Volpini G., Mancianti S., Corazza M., Taccini E., Di Pisa F., Rappuoli R., Del Giudice G. (2004). Therapeutic Vaccination against *Helicobacter pylori* in the Beagle Dog Experimental Model: Safety, Immunogenicity, and Efficacy. *Infection and Immunity* 72(6):3252-3259.
- [18] Malfertheiner P., Schultze V., Rosenkranz B., Kufmann S. H.E., Ulrichs T., Novicki D., Norelli F., Contorni M., Peppoloni S., Berti D., Tornese D., Ganju J., Palla E., Rappuoli R., Scharschmidt B.F., Del Giudice G. (2008). Safety and Immunogenicity of an Intramuscular *Helicobacter pylori* Vaccine in Noninfected Volunteers: A Phase I Study. *Gastroenterology* 135(3):787-795.
- [19] Ghiara P., Rossi M., Marchetti M., Tommaso A., Vindigni C., Ciampolini F., Covacci A., Telford J.L., Magistris M.T., Pizzi M., Rappuoli R., Del Giudice G. (1997). Therapeutic Intra-gastric Vaccination against *Helicobacter pylori* in Mice Eradicates an Otherwise Chronic Infection and Confers Protection against Reinfection. *Infection and Immunity* 65(12):4996-5002.
- [20] Zhou W.-Y., et al. Therapeutic efficacy of a multi-epitope vaccine against *Helicobacter pylori* infection in BALB/c mice model. *Vaccine* (2009), doi:10.1016/j.vaccine.2009.05.009.
- [21] Nyström J. & Svennerholm A.M. (2007). Oral immunization with HpaA affords therapeutic protective immunity against *H. pylori* that is reflected by specific mucosal immune responses. *Vaccine*. 25(14):2591-8.
- [22] Khoder G., Yamaoka Y., Fauchère J.-L., Burucoa C., Atanassov C. (2009). Proteomic *Helicobacter pylori* biomarkers discriminating

- between duodenal ulcer and gastric cancer. *Journal of Chromatography B* 877:1193–1199.
- [23] Voland P., Hafsi N., Zeitner M., Laforsch S., Wagner H., Prinz C. (2003). Antigenic Properties of HpaA and Omp18, Two Outer Membrane Proteins of *Helicobacter pylori*. *Infection and Immunity* 71(7):3837–3843.
- [24] Oleastro M., Monteiro L., Lehours P., Mégraud F., Ménard A. (2006). Identification of Markers for *Helicobacter pylori* Strains Isolated from Children with Peptic Ulcer Disease by Suppressive Subtractive Hybridization. *Infection and Immunity* 74(7): 4064–4074.
- [25] Oleastro M., Monteiro L., Lehours P., Mégraud F., Ménard A. (2006). Identification of Markers for *Helicobacter pylori* Strains Isolated from Children with Peptic Ulcer Disease by Suppressive Subtractive Hybridization. *Infection and Immunity* 74(7): 4064–4074.
- [26] Oleastro M., Cordeiro R., Ferrand J., Nunes B., Lehours P., Carvalho-Oliveira I., Mendes A.I., Penque D., Monteiro L., Mégraud F., Ménard A. (2008). Evaluation of the Clinical Significance of homB, a Novel Candidate Marker of *Helicobacter pylori* Strains Associated with Peptic Ulcer Disease. *JID* 198.
- [27] Oleastro M., Cordeiro R., Yamaoka Y., Queiroz D., Mégraud F., Monteiro L., Ménard A. (2009). Disease association with two *Helicobacter pylori* duplicate outer membrane protein genes, homB and homA. *Gut Pathogens* 2009, 1:12.
- [28] Go M.F., Kapur V., Graham D.Y., Musser J.M. (1996). Population Genetic Analysis of *Helicobacter pylori* by Multilocus Enzyme Electrophoresis: Extensive Allelic Diversity and Recombinational Population Structure. *J. Bacteriol.* 178(13), 3934–3938.
- [29] Suerbaum S., Smith J.M., Bapumia K., Morelli G., Smith N.H., Kunstmann E., Dyrek I., Achtman M. (1998). Free recombination within *Helicobacter pylori* (nucleotide sequencing/horizontal genetic exchange/evolution/ linkage equilibrium). *Proc. Natl. Acad. Sci. USA* 95:12619–12624.
- [30] Kansau I., Raymond J., Bingen E., Courcoux P., Kalach N., Bergeret M., Braimi N., Dupont C., Labigne A. (1996). Genotyping of *Helicobacter pylori* isolates by sequencing of PCR products and comparison with the RAPD technique. *Res. Microbiol.* 147, 661-669.
- [31] Pan Z.-J., Van Der Hulst R.W.M., Feller M., Xiao S.-D., Tytgat G.N.J., Dankert J., Van Der Ende A. (1997). Equally High Prevalences of Infection with cagA-Positive *Helicobacter pylori* in Chinese Patients with Peptic Ulcer Disease and Those with Chronic Gastritis-Associated Dyspepsia. *Journal of Clinical Microbiology* 35(6):1344-1347.