Protein NMR

International Edition: DOI: 10.1002/anie.201612518
German Edition: DOI: 10.1002/ange.201612518

Structure of the Complex between a Heparan Sulfate Octasaccharide and Mycobacterial Heparin-Binding Hemagglutinin

Teng-Yi Huang[+], Deli Irene[+], Medel Manuel L. Zulueta, Tzu-Jui Tai, Shih-Han Lain, Cheng-Po Cheng, Ping-Xi Tsai, Shu-Yi Lin, Zhi-Geng Chen, Chiao-Chu Ku, Chwan-Deng Hsiao, Chia-Lin Chyan, and Shang-Cheng Hung**

Dedicated to Professor Chun-Chen Liao on the occasion of his 75th birthday

Abstract: *Heparin-binding hemagglutinin (HBHA) is a 199 amino acid virulence factor at the envelope of Mycobacterium tuberculosis that contributes to latent tuberculosis. The binding of HBHA to respiratory epithelial cells, which leads to extrapulmonary dissemination of the pathogen, is mediated by cell-surface heparan sulfate (HS). We report the structural characterization of the HBHA/HS complex by NMR spectroscopy. To develop a model for the molecular recognition, the first chemically synthesized uniformly ^{13}C- and ^{15}N-labeled HS octasaccharide and a uniformly ^{13}C- and ^{15}N-labeled form of HBHA were prepared. Residues 180–195 at the C-terminal region of HBHA show large chemical shift perturbation upon association with the octasaccharide. Molecular dynamics simulations conforming to the multidimensional NMR data revealed key electrostatic and even hydrophobic interactions between the binding partners that may aid in the development of agents targeting the binding event.*

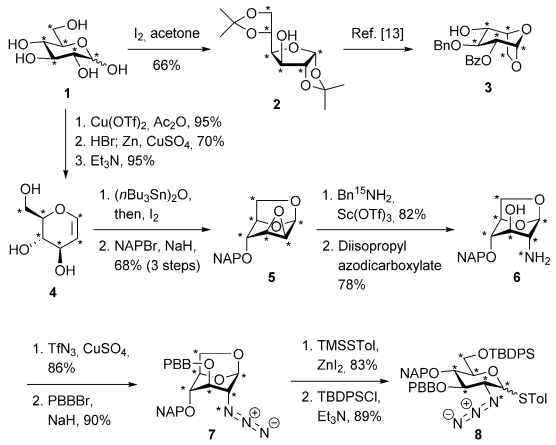
Tuberculosis is a major global threat with over a million annual deaths and a high incidence of drug resistance.[1] The causative agent, *Mycobacterium tuberculosis*, invades alveolar macrophages and can survive and evade the host defenses. It is a terrible pulmonary disease in active form, but most infections lead to latent tuberculosis with no clinical expression.[2] Critical for enabling latency is heparin-binding hemagglutinin (HBHA). This mycobacterial envelope protein is essential for binding to heparan sulfate (HS) at the surface of respiratory epithelial cells, thereby promoting extrapulmonary dissemination.[3] HBHA contains a transmembrane domain (residues 12–29), a coiled-coil domain (residues 29–109), and a C-terminal region (residues 160–199) rich in lysine, proline, and alanine residues.[4] The coiled-coil domain acts as a trigger sequence for protein dimerization and promotes bacterial agglutination.[5]

HS is a polysaccharide component of proteoglycans found on cell surfaces and in the extracellular matrix.[6] It holds alternating (1→4)-linked α-D-glucosamine (GlcN) and either β-D-glucuronic acid (GlcA) or α-L-iduronic acid (IdoA) residues. O- and N-sulfonations occur to variable extents, which, together with the carboxylate group, provide a polyanionic chain and a complex functional-group pattern.[7] These properties facilitate association with many circulating proteins, including growth factors, chemokines, and microbial proteins such as HBHA.[8] Heparinase treatment[9] and anti-HBHA monoclonal antibodies[10] disrupt the attachment of HBHA to epithelial cells. A deep understanding of this recognition process could thus inspire new therapeutic strategies against tuberculosis.

Electrostatic forces likely feature prominently in the interactions between HS and HBHA, as observed with many other HS-binding proteins.[6] To investigate the binding process and avoid heterogeneous isolates from natural sources, we previously synthesized HS oligosaccharides bearing 2-N- and 6-O-sulfonated GlcN and 2-O-sulfonated IdoA.[11] Such units are abundant in the sulfated regions of HS and are also the major components of heparin, a related sugar that is mainly localized within mastocytes. Using these compounds, the hexasaccharide was identified as the minimum length required for binding to HBHA. The octasaccharide analogue provided a dissociation constant (K_D) of 4.2 μM in an entropically driven 1:1 interaction with a truncated form of HBHA (i.e., $HBHA_{60\text{-}199}$), thus curiously suggesting that hydrophobic interactions could be major contributors to the binding. Mindful that the seemingly unstructured nature of the C-terminal region[5a] may deter X-ray characterization, we turned to NMR spectroscopy to probe the complexation process at the molecular level. To address the severe resonance overlaps that have historically hindered NMR interpretations, the first chemical synthesis of a uniformly ^{13}C- and ^{15}N-labeled HS octasaccharide (hereafter referred to as HS8) was carried out. Similar uniform labeling was also implemented on a form of HBHA expressed in *Escherichia coli* with ^{13}C-labeled glucose and $^{15}NH_4Cl$ as the sources of carbon and nitrogen atoms, respectively.

[*] Dr. T.-Y. Huang,[+] Dr. M. M. L. Zulueta, Dr. C.-P. Cheng, Dr. P.-X. Tsai, Dr. S.-Y. Lin, Dr. Z.-G. Chen, Prof. Dr. S.-C. Hung
Genomics Research Center, Academia Sinica
No. 128, Section 2, Academia Road, Taipei 115 (Taiwan)
E-mail: schung@gate.sinica.edu.tw
Dr. D. Irene,[+] T.-J. Tai, S.-H. Lain, Prof. Dr. C.-L. Chyan
Department of Chemistry, National Dong Hwa University
No. 1, Section 2, Da Hsueh Road, Shoufeng, Hualien 974 (Taiwan)
E-mail: chyan@gms.ndhu.edu.tw
C.-C. Ku, Prof. Dr. C.-D. Hsiao
Institute of Molecular Biology, Academia Sinica
No. 128, Section 2, Academia Road, Taipei 115 (Taiwan)
[+] These authors contributed equally to this work.
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: http://dx.doi.org/10.1002/anie.201612518.

To prepare the desired labeled HS8, a uniformly ^{13}C-labeled D-glucose (**1**) was transformed into the crucial anhydro-L-idose and GlcN building blocks **3** and **8**, respectively (Scheme 1). Both **3** and **8** were designed to offer excellent stereoselectivity during glycosylations and permit the final functionalizations.[11,12] The synthesis of **3** began with

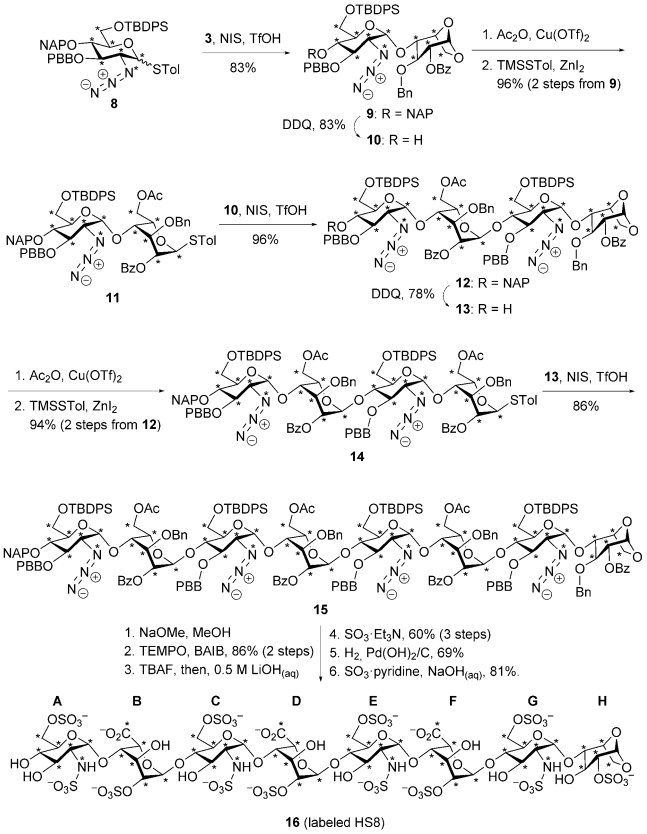


Scheme 1. Preparation of the monosaccharide building blocks. Asterisks indicate ^{13}C- or ^{15}N-labeled atoms. Bn = benzyl, Bz = benzoyl, NAP = 2-naphthylmethyl, PBB = *p*-bromobenzyl, TBDPS = *tert*-butyldiphenylsilyl, Tf = triflyl, TMS = trimethylsilyl, Tol = *p*-tolyl.

the formation of the diacetonide **2**. The transformation then followed the typical method[13] of C5-epimerization and 1,6-anhydroidose formation, thereby ultimately generating the desired 4-alcohol **3**. Introducing the ^{15}N label is the major concern for the GlcN synthesis. Inspired by an early report,[14] we accessed this unit through epoxide **5**. After considering some ^{15}N sources such as azide (undesirable 1,3-dipolar property) or phthalimide (less than satisfactory yield), we settled on ^{15}N-labeled benzylamine for epoxide opening under $Sc(OTf)_3$ catalysis. Subsequently, the resulting ^{15}N-labeled amine **6** was converted into the azide **7**. Simultaneous thioglycosylation and anhydro-ring opening, followed by 6-O-silylation led to the fully functionalized thioglycoside **8**.

We then shifted to oligosaccharide assembly and functional-group transformations (Scheme 2). Activation of thioglycoside **8** in the presence of **3** supplied the key disaccharide **9**, which was transformed into the corresponding acceptor **10** and thioglycoside donor **11** for the next round of glycosylation. Convergent [2+2] and [4+4] glycosylations effectively generated the fully protected octasaccharide **15**. Partial deblocking, selective oxidation, O-sulfonation, global deprotection, and final N-sulfonation delivered the target analogue **16** (labeled HS8).

Circular dichroism (CD) spectroscopy and isothermal titration calorimetry (ITC) gave preliminary insight into the complexation process. Addition of HS8 to full-length HBHA changed the CD absorptivities at 208 and 222 nm (Figure S1 in the Supporting Information). Analysis[15] showed that, at the expense of some β-sheet residues, the α-helical content of full-length HBHA in the presence of equimolar HS8



Scheme 2. Preparation of the uniformly ^{13}C- and ^{15}N-labeled HS8. Ac = acetyl, BAIB = bis(acetoxy)iodobenzene, DDQ = 2,3-dichloro-5,6-dicyano-1,4-benzoquinone, NIS = *N*-iodosuccinimide, TBAF = tetrabutylammonium fluoride, TEMPO = 2,2,6,6-tetramethyl-1-piperidinyloxyl free radical.

increased by 6%, roughly equivalent to a 12-residue α-helical structure (Table S1 in the Supporting Information). Conversely, no noteworthy changes occurred in the CD spectra when HS8 was added to the truncated forms $HBHA_{60\text{-}199}$, $HBHA_{110\text{-}199}$, and $HBHA_{1\text{-}160}$ (the subscripts are amino acid positions). The thermodynamic parameters for the complexation of HS8 with $HBHA_{87\text{-}199}$, $HBHA_{110\text{-}199}$, and $HBHA_{160\text{-}199}$ were also acquired by ITC (Figure S2 and Table S2). As with $HBHA_{60\text{-}199}$,[11] the interactions were endothermic and largely stabilized by entropic factors, with comparable K_D values in the low micromolar range (3.18–7.46 μM). HS8 has a binding stoichiometry of around 1:1 with both $HBHA_{87\text{-}199}$ and $HBHA_{110\text{-}199}$, but a 1:2 sugar-to-protein binding ratio was recorded with $HBHA_{160\text{-}199}$. These data indicate that HS8 binds to the C-terminal region and induces a conformational change in HBHA, increasing the helical content within residues 1–60.

With adequate water solubility and reasonable affinity to HS8, $HBHA_{110\text{-}199}$ was chosen as a surrogate in our NMR characterization. The ^{13}C- and ^{15}N-labeled $HBHA_{110\text{-}199}$ enabled resonance assignments of the free protein through a series of multidimensional NMR experiments (see the Supporting Information).[16] Aliphatic side-chain proton and carbon chemical shifts were assigned by through-bond correlations and then confirmed by nuclear Overhauser

effects (NOEs) in ^{13}C- and ^{15}N-separated NOE spectroscopy (NOESY)-heteronuclear single quantum correlation (HSQC) spectra. A similar analysis for bound $HBHA_{110\text{-}199}$ was performed using a sample containing labeled $HBHA_{110\text{-}199}$ and unlabeled HS8. The chemical shift index algorithm[17] confirmed that free $HBHA_{110\text{-}199}$ is mostly unstructured, and no secondary structure was induced upon association with HS8. The solution structures of free and bound $HBHA_{110\text{-}199}$ were calculated by CYANA.[18] The majority of the NOE cross-peaks were from intra-residue and sequential proton pairs, and medium-range NOEs were found in three regions, indicating short α-helical structures (within residues 116–120, 128–138, and 146–150).

Chemical-shift perturbation was used to map the interaction site and to calculate K_D by regression analysis. The NMR spectra of a series of complexes of labeled $HBHA_{110\text{-}199}$ with unlabeled HS8 in different stoichiometric ratios were acquired, taking note of the perturbed cross-peaks of $HBHA_{110\text{-}199}$ (Figure 1a and Figure S3). The weighted chemical-shift perturbation for each residue ($\Delta\delta_{residue}$) was calculated (see the Supporting Information). The residues K180, A181, A182, A183, K185, A186, K190, A191, A192, A193, K194, and K195 were the most perturbed, with $\Delta\delta_{residue}$ values more than three times the average (0.019 ppm). Situated near the end of the C-terminal region of HBHA, these residues are likely in proximity to the interaction surface or undergo major structural readjustment upon binding. The close match of the perturbed amino acids with that of a previous NMR analysis utilizing a 14-mer heparin sample and a chimeric protein[4b] indicates their pivotal roles. The cross-peak movements with increasing concentrations of HS8 suggest fast ligand exchange (Figure 1b and Figure S4). Analysis of the tracked cross-peaks shows that HS8 has 1:1 stoichiometry with $HBHA_{110\text{-}199}$, with an average K_D of 4.49 μM (Table S3), which is consistent with ITC data.

Resonance assignments of free and bound HS8 were obtained by correlating intra- and inter-residue through-bond and through-space connectivities (Figure S5–S8). The NMR spectra for bound HS8 were acquired from a sample containing labeled HS8 and unlabeled $HBHA_{110\text{-}199}$. The chemical shifts of residues A, F, and G are well-dispersed, whereas residues B and C are degenerate to residues D and E, respectively (Table S4, see compound **16** in Scheme 2 for sugar residue notation). The majority of the HS8 cross-peaks show only minor or no deviation upon association with $HBHA_{110\text{-}199}$. Coupling constants (Table S5) were used to elucidate the ring conformations[19] and to derive the dihedral angles according to the Karplus correlation.[20] While all of the GlcN residues were in the $^4C_1$ conformation, the IdoA residues appeared as mixtures of ^2S_O and $^1C_4$ in free HS8.[19–21] Upon interaction with HBHA, a preference for the ^2S_O conformation was observed.[22] This was confirmed by the strong NOE intensities between H2 and H5 of the IdoA residues in the bound form; the corresponding free-form NOEs were much weaker (Figure S8). Consistent with restricted motion, more NOEs were found for bound than free HS8.

Structural modeling of the sugar/protein complex was performed using restrained molecular dynamics with distance

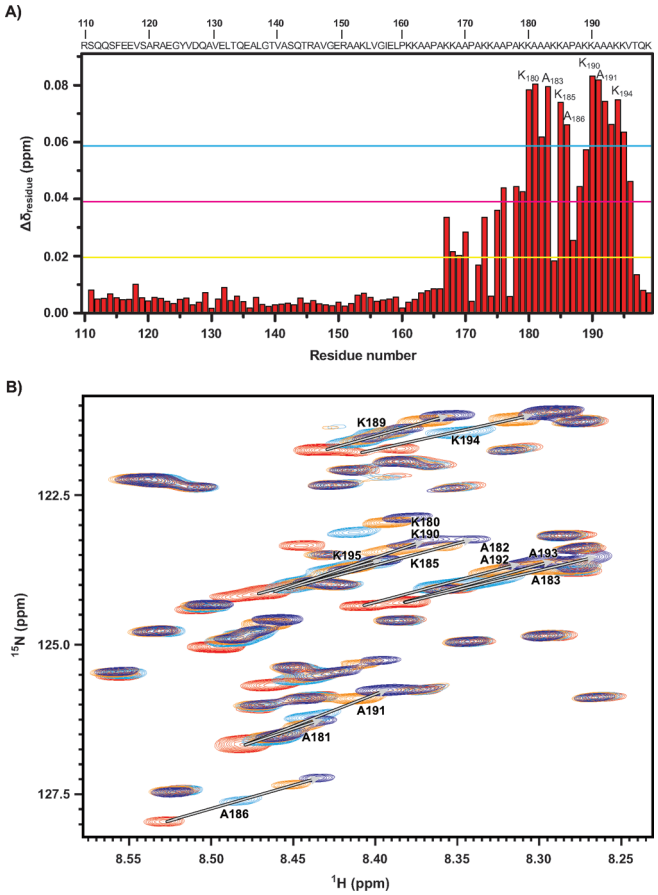


Figure 1. Perturbations in $HBHA_{110\text{-}199}$ upon complexation with HS8. A) The weighted chemical-shift perturbation ($\Delta\delta_{residue}$) of $HBHA_{110\text{-}199}$ upon association with HS8 plotted against residue number. The yellow line corresponds to the average $\Delta\delta_{residue}$ (0.019 ppm); two- and threefold averaged $\Delta\delta_{residue}$ are indicated in magenta and cyan, respectively. B) Overlaid ^1H-^{15}N HSQC spectra showing the movement of the cross-peaks of key residues of labeled $HBHA_{110\text{-}199}$ upon titration with unlabeled HS8. The spectrum of free $HBHA_{110\text{-}199}$ is shown in red, with cyan for the complex of HS8 with $HBHA_{110\text{-}199}$ in a stoichiometric ratio of 0.5, orange for a ratio of 1, and in blue for a ratio of 2.

and dihedral angle restraints from the NMR data (Table S6–S9). Focusing on the lysine ε proton and the alanine β proton, 16 intermolecular NOE restraints were deduced from the ^{13}C-separated NOESY–HSQC spectra of the doubly labeled $HBHA_{110\text{-}119}$ and HS8 (Table S10). Twenty lowest-energy structures from the simulations, which fulfil the binding interface defined by the intermolecular NOEs and chemical-shift deviation experiments, were selected (Figure 2a). The best low-energy structure was chosen for further analysis. In this model, the protein formed three prominent loops spanning residues 181–185, 186–189, and 193–199 in such a way as to produce a shallow binding ridge for HS8. As expected, lysine residues, specifically, K179, K180, K184, K185, K189, and K199, participated in ionic interactions with the sulfonate and carboxylate groups found within residues A–G of HS8 (Figure 2b). Notably, the terminal K199 residue looped around and formed salt bridges with the 2-O-sulfonate group of D (IdoA), the 6-O-sulfonate group of E (GlcN), and the carboxylate group of F (IdoA) residues. Hydrogen

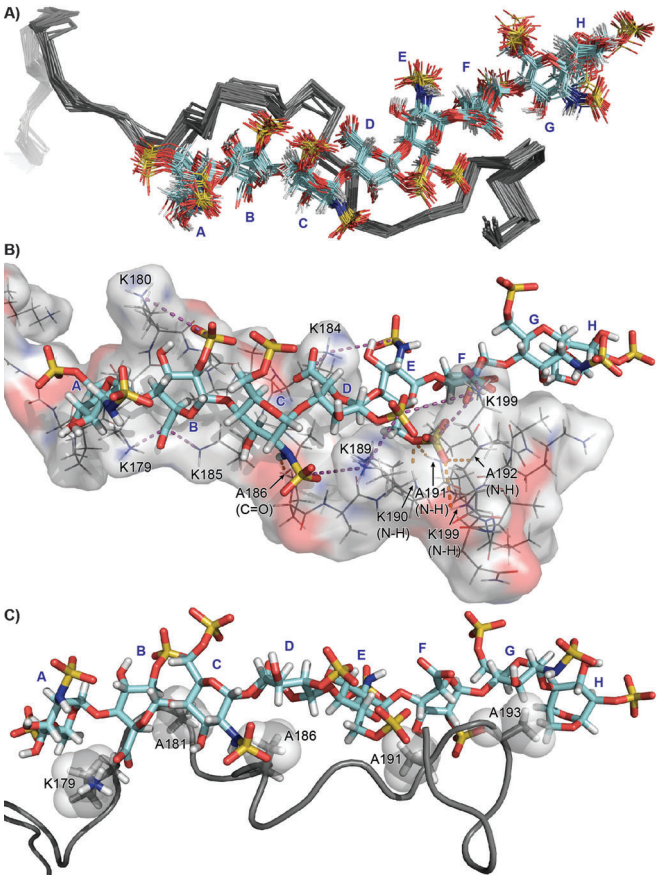


Figure 2. Modeling of the HBHA/HS8 complex, focusing on the interaction site. A) Superposition of the 20 lowest-energy conformers of HBHA and HS8. B) Intermolecular ionic interactions (magenta dashed line) and hydrogen bonding (orange dashed line) with HBHA shown as a transparent surface. C) Hydrophobic interactions involving K179, A181, A186, A191, and A193 with the relevant atoms shown as transparent spheres. HS8 is shown with cyan carbon atom; HBHA carbon atoms and backbone are shown in dark gray. The rest of the model has the color scheme N blue, O red, S yellow, H white.

bonding, mainly through the 6-sulfate group of the E residue of HS8, also helped anchor the sugar to the protein. Judging from the length and pattern of the lysine-rich C-terminal region, a sugar length about twice that of HS8 would probably be accommodated by HBHA. Moreover, the minimal requirement of six sugar units (represented by residues C to H) found by ITC[11] also made sense in this model.

Hydrophobic interactions were implied by the entropically driven binding of HS8 to HBHA. Our model shows that four alanine residues participate in hydrophobic clusters with the sugar rings (Figure 2c and Figure S9). A181 and A186 associate with the B-D rings of the sugar, whereas A191 and A193 engage in close interaction with E-H rings. The key presence of several alanines alongside multiple lysine residues within the HS-binding C-terminal region of HBHA provides an opportunity for this notable hydrophobic interaction to occur.[8] In addition, the aliphatic side-chain component of K179 shows hydrophobic interactions with the A-residue (GlcN) atoms of HS8. These interfaces likely displace ordered water molecules in the sugar to create the entropic effect that drives the binding process. On the other hand, the protein residues that showed highly perturbed chemical-shift values but were not involved in interacting with the sugar (A183, K190, K194, and K195) instead underwent structural rearrangement of the flexible random coil to permit the binding of HS8 to HBHA.

In summary, we have examined the complexation of HS8 with HBHA using CD, ITC, and NMR spectroscopy. The NMR analysis was enhanced by the use of ^{13}C- and ^{15}N-labeled forms of both the sugar and the protein, thereby providing key information that was beneficial in the molecular dynamics simulation of the complex. We noted relevant electrostatic interactions between the lysine side chains and the sulfonate and carboxylate groups of the sugar. The hydrophobic interactions that we suspected to be major factor in the binding were also clarified as coming from certain lysine aliphatic regions and alanine residues populating the C-terminal region. Overall, these molecular details should offer important insight into the binding process and present new opportunities for drug design against latent tuberculosis. Further examination of the binding process and its effect on the conformation of HBHA is in progress.

Acknowledgements

This work was supported by the Ministry of Science and Technology of Taiwan (MOST 104-2113-M-259-006, MOST 104-0210-01-09-02, MOST 104-2628-M-001-001, MOST 105-0210-01-13-01, and MOST 105-2745-M-001-003-ASP), National Dong Hwa University, and Academia Sinica.

Conflict of interest

The authors declare no conflict of interest.

Keywords: carbohydrates · heparan sulfate · heparin-binding hemagglutinin · NMR spectroscopy · protein structures

[1] World Health Organization (WHO), *Global Tuberculosis Report*, 20th ed., WHO Press, Geneva, **2015**.
[2] A. Koul, E. Arnoult, N. Lounis, J. Guillemont, K. Andries, *Nature* **2011**, *469*, 483–490.
[3] a) K. Pethe, S. Alonso, F. Biet, G. Delogu, M. J. Brennan, C. Locht, F. D. Menozzi, *Nature* **2001**, *412*, 190–194; b) N. Krishnan, B. D. Robertson, G. Thwaites, *Tuberculosis* **2010**, *90*, 361–366.
[4] a) G. Delogu, M. J. Brennan, *J. Bacteriol.* **1999**, *181*, 7464–7469; b) P. Lebrun, D. Raze, B. Fritzinger, J.-M. Wieruszeski, F. Biet, A. Dose, M. Carpentier, D. Schwarzer, F. Allain, G. Lippens, C. Locht, *PLoS ONE* **2012**, *7*, e32421.
[5] a) J. V. Lomino, A. Tripathy, M. R. Redinbo, *J. Bacteriol.* **2011**, *193*, 2089–2096; b) C. Esposito, M. Cantisani, G. D'Auria, L. Falcigno, E. Pedone, S. Galdiero, R. Berisio, *FEBS Lett.* **2012**, *586*, 659–667.

[6] D. Xu, J. D. Esko, *Annu. Rev. Biochem.* **2014**, *83*, 129 – 157.
[7] M. M. L. Zulueta, S.-Y. Lin, Y.-P. Hu, S.-C. Hung, *Curr. Opin. Chem. Biol.* **2013**, *17*, 1023 – 1029.
[8] N. S. Gandhi, R. L. Mancera, *Chem. Biol. Drug Des.* **2008**, *72*, 455 – 482.
[9] F. D. Menozzi, V. M. Reddy, D. Cayet, D. Raze, A.-S. Debrie, M.-P. Dehouck, R. Cecchelli, C. Locht, *Microbes Infect.* **2006**, *8*, 1 – 9.
[10] K. Pethe, V. Puech, M. Daffé, C. Josenhans, H. Drobecq, C. Locht, F. D. Menozzi, *Mol. Microbiol.* **2001**, *39*, 89 – 99.
[11] M. M. L. Zulueta, S.-Y. Lin, Y.-T. Lin, C.-J. Huang, C.-C. Wang, C.-C. Ku, Z. Shi, C.-L. Chyan, D. Irene, L.-H. Lim, T.-I. Tsai, Y.-P. Hu, S. D. Arco, C.-H. Wong, S.-C. Hung, *J. Am. Chem. Soc.* **2012**, *134*, 8988 – 8995.
[12] Y.-P. Hu, S.-Y. Lin, C.-Y. Huang, M. M. L. Zulueta, J.-Y. Liu, W. Chang, S.-C. Hung, *Nat. Chem.* **2011**, *3*, 557 – 563.
[13] S.-C. Hung, X.-A. Lu, J.-C. Lee, M. D.-T. Chang, S.-l. Fang, T.-c. Fan, M. M. L. Zulueta, Y.-Q. Zhong, *Org. Biomol. Chem.* **2012**, *10*, 760 – 772.
[14] S. Arndt, L. C. Hsieh-Wilson, *Org. Lett.* **2003**, *5*, 4179 – 4182.
[15] a) L. Whitmore, B. A. Wallace, *Nucleic Acids Res.* **2004**, *32*, W668 – W673; b) L. Whitmore, B. A. Wallace, *Biopolymers* **2008**, *89*, 392 – 400.
[16] The complete ^{1}H, ^{13}C, and ^{15}N chemical-shift assignments the free and bound $HBHA_{110-199}$ have been deposited in the Biological Magnetic Resonance Bank (BMRB) with accession numbers 26887 and 26888, respectively.
[17] D. S. Wishart, B. D. Sykes, *J. Biomol. NMR* **1994**, *4*, 171 – 180.
[18] P. Güntert, L. Buchner, *J. Biomol. NMR* **2015**, *62*, 453 – 471.
[19] P.-H. Hsieh, D. F. Thieker, M. Guerrini, R. J. Woods, J. Liu, *Sci. Rep.* **2016**, *6*, 29602.
[20] M. Hricovíni, F. Bízik, *Carbohydr. Res.* **2007**, *342*, 779 – 783.
[21] M. Hricovíni, P.-A. Driguez, O. L. Malkina, *J. Phys. Chem. B* **2014**, *118*, 11931 – 11942.
[22] a) M. Guerrini, S. Guglieri, D. Beccati, G. Torri, C. Viskov, P. Mourier, *Biochem. J.* **2006**, *399*, 191 – 198; b) Z. Zhang, S. A. McCallum, J. Xie, L. Nieto, F. Corzana, J. Jiménez-Barbero, M. Chen, J. Liu, R. J. Linhardt, *J. Am. Chem. Soc.* **2008**, *130*, 12998 – 13007; c) L. Nieto, Á. Canales, G. Giménez-Gallego, P. M. Nieto, J. Jiménez-Barbero, *Chem. Eur. J.* **2011**, *17*, 11204 – 11209; d) L. Nieto, Á. Canales, I. S. Fernández, E. Santillana, R. González-Corrochano, M. Redondo-Horcajo, F. J. Cañada, P. Nieto, M. Martín-Lomas, G. Giménez-Gallego, J. Jiménez-Barbero, *ChemBioChem* **2013**, *14*, 1732 – 1744.

Manuscript received: December 24, 2016
Revised: February 11, 2016
Final Article published: ■■ ■■, ■■■■

