

*babA* alleles reflects long-term adaptation to the types of receptors available in local populations, punctuated by fine-tuning of adherence during persistent infection of any individual host. Although adherence should benefit *H. pylori* by allowing better access to nutrients and delivery of effector molecules, tight adherence may be deleterious when host responses are robust. Recently, *H. pylori* infection and gastritis were found to promote gastric mucosal expression of inflammation-associated sialyl-Lewis x/a antigens (3), in competition with the fucosylated antigens that were studied here (10). Changes in BabA adhesins that help strains adapt to host gastric environments could arise by point mutation or short patch recombination between strains, or between divergent alleles in the same strain as illustrated by the ALeb binding transformant P445G [(11–13) and this study]. Such new *babA* alleles will often differ from their parents in affinity, and those that are best suited to the local gastric environment (whether they are of higher or lower affinity) will be selected. Such flexibility should help to ensure rapid adaptation of *H. pylori* populations to the glyco-phenotype and host response of each infected person.

The postulated cycles of selection for decreased and then increased adherence during chronic infection and transmission from one person to another should result in retention of the ALeb, BLeb, and Leb generalist binding modes in most human populations because of the abundance of A, B, and O blood groups in them. In explaining the abundance of Leb-only specialist strains in South America, we note that Amerindians of this region are unique in being almost entirely of blood group O phenotype (14). One might invoke the idea of selection for dedicated specialists in this population. However, because the distribution of Leb affinities of Peruvian specialists is similar to those of generalists everywhere, we prefer an alternative explanation, which invokes recurrent cycles of selection for loss and restoration of binding activity: Only restoration of Leb binding activity would be selected for in any uniformly blood group O population; thus generalist binding would be lost by attrition.

Most alleles of housekeeping genes in the Peruvian *H. pylori* strains studied here were closely related to those found in Spanish strains but not those of Asian strains (15, 16). This implies descent of these Peruvian strains mainly from European strains (16). If this is correct, most Peruvian specialist *babA* alleles may have arisen by mutation and/or recombination over only the last ~500 years. We propose that such rapid evolvability of the BabA adhesin in response to host mucosal glycosylation patterns fine-tunes *H. pylori* strains to their individual hosts, helps them to avoid the most deleterious of host responses, and contributes importantly to the extraordinary chronicity of human *H. pylori* infection worldwide.

# References and Notes

1. R. M. Peek Jr., M. J. Blaser, *Nat. Rev. Cancer* **2**, 28 (2002).
2. T. Borén, P. Falk, K. A. Roth, G. Larson, S. Normark, *Science* **262**, 1892 (1993).
3. J. Mahdavi et al., *Science* **297**, 573 (2002).
4. H. Clausen, S. i. Hakomori, *Vox Sang.* **56**, 1 (1989).
5. C. A. Clarke et al., *BMJ* **2**, 643 (1955).
6. D. Ilver et al., *Science* **279**, 373 (1998).
7. M. Gerhard et al., *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12778 (1999).
8. Materials and methods and all radioimmunoassay (RIA) and Scatchard data are available as supporting materials on Science Online.
9. M. Aspholm-Hurtig et al., unpublished data.
10. J. R. Brown, M. M. Fuster, T. Whisenant, J. D. Esko, *J. Biol. Chem.* **278**, 23352 (2003).
11. D. T. Pride, M. J. Blaser, *J. Mol. Biol.* **316**, 629 (2002).
12. D. Falush et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15056 (2001).
13. J. V. Solnick et al., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2106 (2004).
14. A. E. Mourant, A. C. Kōpéc, K. Domaniewska-Sobczak, in *The Distribution of Human Blood Groups and other Polymorphisms* (Oxford Univ. Press, London, 1976).
15. G. Dalilide et al., in preparation.
16. D. Kersulyte et al., *J. Bacteriol.* **182**, 3210 (2000).
17. We thank P. Falk for valuable discussions, F. Lindh, IsoSepAB (Tullinge, Sweden), and T. Norberg for synthesis of the A7 conjugate, A. Parkinson for Alaska Natives strains, P.-G. Nyholm for structural predictions of blood group antigens, I. Sjöström for techni-

cal assistance, and A. Kidd for proofreading. This work was supported by the Mizutani Foundation Glycoscience Award 2003, the Commission of the European Union (QLK2-2001-01216-ADRI), Umeå University Biotechnology Fund, the County Council of Västerbotten and Stiftelsen JC Kempeoch Seth M Kempes Minne (T.B.), the Swedish Medical Research Council [11218 (T.B.), 12628 (S.T.), and 7902 (I.C.)], the Swedish Cancer Society [4101-B00-03XAB (T.B.) and 4128-B99-02XAB (S.T.)], Swedish Foundation for Strategic Research (SSF) programs [Glycoconjugates in Biological Systems (N.R., T.B., and S.T.) and Infection and Vaccinology (M.A.-H. and T.B.)], the Swedish Medical Society (S.T.), the Swedish Society for Medical Research (S.V. and M.A.-H.), the J. C. Kempe Memorial Foundation, the Royal Swedish Academy of Sciences (M.A.-H.), and NIH [grants RO1 AI38166, RO3 AI49161, RO1 DK53727, RO1 DK63041 (D.B.), P30 DK52574 (Washington University), and RO1GM62370 (M.B.)]. The 66 *babA* sequences in Fig. 2 have been deposited in GenBank under accession AY648168–AY648233.

## Supporting Online Material

www.sciencemag.org/cgi/content/full/305/5683/519/DC1

Materials and Methods

Figs. S1 to S12

Tables S1 to S8

6 April 2004; accepted 23 June 2004

## A Synthetic Conjugate Polysaccharide Vaccine Against *Haemophilus influenzae* Type b

V. Verez-Bencomo,<sup>1\*</sup> V. Fernández-Santana,<sup>1\*</sup> Eugenio Hardy,<sup>2\*</sup> Maria E. Toledo,<sup>3\*</sup> Maria C. Rodríguez,<sup>1</sup> Lazaro Heynngnezz,<sup>2</sup> Arlene Rodríguez,<sup>2</sup> Alberto Baly,<sup>3</sup> Luis Herrera,<sup>2</sup> Mabel Izquierdo,<sup>2</sup> Annette Villar,<sup>1</sup> Yury Valdés,<sup>1</sup> Karelía Cosme,<sup>2</sup> Mercedes L. Deler,<sup>1</sup> Manuel Montane,<sup>2</sup> Ernesto Garcia,<sup>1</sup> Alexis Ramos,<sup>1</sup> Aristides Aguilar,<sup>2</sup> Ernesto Medina,<sup>2</sup> Gilda Toraño,<sup>3</sup> Iván Sosa,<sup>2</sup> Ibis Hernandez,<sup>3</sup> Raydel Martínez,<sup>3</sup> Alexis Muzachio,<sup>2</sup> Ania Carmenates,<sup>4</sup> Lourdes Costa,<sup>2</sup> Félix Cardoso,<sup>1</sup> Concepción Campa,<sup>5</sup> Manuel Diaz,<sup>3</sup> René Roy<sup>6\*</sup>

Glycoconjugate vaccines provide effective prophylaxis against bacterial infections. To date, however, no commercial vaccine has been available in which the key carbohydrate antigens are produced synthetically. We describe the large-scale synthesis, pharmaceutical development, and clinical evaluation of a conjugate vaccine composed of a synthetic capsular polysaccharide antigen of *Haemophilus influenzae* type b (Hib). The vaccine was evaluated in clinical trials in Cuba and showed long-term protective antibody titers that compared favorably to licensed products prepared with the Hib polysaccharide extracted from bacteria. This demonstrates that access to synthetic complex carbohydrate-based vaccines is feasible and provides a basis for further development of similar approaches for other human pathogens.

*Haemophilus influenzae* type b (Hib) is an important human pathogen and was prevalent in developed countries until the introduction of successful conjugate vaccines during the 1990s (1). However, in developing countries more than 600,000 infant deaths occur annually as a result of Hib-induced pneumonia or meningitis (2). Extensive use of the polysaccharides as vaccines has offered a useful way to protect

adults and older children (3–5), and further improvement in generating long-lasting immunity, especially in infants, has been achieved by covalently coupling the polysaccharide to carrier proteins (6). In fact, the high level of success attained by Hib glycoconjugate vaccine (7) has been quickly followed by similar approaches to meningococcal group C (8) and *Streptococcus pneumoniae* (9). Many candidate vac-

cines against other pathogens using the same principles are currently at different stages of research (10, 11).

The fragment of the Hib capsular polysaccharide used in some of the licensed vaccines can be as short as five ribosylribitol-phosphate repeating units (12). The ability of synthetic carbohydrate chemistry to mimic such fragments has been demonstrated in several laboratories with the use of stepwise multistep preparation (13–15); the resulting synthetic antigens have served as components of candidate vaccines that have proven efficient in generating immunity in animals (16, 17). We set out to develop a synthetic methodology amenable to large-scale good manufacturing practice (GMP) production of antigens by reassembling Hib polysaccharide fragments. The previous process was redesigned to include a synthetic pathway with a reduced number of reaction and chromatography purification steps. We also identified a potentially superior method for oligomerization of the ribosylribitol-phosphate repeating unit, in which the saccharide fragment encompassing the key conformational epitope can be obtained in a single step.

To this end, we selected suitably protected ribitol derivative **1** and ribose acetate **2** because they best fit the criteria for large scale production as synthetic intermediates and could be more readily purified than others derived from D-glucose (Fig. 1). The use of crystalline peracetylated  $\beta$ -D-ribofuranose, **2**, was found to readily glycosidate ribitol derivative **1** to provide **3a** (18). The successful large-scale synthesis of ribosylribitol derivatives **4** and **5** from **3a** as described in Fig. 1 thus represented one of the key features of our strategy.

Although construction of oligomers with controlled numbers of repeating units by solution and solid-phase techniques was possible in small quantities, their large-scale syntheses proved more difficult. To overcome this, we undertook a one-step polycondensation reaction with the use of H-phosphonate chemistry (19). Thus, the phosphate-containing end residue **4** and H-phosphonate derivative **5** were oligomerized in high yield and purity with the use of pivaloyl chloride as a polycondensation reagent

(Fig. 1). Although this reaction is complex, it could be managed to generate the desired oligomers, avoiding several competing side reactions, such as O-acylation.

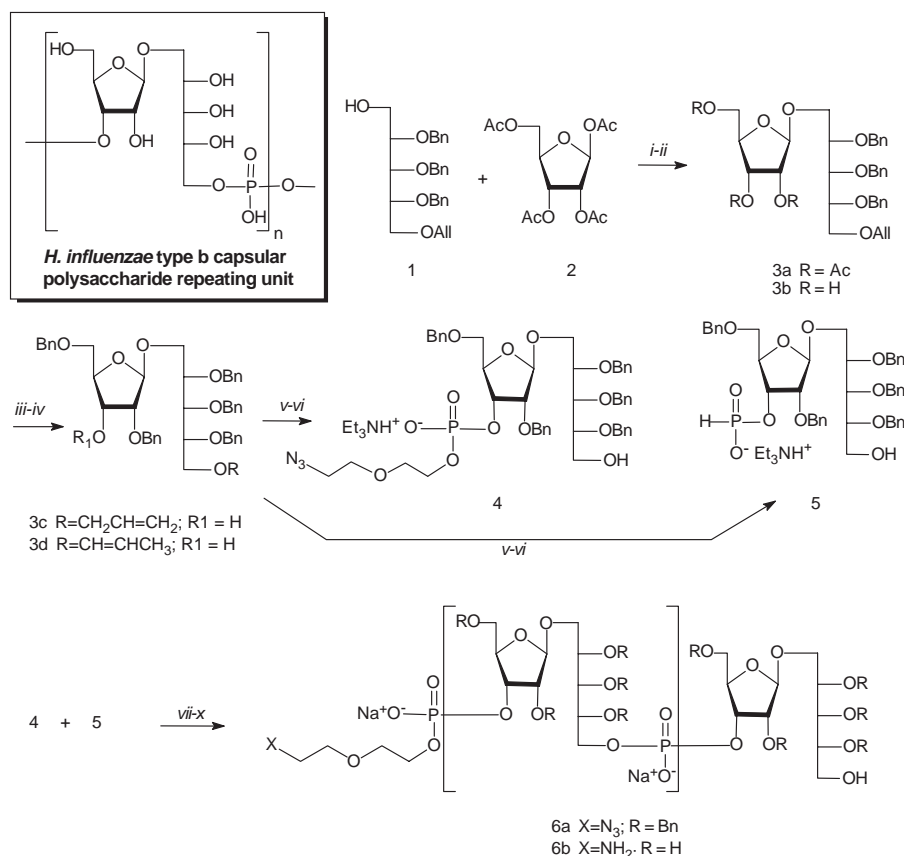
Synthetic oligomers of **6a** with an average of eight repeating units were reproducibly obtained in high yield (80%) after purification by size exclusion chromatography. Deprotection and azide reduction of **6a** to the amine **6b**, which, after treatment with 3-maleimidopropionic acid N-hydroxysuccinimide ester, gave **7** (Fig. 2). The overall process accomplished under GMP proceeded with a high yield and was amenable to a 100-g scale per batch.

A vaccine prototype was subsequently produced by conjugating synthetic antigen **7** with thiolated human serum albumin (HSA) (20). The polyribosylribitol phosphate (sPRP)–HSA conjugate was used for coating enzyme-linked immunosorbent assay (ELISA) plates to screen for anti-Hib activity (21) of sera obtained from rabbits immunized with commercially available vaccines [Vaxem-Hib (Chiron, Emeryville, CA) and Hiberix (Glaxo Smith Kline Biologicals, Rixensart, Belgium)] and human antibodies against Hib obtained from immunized children (Vaxem-Hib). All sera showed equivalent

recognition of sPRP-HSA and natural PRP-HSA conjugates, demonstrating that the synthetic oligosaccharide possessed the relevant antigenic epitopes for antibody binding recognition (22).

As a potentially more relevant protein carrier for synthetic oligosaccharides, a tetanus toxoid (TT) conjugate **9** was evaluated in animals. The glycoconjugate was obtained through the thiolation of TT lysine  $\epsilon$ -amino groups as for HSA, and this sPRP-TT conjugate was immunogenic in rabbits with a wide range of sPRP/protein ratios, inducing a strong and specific antibody response (22).

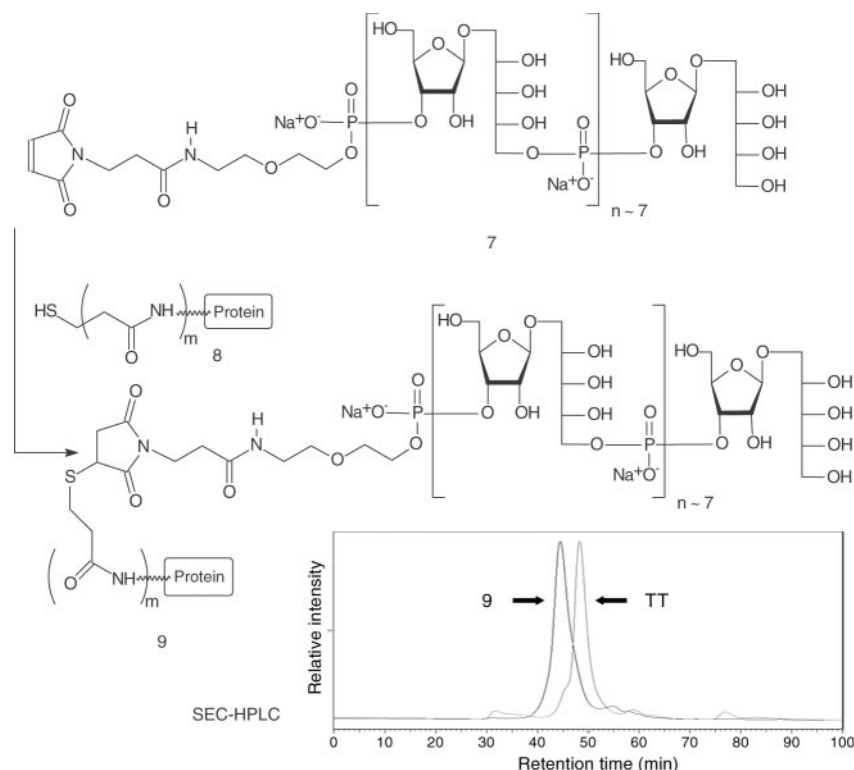
At this point, we identified four key issues that allowed us to accomplish further development of sPRP-TT as a vaccine candidate: (i) a synthetic pathway to disaccharide derivative **5** with only one chromatographic purification step, making it amenable to large-scale GMP production, (ii) a single-step, high-yielding polycondensation reaction between **4** and **5** for the elongation of the oligosaccharide chain, (iii) a method for careful removal of protective groups, yielding highly pure sPRP **7**, and lastly (iv) a conjugation process to TT carrier that incorporated sPRP in good yields.



**Fig. 1.** Synthetic pathway leading to oligomeric polyribosylribitol phosphate **6**. Reagents and conditions (Et, ethyl; Bn, benzyl; Bu, butyl; Ac, acetyl; and Piv, pivaloyl): (i)  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  and  $\text{CH}_2\text{Cl}_2$ ; (ii)  $\text{CH}_3\text{ONa}$  and  $\text{CH}_3\text{OH}$ ; (iii)  $\text{BnCl}$ ,  $\text{Bu}_2\text{SnO}$ ,  $\text{NaH}$ , and  $\text{Bu}_4\text{NI}$ ; (iv)  $\text{tBuOK}$  and dimethyl sulfoxide (DMSO) at  $100^\circ\text{C}$ ; (v)  $\text{PCl}_5$ , imidazole,  $\text{CH}_3\text{CN}$  for **5** and  $\text{N}_3(\text{CH}_2)_2\text{O}(\text{CH}_2)_2\text{OH}$  for **4**, followed by  $\text{I}_2$  oxidation; (vi)  $\text{AcOH-H}_2\text{O}$  at  $80^\circ\text{C}$ ; (vii)  $\text{PivCl}$  and pyridine; (viii)  $\text{I}_2$ , pyridine,  $\text{H}_2\text{O}$  then gel filtration on LH-20; (ix)  $\text{H}_2$ , Pd-C, and  $\text{EtOH-H}_2\text{O-EtOAc-AcOH}$  at 1.5 atm; (x) cation exchange chromatography on Sephadex SP-C25.

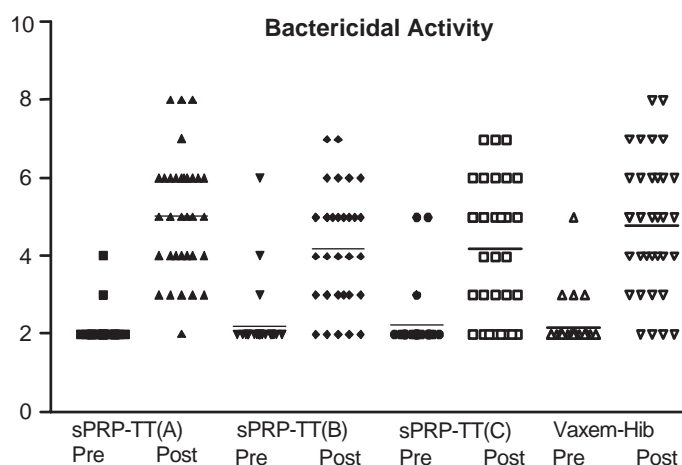
<sup>1</sup>Center for the Study of Synthetic Antigens, Facultad de Química, Universidad de la Habana, Ciudad Habana, Cuba, 10400. <sup>2</sup>Center for Genetic Engineering and Biotechnology, Apdo 6162, Cubanacán, Playa, Ciudad Habana, Cuba 10600. <sup>3</sup>Institute for Tropical Medicine Pedro Kouri, Autopista Novia del Mediodía, Km 6, La Lisa, Apdo 601, Marianao 13, Ciudad Habana, Cuba 11300. <sup>4</sup>Provincial Center for Hygiene and Epidemiology, Camaguey, Cuba. <sup>5</sup>Finlay Institute, Havana, Cuba 11600. <sup>6</sup>Department of Chemistry, Université du Québec à Montréal, Post Office Box 8888, Succursale Centre-Ville, Montreal, Québec H3C 3P8, Canada.

\*To whom correspondence should be addressed. E-mail: vicente@fq.uh.cu (V.V.-B.); violeta@fq.uh.cu (V.F.-S.); eugenio.hardy@cigb.edu.cu (E.H.); mariaeugenia@ipk.sld.cu (M.E.T.); roy.rene@uqam.ca (R.R.).

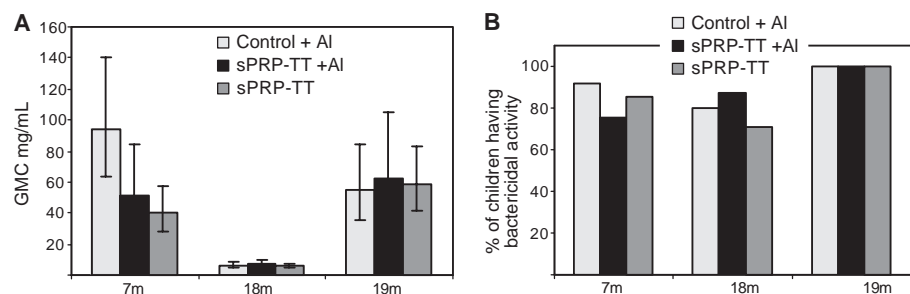


**Fig. 2.** Conjugation of the maleimido-functionalized polyribosylribitol phosphate **7** obtained from **6b** after coupling with 3-maleimidopropionic acid N-hydroxysuccinimide (DMSO, >95% conversion). 1,4-Conjugate addition of thiolated protein **8** onto **7** provided conjugate **9**. The shift in the molecular weight for TT could be observed in size exclusion chromatography–high performance liquid chromatography (TSK-5000-1 column) for conjugate **9** (PRP/TT ratio of 1/2.6).

**Fig. 3.** Bactericidal activity of immune serum against Hib (Eagan strain) obtained during a phase I trial conducted with children 4 to 5 years old in the province of Camaguey, Cuba, in the winter of 2002. The trial was conducted with 133 preschool children volunteers with the informed written consent of both parents. The trial was performed on a double-blind basis and all subjects were divided in four groups. Three groups (A to C) received a single dose of s-PRP-TT without any adjuvant (lots EH1024E, EH1026E, and EH1025E, respectively). The control vaccine (Vaxem-Hib) was administered to another group. The serum was obtained before and 4 weeks after the immunization. The plots are the reciprocal of the dilution killing 50% of the bacteria according to (26, 27).



**Fig. 4.** Geometric means concentration (GMC) of serum IgG antibodies against Hib (**A**) and % of children having bactericidal anti-Hib in their serum (**B**) after the primary series of immunization (7 months), a year after the preimmunization (18 months), and after a booster dose (19 months). A phase I clinical trial was performed with 139 infants in the Camaguey province east of Havana. Infants were enrolled in five groups receiving three doses at 2, 4, and 6 months of age. One group received control vaccine (Vaxem-Hib). Two groups received s-PRP-TT plus aluminum phosphate, and two other groups received s-PRP-TT alone. Blood was drawn at 7, 18, and 19 months for the evaluation of the immune status. At 18 months, all infants received a single booster dose of sPRP-TT irrespective of group. Bars indicate the 95% confidence interval.



The sPRP-TT conjugate vaccine with 10  $\mu$ g of sPRP and a sPRP-to-TT ratio of 1/2.6 by weight was next used as part of a phase I clinical trial in 40 adult volunteers after rigorous toxicological assessment in animals (22). A second phase I trial in adults with the vaccine plus aluminum phosphate was also performed. Both formulations of sPRP-TT were shown to be safe in adults. The average immunoglobulin G (IgG) titers obtained from our synthetic antigens (23) displayed a behavior similar to that of the control vaccine (Vaxem-Hib); the results obtained from this initial evaluation (table S1) prompted the initiation of a full clinical evaluation. All trials were performed as recommended and according to good clinical practice (24). Generally, they were randomized and double-blind and used a control vaccine and at least two different batches of the test vaccine (25).

The next phase I clinical trial consisted of a single vaccine dose in 133 4- to 5-year-old children previously unvaccinated against Hib. This was followed by a phase II trial using 1041 children. A substantial postvaccination increase in the anti-Hib IgG titers was observed in both trials, and a significant increase in the bactericidal activities of the sera (26) was reached after the administration of a single vaccine dose in three groups of children. The similarities of these with previous studies (27) again indicated that the sPRP-TT



vaccine was both as safe and as immunogenic as the commercial control vaccine (Fig. 3).

The above results set the stage for more detailed clinical assessment of the vaccine in a targeted infant population. Thus, a phase I trial was initiated with 139 2-month-old infants who received three vaccine doses scheduled at 2, 4, and 6 months, as recommended for other conjugate anti-Hib vaccines. The test vaccine induced a strong and bactericidal antibody response against Hib in infants (Fig. 4) that fell to values ranging from 5 to 7  $\mu\text{g/mL}$  at 18 months of age but remained at least five times that required for long-term protection (Fig. 4A). A booster dose with sPRP-TT applied to all groups increased the antibody against Hib titers by 10-fold. Thus, the capacity of sPRP-TT to prime an effective immune response against Hib was demonstrated.

In a second phase II trial, a total of 1141 infants distributed in three groups received three doses of either sPRP-TT conjugate, sPRP-TT mixed with aluminum phosphate, or the control vaccine (Vaxem-Hib). Of the test infants, 99.7% reached antibody titers above 1  $\mu\text{g/mL}$ , which is considered appropriate for long-lived protection against Hib (28, 29). The mean IgG anti-PRP titer was 27.4  $\mu\text{g/mL}$  for all infants vaccinated with the sPRP-TT, which is consistent with previously reported clinical trials (between 7.67 and 35  $\mu\text{g/mL}$ ) for anti-Hib vaccines without adjuvant (30, 31).

The present study demonstrates that a synthetic capsular polysaccharide antigen can be produced on a large scale under GMP conditions and used to manufacture an effective vaccine for human use. The resulting conjugate vaccine incorporating a synthetic bacterial carbohydrate antigen was demonstrated to be as safe and immunogenic in humans as already-licensed vaccines incorporating the native polysaccharide (32–34). Access to synthetic complex carbohydrate-based vaccines is therefore feasible and provides an alternative strategy in the fight against Hib infections. It also sets the stage for further development of similar approaches against other human pathogens.

# References and Notes

1. J. B. Robbins, R. Schneerson, P. Anderson, D. H. Schmidt, *JAMA* **276**, 1181 (1996).
2. H. Peltola, *Clin. Microbiol. Rev.* **13**, 302 (2000).
3. P. Anderson, G. Peter, R. B. Johnston Jr., L. H. Wetterlow, D. H. Smith, *J. Clin. Invest.* **5**, 39 (1972).
4. L. P. Rodrigues, R. Schneerson, J. B. Robbins, *J. Immunol.* **107**, 1071 (1971).
5. H. Peltola, H. Kayty, M. Virtanen, P. H. Makela, *N. Engl. J. Med.* **310**, 1561 (1984).
6. R. Schneerson, O. Barrera, A. Sutton, J. B. Robbins, *J. Exp. Med.* **152**, 361 (1980).
7. S. L. Cochi, D. O'Mara, S. R. Preblud, *Pediatrics* **81**, 166 (1988).
8. M. E. Ramsey, N. Andrews, E. B. Kaczmarski, E. Miller, *Lancet* **357**, 195 (2001).
9. S. Black et al., *Pediatr. Infect. Dis. J.* **19**, 187 (2000).
10. H. J. Jennings, R. A. Pon, in *Polysaccharides in Medicinal Applications*, S. Dimitriu, Ed. (Marcel Dekker, New York, 1996), p. 443.
11. J. B. Robbins, R. Schneerson, S. C. Szu, V. Pozsgay, in *Vaccinia, Vaccination and Vaccinology*, Jenner, Pas-

- teur and their Successors, S. Plotkin, B. Fantini, Eds. (Elsevier, Paris, 1996), p. 135.
12. P. Constantino et al., *Vaccine* **17**, 1251 (1999).
13. P. Hoogerhout et al., *Tetrahedron Lett.* **28**, 1553 (1987).
14. A. A. Kandil, N. Chan, P. Chong, M. Klein, *Synlett* **7**, 555 (1992).
15. S. Nilsson, M. Bengtsson, T. Norberg, *J. Carbohydr. Chem.* **11**, 265 (1992).
16. C. C. Peeters et al., *Infect. Immun.* **60**, 1826 (1992).
17. P. Chong et al., *Infect. Immun.* **65**, 4918 (1997).
18. I. Chiu-Machado, J. C. Castro-Palomino, O. Madrazo-Alonso, C. Lopetegui-Palacios, V. Verez-Bencomo, *J. Carbohydr. Chem.* **14**, 551 (1995).
19. A. V. Nikolaev, J. A. Chudek, M. A. J. Fergusson, *Carbohydr. Res.* **272**, 179 (1995).
20. V. Fernandez Santana, R. Gonzalez Lio, J. Sarracent Perez, V. Verez-Bencomo, *Glycoconjugate J.* **15**, 549 (1998).
21. D. C. Phipps et al., *J. Immunol. Methods* **135**, 121 (1990).
22. V. Fernandez-Santana et al., unpublished observations.
23. ELISA was performed according to (21) with the use of Hib reference serum pool for calibration, Center for Biologics Evaluator and Research, Food and Drug Administration.
24. P. Farrington, E. Miller, in *Vaccine Protocols*, vol. 87 of *Methods in Molecular Medicine*, A. Robinson, M. J. Hudson, M. P. Cranage, Eds. (Humana, Totowa, NJ, 2003), p. 335.
25. Materials and methods are available on Science Online.
26. Y. Schlesinger, D. M. Granoff, *JAMA* **267**, 1489 (1992).

27. S. Romero-Steiner et al., *Clin. Diagn. Lab. Immunol.* **8**, 1115 (2001).
28. H. Kayhty, H. Peltola, V. Karamko, P. H. Makela, *J. Infect. Dis.* **147**, 1100 (1983).
29. J. Amir, X. Liang, D. M. Granoff, *Pediatr. Res.* **27**, 358 (1990).
30. S. Holmes et al., *Am. J. Dis. Child.* **147**, 832 (1993).
31. D. M. Granoff, S. J. Holmes, *Vaccine* **9** (suppl.), S30 (1991).
32. P. Anderson, *Infect. Immun.* **39**, 233 (1983).
33. C. Chu, R. Schneerson, J. B. Robbins, S. C. Rastorgi, *Infect. Immun.* **40**, 245 (1983).
34. S. Marburg et al., *J. Am. Chem. Soc.* **108**, 5282 (1986).
35. We would like to thank the World Health Organization, the Pan-American Health Organization, and the following Cuban institutions: State Council, Ministers of Science Technology and Environment, and Ministry of Health. We are particularly thankful to J. M. Miyar, M. C. Santana, L. Yañez, J. L. DiFabio, M. Beurret, and C. Jones for their fundamental contributions and the many laboratory assistants, medical doctors, and nurses that were involved in the project over the years. C. H. Fox critically reviewed the manuscript.

# Supporting Online Material

www.sciencemag.org/cgi/content/full/305/5683/522/DC1

Materials and Methods

Fig. S1

Table S1

31 December 2003; accepted 23 June 2004

## Large-Scale Copy Number Polymorphism in the Human Genome

Jonathan Sebat,<sup>1</sup> B. Lakshmi,<sup>1</sup> Jennifer Troge,<sup>1</sup> Joan Alexander,<sup>1</sup> Janet Young,<sup>2</sup> Pär Lundin,<sup>3</sup> Susanne Månér,<sup>3</sup> Hillary Massa,<sup>2</sup> Megan Walker,<sup>2</sup> Maoyen Chi,<sup>1</sup> Nicholas Navin,<sup>1</sup> Robert Lucito,<sup>1</sup> John Healy,<sup>1</sup> James Hicks,<sup>1</sup> Kenny Ye,<sup>4</sup> Andrew Reiner,<sup>1</sup> T. Conrad Gilliam,<sup>5</sup> Barbara Trask,<sup>2</sup> Nick Patterson,<sup>6</sup> Anders Zetterberg,<sup>3</sup> Michael Wigler<sup>1\*</sup>

The extent to which large duplications and deletions contribute to human genetic variation and diversity is unknown. Here, we show that large-scale copy number polymorphisms (CNPs) (about 100 kilobases and greater) contribute substantially to genomic variation between normal humans. Representational oligonucleotide microarray analysis of 20 individuals revealed a total of 221 copy number differences representing 76 unique CNPs. On average, individuals differed by 11 CNPs, and the average length of a CNP interval was 465 kilobases. We observed copy number variation of 70 different genes within CNP intervals, including genes involved in neurological function, regulation of cell growth, regulation of metabolism, and several genes known to be associated with disease.

Many of the genetic differences between humans and other primates are a result of large duplications and deletions (1–3). From these observations, it is reasonable to expect that differences in gene copy number could be a significant source of genetic variation between humans. A few examples of large duplication polymorphisms have been reported (4). However, because of previous limitations in the power to determine DNA copy number at high resolution throughout the genome, the extent to which copy number polymorphisms (CNPs) contribute to human genetic diversity is unknown.

In our previous studies of human cancer with the use of representational oligonucleotide microarray analysis (ROMA), we have detected many genomic amplifications and deletions in tumor genomes when analyzed in comparison to an unrelated normal genome (5), but some of these genetic differences could be due to germline CNPs. To correctly interpret genomic data relating to cancer and other diseases, we must distinguish abnormal genetic lesions from normal CNPs.

We used ROMA to investigate the extent of copy number variation between normal