Synthesis and Evaluation of Glycomimetics:

Tool Compounds Binding to the FimH Adhesin for Analytical Applications

and

New Antagonists of the PapG-II Adhesin

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Giulio Navarra
aus Rom, Italien

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von

Prof. Dr. Beat Ernst, *Departement für Pharmazeutische Wissenschaften, Universität Basel*

Prof. Dr. Peter Kolb, *Departement für Pharmazeutische Chemie, Philipps-Universität Marburg*

Basel, den 8. Dezember 2015

Dekan

Prof. Dr. Jörg Schibler

Acknowledgements

I would like to express my deep gratitude to Prof. Dr. Beat Ernst for his patient guidance, constant motivation and support, and constructive critiques during my doctoral studies. It has been a great honor to be part of his research group.

I would also like to thank Prof. Dr. Peter Kolb for being the co-referee of my dissertation.

I wish to acknowledge the great support provided by Prof. Dr. Peter Kolb during my short-term scientific mission in his group in Marburg, Germany.

I am particularly grateful for the technical assistance given by Dr. Brigitte Fiege, Dr. Roman Jakob, Dr. Oliver Schwardt. It was a great pleasure to learn from them and to collaborate.

I would like to offer my special thanks to Dr. Mathias Wipf, Dr. Said Rabbani, Pascal Zihlmann, Dr. Martin Smiesko, Dr. Meike Scharenberg, Dr. Roland Preston, Ralph Stoop, Dr. Alexey Tarasov, Dr. Paolo Livi, Dr. Jacqueline Bezençon, Tobias Mühlethaler, Lisa Beretta Piccoli for the fruitful collaborations in several research projects.

I am grateful to all current and past members of the Institute of Molecular Pharmacy for the great work atmosphere, the high quality of the research, and friendly helpfulness. Special thanks should be given to my colleagues from lab.4014 for the enjoyable environment and for their help. I am particularly grateful to Wojciech Schönemann for his empathy and friendship since the very beginning of my studies.

I am thankful to Bea Wagner, Claudia Huber, and Gabi Lichtenhahn for their technical and administrative support.

I would like to thank Yvonne Götz, Christian Peter, Marco Donnicola, and Anna Eitler for their friendship and for the help in settling in Basel and in learning German.

I would like to acknowledge the constant support and encouragement provided by my family.

Last but not least, I sincerely thank my dearest wife Valentina Todino for having joined me in Basel and for her continual support.

Abstract

The spreading of bacterial resistance is promoting global research efforts toward the development of new therapeutic alternatives. Antivirulence therapy seems to be a valid, new avenue for discovering innovative medicines. In this context, anti-adhesive drugs, which block the first step of bacterial colonization of the host's tissues are particularly attractive, as they do not kill the pathogens, and thus do not contribute to the selection of resistant strains.

Urinary tract infections (UTIs) are among the most frequent reasons for antibiotic intake, thus playing a pivotal role in spreading bacterial resistance. Moreover, their recurrent nature reduces consistently patient's quality of life. As the most common pathogen involved in UTIs is *E. coli* (about 80% in otherwise healthy patients), an anti-adhesive therapy against it would be highly valuable. *E. coli* uses filamentous structures called pili to adhere to the host's tissues. In UTIs concerning the lower urinary tract (cystitis), type 1 pili are mainly involved. At the tip of type 1 pili, the lectin FimH is expressed, which recognizes mannosylated glycoproteins, abundant in the urinary bladder. A large body of literature is dedicated to antagonizing FimH. Despite nanomolar antagonists have been long discovered, suitable clinical candidates are lacking. One important determinant for a successful drug is the target occupancy time. Using surface plasmon resonance, we demonstrated that our lead structures have excellent kinetic profiles, when tested against the FimH lectin domain (*paper 1*). However, one crucial limiting factor was the poor pharmacokinetic profile of these antagonists. We therefore successfully tailored the physicochemical properties of a set of promising lead structures (*paper 2*).

In order to support therapy, a detection system for FimH-expressing *E. coli* is of great importance. Biosensors offer several advantages, including reliability, low cost, and ease of use. Using a FimH antagonist as recognition element and FimH as analyte, we developed a FimH sensitive biosensor, providing the first proof of concept of label-free detection of a pathologically relevant protein, by field-effect, silicon nanoribbons-based sensors (SiNR-BioFET, *paper 4*).

However, most research efforts have until very recently focused on the isolated lectin domain of FimH, which exists in a high-affinity state. To finally clarify if the high-affinity state is the appropriate therapeutic target, a study based on crystallography, molecular dynamics, and kinetics was undertaken on the full-length FimH protein, which exists prevalently in a low-affinity state (*paper 3*). The results support the use of the full-length protein as the most appropriate model for anti-adhesive therapy, thus opening a completely new research path for medicinal chemistry studies.

In UTIs involving the human upper urinary tract (pyelonephritis), *E. coli* type P pili have been shown to play an important role. The adhesive properties of these pili arise from the PapG-II adhesin, which recognizes the tetrasaccharide epitope of tetraosyl galactosyl globosides (GbO4). Although the incidence of upper UTIs as compared to cystitis is rather low, the risk of serious organ damage is high. Moreover, the increasing frequency of resistant strains requires new therapeutic alternatives. Medicinal chemistry has so far focused on the modification of the minimal binding epitope, i.e. Galα(1→4)Gal. However, the best published lead compound exhibits affinity only in the mid-micromolar range. Based on a critical analysis of the present literature on antagonists of PapG-II and of the closely related PapG-I, a new, not yet explored sub-binding site was identified and explored. Disappointingly, no improvement in affinity could be achieved (*chapter 3.2.2*), confirming the challenging nature of the target.

Fragment-based approaches have been shown to have a great potential for hard-to-drug targets. In our group, second-site ligand search using fragments had been successfully applied on other lectin targets. The same strategy was applied to PapG-II, albeit with scarce success (*manuscript 1*).

The observation that the hexasaccharide epitope of the sialosyl galactosyl globoside shows 5-fold increased affinity for PapG-II as compared to the epitope of GbO4, from which it differs by an added disaccharid units at the non-reducing end, led us to study the details of the interaction. Crystallographic and thermodynamic investigations suggested that the improvement in affinity arises from an entropic contribution, due to the non-binding, terminal saccharidic units (*paper 5*).

The data collected during the development of this thesis added important information on PapG-II and will assist further medicinal chemistry research toward the development of high-affinity antagonists.

Abbreviations

$[\alpha]_D^{20}$	Optical rotation at λ=589 nm
°C	Degree Celsius
4-NH_2-TEMPO	4-Amino-2,2,6,6-tetramethylpiperidine-1-oxyl
ABTS	2,2'-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt
Ac	Acetyl
ACN	Acetonitrile
AcOH	Acetic acid
ADMET	Adsorption, distribution, metabolism, elimination, toxicity
ALD	Atomic layer deposition
AllBr	Allyl bromide
Ar	Aromatic
AUC	Area under curve
$BF_3 \cdot Et_2O$	Boron trifluoride ethyl etherate
BioFET	Biosensor based on field-effect transistors
Bn	Benzyl
BnBr	Benzyl bromide
Boc	*tert*-Butyloxycarbonyl
BOX	Buried oxide
BSA	Bovine serum albumin
Bz	Benzoyl
BzCl	Benzoyl chloride
Calcd	Calculated
Cer	Ceramide
CES	Carboxylesterase
CFU	Colony-forming unit
CIP	Ciprofloxacin
CL_{tot}	Total clearance
C_{max}	Maximal concentration
CMOS	Complementary metal-oxide semiconductor
COMU	1-[(1-(Cyano-2-ethoxy-2-oxoethylideneaminooxy)dimethylamino morpholinomethylene)]methanaminiumhexafluorophosphate
ConA	Concanavalin A
Contd	Continued
COSY	Correlation spectroscopy
Cpd	Compound
CRD	Carbohydrate recognition domain
CTD	C-terminal Domain
DCE	Dichloroethane
DCM	Dichloromethane
DI	Distilled
DIC	*N,N'*-Diisopropylcarbodiimide
DIPEA	Diisopropylethylamine
DMAP	4-(Dimethylamino)pyridine
DMBA	1,3-dimethylbarbituric acid

DME	Dimethoxyethane
DMEM	Dulbecco's modified eagle medium
DMF	*N*,*N*-dimethylformamide
DMSO	Dimethylsulfoxide
$DMSO_{d6}$	Hexadeuterodimethyl sulfoxide
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's phosphate buffer saline
Dppf	1,1'-Bis(diphenylphosphino)ferrocene
DSC	Donor-strand Complementation
DSE	Donor-strand Exchange
EA	Ethanolamine
EBL	Electron beam litography
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	Ethylenediaminetetraacetic acid
ELSD	Evaporative light scattering
EM	Electron microscope
ESI-MS	Electrospray ionization mass spectrometry
Et_2O	Diethyl ether
EtOAc	Ethyl acetate
EtOH	Ethanol
FBDD	Fragment-based drug design
FBS	Fetal bovine serum
FCS	Fetal calf serum
FET	Field effect transistor
FITC	Fluorescein isothiocyanante
FP	Fluorescence polarization
FSC	Forward scatter
Gal	D-galactose
GalNAc	D-*N*-acetylgalactosamine
GbO3	Globotriasyl ceramide
GbO4	Globotetraosyl ceramide
GbO5	Globopentaosyl ceramide
Glc	D-glucose
HBTU	*N*,*N*,*N*',*N*'-Tetramethyl-*O*-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
$HEPES_{d18}$	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid-d_{18}
HIV	Human immunodeficiency virus
HM	Heptyl α-D-mannoside
HMBC	Heteronuclear multiple-bond correlation
HOBt	1-hydroxybenzotriazole
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single quantum coherence
IBC	Intracellular bacterial colony
IC_{50}	Median inhibition cocentration
IPGT	Isopropyl β-D-thiogalactopyranoside
ISFET	Ion-sensitive field effect transistor
ITC	Isothermal titration calorimetry
K_D	Equilibrium dissociation constant
k_{off}	Dissociation rate constant

k_{on}	Association rate constant
LB	Luria-Bertani
LC-MS	Liquid chromatography - mass spectrometry
LOD	Limit of detection
MAC_{90}	Minimal antiadhesion concentration to inhibit 90% adhesion
MAG	Myelin-associated glycoprotein
Man	D-mannose
MBP	Mannose-binding protein
MDR	Multidrug resistant
MEM-NEAA	Minimum essential medium - non-essential amino acids
MeOH	Methanol
MeONa	Sodium methoxide
MFI	Mean fluorescence intensity
MHC	Major histocompatibility complex
MHDA	ω-Mercaptohexadecanoic acid
MOPS	3-(*N*-morpholino)propanesulfonic acid
MPD	Methyl-2,4-pentanediol
MPLC	Medium pressure liquid chromatography
MWCO	Molecular weight cutoff
n-BuLi	*n*-butyllithium
NADPH	Nicotinamide adenine dinucleotide phosphate
NHS	*N*-hydroxysuccinimide
Ni-NTA	Nickel-nitrilotriacetic acid
NIS	*N*-iodosuccinimide
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NTe	N-terminal Domain
OD_{600}	Optical Density at 600 nm
PAA	Polyacrylamide
PAMPA	Parallel artificial membrane permeation assay
$PapG\text{-}II_{LD}$	Lectin domain of PapG-II protein
P_{app}	Apparent permeability
PBS	Phosphate buffer saline
PD	Pharmacodynamic
Pd/C	Palladium on charcoal
PDB	Protein data bank
PDMS	Polydimethylsiloxane
PE	Petrol ether
P_e	Effective permeability
PEG	Polyethylene glycol
PK	Pharmacokinetic
PMB	4-Methoxybenzyl
Po	Per os
PPB	Plasma protein binding
PRE	Paramagnetic relaxation enhancement
PTFE	Polytetrafluoroethylene
PyBOP	(Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate)
QSAR	Quantitative structure-activity relationship
r.t.	Room temperature

rIC_{50}	Relative median inhibition cocentration
RLM	Rat liver microsomes
RP-C18	Reverse phase silica gel
SAM	Self-assembled monolayer
SAR	Structure-activity relationship
Satd	Saturated
SBDD	Structure-based drug design
Sc	Subcutaneous
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SGG	Sialosyl galactosyl globoside
SiNR	Silicon nanoribbons
SiNW	Silicon nanowire
SMX	Sulfamethoxazole
SNR	Signal-to-noise ratio
SOI	Silicon on insulator
SPR	Surface plasmon resonance
SSC	Side scatter
STD	Saturation transfer difference spectroscopy
$t_{1/2}$	Half-life
TBABr	Tetrabutylammonium bromide
TBAF	Tetrabutylammonium fluoride
TBDMS	*tert*-Butyldimethylsilyl
TBDPSCl	*tert*-Butyldiphenylsilyl Chloride
TCR	T-cell Receptor
TEER	Transepitelial electrical resistance
TEMPO	2,2,6,6-Tetramethylpiperidine-*N*-oxyl
TEMPO-COOH	4-Carboxy-2,2,6,6-tetramethylpiperidine-1-oxyl
TFA	Trifluoroacetic acid
TfN_3	Trifluoromethanesulfonyl azide
THF	Tetrahydrofuran
TLC	Thin layer chromatography
TMAH	Tetramethylammonium hydroxide
TMP	Trimethoprim
TMS	Trimethylsilyl
TMSOTf	Trimethylsilyl trifluoromethanesulfonate
TOCSY	Total correlation spectroscopy
Triflyl	Trifluoromethanesulfonyl
Tris	2-Amino-2-(hydroxymethyl)-1,3-propanediol
TsOH/pTsOH	*p*-Toluene sulfonic acid
TSP_{d4}	2,2,3,3-Tetradeutero 3-(Trimethylsilyl)propionic acid sodium salt
UPEC	Uropathogenic *Escherichia coli*
UPIa	Uroplakin Ia
UTI	Urinary tract infection
UV	Ultraviolet light
V_z	Volume of distribution in terminal phase

Table of Contents

1. Urinary Tract Infections (UTIs)

1.1 Introduction

The urinary tract infections (UTI) are among the most common bacterial infections and are classified as lower or upper, depending on the affected organs (urethra and the urinary bladder or the kidneys), and either as uncomplicated or complicated. Uncomplicated UTIs concern non-pregnant, non-instrumented hosts, without structural or functional abnormalities of the urinary tract. All other UTIs are considered complicated [1].

UTIs affect prevalently women. It was estimated that about 60% of all women experience a UTI, with 5% of them suffering recurrent episodes at some point during their life [2,3]. Among young women with a first UTI, 21% suffer a second infection within 6 months [4]. In men, UTIs are most often associated with other comorbidities or abnormalities [5]. Most UTIs resolve within a few days and are rarely associated with long-term sequelae [6-8]. However, the very high incidence and the propensity to recur result in very high annual direct and indirect costs (1.6 billion dollars only in the USA, in 1995 [3]). Moreover, the recurrence in some patients contributes importantly to the reduction of life quality.

The most common cause of uncomplicated UTI are uropathogenic *E. coli* (UPEC), which in otherwise healthy women aged 18-39 years accounts for 80% of infections [9]. Complicated UTIs are caused by a broader spectrum of microorganism, due to host factors that can allow less virulent organisms to colonize the urinary tract [10].

Common ways to induce a UTI are the movement of pathogens from the gut [11] or vagina [10] to the urethra, sexual direct or indirect transmission, and at least in case of nosocomial infections, contact with contaminated surfaces (e.g. personnel's hands) [1,12]. Some studies suggest the possibility of water, food or person-to-person transmission during travels [13].

The treatment of UTIs relies intensively on antibiotics [1]. Due to the very high incidence, acute uncomplicated cystitis is still the most common reason for antibiotic prescription [14-16]. Considering also the recurrent nature of UTIs, it is not surprising that their treatment plays an important role in selecting resistant organisms, not only among uropathogenic ones, but also in the normal macrobiota [17,18].

Despite the rather large arsenal of antibiotics, the rapid spread of resistance requires innovative solutions [19,20]. Vaccines are emerging as a viable alternative, although hurdles as pathogens heterogeneity and low immunogenicity have to be addressed. A different and appealing strategy involves the inhibition of pathogen's adhesion to the host. Uropathogenic *E. coli* adhere to target tissues thanks to organelles called pili [21]. Type 1 pili, expressing the FimH adhesin, are playing a crucial role in bladder invasion [22] whereas the P pili, expressing the PapG adhesin are highly correlated with kidney infections [23-26]. Therefore, two promising therapeutic strategies are currently under development, namely the inhibition of pilus biogenesis with small molecules (pilicides) [27-30], and the inhibition of the pilus adhesive properties (antiadhesives) [31].

2. The FimH Adhesin and Its Role in UTIs

2.1 Introduction

2.1.1 Infection Cycle

The infection cycle has been characterized in animal models of UTI and in human tissue [32,33]. The type 1 pili, expressed in almost every isolate from patients with a UTI [34], are involved in the initial step of the infection. They bind to mannosylated glycoproteins on the surface of the bladder epithelium. The main receptor in humans is uroplakin-Ia (UPIa) [35]. The strong adhesion to the epithelium prevents bacteria from being washed off by the urine flow and is followed by invasion of the urothelial cells [36]. Once inside the cells, bacteria start to replicate and form intracellular bacterial communities (IBCs), biofilm-like structures that protect them from the host's immune response [32]. Although cell-death and exfoliation reduce the number of infecting bacteria, this mechanism also leaves uncovered the immature cells in the deeper layers of the urothelium, which can also be invaded [32].

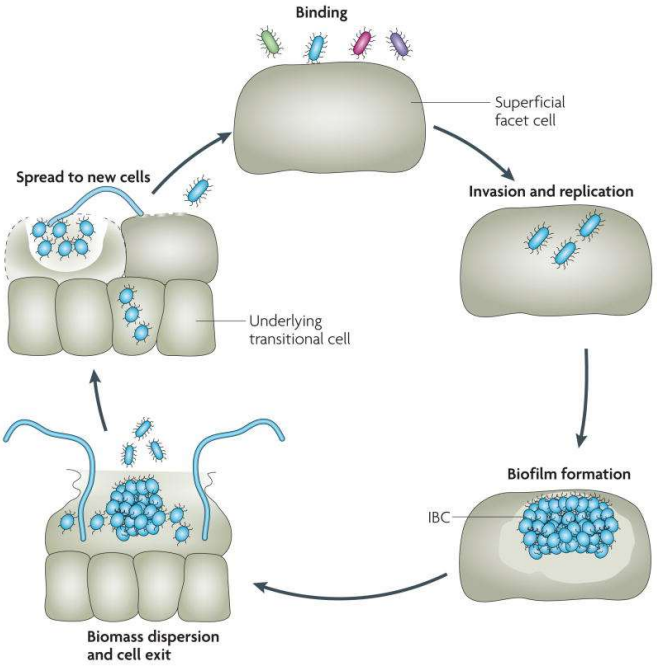


Figure 1. The infection cycle of uropathogenic E. coli in the urinary bladder (adopted from ref. [103] with permission).

After sufficient maturation the IBCs can evade from the cells and further disseminate infective bacteria. Single rod-shaped cells as well as filamentous colonies are released [36]. Part of the bacteria form quiescent colonies - intrinsically less sensitive to antibiotics - that are probably responsible for recurrent infections (figure 1) [36].

2.1.2 Type 1 pilus

The type 1 pilus (fimbria) is a highly specialized and efficient adhesion organelle. On their surface, bacteria commonly bear 200 to 500 type 1 pili [37]. From a structural point of view, a pilus is composed of a rod and a tip fibrillum; the former consists of 500-3000 copies of the protein FimA arranged in a right-handed helix, the latter of one copy of the protein FimH and several copies of the subunits FimG and FimF. The whole structure measures about 7 nm in diameter and 1-2 µm in length (figure 2) [38,39]. The type-1 pilus biogenesis follows the chaperone/usher pathway [40-43]. Each subunit shows an incomplete immunoglobuline-like fold (the so-called "pilin" fold), in which the 7th, C-terminal β-sheet is missing, leaving a hydrophobic pocket uncovered. In the assembled pilus, this space is filled by the N-terminal domain (Nte) of the following subunit [42], in a process called Donor Strand Complementation (DSC) [44,45]. The whole process of pilus assembly has been recently reviewed [43].

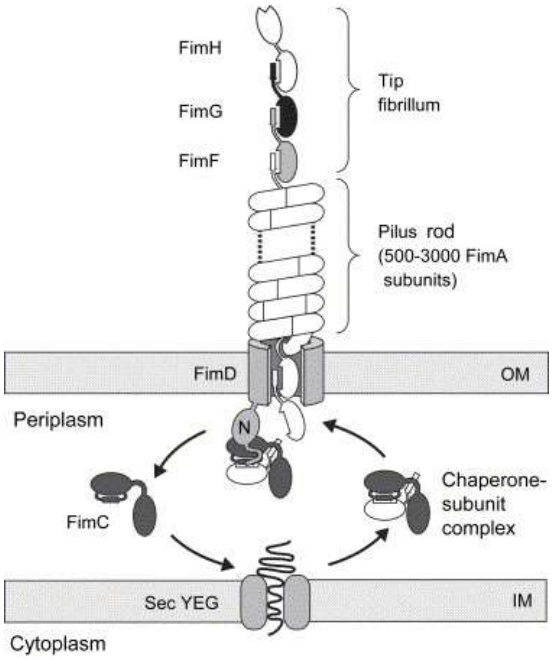


Figure 2. Representation of the type 1 pilus and its assembly. Top part: assembled pilus. Bottom part: chaperone-usher machinery (adopted from ref. [39] with permission).

2.1.3 FimH Adhesin

The FimH adhesin at the tip of type 1 pili is responsible for the adhesion to the host's tissues [46]. It consists of two domains presenting Ig-like folds, the so-called lectin domain and pilin domain (figure 3).

The former is responsible for the adhesive properties; the latter connects FimH to the FimG subunit [45].

The crystal structure of the FimH:FimC complex was published in 1999 [45], and shortly after the first FimH:FimC:mannoside complex, which furnished valuable information on the binding site [47]. FimC was necessary to keep FimH in the active conformation. A step forward was made when the lectin domain of FimH alone — a truncated version of FimH locked in the high-affinity state — was crystallized [48], facilitating the discovery of high-affinity antagonists.

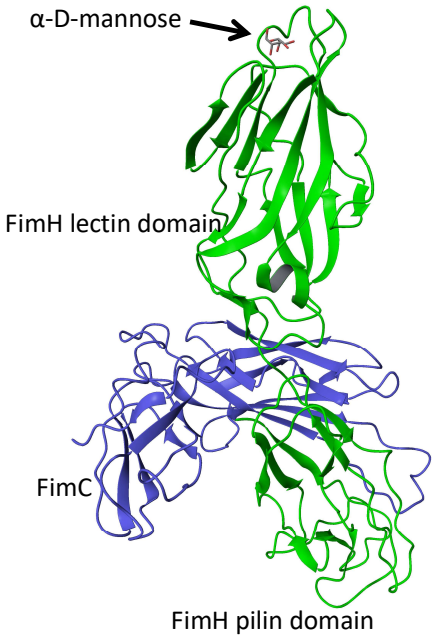


Figure 3. Crystal structure of the FimH:FimC:mannose ternary complex (PDB code 1KLF) [47]. FimH is represented in green, FimC in orchid. The mannose is depicted in sticks, with oxygen atoms in red and carbon atoms in gray.

The main natural target of FimH in the human urinary tract are uroplakins, especially UPIa [35,49], on which a heterogeneous pattern composed of moieties ranging from Man6GlcNAc2 to Man9GlcNAc2 residues was evidenced [35]. Other targets are the surface-expressed integrins α3 and β1 [50,51], and the Tamm-Horsfall protein (THP) [52], also decorated with high-mannose glycans. The latter is a soluble glycoprotein, which acts as a natural antagonist of FimH [53,54].

The FimH adhesin is able to increase its affinity for mannosylated substrates under tensile mechanical force [55], a property called "catch-bond" behavior [56]. The details of the allosteric regulation mechanism were recently disclosed by solving the crystal structure of the full-length FimH protein in the context of the whole fimbrial tip [57].

The majority of E. coli expresses low-affinity FimH adhesins presenting catch-bond behavior, suggesting an evolutionary advantage [58]. Probably this property favors rapid host colonization and bacterial survival. It was suggested that in fact low-affinity variants could be resistant to soluble antagonists [59] and able to colonize surfaces more rapidly [60].

2.1.4 FimH antagonists

The idea of blocking the interaction between the FimH adhesin and its natural target inspired a number of medicinal chemistry studies. Owing to the relatively high affinity (millimolar–micromolar range) of the rather simple methyl α-D-mannoside for FimH [48,61,62], most efforts were directed towards the modification of the aglycone part. Already in the 1980s, aromatic aglycones were known to enhance the inhibitory potency of mannose on yeast agglutination by E. coli, by a factor of 400 to 1000 [63]. In 2005, Bouckaert et al. reported the 5 nM antagonist n-heptyl mannopyranoside [48]. Later, the higher potential of mannopyranosides with extended aromatic portions was recognized [64], resulting in a number of biphenyl [64-66], indolinylphenyl [67], and squaric acid derivatives [68-70].

2.1.5 Fluorescence Polarization

The first report on fluorescence polarization in liquid samples was published by Perrin in 1926 [71]. When polarized light is passed through a solution of a fluorescent molecule, the solution emits only partially polarized light. The depolarization is due to a number of factors, mainly rotational diffusion and fluorescence lifetime. Molecules with high rotational diffusion during the lifetime of the fluorescence will lose polarization faster. It follows that small and flexible molecules will experience more depolarization than larger and stiffer ones [72]. The polarization P is expressed as:

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \qquad \text{eq. 1}$$

The measured polarization is related to the tumbling of the fluorescent molecules in solution and can be used to determine interactions of small ligands with large receptors. A fluorescent small molecule probe in solution exhibits low polarization value. Upon binding to a macromolecular target (i.e. a protein), the value increases, as

a consequence of the slower tumbling. A great advantage of FP over other traditional binding assays (radiolabeling, ELISA, etc.) is that no separation step and no immobilization is needed. The measurement is done at the equilibrium, the crucial parameter being the fraction of fluorescent ligand bound to the target. The mathematical resolution of free from bound fluorescent probe was derived by Weber in 1952 [73], based on the additivity of the polarization of different species:

$$\left(\frac{1}{P_{obs}} - \frac{1}{3}\right)^{-1} = \sum f_i \left(\frac{1}{P_i} - \frac{1}{3}\right)^{-1} \qquad \text{eq. 2}$$

However, changes in the intensity of the fluorescence emitted by the fluorophore in the free and bound states must be corrected by the term g, as in equation 3 [74]:

$$f_b = \frac{(3-P_b)(P_{obs}-P_f)}{(3-P_{obs})(P_b-P_f)+(g-1)(3-P_f)(P_b-P_{obs})} \qquad \text{eq. 3}$$

The value for fb can then be used to obtain the KD for the ligand:target interaction:

$$f_b = \frac{[\text{Protein}]}{[\text{Protein}]+K_D} \qquad \text{eq. 4}$$

A plot of fb vs. concentration of protein can be fitted with equation 4. The FP measurement can be designed as competitive binding assay, in which a known fluorescently-labeled molecule is titrated with a competitor ligand [75,76]. The obvious advantage lies in the possibility to screen a number of competitors with a single fluorescently-labeled, known binder. Thanks to the improvements in instrumentation, FP assays are now widely applied in medicinal chemistry and life sciences, and offer the advantages of high-throughput, homogeneity ("mix and measure assay"), automation, reliability, reproducibility [72]. Because of that, we envisioned the application of a competitive FP assay for screening FimH antagonists [75]. For the establishment of a competitive FP assay, several parameters must be considered.

2.1.5.1 Probe design

The establishment of a competitive FP assay starts from the design of a suitable fluorescently-labeled binder. First of all, the fluorescent lifetime τ must be long enough for observing a difference between bound and free states (equation 5) [72].

$$\frac{1}{P} - \frac{1}{3} = \left(\frac{1}{P_0} - \frac{1}{3}\right)\left(1 + \frac{3\tau}{\rho}\right) \qquad \text{eq. 5}$$

Secondly, the fluorescent probe should not retain too much flexibility when bound to the test molecule, because residual local motion results in lower polarization. In the case of a competitive FP assay, the probe must be linked to a known binder, without changing its binding mode and ideally without influencing the affinity [72,77]. Moreover, neither the linking, nor the complex formation with the target should result in complete fluorescence quenching.

2.1.5.2 Assay design

After a suitable reporter compound has been designed, appropriate assay conditions have to be identified. This involves finding the proper probe and protein concentrations, buffer composition, plate type, etc. [76,77]. In general, it is important that the probe concentration does not exceed too much 2KD, to avoid stoichiometric titration, and that the fraction of bound probe is in the range 0.5-0.8 [76].
The assay we developed for the screening of FimH antagonist is detailed in paper 2.

2.1.6 Nanowires-based Field-Effect Biosensors

The sensitive and specific detection and quantification of biological and chemical species are crucial in a number of fields, including research in life science, healthcare, and medicinal chemistry. Most standard techniques involve optical readouts of fluorescently-labeled analytes [78,79]. As this commonly involves sample manipulation and often target labeling, i.e. a chemical modification that could influence target's properties, these methods are not ideal [80].
An alternative approach is the use of biosensors [81], devices that are "capable of providing specific quantitative or semi-quantitative information using a biological recognition element retained in direct spatial contact with an electrochemical transduction element" [82]. Field-effect based biosensors (BioFETs) based on silicon nanowires are promising tools for label-free detection of biomolecules, as they have shown high sensitivity and scalability. Thanks to the recent advances in the fabrication processes, device characteristics can be controlled at the micro/nano-scale, thus achieving multiplexing and selective addressing. Analytical devices relying on

this technology could perform point-of-care multicomponent analysis with minimal sample consumption.

A BioFET consists of the following parts: a source (S), a drain (D), and a channel connecting S and D, composed of doped semiconducting materials (traditionally, S and D have opposite doping than the channel, but this is not always the case, as shown in manuscript 2); a gate insulator covering the channel and composed of SiO2, Si_3N_4, Al_2O_3, or other insulating materials, which contacts the liquid sample; a gating circuit, connected to a reference electrode; a functionalization layer, consisting of a biomolecule and interacting with the analyte [83]. A BioFET is therefore conceptually an ion-sensitive field-effect transistor (ISFET) interfaced with a biomolecule that recognizes biological analytes. In an ISFET, when a sufficient voltage is applied to the gate, an inversion layer is induced in the channel, thus influencing its conductance and therefore the current flowing from S to D (figure 4). The presence of charges close to the surface can modify the surface potential (and consequently the number of charge transporters in the inversion layer), generating a measurable electric signal. It follows that every (bio-)chemical reaction that modifies the charges at the interface gate oxide/liquid sample can in theory be measured by an ISFET interfaced with a bioreceptor. Many examples of BioFETs are described in literature, ranging from enzyme- (EnFET) to antibody-modified sensors (ImmunoFET), and from DNA- (GenFET) to cell- and even beetle-modified ones [83]. More recently, sensors based on nano-sized structures, such as nanowires [84] (SiNW-BioFET) and nanoribbons [85] (SiNR-BioFET) have been produced. The main advantages of this new generation of sensors are the readily conceivable massive production [86], the ultra-high sensitivity, and the possibility of multiplexing [87-90]. Real-time and label-free detection of DNA [91-94], RNA [95], proteins [96] (including cancer markers [97]), and viruses [98,99] has been demonstrated.

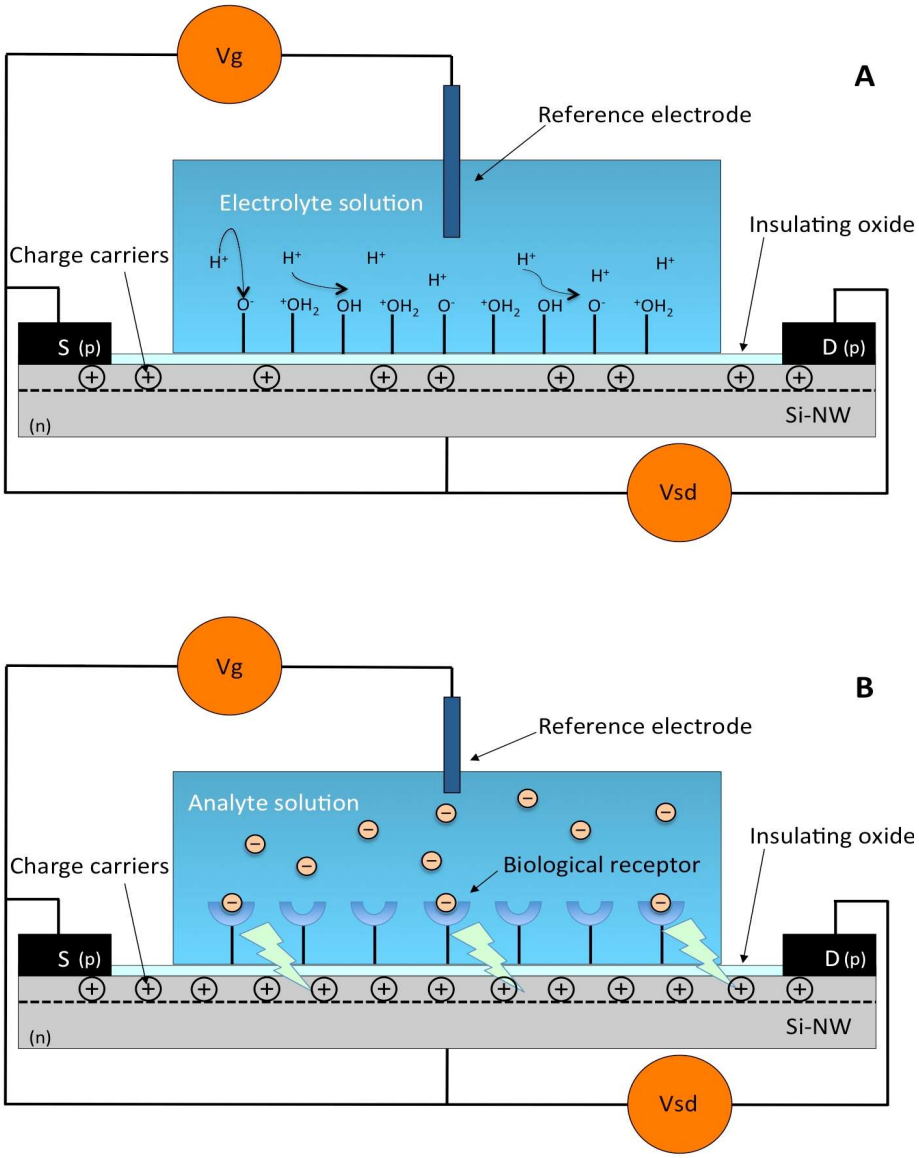


Figure 4. The schemes of a pH-sensitive ISFET (A) and a BioFET (B) with n-type channels and p-type source and drain are depicted, working in inversion mode. The binding of positively charged species reduces the number of charge carriers, thus reducing the conductance (A). Negatively charged species have opposite effect (B).

So far, studies on quantifying binding affinities and kinetic data have primarily focused on DNA [91] and biotin-streptavidin interactions [100]. However, as a model system for protein-small molecule interactions and for the benchmarking of the limit of detection (LOD), the use of the biotin-streptavidin pair is questionable, due to the extremely strong binding (dissociation constant $K_D \approx 10{-}14$ M, one of the strongest non-covalent interactions known in nature) [101].

In our work (manuscript 2), we demonstrate real-time, label-free detection of the clinically relevant protein FimH with gold-coated silicon nanoribbons (SiNR). SiNR

are easier than SiNW to produce and have shown comparable sensitivity [85]. Our results are an important step toward the study of protein-ligand interactions by FET nanosensors, and demonstrate that SiNR-BioFETs are excellent candidates to compete with surface plasmon resonance, the golden standard for such application [102].

2.1.7 Aims of this Project

This project is aiming at the development of FimH antagonists with excellent pharmacodynamic and pharmacokinetic profiles for treating UTIs in humans. This thesis' aims in the context of this project were: 1) the contribution to the kinetic characterization of FimH interaction with FimH antagonists (paper 1); 2) the design of a fluorescently-labeled compound with suitable properties for FP-based high-throughput screening of new FimH antagonists (paper 2) and which was used for the characterization of FimH high- and low-affinity states (paper 3); 3) the demonstration of the proof-of-concept of direct protein measurement on nanoribbons-based sensors, using FimH as analyte and a FimH antagonist as recognition element (paper 4).

References

[1] B. Foxman. The Epidemiology of Urinary Tract Infection. *Nature Reviews* **2010**, *7*(12), 653-660.

[2] L. E. Nicolle. Uncomplicated urinary tract infection in adults including uncomplicated pyelonephritis. *Urol. Clin. North Am.* **2008**, *35*(1), 1-12.

[3] B. Foxman, R. Barlow, H. D'Arcy, B. Gillespie, J. D. Sobel. Urinary tract infection: self-reported incidence and associated costs. *Ann. Epidemiol.* **2000**, *10*(8), 509-515.

[4] B. Foxman, B. Gillespie, J. Koopman, L. Zhang, K. Palin, P. Tallman, J. V. Marsch, S. Spear, J. D. Sobel, M. J. Marty, C. F. Marrs. Risk factors for second urinary tract infection among college women. *Am J Epidemiol* **2000**, *151*, 1194–205.

[5] E. Hummers-Pradier, A. M. Ohse, M. Koch, W. R. Heizmann, M. M. Kochen. Urinary tract infection in men. *Int. J. Clin. Pharm. Ther.* **2004**, *42*(7), 360-366.

[6] B. Foxman. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Am. J. Med.* **2002**, *113*(1), 5S – 13S.

[7] T. C. Christiaens, M. De Meyere, G. Verschraegen, W. Peersman, S. Heytens, J. M. De Maeseneer. Randomised controlled trial of nitrofurantoin versus placebo in the treatment of uncomplicated urinary tract infection in adult women. *Br. J. Gen. Pract.* **2002**, *52*(482), 729-734.

[8] S. A. Ferry, S. E. Holm, H. Stenlund, R. Lundholm, T. J. Monsen. Clinical and bacteriological outcome of different doses and duration of pivmecillinam compared with placebo therapy of uncomplicated lower urinary tract infection in women: the LUTIW project. *Scand. J. Prim. Health Care* **2007**, *25*(1), 49-57.

[9] W. E. Stamm. Scientific and clinical challenges in the management of urinary tract infections. *An. J. Med.* **2002**, *113*(Suppl 1A), 1S-4S.

[10] R. Allan. The etiology of urinary tract infection: traditional and emerging pathogens. *Dis. Mon.* **2003**, *49*(2), 71-82.

[11] G. S. Goetz, G. S. Mahmood, S. J. Hultgren, M. J. Engle, K. Dodson, D. H. Alpers. Binding of pili from uropathogenic Escherichia coli to membranes secreted by human colonocytes and enterocytes. *Infect. Immun.* **1999,** *67*(11), 6161-6163.

[12] D. R. Schaberg, R. A. Weinstein, W. E. Stamm. Epidemics of nosocomial urinary tract infection caused by multiply resistant Gram-negative bacilli: epidemiology and control. *J. Infect. Dis.* **1976**, *133*(3), 363-366.

[13] J. T. Freeman, S. J. McBride, H. Heffernan, T. Bathgate, C. Pope, R. B. Ellis-Pegler. Community-onset genitourinary tract infection due to CTX-M-15-producing *Escherichia coli* among travelers to the Indian subcontinent in New Zealand. *Clin. Infect. Dis.* **2008**, *47*(5), 689-692.

[14] K. Gupta, T. M. Hooton, K. G. Naber, B. Wullt, R. Colgan, L. G. Miller, G. J. Moran, L. E. Nicolle, R. Raz, A. J. Schaeffer, D. E. Soper. International clinical practice guidelines for the treatment of acute uncomplicated cystitis and pyelonephritis in women: A 2010 update by the Infectious Diseases Society of America and the European Society for Microbiology and Infectious Diseases. *Clin. Infect. Dis.* **2011**, *52*(5), e103–e120.

[15] M. Totsika,D. G. Moriel. Uropathogenic Escherichia coli mediated urinary tract infection. *Curr. Drug Targets* **2012**, *13*(11), 1386-1399.

[16] A. E. Barber, J. P. Norton, A. M. Spivak, M. A. Mulvey. Urinary tract infections: current and emerging management strategies. *Clin. Infect. Dis.* **2013**, *57*(5), 719–24.

[17] K. Gupta, D. Scholes, W. E. Stamm. Increasing prevalence of antimicrobial resistance among uropathogens causing acute uncomplicated cystitis in women. *JAMA* **1999**, 281(8), 736-738

[18] G. C. Schito, K. G. Naber, H. Botto, J. Palou, T. Mazzei, L. Gualco, A. Marchese. The ARESC study: an international survey on the antimicrobial resistance of pathogens involved in uncomplicated urinary tract infections. *Int. J. Antimic. Ag.* **2009**, *34*(5), 407-413.

[19] A. R. Brumbaugh, H. L. T. Mobley. Preventing urinary tract infection: progress toward an effective Escherichia coli vaccine. *Expert Rev. Vaccines* **2012**, *11*(6), 663-676.

[20] D. G. Moriel, M. A. Schembri. Vaccination approaches for the prevention of urinary tract infection. *Curr. Pharm. Biotec.* **2013**, *14*(11), 976-974.

[21] J. R. Johnson Virulence factors in Escherichia coli urinary tract infection. *Clin. Microb. Rev.* **1991**, *4*(1), 80-128.

[22] M. A. Mulvey, J. D. Schilling, J. J. Martinez, S. J. Hultgren. Bad bugs and beleaguered bladders: interplay between uropathogenic Escherichia coli and innate host defenses. *P. Natl. Acad. Sci. USA* **2000**, *97*(16), 8829-8835.

[23] I. M. Johanson, K. Plos, B. I. Marklund, C. Svanborg. *Pap, papG* and *prsG* DNA sequences in *Escherichia coli* from the fecal flora and the urinary tract. *Microb. Pathogenesis* **1993**, *15*(2), 121-129.

[24] J. A. Roberts, B. I. Marklund, D. Ilver, D. Haslam, M. B. Kaack, G. Baskin, M. Louis, R. Möllby, J. Winberg, S. Normark. The Gal(alpha 1-4)Gal-specific tip adhesin of Escherichia coli P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. *P. Natl. Acad. USA* **1994**, *91*(25), 11889-11893.

[25] C. C. Tseng, J. J. Huang, M. C. Wang, A. B. Wu, W. C. Ko, W. C. Chen, J. J. Wu. PapG II adhesin in the establishment and persistence of *Escherichia coli* infection in mouse kidneys. *Kidney int.* **2007**, *71*(8), 764-770.

[26] J. R. Johnson. *papG* Alleles among *Escherichia coli* Strains Causing Urosepsis: Associations with Other Bacterial Characteristics and Host Compromise. *Infect. Immun.* **1998**, *66*(9), 4568-4571.

[27] J. S. Pinkner, H. Remaut, F. Buelens, E. Miller, V. Aberg, N. Pemberton, M. Hedenström, A. Larsson, P. Seed, G. Waksman, S. J. Hultgren, F. Almqvist. Rationally designed small compounds inhibit pilus biogenesis in uropathogenic bacteria. *P. Natl. Acad. Sci. USA* **2006**, *103*(47), 17897-17902.

[28] L. Cegelski, J. S. Pinkner, N. D. Hammer, C. K. Cusumano, C. S. Hung, E. Chorell, V. Åberg, J. N. Walker, P. C. Seed, F. Almqvist, M. R. Chapman, S. J. Hultgren. Small-molecule inhibitors target *Escherichia coli* amyloid biogenesis and biofilm formation. *Nat. Chem. Biol.* **2009**, *5*(12), 913-919.

[29] E. Chorell, C. Bengtsson, T. Sainte-Luce Banchelin, P. Das, H. Uvell, A. K. Sinha, J. S. Pinkner, S. J. Hultgren, F. Almqvist. Synthesis and application of a bromomethyl substituted scaffold to be used for efficient optimization of anti-virulence activity. *Europ. J. Med. Chem.* **2011**, *46*(4), 1103-1116.

[30] R. Piatek, B. Zalewska-Piatek, K. Dzierzbicka, S. Makowiec, J. Pilipczuk, K. Szemiako, A. Cyranka-Czaja, M. Wojciechowski. Pilicides inhibit the FGL chaperone/usher assisted biogenesis of the Dr fimbrial polyadhesin from uropathogenic *Escherichia coli. BMC Microb.* **2013**, *13*, 131-142.

[31] M. Aronson, O. Medalia, L. Schori, D. Mirelman, N. Sharon, I. Ofek. Prevention of colonization oft he urinary tract of mice with Escherichia coli by blocking of bacterial adherence with methyl-α-D-mannopyranoside. *J. Infect. Dis.* **1979**, *139*(3), 329-332.

[32] G. G. Anderson, K. W. Dodson, T. M. Hooton, S. J. Hultgren. Intracellular bacterial communities of uropathogenic Escherichia coli in urinary tract pathogenesis. *Trends Microb.* **2004**, *12*(9), 424-430.

[33] J. J. Martinez, M. A. Mulvey. Type 1 pilus-mediated bacterial invasion of bladder epithelial cells. *EMBO* **2000**, *19*(12), 2803-2812.

[34] C. K. Garofalo, T. M. Hooton, S. M. Martin, W. E. Stamm, J. J. Palermo, J. I. Gordon, S. J. Hultgren. Escherichia coli from urine of female patients with urinary tract infections is competent for intracellular bacterial community formation. *Infect. Immun.* **2007**, *75*(1), 52-60.

[35] B. Xie, G. C. Zhou, C. Shiu-Yung, E. Shapiro, X. P. Kong, X. R. Wu, T. T. Sun, C. E. Costello. Distinct glycan structures of uroplakins Ia and Ib: structural basis for the selective binding of FimH adhesin to uroplakin Ia. *J. Biol. Chem.* **2006**, *281*(21), 14644-14653.

[36] S. S. Justice, C. Hung, J. A. Theriot, D. A. Fletcher, G. G. Anderson, M. J. Footer, S. J. Hultgren. Differentiation and developmental pathways of uropathogenic Escherichia coli in urinary tract pathogenesis. *P. Natl. Acad. Sci. Usa* **2004**, *101*(5), 1333-1338.

[37] N. Sharon. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Bioch. Bioph. Acta* **2006**, *1760*(4), 527-537.

[38] E. Hahn, P. Wild, U. Hermanns, P. Sebbel, R. Glockshuber, M. Häner, N. Taschner, P. Burkhard, U. Aebi, S. A. Müller. Exploring the 3D Molecular Architecture of Escherichia coli Type 1 Pili. *J. Mol. Biol.* **2002**, *323*(5), 845-857.

[39] G. Capitani, O. Eidam, R. Glockshuber, M. G. Grütter. Structural and functional insights into the assembly of type 1 pili from *Escherichia coli. Microbes Infect*. **2006**, *8*(8), 2284-2290.

[40] G. Waksman, S. J. Hultgren. Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat. Rev. Microbiol.* **2009**, *7*(11), 765-774.

[41] D. L. Hung, S. J. Hultgren. Pilus biogenesis via the chaperone/usher pathway: an integration of structure and function. *J. Struct. Biol.* **1998**, *124*(2-3), 201-220.

[42] F. G. Sauer, M. Barnhart, D. Choudhury, S. D. Knight, G. Waksman, S. J. Hultgren. Chaperone-assisted pilus assembly and bacterial attachment. *Curr. Opin. Struct. Biol.* **2000**, *10*(5), 548-556.

[43] S. Geibel, G. Waksman. The molecular dissection of the chaperone-usher pathway. *Curr. Opin. Struct. Biol.* **2014**, *1843*(8), 1559-1567.

[44] F. G. Sauer, K. Futterer, J. S. Pinkner, K. W. Dodson, S. J. Hultgren, G. Waksman. Structural basis of chaperone function and pilus biogenesis. *Science* **1999**, *285*(5430), 1058–1061.

[45] D. Choudhury, A. Thompson, V. Stojanoff, S. Langermann, J. Pinkner, S. J. Hultgren, S. D. Knight. X-ray structure of the FimC–FimH chaperone–adhesin complex from uropathogenic Escherichia coli. *Science* **1999**, *285*(5430), 1061–1066.

[46] K. A. Krogfelt, H. Bergmans, P. Klemm. Direct evidence that the FimH protein is the mannose-specific adhesin of Escherichia coli type 1 fimbriae. *Infect. Immun.* **1990**, *58*(6), 1995-1998.

[47] C. S. Hung, J. Bouckaert, D. Hung, J. Pinkner, C. Widberg, A. DeFusco, C. G. Auguste, R. Strouse, S. Langermann, G. Waksman, S. J. Hultgren. Structural basis of tropism of Escherichia coli to the bladder during urinary tract infection. *Mol. Microb.* **2002**, *44*(4), 903-915.

[48] J. Bouckaert, J. Berglund, M. Schembri, E. De Genst, L. Cools, M. Wuhrer, C. S. Hung, J. Pinkner, R. Slättegård, A. Zavialov, D. Choudhury, S. Langermann, S. J. Hultgren, L. Wyns, P. Klemm, S. Oscarson, S. D. Knight, H. De Greve. Receptor binding studies disclose a novel class of high-affinity inhibitors of the Escherichia coli FimH adhesin. *Mol Microb.* **2005**, *55*(2), 441-455.

[49] W. R. Wu, T. T. Sun, J. J. Medina In vitro binding of type 1-fimbriated Escherichia coli to uroplakins Ia and Ib: relation to urinary tract infections. *P. Natl. Acad. Sci. Usa* **1996**, *93*(18), 9630-9635.

[50] D. S. Eto, T. A. Jones, J. L. Sundsbak M. A. Mulvey. Integrin-mediated host cell invasion by type 1-piliated uropathogenic Escherichia coli. *PLoS Pathog.* **2007**, *3*(7), e100.

[51] A. Litynska, E. Pochec, D. Hoja-Lukowicz, E. Kremser, P. Laidler, A. Amoresano, C. Monti. The structure of the oligosaccharides of alpha3beta1 integrin from human ureter epithelium (HCV29) cell line. *Acta Biochim. Pol.* **2002**, *49*(2), 491-500.

[52] J. Pak, Y. Pu, Z. T. Zhang, D. L. Hasty, X. R. Wu. Tamm-Horsfall Protein Binds to Type 1 Fimbriated Escherichia coli and Prevents *E. coli* from Binding to Uroplakin Ia and Ib Receptors. *J. Biol. Chem.* **2001**, *276*(13), 9924-9930

[53] M. D. Säemann, T. Weichhart, W. H. Hörl,G. J. Zlabinger. Tamm-Horsfall protein: a multilayered defence molecule against urinary tract infection. *Eur. J. Clin. Invest.* **2005**, *35*(4), 227-235.

[54] F. Serafini-Cessi, A. Monti, D. Cavallone. N-Glycans carried by Tamm-Horsfall glycoprotein have a crucial role in the defense against urinary tract diseases. *Glycoconj. J.* **2005**, *22*(7-9), 383-394.

[55] W. E. Thomas, E. Trintchina, M. Forero, V. Vogel, E. V. Sokurenko. Bacterial Adhesion to Target Cells Enhanced by Shear Force. *Cell* **2002**, *109*(7), 913 - 923.

[56] M. Dembo, D. C. Torney, K. Saxman, D. Hammer. The Reaction-Limited Kinetics of Membrane-to-Surface Adhesion and Detachment. *Proc. R. Soc. Lond. B Biol. Sc.* **1988**, *234*(1274), 55-83.

[57] I. Le Trong, P. Aprikian, B. A. Kidd, M. Forero-Shelton, V. Tchesnokova,P. Rajagopal, V. Rodriguez, G. Interlandi, R. Klevit, V. Vogel, R. E. Stenkamp, E. V. Sokurenko, W. E. Thomas. Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like beta sheet twisting. *Cell* **2010**, *141*(4), 645-655.

[58] E. V. Sokurenko, M. Feldgarden, E. Trintchina, S. J. Weissman, S. Avagyan, S. Chattopadhyay, J. R. Johnson, D. E. Dykhuizent. Selection footprint in the FimH adhesin shows pathoadaptive niche differentiation in Escherichia coli. *Mol. Biol. Ev.* **2004**, *21*(7), 1371-1383.

[59] L. M. Nilsson, W. E. Thomas, E. V. Sokurenko, V. Vogel. Elevated shear stress protects Escherichia coli cells adhering to surfaces via catch bonds from detachment by soluble inhibitors. *Appl. Env. Microbial.* **2006**, *72*(4), 3005-3010.

[60] B. N. Anderson, A. M. Ding, L. M. Nilsson, K. Kusuma, V. Tchesnokova, V. Vogel, E. V. Sokurenko, W. E. Thomas. Weak rolling adhesion enhances bacterial surface colonization. *J. Bacteriol.* **2007**, *189*(5), 1794-1802.

[61] S. Rabbani, X. Jiang, O. Schwardt, B. Ernst. Expression of the carbohydrate recognition domain of FimH and development of a competitive binding assay. Anal. *Biochem.* **2012**, *407*(2), 188-195.

[62] A. K. Horst, S. Kötter, T. K. Lindsorst, A. Ludwig, E. Brandt, C. Wagener. Binding inhibition of type 1 fimbriae to human granulocytes: a flow cytometric inhibition assay using trivalent cluster mannosides. *Med. Microbiol. Immunol.* **2001**, *190*(3), 145-149.

[63] N. Firon, S. Ashkenazi, D. Mirelman, I. Ofek, N. Sharon. Aromatic alpha-glycosides of mannose are powerful inhibitors of the adherence of type 1 fimbriated Escherichia coli to yeast and intestinal epithelial cells. *Infect. Immun.* **1987**, *55*(2), 472-476.

[64] Z. Han, J. S. Pinkner, B. Ford, R. Obermann, W. Nolan, S. A. Wildman, D. Hobbs, T. Ellenberger, C. K. Cusumano, S. J. Hultgren, J. W. Janetka. Structure-based drug design and optimization of mannoside bacterial FimH antagonists. *J. Med. Chem.* **2010**, *53*(12), 4779-4792.

[65] L. Pang, S. Kleeb, K. Lemme, S. Rabbani, M. Scharenberg, A. Zalewski, F. Schädler, O. Schwardt, B. Ernst. FimH antagonists: structure-activity and structure-property relationships for biphenyl α-D-mannopyranosides. *ChemMedChem* **2012**, *7*(8), 1404-1422.

[66] Z. Han, J. S. Pinkner, B. Ford, E. Chorell, J. M. Crowley, C. K. Cusumano, S. Campbell, J. P. Henderson, S. J. Hultgren, J. W. Janetka. Lead Optimization Studies on FimH Antagonists: Discovery of Potent and Orally Bioavailable Ortho-Substituted Biphenyl Mannosides. *J. Med. Chem.* **2012,** *55*(8), 3945-3959.

[67] X. Jiang, D. Abgottspon, S. Kleeb, S. Rabbani, M. Scharenberg, M. Wittwer, M. Haug, O. Schwardt, B. Ernst. Antiadhesion Therapy for Urinary Tract Infections—A Balanced PK/PD Profile Proved To Be Key for Success. *J. Med. Chem* **2012**, *55*(10), 4700-4713.

[68] O. Sperling, A. Fuchs, T. K. Lindhorst. Evaluation of the carbohydrate recognition domain of the bacterial adhesin FimH: design, synthesis and binding properties of mannoside ligands. *Org. Biom. Chem.* **2006**, *4*(21), 3913-3922.

[69] O. Sperling, M. Dubber, T. K. Lindhorst. Functionalization of oligosaccharide mimetics and multimerization using squaric diester-mediated coupling. *Carbohyd. Res.* **2007**, *342*(5), 696-703.

[70] C. Grabosch, M. Hartmann, J. Schmidt-Lassen, T. K. Lindhorst. Squaric Acid Monoamide Mannosides as Ligands for the Bacterial Lectin FimH: Covalent Inhibition or Not? *ChemBioChem* **2011**, *12*(7), 1066-1074.

[71] F. Perrin. Polarisation de la lumière de fluorescence. Vie moyenne des molécules dans l'etat excité. *J. Phys. Radium* **1926**, *7*(12), 390-401.

[72] D. Jameson, J. A. Ross. Fluorescence Polarisation/Anisotropy in Diagnostics and Imaging. *Chem. Rev.* **2010**, *110*(5), 2685-2780.

[73] G. Weber. Polarization of the fluorescence of macromolecules. I. Theory and experimental method. *Biochem J.* **1952**, *51*(2), 145–155 .

[74] Y. Engelborghs, A. J. W. G. Visser. Fluorescence Spectroscopy and Microscopy: Methods in Molecular Biology, vol. 1076, Springer Protocols **2013**.

[75] S. Kleeb, L. Pang, K. Mayer, D. Eris, A. Sigl, R. C. Preston, P. Zihlmann, T. Sharpe, R. P. Jakob, D. Abgottspon, A. S. Hutter, M. Scharenberg, X. Jiang, G. Navarra, S. Rabbani, M. Smiesko, N. Lüdin, J. Bezençon, O. Schwardt, T. Maier, B. Ernst. FimH Antagonists: Bioisosteres To Improve the in Vitro and in Vivo PK/PD Profile. *J. Med. Chem.* **2015**, *58*(5), 2221-2239.

[76] X. Huang. Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. *J. Biomol. Screen.* **2003**, *8H*(1), 34-38.

[77] P. Wu, M. Brasseur, U. Schindler. A high-throughput STAT binding assay using fluorescence polarization. *Anal. Biochem.* **1997**, *249*(1), 29-36.

[78] S.S. Iqbal, M. W. Mayo, J. G. Bruno, B. V. Bronk, C. A. Batt, J. P. Chambers. A review of molecular recognition technologies for detection of biological threat agents. *Biosens. Bioelectron.* **2000**, *15*(11-12), 549–578.

[79] A. Ibraheem R. E. Campbell. Designs and applications of fluorescent protein-based biosensors. *Curr. Opin. Chem. Biol.* **2010**, *14*(1), 30–36.

[80] N. K. Rajan, X. Duan, M. A. Reed. Performance limitations for nanowire/nanoribbon biosensors. *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* **2013**, *5*(6), 629–645.

[81] L. C. Clark, C. Lyons. Electrode Systems For Continuous Monitoring In Cardiovascular Surgery. *Ann. N.Y. Acad. Sci.* **1962**, *102*, 29–45.

[82] D. R. Thévenot, K. Toth, R. A. Durst, G. S. Wilson. Electrochemical biosensors: recommended definitions and classification. *Biosens. Bioelectron.* **2001**, *16*(-2), 121–131.

[83] M. J. Schöning, A. Poghossian. Recent advances in biologically sensitive field-effect transistors (BioFETs). Analyst, **2002**, *127*(9), 1137–1151.

[84] Y. Cui, Q. Wei, H. Park, C. M. Lieber. Nanowire Nanosensors for Highly Sensitive and Selective Detection of Biological and Chemical Species. *Science* **2001**, *293*(5533), 1289–1292.

[85] Niklas Elfström, Amelie Eriksson Karlström, and Jan Linnros. Silicon Nanoribbons for Electrical Detection of Biomolecules. *Nano Lett.* **2008**, *8*(3), 945-949.

[86] K. I. Chen, B. R. Li, Y. T. Chen. Silicon nanowire field-effect transistor-based biosensors for biomedical diagnosis and cellular recording investigation. *Nano Today* **2011**, *6*(2), 131–154.

[87] M. M. C. Cheng, G. Cuda G, Y. L. Bunimovich, M. Gaspari, J. R. Heath, H. D. Hill, C. A. Mirkin, A. J. Nijdam, R. Terracciano, T. Thundat, M. Ferrari. Nanotechnologies for biomolecular detection and medical diagnostics. *Curr. Opin. Chem. Biol.* **2006**, *10*(1), 11–19.

[88] F. Patolsky, G. Zheng, C. M. Lieber. Nanowire sensors for medicine and the lifesciences. *Nanomedicine* **2006**, *1*(1), 51–65.

[89] Z. Li, Y. Chen, X. Li, T. I. Kamins, K. Nauka, R. S. Williams. Sequence-specific label-free DNA sensors based on silicon nanowires. *Nano Lett.* **2004**, *4*(2), 245–247.

[90] E. Stern, J. F. Klemic, D. A. Routenberg, P. N. Wyrembak, D. B. Turner-Evans, A. D. Hamilton, D. A. LaVan, T. M. Fahmy, M. A. Reed. Label-free immunodetection with CMOS-compatible semiconducting nanowires. *Nature* **2007**, *445*, 519–522.

[91] Y. L. Bunimovich, Y. S. Shin, W. S. Yeo, M. Amori, G. Kwong, J. R. Heath. Quantitative Real-Time Measurements of DNA Hybridization with Alkylated Nonoxidized Silicon Nanowires in Electrolyte Solution. *J. Am. Chem. Soc.* **2006**, *128*(50), 16323–16331.

[92] J. Hahm, C. M. Lieber. Direct Ultrasensitive Electrical Detection of DNA and DNA Sequence Variations Using Nanowire Nanosensors. *Nano Lett.* **2004**, *4*(1), 51–54.

[93] A. Vacic, J. M. Criscione, N. K. Rajan, E. Stern, T. M. Fahmy, M. A. Reed. Determination of Molecular Configuration by Debye Length Modulation. *J. Am. Chem. Soc.* **2011**, *133*(35), 13886–13889.

[94] C. W. Wang, C. Y. Pan, H. C. Wu, P. Y. Shih, C. C. Tsai, K. T. Liao, L. L. Lu, W. H. Hsieh, C. D. Chen, Y. T. Chen. In Situ Detection of Chromogranin A Released from Living Neurons with a Single-Walled Carbon-Nanotube Field-Effect Transistor. *Small* **2007**, *3*(8), 1350–1355.

[95] G. J. Zhang, J. H. Chua, R. E. Chee, A. Agarwal, S. M. Wong. Label-free direct detection of MiRNAs with silicon nanowire biosensors. *Biosens. Bioelectron.* **2009**, *24*(8), 2504–2508.

[96] T. W. Lin, P. J. Hsieh, C. L. Lin, Y. Y. Fang, J. X. Yang, C. C. Tsai, P. L. Chiang, C. Y. Pan, Y. T. Chen. Label-free detection of protein-protein interactions using a calmodulin-modified nanowire transistor. *P. Natl. Acad. Sci. USA* **2010**, *107*(3), 1047–1052.

[97] G. F. Zheng, F. Patolsky, Y. Cui, W. U. Wang, C. M. Lieber. Multiplexed electrical detection of cancer markers with nanowire sensor arrays. *Nat. Biotechnol.* **2005**, *23*(10), 1294–1301.

[98] P. L. Chiang, T. C. Chou, T. H. Wu, C. C. Li, C. D. Liao, J. Y. Lin, M. H. Tsai, C. C. Tsai, C. J. Sun, C. H. Wang, J. M. Fang, Y. T. Chen. Nanowire Transistor-Based Ultrasensitive Virus Detection with Reversible Surface Functionalization. *Chem.-Asian. J.* **2012**, *7*(9), 2073–2079.

[99] F. Patolsky, G. Zheng, O. Hayden, M. Lakadamyali, X. Zhuang, C. M. Lieber. Electrical detection of single viruses. *P. Natl. Acad. Sci. USA* **2004**, *101*(39), 14017–14022.

[100] X. Duan, Y. Li, N. K. Rajan, D. A. Routenberg, Y. Modis, M. A. Reed. Quantification of the Affinities and Kinetics of Protein Interactions Using Silicon Nanowire Biosensors. *Nat. Nanotechnol.* **2012**, *7*, 401–407.

[101] J. DeChancie, K. N. Houk. The Origins of Femtomolar Protein-Ligand Binding: Hydrogen-Bond Cooperativity and Desolvation Energetics in the Biotin-(Strept)Avidin Binding Site. *J. Am. Chem. Soc.* **2007**, *129*(17), 5419–5429.

[102] M. A. Cooper. Optical biosensors in drug discovery. *Nat. Rev. Drug Discov.* **2002**, *1*(7), 515–528.

[103] L. Cegelski, G. R. Marshall, G. R. Eldridge, S. J. Hultgren. The biology and future prospects of antivirulence therapies. *Nat. Rev. Microb.* **2008**, *6*(1), 17-27.

2.2 Results

2.2.1 Outline

For a successful therapeutic application of FimH antagonists, several parameters have to be adjusted, among them the receptor occupancy time is of paramount importance. By means of surface plasmon resonance experiments, the dissociation rate constants k_{off} for biphenyl and indolyl α-D-mannosides were measured. The unexpectedly long occupancy times for all tested compounds suggest a high potential for *in vivo* treatment of UTIs (*paper 1*). However, the pharmacokinetic profiles of these molecules were not optimal. In fact, for oral treatment good water solubility, permeability, and limited first-pass metabolism are very important. Moreover, in order to reach the target renal excretion of the non-metabolized molecule must also be achieved. By applying bioisosteric substitutions on the lead compounds **1** and **2**, a suitable candidate was developed (compound **3**), which demonstrated high efficacy in an *in vivo* model of UTI (*paper 2*).

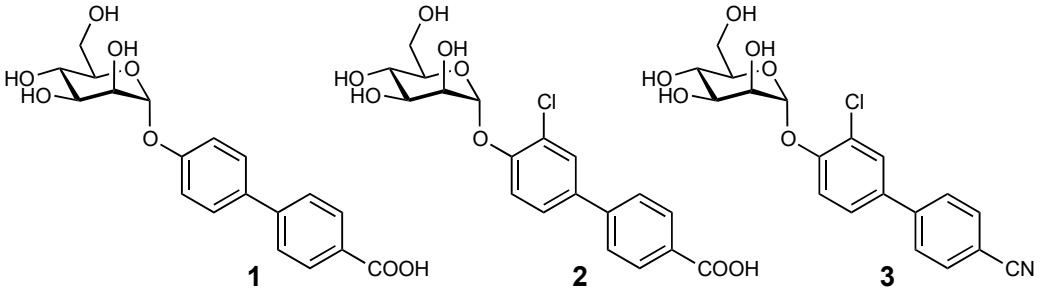


Figure 1. Structures of the lead compounds that were modified by bioisosteric substitutions of the carboxy group (compounds **1** and **2**) and of the optimized candidate (compound **3**).

The FimH protein exists in different affinity stages. Very promising results from our and other groups in targeting FimH were generated studying a truncated version of the FimH protein ($FimH_{LD}$) as a model system, which exists only in high affinity state. However, more studies were needed to understand which affinity state is the most therapeutically relevant *in vivo*. With the help of crystallographic and kinetic data, molecular dynamic simulations, and cell-tracking experiments we confirmed that the

low affinity state is the most relevant target for medicinal chemistry (*paper 3*), thus opening a new avenue for the development of FimH antagonists.

However, the $FimH_{LD}$ is an excellent study system for long-lived interactions involving carbohydrates. Moreover, FimH represents a pathologically relevant protein. The available compound **2** and $FimH_{LD}$ were selected as a model for the development of a label-free protein detection system based on silicon nanoribbons configured as field-effect transistors (SiNR-BioFET). In this area most research had focused so far on biotin-streptavidin as a model system, a rather questionable choice, due to the extremely low K_D. Our results constitute the first successful proof-of-concept for the detection of a pathologically relevant protein by SiNR-BioFETs (*paper 4*).

2.2.2 Paper 1

Kinetic Properties of Carbohydrate–Lectin Interactions: FimH Antagonists

This paper describes the kinetic characterization of FimH antagonists. The binding to FimH of a set of biphenyl and indolyl α-D-mannosides was measured by surface plasmon resonance. The results highlighted the unexpectedly long off-rate constants exhibited by the test molecules after binding to the lectin domain of FimH, suggesting high potential for in vivo treatment of UTIs.

Contribution to the project:

Giulio Navarra synthesized compounds 2 and 3b.

This paper was published in ChemBioChem:

Meike Scharenberg, Xiaohua Jiang, Lijuan Pang, Giulio Navarra, Said Rabbani, Florian Binder, Oliver Schwardt, and Beat Ernst

CHEMMEDCHEM
COMMUNICATIONS

ChemPubSoc Europe

DOI: 10.1002/cmdc.201300349

Kinetic Properties of Carbohydrate–Lectin Interactions: FimH Antagonists

Meike Scharenberg, Xiaohua Jiang, Lijuan Pang, Giulio Navarra, Said Rabbani, Florian Binder, Oliver Schwardt, and Beat Ernst*[a]

The lectin FimH is terminally expressed on type 1 pili of uropathogenic Escherichia coli (UPEC), which is the main cause of urinary tract infections (UTIs). FimH enables bacterial adhesion to urothelial cells, the initial step of infection. Various mannose derivatives have been shown to antagonize FimH and are therefore considered to be promising therapeutic agents for the treatment of UTIs. As part of the preclinical development process, when the kinetic properties of FimH antagonists were examined by surface plasmon resonance, extremely low dissociation rates (k_{off}) were found, which is uncommon for carbohydrate–lectin interactions. As a consequence, the corresponding half-lives ($t_{1/2}$) of the FimH antagonist complexes are above 3.6 h. For a therapeutic application, extended $t_{1/2}$ values are a prerequisite for success, since the target occupancy time directly influences the in vivo drug efficacy. The long $t_{1/2}$ value of the tested FimH antagonists further confirms their drug-like properties and their high therapeutic potential.

Urinary tract infections (UTIs) are among the most prevalent infections and affect millions of people each year. In 70–95% of all cases, the UTI is caused by uropathogenic Escherichia coli (UPEC).[1] These bacteria express type 1 pili with a terminally located adhesive protein called FimH. FimH-mediated adhesion to the surface of urothelial cells by binding to oligomannoside residues of the glycoprotein uroplakin Ia (UPIa)[2–5] is a prerequisite for the invasion of the host cells leading to a UTI.[2,3] Therefore, efforts have been made to identify orally available FimH antagonists to interfere with the attachment of UPEC to urothelial cells. From these studies, a-d-mannopyranosides have emerged providing a novel therapeutic opportunity for prevention and treatment of UTIs as an alternative to antibiotics.[6–8] To date, several mannose-based FimH antagonists have been validated in various in vitro and in vivo studies.[9–20]

As part of their pharmacodynamic characterization, not only equilibrium dissociation constants (K_D) or half-maximal inhibitory concentrations (IC_{50}) but also the kinetics of the binding process are studied.[21–23] One crucial factor for a sustained in vivo efficacy is the half-life ($t_{1/2}$) of the drug–receptor complex, especially when drugs compete with endogenous ligands.

[a] Dr. M. Scharenberg, Dr. X. Jiang, L. Pang, G. Navarra, Dr. S. Rabbani, Dr. F. Binder, Dr. O. Schwardt, Prof. Dr. B. Ernst
Institute of Molecular Pharmacy, Pharmacenter, University of Basel
Klingelbergstrasse 50, 4056 Basel (Switzerland)
E-mail: beat.ernst@unibas.ch

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cmdc.201300349.

The $t_{1/2}$ of a drug–receptor complex depends on the dissociation rate (k_{off}). Slow off-rates are beneficial for the in vivo efficacy, as prolonged occupancy of the target by the drug results in an extended duration of the pharmacological effect. Consequently, lower drug concentrations are required to obtain high efficacy, decreasing the risk of off-target toxicity.[21–23] The importance of long target occupancy is reflected in the long $t_{1/2}$ of many drugs reaching the market, such as the HIV-1 protease inhibitor Darunavir ($t_{1/2}$ > 240 h),[24] the CCR5 receptor antagonist Maraviroc ($t_{1/2}$ = 10.5 h),[25] or the viral neuraminidase inhibitor Zanamivir ($t_{1/2}$ > 33 min), which was developed from a carbohydrate-based lead structure.[26]

For carbohydrate–lectin interactions, only a few studies describing the kinetic properties are available. For the lectins, myelin-associated glycoprotein (MAG),[27,28] E-, L- and P-selectin,[29–31] galectin-1 and -3,[32] mannose-binding protein (MBP),[33] concanavalin A (ConA),[34] and calreticulin[35] surface plasmon resonance (SPR) experiments revealed fast association and dissociation kinetics with k_{off} rates between 2.6×10^{-3} and > 10 s^{-1}, resulting in short $t_{1/2}$ values ranging from 266 to 0.07 seconds (Table 1). These fast binding kinetics, typical for carbohydrate–lectin interactions, strongly hamper the development of carbohydrate-derived drugs. The determination of the kinetic parameters of FimH antagonists is therefore of utmost importance for successful lead optimization.

Beside K_D values, dissociation rates (k_{off}) of the complex between the antagonist and the target protein FimH are of special interest. To study these parameters, SPR is widely applied, including for carbohydrate–lectin[37] and carbohydrate–antibody[36] interactions. For the lectin domain of FimH, different affinity states are known.[38] In this study, the lectin domain in the high-affinity state was used.[39] Immobilization attempts by standard amine coupling failed, presumably due to accessible amino groups in and close to the ligand binding site. Thus, the N-terminal phenylalanine is part of the binding site. Immobilization via a C-terminal His-tag onto a nickel(II)–nitrilotriacetate (Ni-NTA) chip or indirect coupling via an anti-His-tag antibody failed due to instability of the base line, resulting from a slow detachment of the noncovalently immobilized FimH. Furthermore, harsh regenerating conditions (50 mm NaOH), necessary for the dissociation of the antagonist–lectin complex, caused the inactivation of the protein. Consequently, we immobilized FimH antagonists functionalized with an amino- (1 and 2) or N-hydroxy- (3a,b) succinimide via an amine-coupling procedure on CM4 dextran sensor surface chips (Scheme 1).

To determine the kinetic parameters of the FimH–antagonist interaction, a direct binding assay was performed. FimH was

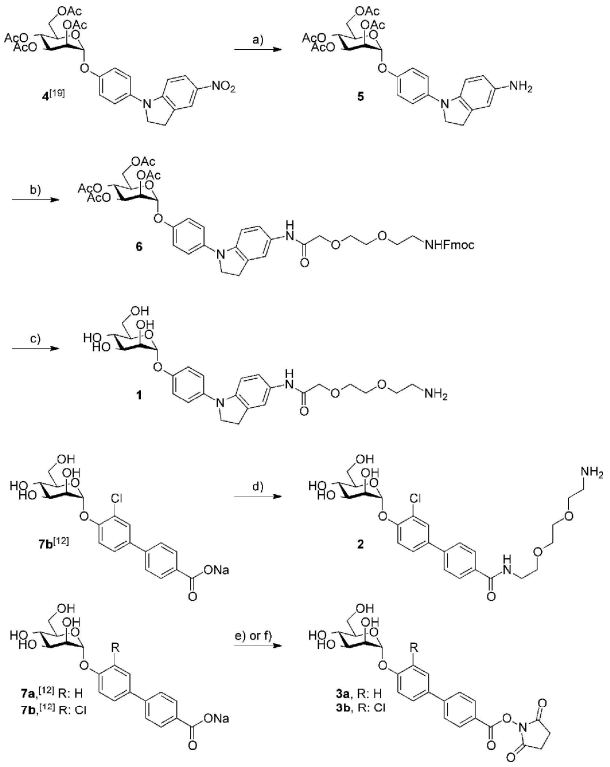
Table 1. Typical kinetic binding parameters of carbohydrate–protein interactions.

Protein	Ligand	K_D [mM]	k_{on} [m^{-1}/s^{-1}]	k_{off} [s^{-1}]	$t_{1/2}$ [s]	Ref.
L-selectin	GlyCAM1	108	> 10^5	> 10	0.07	[30]
E-selectin	ESL 1	62	4″ 10^4	3.0	0.2	[29]
P-selectin	PSGL 1	0.3	4″ 10^6	1.4	0.5	[31]
GSLA-2 mAB	sialyl Lewis[a]	4.3	1.1″ 10^5	8″ 10^{-1}	0.9	[36]
MAG	d-Neu5Ac derivative	2.8	3.5″ 10^5	0.8″ 10^{-1}	0.9	[27]
Galectin-1	d-Lactose derivative	1010	1.9″ 10^2	2.1″ 10^{-1}	3.3	[32]
Galectin-3	d-Lactose derivative	280	7.3″ 10^2	2.0″ 10^{-1}	3.4	[32]
CG-1A (avian galectin)	d-Lactose derivative	83.5	2.5″ 10^3	2.1″ 10^{-1}	3.3	[32]
Calreticulin	Glc_1Man_9-$GlcNAc_2$	2	3.9″ 10^4	8″ 10^{-2}	8.6	[35]
Con A	d-Man derivative	65	1.43″ 10^2	9.4″ 10^{-3}	73.7	[34]
MBP	d-Man_{16}/BSA	13.3	3.47″ 10^4	2.6″ 10^{-3}	266.6	[33]

Abbreviations: myelin-associated glycoprotein (MAG); concanavalin A (Con A); mannose-binding protein (MBP).

passed at concentrations between 0–200 nm over the flow cells (CM4 chip) equipped with covalently linked antagonists (Figure 1a). A reference cell without antagonist but treated with N-hydroxysuccinimide (NHS)/N-(3-dimethylaminopropyl)-N-ethylcarbodiimide hydrochloride (EDC) and ethanolamine (EA) was used to account for nonspecific binding of the protein to the dextran matrix. The kinetic parameters k_{on} and k_{off} were obtained by applying a global fit to the sensorgrams, using a 1:1 (Langmuir-type) binding-model (Scrubber 2.0c). The fitted sensorgrams of compounds 1, 2, 3a and 3b are shown in Figure 1b.

Mass transfer limitations, which might occur when using proteins as analytes (FimH: MW= 18.6 kD), can falsify the measured kinetic parameters. They depend on the cell dimension, the flow rate, and the diffusion coefficient of the analyte. Proteins having smaller diffusion coefficients than low-molecular-weight compounds are prone to show mass transfer limitations. To rule out these limitations, we used high flow rates (20–30 mL min^{-1}) and a low surface antagonist density (usage of CM4 chips instead of CM5 chips). Furthermore, we immobilized antagonist 1 at three different immobilization levels (differ-



Scheme 1. Synthesis of the amino- or N-hydroxyl-succinimide-functionalized FimH antagonists 1–3. Reagents and conditions: a) H_2 (1 atm), PtO_2, morpholine, EtOAc/MeOH (1:1), RT, 1 h (97%); b) 8-(Fmoc-amino)-3,6-dioxa-octanoic acid, PyBOP, DIPEA, DMF, RT, overnight; c) NaOMe, MeOH, RT, 2 h (47% over two steps); d) EBE, COMU, DIPEA, DMF, 0 8C! RT, overnight (20%); e) EDC, NHS, H_2O, RT, 30 min (3a: 98%); f) EDC, NHS, MES buffer (pH 5.6), RT, 30 min (3b: 99%).

CHEMMEDCHEM
COMMUNICATIONS
www.chemmedchem.org

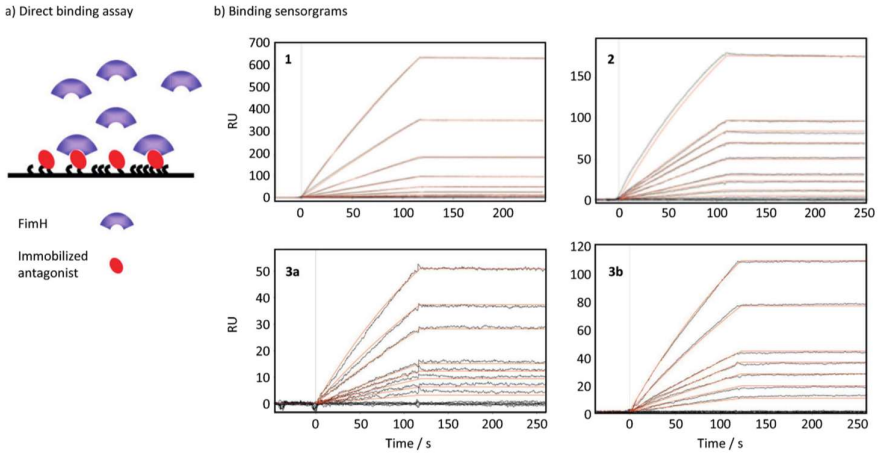


Figure 1. a) Schematic representation of a direct binding assay format. FimH binds to test compound (i.e., **1**, **2**, **3a**,**b**) immobilized on the chip. b) Sensorgrams obtained by kinetic fits of FimH binding to immobilized antagonists **1**, **2**, **3a** and **3b**. Solutions of FimH ranging between 0–200 nM were passed over the surface. For the fitting of the sensorgrams, Scrubber 2.0c was applied.

ent ratios between **1** and EA used for the immobilization) on the same sensor chip. FimH was screened simultaneously on all three surfaces, and the kinetic parameters and affinities were evaluated. Since all three surfaces showed similar kinetic rates and affinities, mass transfer effects are negligible. The obtained kinetic parameters are summarized in Table 2.

Association rates (k_{on}) between 1.4×10^4 and $4.8\times10^4\,M^{-1}s^{-1}$ were obtained and are in the expected range for low-molecular-weight compounds. The dissociation rates (k_{off}) were $5.2\times10^{-5}\,s^{-1}$ for **1**, $3.5\times10^{-5}\,s^{-1}$ for **2**, and $2.0\times10^{-5}\,s^{-1}$ for **3a**. For compound **3b**, the detection limit of k_{off} ($<10^{-6}\,s^{-1}$) was reached and consequently the K_D value was not determinable. The small dissociation constants resulted in $t_{1/2}$ values between 3.6 h and >19 h, representing extraordinary long $t_{1/2}$ values for carbohydrate–lectin interactions, which are usually in the range of seconds (see Table 1). Consistent with the long $t_{1/2}$ values, the equilibrium state of the interactions can only be reached after an extended period of time, conditions that were not applicable to our SPR experiments. The affinity (K_D) of the antagonists was therefore calculated from the kinetic parameters ($K_D=k_{off}/k_{on}$) and not determined from steady state measurements. As expected,[12,19] affinities in the low nanomolar range (0.7–3.5 nM) were obtained. Furthermore, the three-fold higher affinity of compound **2** compared with compound **3a** is consistent with published data, confirming that a chloro substituent in the *ortho* position of the aromatic ring adjacent to the anomeric center enhances binding affinity to FimH.[12] The off-rate of compound **2** was already close to the detection limit of the method applied. Compound **3b**, which differs from **2** only by a shorter linker length, did not reveal reliable kinetic data, due to its immeasurable k_{off} value ($\leq10^{-6}\,s^{-1}$), indicating that the linker length presumably has a small effect on the binding affinity.

Table 2. Kinetic binding parameters for the interaction of FimH with antagonists **1**, **2** and **3a**,**b**.

Compd	Ligand/EA ratio[a]	K_D [nM]	k_{on} [$10^4\,M^{-1}/s^{-1}$]	k_{off} [$10^{-5}\,s^{-1}$]	$t_{1/2}$ [h]
1	1:0 (high)	3.5	1.4	5.2	3.7
1	1:10 (middle)	2.5	2.0	5.1	3.8
1	1:100 (low)	2.0	2.6	5.3	3.6
2	1:0	0.7	4.8	3.5	5.5
3a	1:0	2.3	1.1	2.0	9.6
3b	1:0	n.d.[b]	1.4	$\leq10^{-6}$[c]	>19

[a] Molar ratio of ligand/ethanolamine (EA); density is given in parentheses; [b] not determinable (n.d.) due to $k_{off}\leq10^{-6}\,s^{-1}$; [c] k_{off} value for **3b** out of limit.

FimH can exhibit two conformations, a low-affinity conformation and a high-affinity conformation. The switch from the low-affinity state to the high-affinity state can be triggered by applying a mechanical force along the molecule. This behavior is characteristic for the catch-bond mechanism found for the FimH–ligand interaction.[38] It enables the bacteria to firmly attach to oligomannosides on bladder epithelial cells, even under the harsh conditions of the urinary tract (i.e., flow of urine). In the high-affinity state, the binding site of FimH forms a deep, narrow, and negatively charged pocket, unlike mammalian lectins, often characterized by shallow and water-accessible binding sites (see for example, selectins,[40] galectins,[41]

CHEMMEDCHEM
COMMUNICATIONS
www.chemmedchem.org

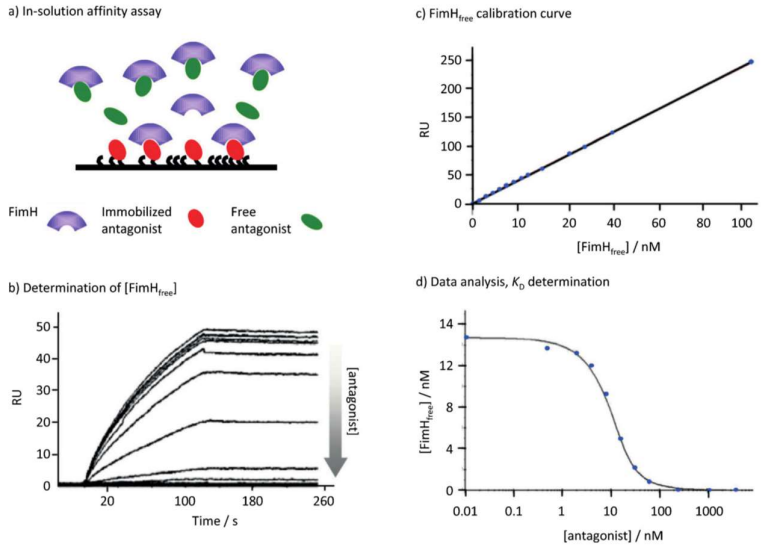


Figure 2. Results from the in-solution affinity assay with *n*-heptyl α-D-mannopyranoside **8** (see Table 3, entry 1). a) Schematic representation of the in-solution affinity assay. FimH$_{free}$ after equilibration binds to **1**, which is immobilized on the chip. b) Sensorgrams obtained after passing over the equilibrated mixtures of compound **8** and FimH. c) FimH calibration curve (FimH: 0–120 nM) and d) K_D determination by in-solution affinity-fit algorithm of the BiaEvaluation software.

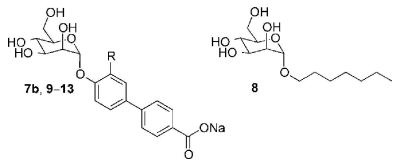
mannose-binding protein[42] or DC-SIGN).[43] The hydrophilic side chains of amino acids lining the FimH binding pocket establish a perfect network of hydrogen bonds with the hydroxy groups of α-D-mannopyranosides.[44] Consequently, the slow dissociation of the carbohydrate–FimH complex found in this study can be explained by the binding mode of FimH ligands to the high-affinity FimH conformation. As the pathogenicity of the bacteria depends on the interaction between FimH and its physiological ligand on the urothelial surface, a long $t_{1/2}$ value for FimH antagonists is of utmost importance for successful treatment.

Due to the high affinity of FimH antagonists and their small dissociation rate, in-solution affinity experiments (as described by Durka et al. in Ref. [45]) can be applied to determine K_D values of antagonists. For these experiments, we used the CM4 sensor chip coated with compound **1** (chip 1). For an accurate determination of K_D values, a constant concentration of FimH (10–15 nM) in the range of the K_D value to prevent stochiometric titration conditions was equilibrated with a dilution series of the antagonists **7b**–**13**.[46] After equilibration, the unbound FimH (FimH$_{free}$) binds to the immobilized antagonist **1** (Figure 2a) and can therefore be determined by SPR (Figure 2b) using a calibration curve (Figure 2c). Finally, [FimH$_{free}$] was plotted versus the antagonist concentration, and the curve was fitted with the in-solution affinity-fit algorithm of the BiaEvaluation software (Figure 2d). The obtained K_D values are summarized in Table 3.

n-Heptyl α-D-mannopyranoside **8** and the biphenyl-substituted mannose antagonists **7b** and **9**–**13** with different substitutions in the *ortho* position of the aromatic ring adjacent to the anomeric position showed affinities in the low nanomolar range, which are in good agreement with data obtained from isothermal titration calorimetry (ITC) experiments.[46] To further validate the assay, we additionally tested compound **7b** on a chip functionalized with compound **2** (chip 2), and we obtained a similar K_D value (0.5 vs 0.7 nM). Furthermore, the K_D value of **7b** determined by the in-solution affinity approach is equal to the K_D value of **2** found by a direct binding assay ($K_D = 0.7$ nM; Table 2). Compounds **2** and **7b** share identical structure with the only difference that **2** was immobilized on the chip via a linker (for the direct binding approach) whereas the K_D for **7b** was determined by the in-solution affinity assay. The comparable affinities derived from the two approaches confirm that the attachment of a linker and the immobilization process do not significantly influence the affinity of the antagonist.

In conclusion, for most medical applications, half-lives ($t_{1/2}$) of the drug–target complex of several tens of minutes or even hours are of utmost importance, since long $t_{1/2}$ values translate into higher in vivo efficacies and decrease adverse side effects

Table 3. K_D values of FimH antagonists[46] determined by an in-solution affinity assay.



Entry	Compd	R	K_D [nm][a]	rK_D[b]
1	8	–	5.21 ± 1.6	1
2	7b	Cl	0.71 ± 0.01 0.5 ± 0.05[c]	0.13 0.10[c]
3	9	F	1.63 ± 0.9	0.32
4	10	OCH_3	1.08 ± 0.18	0.21
5	11	CH_3	0.44 ± 0.22	0.09
6	12	Cyclopropyl	2.19 ± 0.1	0.42
7	13	CF_3	1.1 ± 0.34	0.21

[a] Determined on chip 1 (CM4 chip functionalized with compound 1). Data represent the mean ± SD of n = 3 independent experiments performed in triplicate. [b] The relative K_D (rK_D) values were calculated by dividing the K_D of the antagonist of interest by the K_D of n-heptyl a-d-mannopyranoside (8). [c] Determined on chip 2 (CM4 chip functionalized with compound 2, n = 3).

resulting from off-target toxicity.[21,22] However, carbohydrate–lectin interactions often exhibit low affinities and fast off-rates—properties that hamper the development of carbohydrate-derived drugs. Therefore, as part of the preclinical development process of FimH antagonists, we examined their kinetic characteristics by surface plasmon resonance (SPR). In this study, the lectin domain of FimH in the high-affinity state was used. The surprisingly small dissociation rates for FimH–antagonist complexes resulting in long $t_{1/2}$ values in the range of several hours (> 3.6 h) are indicators for high in vivo efficacy. This is a further indication that the corresponding ester prodrugs not only have a beneficial pharmacokinetic profile,[46] but also fulfill the pharmacodynamic requirements for therapeutic application, that is, high affinity and long residence time. However, whether the investigated high-affinity state of FimH investigated in our study[39] is the only pathophysiologically relevant state remains to be demonstrated. Therefore, studies with FimH lectin in other affinity states are planned.

Experimental Section

Synthesis: For synthesis and spectroscopic details of antagonists 1, 2 and 3 a,b, see the Supporting Information. The synthesis and characterization of compounds 9–13 are described in Ref. [46].

Chemical abbreviations used in Scheme 1 and elsewhere: (benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyBOP), (1-cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino-morpholino-carbenium hexafluorophosphate (COMU), N,N-diisopropylethylamine (DIPEA), N-(3-dimethylaminopropyl)-N-ethylcarbodiimide hydrochloride (EDC), 2,2'-(ethylenedioxy)bis(ethylamine) (EBE), ethanolamine (EA), 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), N-hydroxysuccinimide (NHS), 2-(N-morpholino)ethanesulfonic acid (MES).

FimH-CRD-6His protein expression: The FimH carbohydrate recognition domain (CRD) with a thrombin cleavage site linked to a 6His tag (FimH-CRD-Th-6His) was expressed in E coli strain HM125 and purified by affinity chromatography as described by Rabbani et al.[39]

Surface plasmon resonance (SPR) analysis: SPR measurements were performed on a Biacore 3000 SPR-based optical biosensor (Biacore, GE Healthcare, Uppsala, Sweden). Sensor chips (CM4), immobilization kits, maintenance supply and HBS-P buffer were purchased from GE Healthcare (Uppsala). The amino-functionalized monovalent compounds 1 and 2 were covalently attached to the activated dextrane matrix on CM4 chips by the standard amine-coupling method (GE Healthcare, Uppsala). The surface was activated by NHS and EDC. To obtain different ligand densities on the chip, compound 1 (1 mm in borate buffer) was mixed with EA (1 m) in different molar ratios prior to the coupling process (1/EA ratios: 1:0, 1:10, and 1:100). Pure compound 2 (1 mm in borate buffer) was coupled to the chip. After coupling, the matrix was capped with EA. For the coupling of the N-hydroxy-succinimide-functionalized compounds 3 and 4, the free carboxyl groups on the chip were activated with NHS and EDC and reacted with 1,2-diaminoethane (0.1 m in borate buffer, pH 8.5) to give free amino groups. The next steps were followed as described above. A reference cell without immobilized ligand was prepared and the system equilibrated with HEPES-buffered saline (HBS)-P buffer (10 mm HEPES, 150 mm NaCl, 0.005 % P20, pH 7.4). The activity of the chips was confirmed by FimH binding at a constant concentration of 50 nm in HBS-P buffer. All binding experiments were performed at 25 8C at a flow rate of 20 or 30 mL min^{-1} using HBS-P buffer. For kinetic studies, a dilution series of FimH with concentrations ranging from 0–200 nm in HBS-P buffer was used. Contact time was 120 s, and the dissociation time was 1200 s. The surface was regenerated with a single injection of 50 mm NaOH for 120 s. Data processing as well as k_{on}, k_{off}, and K_D determinations were accomplished with the Scrubber software (BioLogic Software, Version 2.0c, Campbell, Australia). Double referencing (subtraction of reference and blank injection) was applied to correct for bulk effects and other systematic artifacts.

In-solution affinity inhibition experiments: FimH (10–15 nm in HBS-P buffer) was inhibited with a series of test compound solutions of increasing concentrations (0–1 mm) overnight at RT to allow the equilibration of the system. The mixtures were run over a sensor surface coated with compound 1, and the resonance units (RU) after an association time of 110 s were detected. The non-inhibited FimH concentration ([$FimH_{free}$]) in the protein–compound mixtures was determined by means of a FimH calibration curve using free FimH concentrations ranging from 0–120 nm. The K_D values of the compounds were calculated by plotting [$FimH_{free}$] versus [compound], and fitting the curve with the in-solution affinity fit.

Acknowledgements

The authors gratefully acknowledge the financial support for G.N. by Nano-Tera (grant no. NT.ch 611_61).

CHEMMEDCHEM
COMMUNICATIONS
www.chemmedchem.org

Keywords: antagonists · carbohydrate–lectin interactions · FimH · kinetics · urinary tract infections · uropathogenic *Escherichia coli*

[1] A. Ronald, *Am. J. Med.* **2002**, *113*, 14S–19S.
[2] M. A. Mulvey, *Cell. Microbiol.* **2002**, *4*, 257–271.
[3] T. J. Wiles, R. R. Kulesus, M. A. Mulvey, *Exp. Mol. Pathol.* **2008**, *85*, 11–19.
[4] B. Xie, G. Zhou, S. Y. Chan, E. Shapiro, X. P. Kong, X. R. Wu, T. T. Sun, C. E. Costello, *J. Biol. Chem.* **2006**, *281*, 14644–14653.
[5] G. Zhou, W. J. Mo, P. Sebbel, G. W. Min, T. A. Neubert, R. Glockshuber, X. R. Wu, T. T. Sun, X. P. Kong, *Cell Sci.* **2001**, *114*, 4095–4103.
[6] I. Ofek, D. L. Hasy, N. Sharon, *FEMS Immunol. Med. Microbiol.* **2003**, *38*, 181–191.
[7] N. Sharon, *Biochim. Biophys. Acta Gen. Subj.* **2006**, *1760*, 527–537.
[8] B. Ernst, J. L. Magnani, *Nat. Rev. Drug Discovery* **2009**, *8*, 661–677.
[9] J. Bouckaert, J. Berglund, M. Schembri, E. De Genst, L. Cools, M. Wuhrer, C. S. Hung, J. Pinkner, R. Slättegård, A. Zavialov, D. Choudhury, S. Langermann, S. J. Hultgren, L. Wyns, P. Klemm, S. Oscarson, S. D. Knight, H. De Greve, *Mol. Microbiol.* **2005**, *55*, 441–455.
[10] T. K. Lindhorst, C. Kieburg, U. Krallmann-Wenzel, *Glycoconjugate J.* **1998**, *15*, 605–613.
[11] O. Sperling, A. Fuchs, T. K. Lindhorst, *Org. Biomol. Chem.* **2006**, *4*, 3913–3922.
[12] T. Klein, D. Abgottspon, M. Wittwer, S. Rabbani, J. Herold, X. Jiang, S. Kleeb, C. Lüthi, M. Scharenberg, J. Bezençon, E. Gubler, L. Pang, M. Smiesko, B. Cutting, O. Schwardt, B. Ernst, *J. Med. Chem.* **2010**, *53*, 8627–8641.
[13] D. Abgottspon, G. Roelli, L. Hosch, A. Steinhuber, X. Jiang, O. Schwardt, B. Cutting, M. Smiesko, U. Jenal, B. Ernst, A. Trampuz, *J. Microbiol. Methods* **2010**, *82*, 249–255.
[14] Z. Han, J. S. Pinkner, B. Ford, R. Obermann, W. Nolan, S. A. Wildman, D. Hobbs, T. Ellenberger, C. K. Cusumano, S. J. Hultgren, J. W. Janetka, *J. Med. Chem.* **2010**, *53*, 4779–4792.
[15] A. Wellens, C. Garofalo, H. Nguyen, N. Van Gerven, R. Slättegård, J.-P. Hernalsteens, L. Wyns, S. Oscarson, H. De Greve, S. Hultgren, J. Bouckaert, *PLoS One* **2008**, *3*, e2040.
[16] C. S. Edén, R. Freter, L. Hagberg, R. Hull, S. Hull, H. Leffler, G. Schoolnik, *Nature* **1982**, *298*, 560–562.
[17] M. Aronson, O. Medalia, L. Schori, D. Mirelman, N. Sharon, I. Ofek, *J. Infect. Dis.* **1979**, *139*, 329–332.
[18] M. Scharenberg, D. Abgottspon, E. Cicek, X. Jiang, O. Schwardt, S. Rabbani, B. Ernst, *Assay Drug Dev. Technol.* **2011**, *9*, 455–464.
[19] X. Jiang, D. Abgottspon, S. Kleeb, S. Rabbani, M. Scharenberg, M. Wittwer, M. Haug, O. Schwardt, B. Ernst, *J. Med. Chem.* **2012**, *55*, 4700–4713.
[20] M. Hartmann, T. K. Lindhorst, *Eur. J. Org. Chem.* **2011**, 3583–3609.
[21] R. A. Copeland, D. L. Pompliano, T. D. Meek, *Nat. Rev. Drug Discovery* **2006**, *5*, 730–739.
[22] D. C. Swinney, *Nat. Rev. Drug Discovery* **2004**, *3*, 801–808.
[23] H. Lu, P. J. Tonge, *Curr. Opin. Chem. Biol.* **2010**, *14*, 467–474.
[24] C. Napier, H. Sale, M. Mosley, G. Rickett, P. Dorr, R. Mansfield, M. Holbrook, *Biochem. Pharmacol.* **2005**, *71*, 163–172.
[25] I. Dierynck, M. De Wit, E. Gustin, I. Keuleers, J. Vandersmissen, S. Hallenberger, K. Hertogs, *J. Virol.* **2007**, *81*, 13845–13851.
[26] W. M. Kati, D. Montgomery, R. Carrick, L. Gubareva, C. Maring, K. McDaniel, K. Steffy, A. Molla, F. Hayden, D. Kempf, W. Kohlbrenner, *Antimicrob. Agents Chemother.* **2002**, *46*, 1014–1021.
[27] O. Schwardt, H. Gäthje, A. Vedani, S. Mesch, G.-P. Gao, M. Spreafico, J. von Orelli, S. Kelm, B. Ernst, *J. Med. Chem.* **2009**, *52*, 989–1004.
[28] S. Mesch, K. Lemme, H. Koliwer-Brandl, D. S. Strasser, O. Schwardt, S. Kelm, B. Ernst, *Carbohydr. Res.* **2010**, *345*, 1348–1359.
[29] M. K. Wild, M. C. Huang, U. Schulze-Horsel, P. A. van der Merwe, D. Vestweber, *J. Biol. Chem.* **2001**, *276*, 31602–31612.
[30] M. W. Nicholson, A. N. Barclay, M. S. Singer, S. D. Rosen, P. A. van der Merwe, *J. Biol. Chem.* **1998**, *273*, 763–770.
[31] P. Mehta, R. D. Cummings, R. P. McEver, *J. Biol. Chem.* **1998**, *273*, 32506–32513.
[32] G. J. Muñoz, J. I. Santos, A. Ardá, S. André, H.-J. Gabius, J. V. Sinisterra, J. Jiménez-Barbero, M. J. Hernáiz, *Org. Biomol. Chem.* **2010**, *8*, 2986–2992.
[33] T. Terada, M. Nishikawa, F. Yamashita, M. Hashida, *Int. J. Pharm.* **2006**, *316*, 117–123.
[34] B. N. Murthy, S. Sinha, A. Surolia, S. S. Indi, N. Jayaraman, *Glycoconjugate J.* **2008**, *25*, 313–321.
[35] A. R. Patil, C. J. Thomas, A. Surolia, *J. Biol. Chem.* **2000**, *275*, 24348–24356.
[36] L. Herfurth, B. Ernst, B. Wagner, D. Ricklin, D. S. Strasser, J. L. Magnani, A. J. Benie, T. Peters, *J. Med. Chem.* **2005**, *48*, 6879–6886.
[37] E. Duverger, N. Frison, A. C. Roche, M. Monsigny, *Biochimie* **2003**, *85*, 167–179.
[38] I. Le Trong, P. Aprikian, B. A. Kidd, M. Forero-Shelton, V. Tchesnokova, P. Rajagopal, V. Rodriguez, G. Interlandi, R. Klevit, V. Vogel, R. E. Stenkamp, E. V. Sokurenko, W. E. Thomas, *Cell* **2010**, *141*, 645–655.
[39] S. Rabbani, X. Jiang, O. Schwardt, B. Ernst, *Anal. Biochem.* **2010**, *407*, 188–195.
[40] a) W. S. Somers, J. Tang, G. D. Shaw, R. T. Camphausen, *Cell* **2000**, *103*, 467–479; b) B. J. Graves, R. L. Crowther, C. Chandran, J. M. Rumberger, S. Li, K. S. Huang, D. H. Presky, P. C. Famillett, B. A. Wolitzky, D. K. Burns, *Nature* **1994**, *367*, 532–538.
[41] P. M. Collins, C. T. Öberg, H. Leffler, U. J. Nilsson, H. Blanchard, *Chem. Biol. Drug Des.* **2012**, *79*, 339–346.
[42] K. S. Kenneth, K. Drickamer, W. I. Weis, *J. Biol. Chem.* **1996**, *271*, 663–674.
[43] Y. Guo, H. Feinberg, E. Conroy, D. Mitchell, R. Alvarez, O. Blixt, M. Taylor, W. I. Weis, K. Drickamer, *Nat. Struct. Mol. Biol.* **2004**, *11*, 591–598.
[44] C. Hung, J. Bouckaert, D. L. Hung, J. Pinkner, C. Widberg, A. DeFusco, G. Auguste, R. Strouse, S. Langermann, S. J. Hultgren, *Abstr. Gen. Meeting Am. Soc. Microbiol.* **2002**, *102*, 41.
[45] M. Durka, K. Buffet, J. Iehl, M. Holler, J. F. Nierengarten, J. Taganna, J. Bouckaert, S. P. Vincent, *Chem. Commun.* **2011**, *47*, 1321–1323.
[46] L. Pang, S. Kleeb, K. Lemme, S. Rabbani, M. Scharenberg, A. Zalewski, F. Schädler, O. Schwardt, B. Ernst, *ChemMedChem* **2012**, *7*, 1404–1422.

Received: August 29, 2013
Published online on December 2, 2013

2.2.3 Paper 2

FimH Antagonists: Bioisosteres To Improve the in Vitro and in Vivo PK/PD Profile

In this paper the optimization of the pharmacokinetic profile of FimH antagonists based on the biphenyl α-D-mannoside structure, by means of bioisosteric substitutions, is described. For the rapid and reliable evaluation of the binding affinity of the new compounds, a fluorescence polarization assay was developed. The molecule exhibiting the best parameters for oral administration was tested in an *in vivo* model of UTI and showed high efficacy, reducing the bacterial count in the bladder by about 1000-fold.

Contribution to the project:
Giulio Navarra designed, synthesized, and characterized the fluorescent compounds **22**, **23**, and **24**, and wrote the corresponding parts of the experimental section.

This paper was published in the *Journal of Medicinal Chemistry*:

Simon Kleeb, Lijuan Pang, Katharina Mayer, Deniz Eris, Anja Sigl, Roland C. Preston, Pascal Zihlmann, Timothy Sharpe, Roman P. Jakob, Daniela Abgottspon, Aline S. Hutter, Meike Scharenberg, Xiaohua Jiang, Giulio Navarra, Said Rabbani, Martin Smiesko, Nathalie Lüdin, Jacqueline Bezençon, Oliver Schwardt, Timm Maier, and Beat Ernst

Journal of Medicinal Chemistry

Article

pubs.acs.org/jmc

FimH Antagonists: Bioisosteres To Improve the in Vitro and in Vivo PK/PD Profile

Simon Kleeb,[†,∥] Lijuan Pang,[†,∥] Katharina Mayer,[†,∥] Deniz Eris,[†,∥] Anja Sigl,[†,∥] Roland C. Preston,[†] Pascal Zihlmann,[†] Timothy Sharpe,[§] Roman P. Jakob,[‡] Daniela Abgottspon,[†] Aline S. Hutter,[†] Meike Scharenberg,[†] Xiaohua Jiang,[†] Giulio Navarra,[†] Said Rabbani,[†] Martin Smiesko,[†] Nathalie Lüdin,[†] Jacqueline Bezençon,[†] Oliver Schwardt,[†] Timm Maier,[‡] and Beat Ernst*[,†]

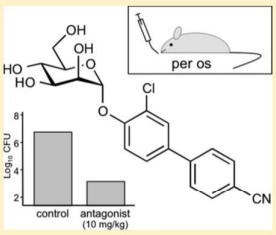
[†]Institute of Molecular Pharmacy, Pharmacenter, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland

[‡]Structural Biology, Biocenter, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

[§]Biophysical Facility, Biocenter, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Supporting Information

ABSTRACT: Urinary tract infections (UTIs), predominantly caused by uropathogenic *Escherichia coli* (UPEC), belong to the most prevalent infectious diseases worldwide. The attachment of UPEC to host cells is mediated by FimH, a mannose-binding adhesin at the tip of bacterial type 1 pili. To date, UTIs are mainly treated with antibiotics, leading to the ubiquitous problem of increasing resistance against most of the currently available antimicrobials. Therefore, new treatment strategies are urgently needed. Here, we describe the development of an orally available FimH antagonist. Starting from the carboxylate substituted biphenyl α-D-mannoside **9**, affinity and the relevant pharmacokinetic parameters (solubility, permeability, renal excretion) were substantially improved by a bioisosteric approach. With 3′-chloro-4′-(α-D-mannopyranosyloxy)biphenyl-4-carbonitrile (**10j**) a FimH antagonist with an optimal in vitro PK/PD profile was identified. Orally applied, **10j** was effective in a mouse model of UTI by reducing the bacterial load in the bladder by about 1000-fold.



INTRODUCTION

Urinary tract infection (UTI) is one of the most frequent infectious diseases worldwide and affects millions of people every year.[1] In more than 70% of the reported cases, uropathogenic *Escherichia coli* (UPEC) is the causal pathogen.[2] Acute, uncomplicated lower urinary tract infection, commonly referred to as cystitis, requires an antibiotic treatment for symptom relief (i.e., reduction of dysuria, frequent and urgent urination, bacteriuria, pyuria) and for prevention of more devastating or even life threatening complications like pyelonephritis and urosepsis.[3,4] However, the repeated use of antibacterial chemotherapeutics provokes antimicrobial resistance leading to treatment failure.[5] Hence, a new approach for the prevention and treatment of UTI with orally applicable therapeutics is urgently needed.[6]

UPEC undergo a well-defined infection cycle within the host.[7] The key step in pathogenesis is bacterial adhesion to the epithelial cells in the lower urinary tract.[8] This interaction prevents UPEC from clearance by the bulk flow of urine and enables the bacteria to colonize the epithelial cells. The adhesion is mediated by the virulence factor FimH located at the tip of bacterial type 1 pili.[9,10] FimH consists of two immunoglobulin-like domains: the N-terminal lectin domain and (connected by a short linker) the C-terminal pilin domain.[11] The lectin domain encloses the carbohydrate recognition domain (CRD) that binds to the oligomannosides of the glycoprotein uroplakin Ia on the epithelial cell surface.[12] The pilin domain anchors the adhesin to the pilus and regulates the switch between two conformational states of the CRD with high and low affinity for mannosides, respectively.

More than 3 decades ago, Sharon and co-workers described various oligomannosides and aryl α-D-mannosides as potential antagonists of the FimH-mediated bacterial adhesion.[13,14] However, only weak interactions in the milli- to micromolar range were observed. In recent years, several high-affinity monovalent mannose-based FimH antagonists with various aglycones like *n*-alkyl,[15] phenyl,[16] dioxocyclobutenylaminophenyl,[17] umbelliferyl,[16] biphenyl,[18−22] indol(in)-ylphenyl,[23] triazolyl,[24] and thiazolylamino[25] have been reported. In addition, different multivalent presentations of the mannose have been synthesized[26−32] and a heptavalent presentation of *n*-heptyl α-D-mannoside (**1**) tethered to β-cyclodextin proved to be highly effective when applied together with the UTI89 bacterial strain through a catheter into the bladder of C3H/HeN mice.[32] Importantly, adverse side effects resulting from nonselective binding of FimH antagonists (they are all α-D-mannopyrano-

Received: October 3, 2014
Published: February 10, 2015

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

Journal of Medicinal Chemistry Article

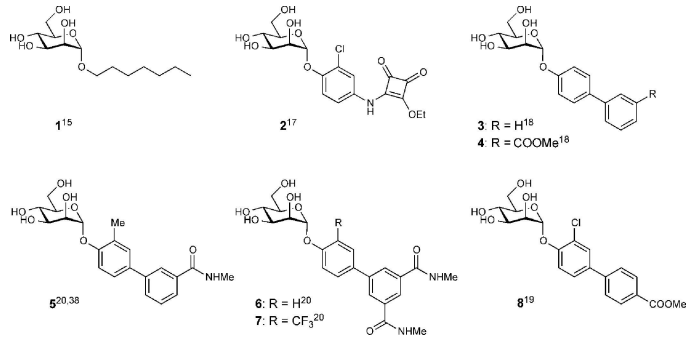


Figure 1. Monovalent FimH antagonists 1−4 acting as reference compounds and 5−8 which have been orally explored in in vivo disease models.

sides) to mannose receptors of the human host system have recently been ruled out.[33]

The high affinities of the monovalent α-D-mannopyranosides are based on optimal interactions with the main structural features of the CRD:[34−37] first, the mannose binding pocket accommodating the mannose moiety by means of an extended hydrogen bond network and, second, the entrance to the binding site composed of three hydrophobic amino acids (Tyr48, Tyr137, and Ile52) and therefore referred to as "tyrosine gate" hosting aliphatic and aromatic aglycones. As an example, n-heptyl α-D-mannopyranoside (1) exhibits nanomolar affinity due to hydrophobic contacts of the alkyl aglycone with the hydrophobic residues of the tyrosine gate.[15] Furthermore, aromatic aglycones, such as present in mannosides 2 and 3 (Figure 1), provide strong π−π stacking interactions with the tyrosine gate. This interaction is further favored by the addition of an electron withdrawing substituent on the terminal ring of the biaryl portion (→4).[18,19]

Recent in vivo PK studies in mice proved the high potential of the biphenyl α-D-mannosides 5−8 for an oral treatment, although high doses (≥50 mg/kg) were necessary to achieve the minimal concentrations required for the antiadhesive effect in the urinary bladder.[19−21] Moreover, the therapeutic effect could only be maintained for a few hours, i.e., 4 h for a po (per os) single-dose application of 7 (50 mg/kg), because of rapid elimination by glomerular filtration and low reabsorption from the primary urine in the renal tubules.[20]

To date, the physicochemical properties affecting the rate of renal excretion, i.e., lipophilicity and plasma protein binding (PPB), or metabolic liabilities promoting nonrenal elimination pathways have been barely investigated for FimH antagonists. The goal of the present study was to optimize the biphenyl α-D-mannoside with respect to oral bioavailability and renal excretion. Starting from antagonist 9[19] (Figure 2), we synthesized new biphenyl derivatives, characterized their affinity to the CRD, structurally investigated their binding mode, and determined physicochemical and pharmacokinetic parameters predictive for intestinal absorption and renal elimination. Furthermore, we determined in vivo PK (pharmacokinetics) of the most promising new antagonists in a mouse model. After oral administration, the compound with the best PK profile proved effective in reducing the bacterial loads upon bladder infection in a mouse model of UTI.

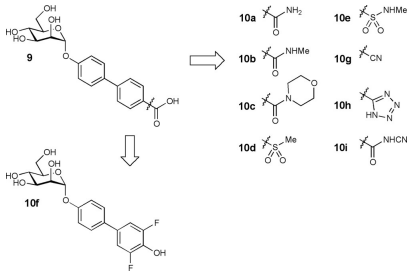


Figure 2. Bioisosteric replacement of the carboxylic acid substituent of biphenyl α-D-mannopyranoside 9.

RESULTS AND DISCUSSION

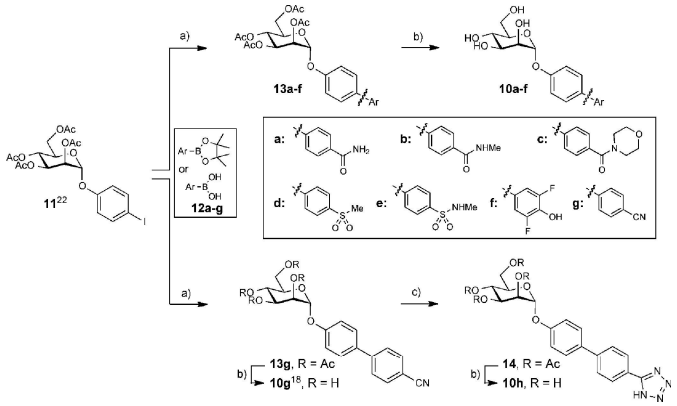
As previously reported, the carboxylate substituent present in the biphenyl mannoside 9 (its electron withdrawing potential being essential for an enhanced drug target interaction) strongly decreases the lipophilicity of the antagonist ($\log D_{7.4} < -1.5$[19]) in comparison to the n-heptyl (→1, log P = 1.7[19]) or the unsubstituted biphenyl aglycone (→3, log P = 2.1[22]). Since low lipophilicity is a major reason for low intestinal absorption and rapid renal excretion of the systemically available antagonist,[19,23] we aspired to improve oral bioavailability as well as renal excretion by replacing the carboxylate in 9 with various bioisosteric groups[39] (Figure 2).

Synthesis. Iodide 11 was prepared from peracetylated mannose and 4-iodophenol in the presence of $BF_3{:}Et_2O$.[22] In a palladium-catalyzed Miyaura−Suzuki coupling[40] with the boronic acid or boronate derivatives 12a−g, the biphenyl derivatives 13a−g were obtained in good to excellent yields. Final deprotection yielded the test compounds 10a−g. When microwave-assisted reaction conditions[41] were utilized, the conversion of arylnitrile 13g to tetrazole 14 proceeded rapidly and with good yield. After deprotection of 14 using Zemplén conditions, the test compound 10h was obtained (Scheme 1).

The cyanobenzamide derivative 10i (Scheme 2) was obtained from 9 by peracetylation (→15) followed by conversion of the

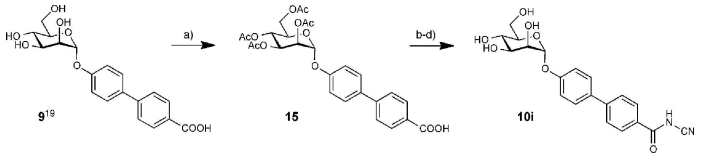
DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

Scheme 1[a]



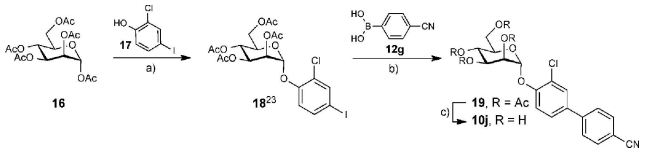
[a](a) $Pd(Cl_2)dppf{\cdot}CH_2Cl_2$, K_3PO_4, DMF, 80 °C, 4 h (13a−g, 44−99%); (b) NaOMe, MeOH, rt, 4 h (10a−h, 29−86%); (c) $TMSN_3$, $Bu_2Sn(O)$, DME, 150 °C, microwave, 10 min (81%).

Scheme 2[a]



[a](a) (i) Ac_2O, DMAP, pyridine, 0 °C to rt, overnight; (ii) sat. $NaHCO_3$ aq, DCM, rt, 2 h (15, 53%); (b) 1-chloro-N,N,2-trimethyl-1-propenylamine, toluene, 0 °C to rt, 2 h; (c) NaH, NH_2CN, DMF, 0 °C to rt, overnight; (d) NaOMe, MeOH, rt, 4 h (10i, 21% for three steps).

Scheme 3[a]



[a](a) $BF_3{\cdot}Et_2O$, CH_2Cl_2, 40 °C (76%); (b) $Pd(Cl_2)dppf{\cdot}CH_2Cl_2$, K_3PO_4, DMF, 80 °C (75%); (c) NaOMe, MeOH, rt, 4 h (48%).

carboxylic acid into its acid chloride with 1-chloro-N,N,2-trimethyl-1-propenylamine.[42] Without isolation, the acid chloride was reacted with sodium hydrogen cyanamide in DMF followed by deacetylation under Zemplén conditions to yield the test compound 10i.

Finally, to further improve the pharmacokinetic properties of mannoside 10g[18] (see Table 3), a chloride substituent was introduced to the ortho-position of the aromatic ring adjacent to the anomeric oxygen. For its synthesis, peracetylated α-D-mannose (16) was coupled with 2-chloro-4-iodophenol (17) using $BF_3{\cdot}Et_2O$ as promotor (→18, 76%). After the introduction of the second aromatic ring by Miyaura−Suzuki coupling (→19, 75%), deprotection yielded mannoside 10j (Scheme 3).

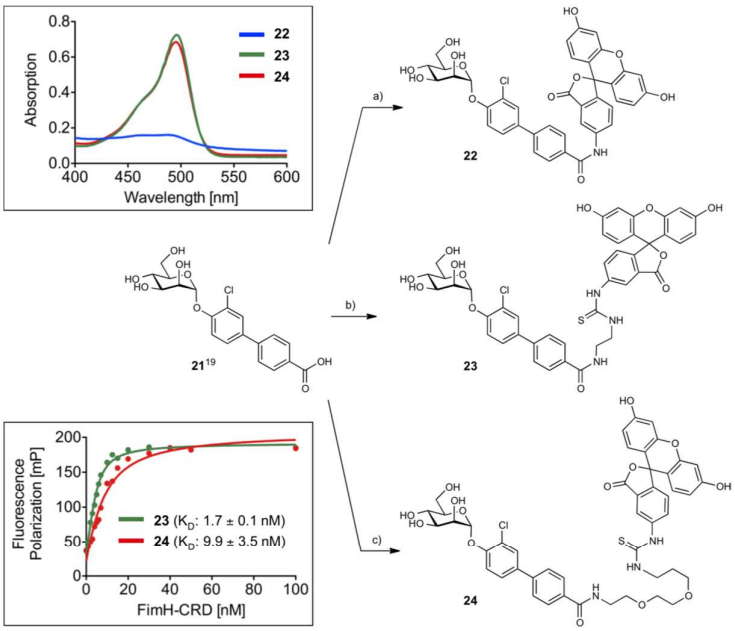
Binding Affinity. The binding affinity of heptyl mannoside 1, the biphenyl mannosides 3, 9, 20,[18] and the bioisosteres 10a−j was determined in a competitive fluorescence polarization assay (FP assay) and with isothermal titration calorimetry (ITC). A protein construct consisting of the CRD with a C-terminal His-tag with a thrombin cleavage site (FimH-CRD-Th-His_6) was used for all experiments.[43]

Competitive Fluorescence Polarization Assay. For the rapid evaluation of binding affinity, we established a competitive

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

Journal of Medicinal Chemistry Article

Scheme 4[a]



[a](a) 1-[(1-(Cyano-2-ethoxy-2-oxoethylideneaminooxy)dimethylaminomorpholinomethylene)]methanaminium hexafluorophosphate (COMU), NEt_3, fluoresceinamine, DMF, rt, 7 h (**22**, 19%); b) (i) DIC, NHS, *N*-Boc-ethylenediamine, DMF, rt, 12 h; (ii) TFA, DCM, rt, 10 min (68% over two steps), (iii) fluorescein isothiocyanate (FITC), NEt_3, DMF, rt, 3 h (**23**, 48%); (c) (i) DIC, NHS, *N*-Boc-PEG2-NH_2, DMF, rt, 14 h; (ii) TFA, DCM, rt, 30 min (62% over two steps); (iii) FITC, DMF, rt (**24**, 65%).

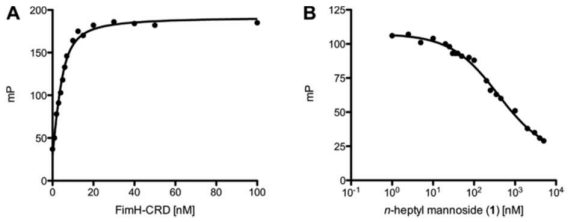


Figure 3. (A) Direct binding curve of the labeled competitor **23** obtained by adding a linear dilution of FimH-CRD (0−100 nM) and a constant concentration of competitor **23** (5 nM). The K_D was determined by fitting the experimental data to a single-site binding fit that accounts for ligand depletion. In three FP based direct binding experiments the K_D of competitor **23** was determined to be 1.7 nM. (B) Inhibition curve of *n*-heptyl mannoside (**1**) from the competitive FP assay. The IC_{50} value was determined by nonlinear least-squares fitting to a standard four-parameter equation. A modified Cheng−Prusoff equation[45] was used to calculate the corresponding K_D value (K_D = 28.3 nM).

binding assay based on fluorescence polarization (FP). Similar formats have been applied before for the detection of carbohydrate−lectin interactions.[18,44] In this assay, the antagonist of interest displaces a fluorescently labeled competitor from the binding site, thereby causing a reduction in fluorescence polarization.[45] To identify the optimal competitor, fluorescein isothiocyanate (FITC) was connected to the FimH ligand **21** by three linkers of different lengths (→**22**−**24**, Scheme 4). For

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

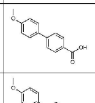
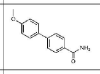
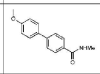
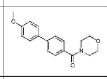
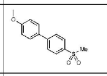
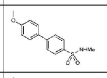
optimal sensitivity and signal-to-noise ratio, three main parameters need to be considered: (i) the affinity of the competitor should not be impaired by the fluorescent label; (ii) the conformational flexibility of the label upon binding of the competitor to the CRD should be low; (iii) the fluorescence properties of the label should not be affected by the connected ligand.[46−48] A change in fluorescence properties was observed for reporter ligand 22 in which the label was linked to the biphenyl agylcone by an amide bond. The absorption spectrum revealed a lack of the characteristic fluorescein absorption peak at 494 nm (Scheme 4), likely due to an extension of the conjugated system to the biphenyl moiety of the ligand. The elongated saturated spacer groups in competitors 23 and 24 ensured that the expected spectral properties of the dye were retained (Scheme 4).

For the determination of their binding affinity, fixed concentrations of the reporter ligands 23 and 24 were incubated for 24 h with a linear dilution of the FimH-CRD (0−100 nM). FP was measured using a plate reader, with polarized excitation at 485 nm and emission at 528 nm measured through appropriately oriented polarizers. Fitting the single-site binding function of Cooper[49] to the observed FP data resulted for compound 23 in a dissociation constant (K_D = 1.7 nM, Figure 3A) similar to that of the unlabeled parent compound 21,[19] whereas 24 showed a 5-fold lower affinity (9.9 nM) (Scheme 4). Therefore, the reporter ligand 23 fulfills all characteristics as an optimal competitor and was used for the FP assay.

For the test compounds 1, 3, 9, 20, and 10a−j, a 24 h incubation time was applied before FP was measured because of the long residence time of FimH antagonists ($t_{1/2}$ > 3.5 h, Figure 3B[50]). The 24 h incubation period was empirically determined to be necessary to reach equilibrium between reporter ligand and compound of interest. IC_{50} values were obtained by nonlinear least-squares regression (standard four-parameter dose−response curve) and converted to K_D values using a modified Cheng−Prusoff equation.[45] This equation accounts for the ligand depletion effect in competitive titrations involving high-affinity interaction partners present in similar concentrations. Under these conditions, the free concentration of an interacting species cannot be assumed to equal the total concentration.

The K_D values determined for the test compounds 1, 3, 9, 20, and 10a−j are summarized in Table 1. Against our expectations, the biphenyl mannosides 3 and 9 exhibit similar affinities (Table 1), despite the presence of an electron withdrawing carboxylate substituent in antagonist 9. According to the crystal structure of FimH cocrystallized with the sulfonamide derivative 10e (Figure 4A), the outer aromatic ring of the biphenyl aglycone forms π−π interactions with the electron rich Tyr48, which is part of the tyrosine gate of FimH.[15] A reduction of electron density of the aglycone by the electron withdrawing carboxylate was expected to enforce these π−π stacking interactions and lead to improved affinity. However, this beneficial effect might be compensated by an entropic penalty originating from the improved π−π stacking to Tyr48 that might lead to the reduced flexibility of both protein and antagonist. Furthermore, a beneficial enthalpy effect might be partially compensated by an enthalpy penalty originating from the desolvation of the charged carboxylate in 9[51] (see also Experimental Section). Although this substituent is solvent exposed, at least a partial desolvation may be necessary upon antagonist binding. To prove this assumption, we replaced the carboxylate by the corresponding methyl ester (→20)[18] in order to reduce the desolvation penalty and, as predicted by the Hammett constant σ_p,[52] to further improve the π−π stacking.

Table 1. Affinities (K_D) of FimH Antagonists to FimH-CRD-Th-His_6[b]

Entry	Compd		Affinity K_D [nM]
1	1		28.3 ± 5.0
2	3		15.1 ± 2.2
3	9		17.9 ± 1.5
4	20		3.6 ± 0.9
5	10a		2.8 ± 0.3
6	10b		2.9 ± 0.5
7	10c		3.0 ± 0.1
8	10d		1.7 ± 0.2
9	10e		2.7 ± 0.4
10	10f		3.7 ± 0.2
11	10g		2.0 ± 0.6
12	10h		5.7 ± 0.1
13	10i		8.4 ± 0.3
14	10j		< 1[a]

[a]The K_D value of 10j was approximated to be in the subnanomolar range. The IC_{50} value obtained in the competitive FP assay was equal to the lowest value that can be resolved by the assay, indicating stoichiometric titration of 10j due to its high affinity. Consequently, its K_D must be below the K_D of competitor 23. [b]Dissociation constants (K_D) were determined in a competitive fluorescence polarization assay.

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

Journal of Medicinal Chemistry Article

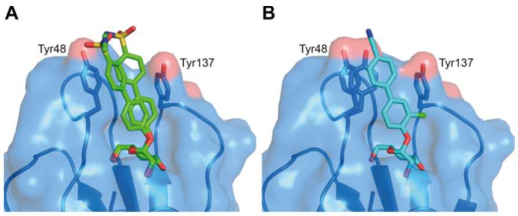


Figure 4. Ligand binding poses determined by X-ray cocrystallization with compounds **10e** resolved to 1.07 Å (A) and **10j** resolved to 1.10 Å (B). The electron density surrounding the aglycone of **10e** indicates flexibility of the aglycone and was modeled in two poses. Both compounds bind in a similar pose with a well-defined hydrogen network surrounding the mannose moiety and $\pi-\pi$ stacking interactions between the second aromatic ring and Tyr48 side chain (A). In contrast, in the FimH-CRD/**10j** structure the amino acid side chain of Y48 can be modeled in two distinct rotamers, suggesting flexibility also of the receptor (B).

Indeed, a 6-fold improvement in affinity was achieved. However, since the methyl ester undergoes rapid enzyme-mediated hydrolysis in vivo,[19] it will not be available at the place of action in the urinary bladder. The methyl ester was therefore replaced by metabolically stable bioisosteres[39] exhibiting comparable electron withdrawing properties[52] (Table 1, entries 5−13). The most potent derivatives **10d**, **10e**, and **10g** showed affinities in the low nanomolar range.

As previously reported,[22] a chloro substituent in the ortho-position of the aromatic ring adjacent to the anomeric oxygen is favorable for affinity and improves the physicochemical properties relevant for oral bioavailability. Indeed, the corresponding antagonist **10j** was the most potent compound tested in this study.

Isothermal Titration Calorimetry (ITC). To further confirm our hypothesis regarding $\pi-\pi$ stacking and desolvation, we performed ITC experiments with the reference compound **1**, the unsubstituted biphenyl mannoside **3**, the carboxylic acid **9**, and the bioisosteres **10b−e,g,j** (Table 2). ITC allows the simultaneous determination of the stoichiometry (N), the change in enthalpy (ΔH) and the dissociation constant (K_D) for ligand−protein binding.[53,54] The reliable determination of these three parameters requires well-defined sigmoidal titration curves characterized by the dimensionless Wiseman parameter c (c = Mt(0) K_D^{-1}, where Mt(0) is the initial macromolecule concentration).[55] To be sure that data can be fitted with confidence, the c-value should be between 1 and 1000 (ideally between 5 and 500),[56] which could be achieved for the antagonists **3** and **9**. For titrations involving low micromolar Mt(0) and interactions in the low nanomolar or picomolar range, as suggested for the bioisosteres **10b−j**, c-values above 1000 were expected. Since these conditions lead to steep titration curves that do not allow the determination of the curve slope representing $1/K_D$, we applied an alternative, competitive format referred to as displacement assay.[57,58] First, FimH-CRD-Th-His$_6$ was preincubated with the low affinity antagonist n-heptyl 2-deoxy-α-D-mannopyranoside (**25**, for synthesis see Supporting Information). The high-affinity bioisosteres of interest were titrated into the protein−ligand complex giving well-defined sigmoidal titration curves.

The resulting K_D values (Table 2) correspond well with the data obtained from the FP assay (Table 1). A comparison of the thermodynamic fingerprints of antagonists **3** and **9** reveals that the more favorable enthalpic contribution resulting from facilitated $\pi-\pi$ stacking leads to a net enthalpy gain ($\Delta\Delta H$ = −3.7 kJ/mol). However, an even greater increase in enthalpy is likely countered by the enthalpy costs for desolvation of the electron withdrawing carboxylate.

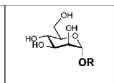
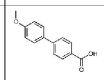
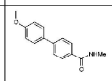
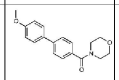
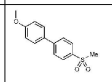
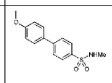
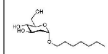
The gain in enthalpy is in turn compensated by an unfavorable entropy ($-T\Delta\Delta S$ = 3.2 kJ/mol) as a result of the reduced flexibility of both the antagonist and the Tyr48 side chain caused by the improved interaction. This is not entirely outweighed by the beneficial entropy contribution related to the partial desolvation of the carboxylate and the related release of water into the bulk. Added together, the enthalpy and entropy contributions of antagonists **3** and **9** result in similar affinities (K_D of 17.7 and 15.0 nM, respectively).

In contrast, the replacement of the carboxylate group by various neutral bioisosteres (entries 4−7) reduces the enthalpy costs for desolvation (see calculated free energies of desolvation, Experimental Section) and therefore leads to a markedly improved enthalpy ($\Delta\Delta H$ from −3.5 to −5.8 kJ/mol). As a result, an up to 5-fold improvement of the K_D values was achieved. Finally, with a cyano substituent (entries 8 and 9), the enthalpy term was further improved ($\Delta\Delta H$ = −3.7 kJ/mol) because of a reduced desolvation penalty and improved $\pi-\pi$ stacking interactions. However, this beneficial component is again partially compensated by a decrease in entropy. This can be attributed, first, to the loss of flexibility of the tightly bound ligand (Figure 4B) and, second, to the smaller surface area of the cyano substituent compared to amide, sulfonamide, and sulfone, which results in a smaller number of water molecules being released to bulk upon binding.

X-ray Crystallography. To determine the binding poses of the bioisosters, we cocrystallized the compounds **10e** and **10j** with the FimH-CRD (Figure 4). Atomic resolution crystal structures were obtained at 1.07 Å (**10e**) and 1.10 Å (**10j**). As observed in previous mannoside cocrystal structures,[15,18,36] the mannose moiety forms an extensive hydrogen bond network to the well-defined binding site with all of its hydroxyl groups. The biphenyl aglycone is located between the tyrosine gate residues (Tyr48/Tyr137). The $\pi-\pi$ stacking of the second aromatic ring of the aglycone to the side chain of Tyr48 contributes most to the interaction energy of the aglycone moiety. Interactions to the Tyr137 side chain on the other hand are only limited. Whereas a previously published crystal structure of a biphenyl mannoside in complex with FimH-CRD suffers from crystal contacts of binding site residues (Tyr48 side chain to backbone oxygen of Val27) possibly causing the distortion of the binding site,[18] the binding sites of our structures are mostly solvent exposed. This

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

Table 2. Thermodynamic Parameters from ITC for Selected FimH Antagonists Binding to FimH-CRD-Th-His$_6$[d]

Entry	Compd		K_D[a] [nM]	ΔG [kJ/mol]	ΔH[a] [kJ/mol]	$-T\Delta S$ [kJ/mol]	n	Type of measurement
1	**1**[b,c]		28.9 (25.8 – 32.3)	-43.0	-50.3 (-50.2 – -50.7)	7.3	1.00	direct
2	**3**[b]		17.7 (14.1 – 22.3)	-44.2	-45.0 (-44.5 – -45.6)	0.8	1.07	direct
3	**9**		15.0 (13.4 – 16.7)	-44.7	-48.7 (-48.4 – -49.0)	4.0	1.05	direct
4	**10b**		4.3 (3.2 – 5.6)	-47.8	-54.5 (-54.1 – -54.9)	6.7	1.02	competitive vs. **25**
5	**10c**		5.0 (3.8 – 6.6)	-47.4	-54.5 (-54.1 – -54.8)	7.1	0.97	competitive vs. **25**
6	**10d**		3.0 (2.1 – 4.2)	-48.7	-52.3 (-51.5 – -53.1)	3.6	0.99	competitive vs. **25**
7	**10e**		3.5 (2.9 – 4.3)	-48.2	-52.2 (-51.6 – -52.8)	3.9	1.06	competitive vs. **25**
8	**10g**		2.8 (2.3 – 3.3)	-48.8	-58.2 (-57.8 – -58.6)	9.4	1.00	competitive vs. **25**
9	**10j**		1.3 (1.1 – 1.6)	-50.7	-60.9 (-60.4 – -61.4)	10.1	1.01	competitive vs. **25**
10	**25**		9'386 (8'555 – 10'287)	-28.7	-19.5 (-19.1 – -20.0)	-9.1	1.00	direct

[a]95% confidence interval from fitting in parentheses. [b]Global fit including two direct titration measurements. [c]ITC data were previously published with an n-value of 0.82.[37] [d]n, stoichiometric correction factor.

revealed the flexibility of the aglycone in the FimH-CRD/ 10e structure, since the electron density toward the solvent-exposed sulfonamide indicates that there is not one single orientation. Therefore, the aglycone was modeled in two distinct poses. In contrast, in the FimH-CRD/ 10j structure the amino acid side chain of Y48 can be modeled in two distinct rotamers, suggesting flexibility also of the receptor.

Physicochemical Properties and in Vitro Pharmacokinetics. Intestinal absorption and renal excretion are prerequisites for a successful oral treatment of UTI with FimH antagonists. Furthermore, reabsorption of antagonist from the renal ultrafiltrate is desirable for maintaining the minimal antiadhesive concentration in the target organ, namely, the bladder, over an extended period of time. To estimate the influence of the bioisostere approach on oral bioavailability and the rate of renal excretion, we determined lipophilicity by means of the octanol−water distribution coefficient (log $D_{7.4}$),[59] aqueous solubility, and membrane permeability in the artificial membrane permeability assay (PAMPA)[60] and the colorectal adenocarcinoma (Caco-2) cell monolayer model.[61]

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

Journal of Medicinal Chemistry Article

Table 3. Physicochemical and in Vitro Pharmacokinetic Parameters[h]

compd	pK_a [a]	log $D_{7.4}$ [b]	solubility [μg/mL]/pH[c]	PAMPA log P_e [cm/s]/pH[d]	Caco-2 P_{app} [10^{-6} cm/s][e] a → b	Caco-2 P_{app} [10^{-6} cm/s][e] b → a	PPB f_b [%][f]	metabolic stability $t_{1/2}$ [min][g]
1		1.65	>3000	−4.89	7.0 ± 0.6	9.4 ± 0.2	81	13
3		2.1 ± 0.1	21 ± 1/7.4	−4.7 ± 0.1/7.4	10.0 ± 0.9	19.0 ± 1.2	93 ± 1	nd
20		2.14	33.8/6.51	−4.7	4.23	nd	93	1.0
9	3.88	<−1.5	>3000/6.61	no permeation	nd	nd	73	>60
10a		0.5 ± 0.1	12 ± 1/7.4	−6.8 ± 0.3/7.4	0.12 ± 0.01	0.61 ± 0.03	nd	nd
10b		0.8 ± 0.0	122 ± 13/7.4	−9.2 ± 1.4/7.4	1.10 ± 0.82	0.87 ± 0.15	nd	nd
10c		0.2 ± 0.1	>250/7.4	−7.8 ± 0.3/7.4	0.18 ± 0.07	1.30 ± 0.03	48 ± 2	>60
10d		0.4 ± 0.0	246 ± 17/7.4	−7.2 ± 0.0/7.4	0.36 ± 0.01	1.76 ± 0.12	99 ± 1	>60
10e		0.7 ± 0.1	>250/7.4	−8.6 ± 0.2/7.4	0.28 ± 0.23	1.82 ± 0.14	>99	>60
10f	6.5	1.1 ± 0.0	>150/3.0	−7.7 ± 0.8/5.0	0.40 ± 0.02	1.90 ± 0.17	nd	nd
			>150/7.4	−8.8 ± 0.1/7.4				
10g		1.4 ± 0.0	186 ± 4/7.6	−5.7 ± 0.0/7.4	2.0 ± 0.1	13.2 ± 2.1	99 ± 0	>60
10h	3.7	−1.4 ± 0.1	11 ± 0/3.0	−9.3 ± 1.4/5.0	0.17 ± 0.00	0.22 ± 0.01	nd	nd
			273 ± 2/7.4	−8.8 ± 1.4/7.4				
10i	2.5	−1.1 ± 0.1	>150/3.0	−6.8 ± 0.2/5.0	0.22 ± 0.14	0.29 ± 0.03	nd	nd
			>150/7.4	−7.0 ± 0.1/7.4				
10j		2.1 ± 0.0	192 ± 5/7.4	−5.2 ± 0.0/7.4	2.2 ± 0.4	22.1 ± 1.5	89 ± 1	>60

[a] pK_a values were determined by NMR spectroscopy. [b] Octanol−water distribution coefficients (log $D_{7.4}$) were determined by a miniaturized shake-flask procedure at pH 7.4. Values represent the mean ± SD of sextuplicate measurements.[59] [c] Kinetic solubility was measured in a 96-well format using the μSOL Explorer solubility analyzer at the indicated pH in triplicate. [d] P_e = effective permeability. Passive permeation through an artificial membrane was determined by the parallel artificial membrane permeation assay (PAMPA). Values represent the mean ± SD of quadruplicate measurements performed at the indicated pH.[60] [e] P_{app} = apparent permeability. Permeation through a Caco-2 cell monolayer was assessed in the absorptive (a → b) and secretory (b → a) directions in triplicate.[61] [f] Plasma protein binding (PPB) was determined by equilibrium dialysis in triplicate.[62] [g] Metabolic stability was determined by incubating the compounds (2 μM) with pooled rat liver microsomes (RLM, 0.5 mg/mL) in the presence of NADPH (1 mM, compounds **1**, **9**, **10c**−**e**,**g**,**j**) or without NADPH (compound **20**).[63] [h] nd = not determined.

Oral Bioavailability. Oral bioavailability of a compound relies on solubility, permeation through the membranes lining the intestine, and stability against first pass metabolism.[64,65] As discussed by Lipinski[66] and Curatolo,[67] dose and permeability define the minimum aqueous solubility required for oral administration. Thus, a dose of 1 mg/kg of a moderately permeable compound requires a solubility of at least 52 μg/mL. Whereas sufficient aqueous solubility (>3000 μg/mL) was reported for *n*-heptyl α-mannopyranoside (**1**),[19] the unsubstituted biphenyl α-D-mannopyranoside **3** and the antagonists bearing a methylcarboxylate, carboxamide, or tetrazole substituent (compounds **20**, **10a**, and **10h**) were found to be scarcely soluble.[22] As proposed by Ishikawa,[68] a possible reason is the apolar and planar aglycone. By contrast, the polar carboxylic acid moiety present in antagonist **9** or the substituents in the bioisosteres **10b**−**j** enhance solubility to 122−273 μg/mL, a level sufficient for in vivo PK studies. For in vivo disease studies, however, dosages of up to 10 mg/kg were foreseen (see below), requiring a solubility of 520 μg/mL.[66,67] For this reason, surfactant Tween 80 (1%) had to be added.

Furthermore, permeability data derived from PAMPA[69] and the Caco-2 model[70] suggest moderate to high permeation of the moderately lipophilic antagonists **1**, **3**, and **20** (log $D_{7.4}$ > 1.6) through the intestinal membranes. The bioisosteres **10a**−**f**,**h**,**i**, although slightly more permeable than the strongly hydrophilic carboxylic acid derivative **9**, show only low values of permeability compared to *n*-heptyl α-D-mannopyranoside (**1**) or the unsubstituted biphenyl mannoside **3**. However, the *p*-cyanobiphenyl derivatives **10g** and **10j** display elevated log $D_{7.4}$ and effective permeability (log P_e) in the range for successful intestinal absorption. Regarding both sufficient aqueous solubility and elevated membrane permeability, the *p*-cyano substituted bioisosteres **10g** and **10j** are thus the most promising candidates for oral absorption. Moreover, combining the bioisosteric replacement with the addition of a chloro substituent in the ortho-position of the aromatic ring adjacent to the anomeric oxygen (→**10j**)[22] resulted in the most advantageous physicochemical profile for oral bioavailability.

Renal Excretion. The rate of renal excretion depends on the rate of glomerular filtration and the propensity to tubular secretion and reabsorption of an antagonist.[71] Only the fraction that is not bound to plasma proteins is expected to enter the glomerular filtrate.[72] Plasma protein binding (PPB) data indicating the fraction bound (f_b) are listed in Table 2.[62] The biphenyls **9** and **10c** were identified as moderate binders to plasma proteins ($f_b \leq 65\%$), which suggests a low impact of PPB on antagonist filtration. The f_b values of the antagonists **1**, **3**, **20**, and **10j** were between 80% and 93%, whereas the bioisosteres **10d**,**e**,**g** showed particularly high protein binding ($f_b \geq 99\%$) implying slow compound entry into the primary urine. However, the kinetic aspects of PPB, that is, association and dissociation rate constants, remain to be determined to quantify precisely the influence of PPB on filtration.[73]

Furthermore, log $D_{7.4}$ was identified as key determinant of tubular reabsorption.[74−76] Accordingly, lipophilic compounds are predominantly reabsorbed from the renal filtrate. Given that renal clearance is the major route of elimination, this will result in a slow but steady excretion into the bladder. In contrast, hydrophilic compounds are poorly reabsorbed and thus quickly renally eliminated, which leads to high initial compound levels in the urine but narrows the time range where the minimal antiadhesive concentration is maintained. Consequently, low log $D_{7.4}$ as shown for the antagonists **9**, **10h**, and **10i** implies low tubular reabsorption and rapid elimination of the filtered molecules by the urine. Otherwise, log $D_{7.4}$ between 0.2 and 0.7, such as determined for the bioisosteres **10a**-**e**, suggests

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

Journal of Medicinal Chemistry Article

Table 4. Pharmacokinetic Parameters Determined after a Single iv Application of Compounds 9, 10c, 10e, and 10j in Female C3H/HeN Mice[a]

	plasma						
compd	C_0 (μg/mL)	dose (mg/kg)	V_z (mL)	$t_{1/2}$ (h)	AUC_{0-inf} (μg·h/mL)	CL_{tot} (mL/h)	urine, C_{max} (μg/mL)
9	40	50	25.2	0.33	23.5	53.1	300
10c	109.7	50	28.3	0.4	25.3	49.4	4611
10e	151.6	50	19.5	1.9	175.1	7.1	387
10j	0.36	0.625	52.8	0.17	0.07	218	10

[a]Values were calculated using PKSolver.[78] C_0, initial concentration; V_z, volume of distribution in terminal phase; AUC, area under the curve; CL_{tot}, total clearance; C_{max}, maximal concentration.

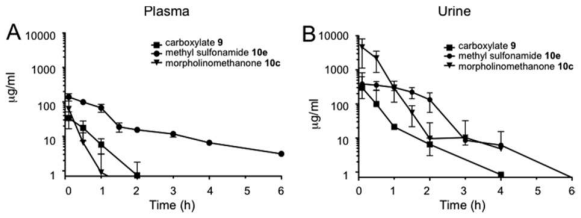


Figure 5. Antagonist concentrations in (A) plasma and (B) urine after a single iv application of **9**, **10c**, and **10e** (50 mg/kg).

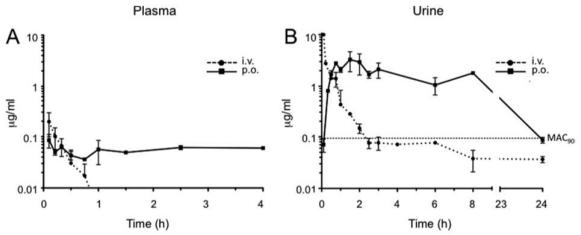


Figure 6. Antagonist concentrations in (A) plasma and (B) urine after a single iv and po application of compound **10j** (iv, 0.625 mg/kg; po, 1.25 mg/kg). MAC_{90} is the minimal antiadhesive concentration to inhibit 90% adhesion (0.094 μg/mL).

increasing propensity to tubular reuptake, whereas log $D_{7.4} > 1$ as shown for heptyl mannoside **1** and the biphenyl mannosides **3**, **20**, **10g**, **10f**, and **10j** is optimal for tubular reabsorption from the glomerular filtrate and thus for slow renal clearance.

Metabolic Stability. Increasing lipophilicity is usually paralleled by increasing susceptibility to metabolism.[77] Liabilities toward metabolic clearance pathways that prevent the intact antagonist from reaching the target in the bladder were therefore of interest. To assess their propensity to cytochrome P450 (CYP450) mediated metabolism, heptyl mannoside **1**, the carboxylic acid derivative **9**, and the bioisosteres **10c**−**e**,**g**,**j** were incubated with rat liver microsomes (RLM, 0.5 mg/mL) in the presence of the cofactor β-nicotinamide adenine dinucleotide phosphate (NADPH).[63] To confirm the high propensity of the methyl ester present in antagonist **20** to carboxylesterase (CES) mediated hydrolysis, this antagonist was incubated with RLM only. The profiles of unchanged compound versus time revealed high susceptibility of heptyl mannoside **1** to CYP450-mediated metabolism ($t_{1/2}$ = 13 min) and rapid hydrolysis of the ester **20** by the hepatic CES ($t_{1/2}$ = 1.0 min). Otherwise, the bioisosteres **10c**−**e**,**g**,**j** were stable against enzyme-mediated bioconversion ($t_{1/2}$ > 60 min), suggesting lower propensity to metabolic, nonrenal elimination pathways.

Considering PPB, lipophilicity, and metabolic stability data, we therefore expected (i) a steady release of compounds **10d**,**e**,**g**,**j** into the bladder because of high PPB decelerating glomerular filtration (**10d**,**e**,**g**) and/or high log $D_{7.4}$ supporting tubular reabsorption (**10g**,**j**), (ii) a fast excretion of antagonists **9** and **10c** via the urine due to low PPB and low log $D_{7.4}$, and (iii) a rapid clearance of heptyl mannoside **1** from the body by renal and metabolic pathways. Compounds featuring high propensity to renal excretion as major route of elimination (**10c**, **10e** and **10j**) were selected for in vivo PK studies in a mouse model.

Pharmacokinetic Studies in C3H/HeN Mice. This first part of our study explored the predicted effects of lipophilicity, PPB, and metabolic stability on antagonist disposition and elimination upon a single dose iv application (50 mg/kg) of compounds **10c** and **10e**. The PK parameters of these applications and those of the previously published carboxylate **9** are summarized in Table 4. The table also contains the results of the iv administration of compound **10j** (0.625 mg/kg).

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

Journal of Medicinal Chemistry Article

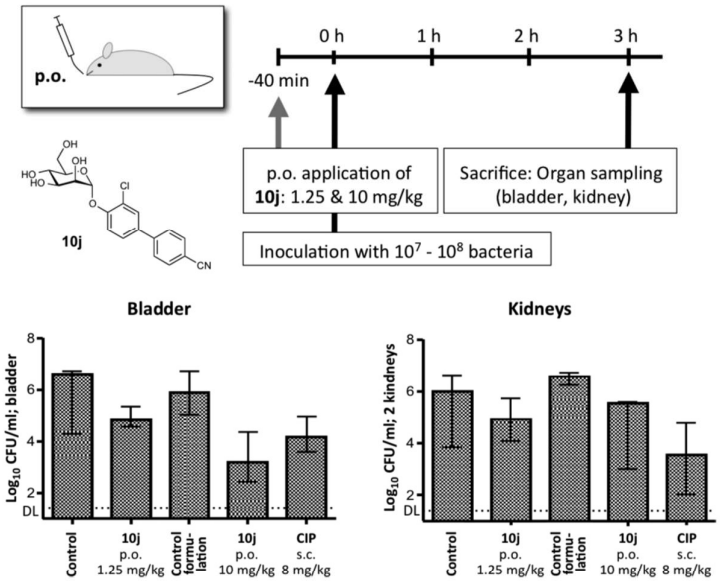


Figure 7. Preventive efficacy of **10j** in the UTI mouse model 3 h after infection. The bars depict the median bacterial load with the interquartile range in the different study groups. Shown are the results of the control group (PBS), control group formulation (5% DMSO in PBS containing 1% Tween 80), and the intervention groups with the preventive applications of either 1.25 or 10 mg/kg **10j** po or 8 mg/kg CIP sc (representing the murine dose equivalent to a human standard dose).[81] DL, detection limit. CFU, colony forming units.

In contrast to the fast plasma clearance of antagonists **9** and **10c** (Figure 5A), the methylsulfonamide bioisostere **10e** attained higher initial concentration in plasma (C_0) and lower total clearance (CL_{tot}). Therefore, it could be detected until 6 h after application, resulting in markedly higher plasma AUC. The observed high C_0 of compound **10e** may be attributed to a small volume of distribution (V_z) resulting from the high PPB ($f_b \geq 99\%$).[72] In urine (Figure 5B), the carboxylic acid **9** and the morpholinomethanone **10c** displayed high levels immediately following administration and a rapid concentration decrease within the first 2 h, reflecting the rapid elimination from plasma. Fast renal excretion as major route of elimination can be rationalized by the physicochemical properties of the antagonists **9** and **10c**, that is, moderate PPB and log $D_{7.4}$, as well as high metabolic stability. Otherwise, the methylsulfonamide bioisostere **10e** showed sustained compound levels in urine over a period of 2 h and subsequent slow decrease until 6 h after administration. This sustained renal excretion is a result of the interplay of the antagonist's elevated PPB and log $D_{7.4}$.

In a second study, the *p*-cyano bioisostere **10j**, characterized by a high oral absorption potential, was administered as a single dose iv (0.625 mg/kg) and po (1.25 mg/kg). The plasma concentration curve upon iv dosing displays a steep decline within the first hour after application, while the po curve shows a prolonged period where absorption and elimination are in equilibrium (Figure 6A). The urine concentration profiles (Figure 6B) parallel the plasma curves obtained by the two modes of application; i.e., high plasma clearance upon iv bolus injection led to high initial antagonist levels in urine and a rapid concentration decline. By contrast, sustained plasma concentrations upon po administration resulted in prolonged urine levels.

As a result, urine concentrations exceed the minimum level required for the antiadhesive effect as estimated from the in vitro cell infection model[79] (minimal antiadhesion concentration,[23] $MAC_{90} = 0.094\ \mu g/mL$) for more than 8 h upon oral single-dose administration (Figure 6B).

Infection Study in C3H/HeN Mice. In a preventive study, six mice were inoculated with UTI89 following an oral application of **10j** (1.25 mg/kg) 40 min prior to infection. Three hours after inoculation, the animals were sacrificed and bladder and kidneys were removed. Organs were homogenized and analyzed for bacterial counts. The effect of the FimH antagonist was compared to a 8 mg/kg dose of ciprofloxacin (CIP), applied subcutaneously (sc) 10 min before infection. CIP is used as standard antibiotic therapy in humans for the treatment of UTI.[80] In mice, the dose of 8 mg/kg sc was shown to mimic the standard human dose regarding peak levels and the AUC_{24} in serum.[81] The median reductions in bacterial counts in mice treated with **10j** and CIP compared to the control group 3 h after infection are displayed in Figure 7.

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

The median value in the untreated control group showed bacterial counts of 6.6 $\log_{10}$ colony forming units (CFU) in the bladder and 6 $\log_{10}$ CFU in the kidneys. After oral application of 1.25 mg/kg **10j**, bacterial loads in the bladder decreased by 1.78 $\log_{10}$ CFU and 1.07 $\log_{10}$ CFU in the kidneys. The lower reduction in the kidneys is most likely due to the differing adhesion mechanisms between bladder and kidneys (type 1 pili vs P-pili), which is not targeted by **10j**.[82] With CIP (8 mg/kg sc) a substantial reduction in both bladder and kidneys (median reductions of 2.44 $\log_{10}$ and 2.47 $\log_{10}$, respectively) was observed. Despite the low oral dose of **10j** (1.25 mg/kg), the approximately 100-fold reduction of CFU in the bladder promised an even higher effect upon dose increase to 10 mg/kg. Since the solubility of **10j** for this increased dose is too low (192 μg/mL), we used 5% DMSO and surfactant Tween 80 (1%) as solubilizer. To effectively compare the effect of a higher dose of **10j**, a control group receiving the formulation only (5% DMSO in PBS containing 1% Tween 80, termed control group formulation) was tested in parallel. When 10 mg/kg **10j** was applied, bacterial loads in the bladder decreased by 2.68 $\log_{10}$ CFU/mL compared to the control group formulation, clearly exceeding the effect of CIP with a reduction of 2.44 $\log_{10}$ CFU/mL. However, only a moderate reduction of 1.04 $\log_{10}$ CFU was achieved in the kidneys.

■ SUMMARY AND CONCLUSION

Recently, numerous monovalent alkyl and aryl α-D-mannopyranosides have been described as potent FimH antagonists. However, most of them suffer from insufficient pharmacokinetic properties, i.e., modest bioavailability and short duration of the therapeutic effect in the bladder, their site of action. As a consequence, high doses at short intervals are required to achieve antiadhesive effects over an extended period of time. Therefore, the goal of the present study was an appropriate optimization of the pharmacokinetic profile of biphenyl α-D-mannopyranosides while keeping their high affinity to the CRD of FimH. The starting point was the biphenylcarboxylate **9** where the critical carboxylate was replaced by bioisosteres.[39,83]

With a series of bioisosteres, a 3- to 5-fold improvement of affinity was achieved compared to **9**. Although binding necessitates only partial desolvation of the carboxylate and its bioisosteric replacements, a reduction of the enthalpy penalty for desolvation[51] was identified as the source of the improved affinity exhibited by the bioisosteres. Thermodynamic evaluation of antagonists **10b**−**e** revealed almost identical enthalpy contribution to binding. However, for antagonists with the *p*-cyano substituent (**10g** and **10j**) an enhancement of up to −8.7 kJ/mol was observed, indicating a reduced desolvation penalty and an improved stacking as derived from the crystal structure of **10j** cocrystallized with the CRD of FimH (Figure 4B). On the other hand, higher affinity originating from a reduction of conformational flexibility of ligand and protein resulted in a concomitant entropy penalty of up to 6.5 kJ/mol.

In addition to the improved pharmacodynamics, the relevant pharmacokinetic parameters (solubility, permeability, renal excretion) were substantially improved. With 3′-chloro-4′-(α-D-mannopyranosyloxy)biphenyl-4-carbonitrile (**10j**), a FimH antagonist with an optimal in vitro PK/PD profile was identified. The *p*-cyano substituent conferred lipophilicity and high binding to plasma proteins, which slowed the rate of renal excretion. Despite higher lipophilicity, antagonist **10j** was insusceptible to CYP450-mediated metabolism and therefore predominantly eliminated via the renal pathway. In vivo experiments confirmed the excellent PK profile of **10j** with steady renal excretion for more than 8 h after oral application (1.25 mg/kg), suggesting a long-lasting antiadhesive effect. Finally, the preventive oral application of **10j** (10 mg/kg) reduced the bacterial load in the bladder by almost 1000-fold 3 h after infection. Although the first 3 h of the infection do not represent the complete infection cycle, they represent the time span of bacteria adhering and invading urothelial cells.[84,85] Nevertheless, the effect of FimH antagonist **10j** within a longer infection time and at higher dosing will be the subject of future investigations.

■ EXPERIMENTAL SECTION

Synthesis. The synthesis of compounds **10a**−**d**, **10f**, **10g**, **10i**, **13a**−**d**, **13f**, **13g**, **15**, **18**, and **25**, including compound characterization data, can be found in the Supporting Information.

General Methods. NMR spectra were recorded on a Bruker Avance DMX-500 (500.1 MHz) spectrometer. Assignment of ^{1}H and ^{13}C NMR spectra was achieved using 2D methods (COSY, HSQC, HMBC). Chemical shifts are expressed in ppm using residual $CHCl_3$, CHD_2OD, or HDO as references. Optical rotations were measured using PerkinElmer polarimeter 341. Electron spray ionization mass spectra were obtained on a Waters micromass ZQ. The LC/HRMS analyses were carried out using a Agilent 1100 LC equipped with a photodiode array detector and a Micromass QTOF I equipped with a 4 GHz digital time converter. Microwave-assisted reactions were carried out with a CEM Discover and Explorer. Reactions were monitored by TLC using glass plates coated with silica gel 60 F_{254} (Merck) and visualized by using UV light and/or by charring with a molybdate solution (a 0.02 M solution of ammonium cerium sulfate dihydrate and ammonium molybdate tetrahydrate in aqueous 10% H_2SO_4). MPLC separations were carried out on a CombiFlash Companion or Rf (Teledyne Isco) equipped with RediSep normal-phase or RP-18 reversed-phase flash columns. LC−MS separations were done on a Waters system equipped with sample manager 2767, pump 2525, PDA 2525, and Micromass ZQ. All compounds used for biological assays are at least of 95% purity based on HPLC analytical results. Commercially available reagents were purchased from Fluka, Aldrich, Alfa Aesar, or abcr GmbH & Co. KG (Germany). Solvents were purchased from Sigma-Aldrich or Acros and were dried prior to use where indicated. Methanol (MeOH) was dried by refluxing with sodium methoxide and distilled immediately before use. Dimethoxyethane (DME) was dried by filtration over Al_2O_3 (Fluka, type 5016 A basic).

4′-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyloxy)-*N*-methylbiphenyl-4-sulfonamide (13e). A Schlenk tube was charged with aryl iodide **11**[22] (116 mg, 0.21 mmol), 4-(*N*-methylsulfamoyl)-phenylboronic acid (**12e**, 50 mg, 0.23 mmol), $Pd(dppf)Cl_2{\cdot}CH_2Cl_2$ (5 mg, 0.006 mmol), K_3PO_4 (67 mg, 0.32 mmol), and a stirring bar. The tube was closed with a rubber septum and was evacuated and flushed with argon. This procedure was repeated once, and then anhydrous DMF (1 mL) was added under a stream of argon. The mixture was degassed in an ultrasonic bath and flushed with argon for 5 min and then stirred at 80 °C overnight. The reaction mixture was cooled to rt, diluted with EtOAc (50 mL), and washed with water (50 mL) and brine (50 mL). The organic layer was dried over Na_2SO_4 and concentrated in vacuo. The residue was purified by MPLC on silica gel (petroleum ether/EtOAc) to afford **13e** (105 mg, 84%) as a white solid. $[\alpha]_D^{20}$ +56.4 (*c* 0.50, MeOH). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.92−7.90 (m, 2H, Ar−H), 7.70−7.68 (m, 2H, Ar−H), 7.57−7.55 (m, 2H, Ar−H), 7.21−7.19 (m, 2H, Ar−H), 5.60−5.57 (m, 2H, H-1, H-3), 5.48 (dd, *J* = 1.8, 3.4 Hz, 1H, H-2), 5.40 (t, *J* = 10.0 Hz, 1H, H-4), 4.38 (dd, *J* = 5.4, 10.8 Hz, 1H, NH), 4.30 (dd, *J* = 4.9, 12.3 Hz, 1H, H-6a), 4.13−4.08 (m, 2H, H-5, H-6b), 2.72 (d, *J* = 5.4 Hz, 3H, NCH_3), 2.22, 2.07, 2.05, 2.04 (4 s, 12H, 4 $COCH_3$). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 170.55, 170.06, 170.03, 169.75 (4 CO), 155.97, 144.81, 137.16, 134.09, 128.62, 127.85, 127.39, 117.01 (Ar−C), 95.78 (C-1), 69.34 (C-5), 69.31 (C-2), 68.81 (C-3), 65.86 (C-4), 62.07 (C-6), 29.44 ($NHCH_3$), 20.92, 20.74, 20.72 (4C, 4 $COCH_3$). ESI-MS *m*/*z*, calcd for $C_{27}H_{31}NNaO_{12}S$ $[M + Na]^+$: 616.1. Found: 616.1.

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

4′-(α-D-Mannopyranosyloxy)-*N*-methylbiphenyl-4-sulfonamide (10e). To a solution of **13e** (40 mg, 0.07 mmol) in dry MeOH (5 mL) was added freshly prepared 1 M NaOMe/MeOH (0.1 equiv) under argon. The mixture was stirred at rt until the reaction was complete (monitored by TLC), then neutralized with Amberlyst-15 (H^+) ion-exchange resin, filtered, and concentrated in vacuo. The residue was purified by MPLC on silica gel (DCM/MeOH, 10:1 to 7:1) to afford **10e** (22 mg, 76%) as white solid. $[\alpha]_D^{20}$ +105.7 (*c* 0.30, MeOH). ^{1}H NMR (500 MHz, CD_3OD): δ = 7.90−7.88 (m, 2H, Ar−H), 7.80−7.79 (m, 2H, Ar−H), 7.66−7.64 (m, 2H, Ar−H), 7.26−7.25 (m, 2H, Ar−H), 5.58 (d, *J* = 1.7 Hz, 1H, H-1), 4.06 (dd, *J* = 1.8, 3.3 Hz, 1H, H-2), 3.96 (dd, *J* = 3.4, 9.5 Hz, 1H, H-3), 3.79−3.74 (m, 3H, H-4, H-6a, H-6b), 3.63 (ddd, *J* = 2.5, 5.2, 9.7 Hz, 1H, H-5), 2.57 (s, 3H, $NHCH_3$). ^{13}C NMR (126 MHz, CD_3OD): δ = 158.34, 146.13, 138.67, 134.55, 129.53, 128.82, 128.21, 118.29 (Ar−C), 100.09 (C-1), 75.53 (C-5), 72.42 (C-3), 71.96 (C-2), 68.32 (C-4), 62.68 (C-6), 29.31 ($NHCH_3$). HRMS *m/z*, calcd for $C_{19}H_{23}NNaO_8S$ [M + Na]$^+$: 448.1037. Found: 448.1038.

5-(4′-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyloxy)-biphenyl-4-yl)-1*H*-tetrazole (14). A Schlenk tube was charged with **13g** (30 mg, 0.06 mmol), trimethylsilyl azide (16 μL, 0.12 mmol), dibutyltin oxide (2 mg, 0.006 mmol), DME (1 mL), and a stirring bar. The mixture was heated to 150 °C for 10 min by microwave irradiation. The reaction mixture was cooled to rt and then concentrated in vacuo. The residue was purified by MPLC on silica gel (DCM/MeOH, 9:1 to 8:1) to afford **14** (26 mg, 81%) as a colorless oil. $[\alpha]_D^{20}$ +56.1 (*c* 0.3, MeOH). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.25−8.15 (m, 2H, Ar−H), 7.75−7.65 (m, 2H, Ar−H), 7.60−7.55 (m, 2H, Ar−H), 7.20−7.17 (m, 2H, Ar−H), 5.64−5.55 (m, 2H, H-1, H-3), 5.49 (dd, *J* = 1.7, 3.3 Hz, 1H, H-2), 5.40 (t, *J* = 10.1 Hz, 1H, H-4), 4.31 (dd, *J* = 5.3, 12.4 Hz, 1H, H-6a), 4.17−4.06 (m, 2H, H-5, H-6b), 2.22, 2.07, 2.06, 2.05 (4 s, 12H, 4 $COCH_3$). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 170.67, 170.14, 170.11, 169.81 (4 CO), 155.61, 128.36, 127.84, 127.49, 116.93 (Ar−C), 95.78 (C-1), 69.36 (C-5), 69.26 (C-2), 68.90 (C-3), 65.89 (C-4), 62.12 (C-6), 20.92, 20.76, 20.73 (4 $COCH_3$). ESI-MS *m/z*, calcd for $C_{27}H_{28}N_4NaO_{10}$ [M + Na]$^+$: 591.2. Found: 591.1.

5-(4′-(α-D-Mannopyranosyloxy)biphenyl-4-yl)-1*H*-tetrazole (10h). Prepared according to the procedure described for **10e** from **14** (26 mg, 0.03 mmol). Yield: 18 mg (quant) as a white solid. $[\alpha]_D^{20}$ +112.1 (*c* 0.1, MeOH/H_2O, 2:1). ^{1}H NMR (500 MHz, CD_3OD): δ = 7.98−7.96 (m, 2H, Ar−H), 7.72−7.71 (m, 2H, Ar−H), 7.58−7.54 (m, 2H, Ar−H), 7.16−7.13 (m, 2H, Ar−H), 5.46 (d, *J* = 1.7 Hz, 1H, H-1), 3.94 (dd, *J* = 1.9, 3.5 Hz, 1H, H-2), 3.83 (dd, *J* = 3.4, 9.5 Hz, 1H, H-3), 3.68−3.61 (m, 3H, H-4, H-6a, H-6b), 3.52 (ddd, *J* = 2.5, 5.4, 9.7 Hz, 1H, H-5). ^{13}C NMR (126 MHz, CD_3OD): δ = 158.19, 145.07, 134.97, 129.29, 128.74, 128.55, 118.26 (Ar−C), 100.13 (C-1), 75.52 (C-5), 72.42 (C-3), 71.98 (C-2), 68.33 (C-4), 62.69 (C-6). HRMS *m/z*, calcd for $C_{19}H_{21}N_4O_6$ [M + H]$^+$: 401.1456. Found: 401.1450.

4′-(2,3,4,6-Tetra-*O*-acetyl-α-D-mannopyranosyloxy)-3′-chloro-biphenyl-4-carbonitrile (19). Prepared according to the procedure described for **13e** from aryl iodide **18**[23] (79 mg, 0.135 mmol), **12g** (22 mg, 0.15 mmol), Pd(dppf)Cl_2·CH_2Cl_2 (3.3 mg, 4 μmol), and K_3PO_4 (57 mg, 0.27 mmol). Yield: 57 mg (75%) as a white solid. $[\alpha]_D^{20}$ +77.7 (*c* 0.5, $CHCl_3$). ^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.72 (d, *J* = 8.3 Hz, 2H, Ar−H), 7.63 (m, 3H, Ar−H), 7.43 (dd, *J* = 2.2, 8.6 Hz, 1H, Ar−H), 7.27 (d, *J* = 8.6 Hz, 1H, Ar−H), 5.64−5.59 (m, 2H, H-1, H-2), 5.54 (dd, *J* = 1.9, 3.2 Hz, 1H, H-3), 5.41 (t, *J* = 10.1 Hz, 1H, H-4), 4.28 (dd, *J* = 5.2, 12.3 Hz, 1H, H-6a), 4.17 (ddd, *J* = 2.1, 5.1, 10.0 Hz, 1H, H-5), 4.10 (dd, *J* = 2.2, 12.3 Hz, 1H, H-6b), 2.21 (s, 3H, $COCH_3$), 2.12- 2.00 (m, 9H, 3 $COCH_3$). ^{13}C NMR (126 MHz, $CDCl_3$): δ = 170.54, 170.08, 169.90, 169.84, (4C, CO) 151.67, 143.61, 135.29, 132.87, 129.41, 127.53, 126.60, 125.20, 118.79, 117.36, 111.47 (Ar−C, CN), 96.72 (C-1), 70.00 (C-5), 69.39 (C-3), 68.82 (C-2), 65.86 (C-4), 62.16 (C-6), 20.98, 20.81, 20.79, 20.78 (4 $COCH_3$). ESI-MS *m/z*, calcd for $C_{27}H_{26}ClNNaO_{10}$ [M + Na]$^+$: 582.1. Found: 582.1.

3′-Chloro-4′-(α-D-mannopyranosyloxy)biphenyl-4-carbonitrile (10j). Prepared according to the procedure described for **10e** from **19** (36 mg, 0.06 mmol). Yield: 12 mg (48%) as a white solid. $[\alpha]_D^{20}$ +109.4 (*c* 0.23, MeOH). ^{1}H NMR (500 MHz, CD_3OD): δ = 7.80−7.72 (m, 5H, Ar−H), 7.59 (dd, *J* = 2.2, 8.6 Hz, 1H, Ar−H), 7.48 (d, *J* = 8.7 Hz, 1H, Ar−H), 5.62 (d, *J* = 1.4 Hz, 1H, H-1), 4.12 (dd, *J* = 1.8, 3.3 Hz, 1H, H-2), 4.00 (dd, *J* = 3.4, 9.5 Hz, 1H, H-3), 3.83−3.68 (m, 3H, H-4, H-6a, H-6b), 3.63 (ddd, *J* = 2.3, 5.4, 9.6 Hz, 1H, H-5). ^{13}C NMR (126 MHz, CD_3OD): δ = 153.65, 145.15, 135.42, 133.86, 129.82, 128.53, 127.87, 125.47, 119.70, 118.59 (Ar−C), 111.97 (CN), 100.66 (C-1), 76.05 (C-5), 72.39 (C-3), 71.80 (C-2), 68.20 (C-4), 62.65 (C-6). IR (KBr), ν = 3400 (OH), 2227 (C≡N), 1606, 1487 (Ar−C=C) cm^{-1}. HRMS *m/z*, calcd for $C_{19}H_{18}ClNNaO_6$ [M + Na]$^+$: 414.0715. Found: 414.0721.

3′-Chloro-*N*-(3′,6′-dihydroxy-3-oxo-3*H*-spiro-[isobenzofuran-1,9′-xanthen]-5-yl)-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxamide (22). Compound **21** (10.0 mg, 0.024 mmol), fluoresceinamine isomer I (12.7 mg, 0.037 mmol), and COMU (20.9 mg, 0.049 mmol) were dissolved in dry DMF (1 mL). Then NEt_3 (10 μL, 0.073 mmol) was added and the mixture was stirred at rt for 7 h. 1 N HCl in DMF was added until acid reaction on pH paper and the mixture was concentrated. The residue was dissolved in DCM/MeOH (3:1) and loaded onto a silica gel column. The complex mixture of compounds was only partially resolved. The fractions containing the product were collected, concentrated, and purified by preparative HPLC (gradient H_2O/MeCN, +0.2% HCO_2H) to afford compound **22** (5 mg, 19%). $[\alpha]_D^{20}$ +21.1 (*c* 0.10, MeOH). ^{1}H NMR (500 MHz, CD_3OD): δ = 8.26 (d, *J* = 8.4 Hz, 2H, Ar−H), 7.88−7.74 (m, 3H, Ar−H), 7.66 (dd, *J* = 2.2, 8.6 Hz, 1H, Ar−H), 7.51 (d, *J* = 8.7 Hz, 1H, Ar−H), 7.29 (dd, *J* = 1.9, 5.3 Hz, 2H, Ar−H), 7.19 (dd, *J* = 2.1, 8.3 Hz, 1H, Ar−H), 7.08−6.99 (m, 2H, Ar−H), 6.95 (d, *J* = 8.7 Hz, 1H, Ar−H), 6.72 (dd, *J* = 5.5, 10.6, Hz, 2H, Ar−H), 6.61 (dd, *J* = 2.3, 8.7 Hz, 1H, Ar−H), 5.65 (s, 1H, H-1), 4.15 (dd, *J* = 1.8, 3.2 Hz, H-2), 4.03 (dd, *J* = 3.4, 9.5, Hz, H-3), 3.87−3.72 (m, 3H, H-4, H-6a, H-6b), 3.65 (m, 1H, H-5). ^{13}C NMR (126 MHz, CD_3OD): δ = 137.50, 136.01, 131.90, 130.24, 130.20, 129.87, 129.24, 128.03, 127.91, 125.79, 125.46, 124.73, 118.99, 118.76, 118.65 (Ar−C), 100.73 (C-1), 76.06 (C-5), 72.42 (C-3), 71.85 (C-2), 68.24 (C-4), 62.69 (C-2). ESI-MS *m/z*, calcd for $C_{39}H_{31}ClNO_{12}$ [M + H]$^+$: 740.2. Found: 740.2.

3′-Chloro-*N*-(2-(3-(3′,6′-dihydroxy-3-oxo-3*H*-spiro-[isobenzofuran-1,9′-xanthen]-5-yl)thioureido)ethyl)-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxamide (23). To a stirred solution of compound **21** (25 mg, 0.061 mmol) in dry DMF (1 mL), NHS (21 mg, 0.183 mmol) was added, followed by DIC (9.2 mg, 0.073 mmol). The mixture was stirred at rt for 2 h. Then *N*-Boc-ethylendiamine (10.7 mg, 0.067 mmol) was added and the reaction was stirred for 10 h. It was then cooled down to 0 °C, diluted with water, and concentrated. Chromatography on silica gel (DCM/MeOH) yielded 23 mg (0.042 mmol, 68%) of *tert*-butyl (3′-chloro-4′-(α-D-mannopyranosyloxy)biphenyl-4-yl-carboxamido)ethyl)carbamate. This product was dissolved in DCM (3 mL), and TFA (1 mL) was added. The solid dissolved during addition of TFA. After 10 min the reaction was complete. The mixture was evaporated, and excess TFA was removed in high vacuum. The intermediate *N*-(2-aminoethyl)-3′-chloro-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxamide TFA salt (23 mg, 0.042 mmol, quant) was used directly in the next step. It was dissolved in dry DMF (0.5 mL), and NEt_3 (12.8 mg, 0.127 mmol) was added. The mixture was cooled to 0 °C. Then FITC (14.8 mg, 0.038 mmol) was added and the mixture was stirred for 3 h in the dark. The mixture was then coevaporated with water, taken up in MeOH/10% aq acetic acid and evaporated. Chromatography on silica gel (DCM/MeOH) yielded compound **23**, contaminated with triethylammonium acetate. The compound was then redissolved in MeOH, and 0.5 N HCl in MeOH was added. The mixture was evaporated and chromatographed on silica gel to yield pure **23** (15 mg, 47%). $[\alpha]_D^{20}$ +12.1 (*c* 0.30, MeOH). ^{1}H NMR (500 MHz, CD_3OD): δ = 8.12 (s, 1H), 7.92 (d, *J* = 8.3 Hz, 2H, Ar−H), 7.70 (dd, *J* = 5.0, 13.1 Hz, 2H, Ar−H), 7.64 (d, *J* = 8.3 Hz, 2H, Ar−H), 7.54 (dd, *J* = 2.2, 8.6 Hz, 1H, Ar−H), 7.46 (d, *J* = 8.7 Hz, 1H, Ar−H), 7.09 (d, *J* = 8.2 Hz, 1H, Ar−H), 6.74 (s, 2H), 6.69 (d, *J* = 1.4 Hz, 2H, Ar−H), 6.55 (d, *J* = 8.4 Hz, 2H, Ar−H), 5.63 (d, *J* = 1.3 Hz, H-1), 4.15 (dd, *J* = 1.8, 3.1 Hz, H-2), 4.03 (dd, *J* = 3.4, 9.5 Hz, H-3), 3.94 (s, 2H, CH_2), 3.86−3.64 (m, 6H, H-4, H-5, H-6, CH_2). ^{13}C NMR (126 MHz, CD_3OD): δ = 153.21, 143.84, 136.41, 129.66, 129.18, 127.76, 127.70, 125.37, 118.64, 103.62 (Ar−C), 100.75 (C-1), 76.00 (C-5), 72.41 (C-3), 71.86 (C-2), 68.24 (C-4), 62.69 (C-6), 40.76 (CH_2). ESI-MS *m/z*, calcd for $C_{42}H_{37}ClN_3O_{12}S$ [M + H]$^+$: 842.2. Found: 842.2.

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

3′-Chloro-*N*-(2-(2-(2-(3-(3′,6′-dihydroxy-3-oxo-3*H*-spiro-[isobenzofuran-1,9′-xanthen]-5-yl)thioureido)ethoxy)ethoxy)-ethyl)-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxamide (24). Compound **21** (280 mg, 0.68 mmol) was dissolved in dry DMF (5 mL) under argon. Then NHS (235 mg, 2.04 mmol) was added, followed by DIC (0.12 mL, 0.78 mmol) and the mixture was stirred at rt for 4 h. Then Boc-PEG2-NH_2 (186 mg, 0.75 mmol) was added, and the mixture was stirred at rt under argon for 10 h. It was then slowly diluted with water and concentrated. The residue was purified by chromatography on silica gel (DCM/MeOH) to give *tert*-butyl (2-(2-(2-(3′-chloro-4′-(α-D-mannopyranosyloxy)biphenyl-4-ylcarboxamido)ethoxy)ethoxy)ethyl)-carbamate (300 mg, 0.468 mmol, 69%). Then the carbamate was suspended in DCM (3 mL), and TFA (1 mL) was added dropwise at rt. After 30 min, the solvents were evaporated and the crude mixture was dissolved in $CHCl_3$/MeOH (6:4, +0.5% conc NH_4OH) and transferred to a silica gel column, eluting with the same solvent mixture, to yield *N*-(2-(2-(2-aminoethoxy)ethoxy)ethyl)-3′-chloro-4′-(α-D-mannopyranosyloxy)biphenyl-4-carboxamide (228 mg, 90%). A fraction of the amine (10 mg, 0.018 mmol) was dissolved in dry DMF (0.5 mL) and cooled to 0 °C. FITC (6.5 mg, 0.017 mmol) was added, and the mixture was stirred for 1 h. The mixture was concentrated and the residue was purified by chromatography on silica (DCM/MeOH) to yield **24** (10 mg, 65%). ^{1}H NMR (500 MHz, CD_3OD): δ = 8.21 (d, *J* = 1.4 Hz, 1H, Ar—H), 7.88 (d, *J* = 8.3 Hz, 2H, Ar—H), 7.68 (d, *J* = 2.2 Hz, 2H, Ar—H), 7.63 (d, *J* = 8.3 Hz, 2H, Ar—H), 7.53 (dd, *J* = 2.2, 8.6 Hz, 1H, Ar—H), 7.43 (d, *J* = 8.7 Hz, 1H, Ar—H), 7.09 (d, *J* = 8.2 Hz, 1H, Ar—H), 6.68 (d, *J* = 2.3 Hz, 2H, Ar—H), 6.65 (dd, *J* = 2.6, 8.6 Hz, 2H, Ar—H), 6.53 (dd, *J* = 1.6, 8.7 Hz, 2H, Ar—H), 5.61 (d, *J* = 1.3 Hz, 1H, H-1), 4.14 (dd, *J* = 1.8, 3.2, Hz, 1H, H-2), 4.03 (dd, *J* = 3.4, 9.5 Hz, 1H, H-3), 3.93—3.53 (m, 16H), 3.37 (s, 2H, NCH_2), 1.30 (s, 2H, CH_2). ^{13}C NMR (126 MHz, CD_3OD): δ = 170.01 (CO), 153.17, 143.72, 136.37, 134.37, 130.39, 129.69, 129.04, 127.78, 127.73, 125.35, 118.60, 103.60 (Ar—C), 100.72 (C-1), 75.97 (C-5), 72.41 (C-3), 71.86, 71.40, 70.59 (5C, C-2, OCH_2), 68.23 (C-4), 62.64 (C-6), 49.88, 45.49, 40.97 (CH_2). ESI-MS *m*/*z*, calcd for $C_{46}H_{45}ClN_3O_{14}S$ $[M + H]^+$: 930.2. Found: 930.4.

Competitive Fluorescence Polarization Assay. *Expression and Purification of CRD of FimH.* A recombinant protein consisting of the CRD of FimH linked to a 6His-tag via a thrombin cleavage site (FimH-CRD-Th-His_6) was expressed in *E. coli* strain HM125 and purified by affinity chromatography as previously described.[43]

K_D Determination of FITC-Labeled Ligands. The functionalized ligands (**23**, **24**) were prepared as a 10 mM stock solution in pure DMSO (Sigma-Aldrich, Buchs, Switzerland). All further dilutions of compounds and FimH-CRD-Th-His_6 protein were prepared in assay buffer (20 mM HEPES, 150 mM NaCl, 50 μg/mL BSA, pH 7.4). BSA was added to the assay buffer to prevent nonspecific binding of protein to the plastic surface. Binding isotherms for the fluorescent ligands were obtained in direct binding studies by adding a constant concentration of ligand (final concentration 5 nM) and a linear dilution of protein (final concentration 0—100 nM) to a final volume of 200 μL in 96-well, black, flat bottom NBS plates (Corning Inc., Corning, NY, USA). After incubation of the plate for 24 h at rt with gentle shaking, the fluorescence polarization was measured with the Synergy H1 hybrid multimode microplate reader (BioTek Instruments Inc., Winooski, VT, USA) with polarized excitation at 485 nm and emission measured at 528 nm through polarizing filters parallel and perpendicularly oriented to the incident polarized light. K_D values were determined by plotting the FP readout as a function of the protein concentration and applying the following single-site binding equation (eq 1) that accounts for ligand depletion:

$$S_{obs} = S_F + (S_B - S_F) \times \left(\frac{C_P + C_L + K_D - \sqrt{(C_P + C_L + K_D)^2 - 4C_PC_L}}{2C_L} \right) \quad (1)$$

where S_{obs} is the observed signal from the ligand, S_F is the signal from free ligand, S_B is the signal from bound ligand, C_P is the total concentration of protein, and C_L is the total concentration of ligand.[49]

K_D Determination of FimH Antagonists. The fluorescently labeled ligand **23** was used for the competitive fluorescence polarization assay. A linear dilution of nonlabeled FimH antagonist with final concentrations ranging from 0 to 10 μM was titrated into 96-well, black, flat-bottom NBS plates (Corning Inc.) to a final volume of 200 μL containing a constant concentration of protein (final concentration 25 nM) and FITC-labeled ligand which was fixed at a higher concentration in competitive binding assays than in direct binding experiments to obtain higher fluorescence intensities (final concentration 20 nM). Prior to measuring the fluorescence polarization, the plates were incubated on a shaker for 24 h at rt until the reaction reached equilibrium. The IC_{50} value was determined with Prism (GraphPad Software Inc., La Jolla, CA, USA) by applying a standard four-parameter IC_{50} function. The obtained IC_{50} values were converted into their corresponding K_D values using the derivation of the Cheng—Prusoff equation.[45] This variation of the Cheng—Prusoff equation is applied to competition assays with tight-binding inhibitors and includes terms to correct for ligand depletion effects. However, the K_D for antagonists having a higher affinity toward FimH than the labeled ligand could not be accurately determined.[45]

Isothermal Titration Calorimetry (ITC). All ITC experiments were performed with the FimH-CRD-Th-His_6 protein using a VP-ITC instrument from MicroCal, Inc. (Malvern Instruments, Worcestershire, U.K.) with a sample cell volume of 1.4523 mL. The measurements were performed with 0—5% DMSO at 25 °C, a stirring speed of 307 rpm, and 10 μcal s^{-1} reference power. The protein samples were dialyzed in assay buffer prior to all experiments. Because of the high protein consumption of ITC, only the experiments for the reference compounds (**1**, **3**, and **25**) were measured in duplicates. Compounds **1**, **3**, **9**, and **25** were measured in a direct fashion by titration of ligand (100—2,000 μM) into protein (8.6—55 μM) with injections of 3—8 μL at intervals of 10 min to ensure nonoverlapping peaks. The quantity c = Mt(0) K_D^{-1}, where Mt(0) is the initial macromolecule concentration, is of importance in titration microcalorimetry. The *c*-values of the direct titrations were below 1000 and thus within the reliable range. For the compounds **10b—e**, **10g**, and **10j** additional competitive ITC experiments were performed because of their high affinity resulting in *c*-values above 1000 for direct titrations. These ligands (600 μM) were titrated into protein (30 μM), which was preincubated with compound **25** (300 μM) resulting in sigmoidal titration curves. Because of slow reaction kinetics, titration intervals of 20 min were used.

Baseline correction and peak integration were performed using the Origin 7 software (OriginLab, Northampton, MA, USA). An initial 2 μL injection was excluded from data analysis. Baseline subtraction and curve-fitting with the three variables N (concentration correction factor), K_D (dissociation constant), and $\Delta H°$ (change in enthalpy) were performed with the SEDPHAT software, version 10.40 (National Institutes of Health).[86] A global fitting analysis was performed for the competition titration (**10b—e**, **10g**, or **10j** competing for the protein binding site with compound **25**) and the direct titration of the competitor (compound **25** binding to protein) to fit for K_D. $\Delta H°$ and N were fitted from direct titrations of **10b—e**, **10g**, or **10j** into protein. For the compounds **3**, **9**, and **25** binding to protein all variables could be determined from a global analysis of the direct titration.

The thermodynamic parameters were calculated with the following equation (eq 2):

$$\Delta G° = \Delta H° - T\Delta S° = RT \ln K_D = -RT \ln K_A \quad (2)$$

where $\Delta G°$, $\Delta H°$, and $\Delta S°$ are the changes in free energy, enthalpy, and entropy of binding, respectively, T is the absolute temperature, and R is the universal gas constant (8.314 J mol^{-1} K^{-1}). The 95% confidence intervals of the measurements were calculated for the two variables K_D and $\Delta H°$ with the one-dimensional error surface projection within the SEDPHAT software.

Calculation of the Free Energy of Desolvation. The three-dimensional representation for each of the aglycons (4-methoxybiphenyl scaffold, Figure 8) was built in the Maestro[87] modeling environment, and the global minimum conformation was identified by performing 500 iterations of the mixed torsional/low-mode conformational sampling in combination with the OPLS-2005 force-field and the implicit solvent model (water) as implemented in the Macromodel

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221—2239

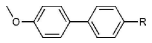


Figure 8. 4-Methoxybiphenyl scaffold of aglycons.

9.9.[88] The global minimum structures were used as input for the AMSOL 7.1 program[89] to obtain the free energy of desolvation ΔG_{des} (Table 5) with the SM5.4A solvation model[90] and the AM1[91] level of theory (used keywords "AM1 SM5.4A SOLVNT=WATER TRUES").

Table 5. Aqueous Free Energy of Desolvation

R	ΔG_{des} [kJ/mol]
neutral	
H	15.6
$CONHCH_3$	39.9
$COOCH_3$	23.0
SO_2NHCH_3	65.5
SO_2CH_3	56.4
4-morpholineamide	45.3
CN	22.0
deprotonated	
COO^-	298.2
SO_2-N^--Me	342.0

Determination of the MAC_{90} by Flow Cytometry. The MAC_{90} was determined in principle as in the previously published flow cytometry assay[79] but with some modifications. The human epithelial bladder carcinoma cell line 5637 (DSMZ, Braunschweig, Germany) was grown in RPMI 1640 medium, supplemented with 10% fetal calf serum (FCS), 100 U/mL penicillin, and 100 µg/mL streptomycin at 37 °C, 5% CO_2. All solutions were purchased from Invitrogen (Basel, Switzerland). The cells were subcultured 1:6 twice per week [using trypsin/EDTA (Sigma-Aldrich) for the detachment]. Two days before infection, 1.8 × 10^5 cells were seeded in each well of a 24-well plate in RPMI 1640 containing 10% FCS without antibiotics. The cell density was approximately (3−5) × 10^5 cells/well at the assay day.

For infection, the GFP-expressing clinical E. coli isolate UTI89[92] (UTI89 wt) and the GFP-expressing FimA-H knockout strain UTI89 ΔfimA-H were used (strains were provided by Prof. Urs Jenal, Biocenter, University of Basel, Switzerland).[79] Bacteria were cultivated at 37 °C in 10 mL Luria−Bertani (LB) broth (Becton, Dickinson and Company) overnight, harvested by centrifugation (3800 rpm, 10 min), and washed three times in phosphate buffered saline (PBS, Sigma-Aldrich), and a bacterial solution of OD_{600} of 0.75 in RPMI + 10% FCS was prepared. For the determination of the MAC_{90} value, the IC_{90}, linear dilutions of the FimH antagonist were prepared in 5% DMSO and PBS. Bacteria and antagonists were preincubated for 10 min at 37 °C, before cells were infected with either only 200 µL of bacterial solution of UTI89 or UTI89 ΔfimA-H (positive and negative controls), or 225 µL of the preincubated bacteria−antagonist mixture. Infection lasted for 1.5 h. During this time infected cells were incubated at 37 °C. Then, cells were washed with PBS and detached from wells by the addition of 150 µL of trypsin and incubation at 37 °C for 10 min, before flushing from wells PBS containing 2% FCS and transferred to tubes. To dilute the trypsin, cells were centrifuged at 13 000 rpm, 1 min, 600 µL of the supernatant was discarded, and the pellet was resuspended in the remaining 300 µL of PBS containing 2% FCS. Samples were stored on ice until measurement. Before analysis with the flow cytometer (Becton Dickinson, FACSCanto II), the samples were gently mixed and filtered using a 35 µm nylon mesh (Corning Life Sciences) to prevent cellular aggregation. Cells were gated with linear scaling for side scatter (SSC) and forward scatter (FSC) and GFP intensity of live cells was evaluated. IC_{90} values were determined by plotting the concentration of the antagonist in a logarithmic mode versus the mean fluorescence intensity (MFI) of living cells and by fitting a dose−response curve (variable slope, four parameters) with the Prism software (GraphPad Prism).

X-ray Analysis of the Antagonists 10e and 10j Cocrystallized with FimH-CRD. FimH-CRD/10e Cocrystallization. Initial FimH-CRD (18 mg/mL in 20 mM HEPES, pH 7.4) crystals were obtained in complex with 4-(5-nitroindolin-1-yl)phenyl α-D-mannopyranoside (5 mM).[23] Crystals were grown in sitting-drop vapor diffusion at 20 °C with 200 nL of protein−antagonist mixture together with 200 nL of precipitant solution in well D3 (0.2 M sodium phosphate monobasic monohydrate, 20% w/v PEG 3,350) of the PEG/Ion HT screen (Hampton Research, CA, USA). Cubic crystals appeared within 1 week, which served as cross-seeding crystals. A solution of FimH-CRD (20 mg/mL) and 10e (5 mM) was mixed with 0.2 M sodium phosphate monobasic monohydrate, 20% w/v PEG 400 with 0.5 µL of each solution. Streak-seeding was performed after 1 day of incubation. Cubic FimH-CRD/10e crystals formed within 24 h. Crystals were flash cooled to 100 K with perfluoropolyether cryo oil (Hampton Research, CA, USA) as cryoprotectant. Data were collected with synchrotron radiation (λ = 0.999 99 Å) at the PXIII beamline, Swiss Light Source, Switzerland.

FimH-CRD/10j Cocrystallization. Cocrystals were initially grown in sitting-drop vapor diffusion at 20 °C with 0.5 µL of a mixture of FimH-CRD (20 mg/mL) and 10j (5 mM) together with 0.5 µL of 0.1 M HEPES, pH 7.5, 2 M ammonium sulfate. Platelike crystals formed within 2 weeks and were used as seeds for subsequent crystallization. Diffraction quality crystals were grown by streak-seeding in 0.5 µL of FimH-CRD (10 mg/mL) with 10j (2.5 mM) and 0.5 µL of 0.1 M HEPES, pH 7.5, 1.25 M ammonium sulfate. The drops were covered with perfluoropolyether cryo oil prior to flash cooling to 100 K. Data were collected with synchrotron radiation (λ = 1.000 03 Å) at the PXIII beamline, Swiss Light Source, Switzerland.

Structure Determination and Refinement. Data were indexed and integrated with the XDS package[93] for the FimH-CRD/10e cocrystal structure, and with mosflm[94] for the FimH-CRD/10j cocrystal structure (Table 6). Scaling was performed with XDS and SCALA included in the CCP4 suite, respectively.[95] Structures were solved by molecular

Table 6. Data Collection and Refinement Statistics for FimH-CRD/10e and FimH-CRD/10j Cocrystals

	FimH-CRD/10e	FimH-CRD/10j
PDB code	4CSS	4CST
space group	$P2_12_12_1$	$P2_12_12_1$
no. of molecules in the asymmetric unit	1	1
Cell Dimensions		
a, b, c (Å)	48.38, 56.23, 61.59	48.84, 55.89, 61.00
α, β, γ (deg)	90, 90, 90	90, 90, 90
Data Collection		
beamline	Swiss Light Source PXIII	Swiss Light Source PXIII
resolution range (Å)[a]	30.0−1.07 (1.13−1.07)	23.5−1.10 (1.12−1.10)
unique observations[a]	72000 (9354)	66470 (2500)
average multiplicity[a]	10.9 (3.7)	5.4 (2.4)
completeness (%)	96.1 (78.0)	97.2 (76.5)
R_{merge}[a]	0.056 (0.57)	0.051 (0.305)
mean I/σ(I)[a]	21.5 (2.22)	15.5 (2.9)
Refinement		
resolution range (Å)	15.7−1.07	23.5−1.10
R, R_{free}	11.2, 13.2	11.4, 13.0
rms deviation from ideal bond length (Å)	0.010	0.010
rms deviation from ideal bond angle (deg)	1.170	1.420

[a]Values in parentheses are for highest-resolution shell.

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

replacement with PHASER[96] using the FimH-CRD-butyl α-D-mannopyranoside complex (PDB code 1UWF) as search model. The structures were iteratively built using the COOT software[97] and refined with the PHENIX software.[98] Geometric restraints for **10e** and **10j** were generated with PRODRG.[99] The models were validated using molprobity.[100] Residues 113−115 were not modeled in the **10e** structure because of disorder. Furthermore, the ligand was modeled in two possible conformations. For both ligands, electron density is reduced on the second aromatic ring because of flexibility of the ligand.

Physicochemical and in Vitro Pharmacokinetic Studies. *Materials.* Dimethyl sulfoxide (DMSO), 1-propanol, 1-octanol, Dulbecco's modified Eagle medium (DMEM)−high glucose, L-glutamine solution, penicillin−streptomycin solution, Dulbecco's phosphate buffered saline (DPBS), trypsin−EDTA solution, magnesium chloride hexahydrate, and reduced nicotinamide adenine dinucleotide phosphate (NADPH) were purchased from Sigma-Aldrich. MEM nonessential amino acid (MEM-NEAA) solution, fetal bovine serum (FBS), and DMEM without sodium pyruvate and phenol red were bought from Invitrogen (Carlsbad, CA, USA). PRISMA HT universal buffer, GIT-0 Lipid Solution, and Acceptor Sink Buffer were ordered from pIon (Woburn, MA, USA). Human plasma was bought from Biopredic (Rennes, France), and acetonitrile (MeCN) and methanol (MeOH) were from Acros Organics (Geel, Belgium). Pooled male rat liver microsomes were purchased from BD Bioscience (Franklin Lakes, NJ, USA). Tris(hydroxymethyl)aminomethane (TRIS) was obtained from AppliChem (Darmstadt, Germany). The Caco-2 cells were kindly provided by Prof. G. Imanidis, FHNW, Muttenz, and originated from the American Type Culture Collection (Rockville, MD, USA).

pK_a. The pK_a values were determined as described elsewhere.[101] In brief, the pH of a sample solution was gradually changed and the chemical shift of protons adjacent to ionizable centers was monitored by ^{1}H nuclear magnetic resonance (NMR) spectroscopy. The shift was plotted against the pH of the respective sample, and the pK_a was read out from the inflection point of the resulting sigmoidal curve.

log $D_{7.4}$. The in silico prediction tool ALOGPS[102] was used to estimate log *P* values of the compounds. Depending on these values, the compounds were classified into three categories: hydrophilic compounds (log *P* below zero), moderately lipophilic compounds (log *P* between zero and one), and lipophilic compounds (log *P* above one). For each category, two different ratios (volume of 1-octanol to volume of buffer) were defined as experimental parameters (Table 7).

Table 7. Compound Classification Based on Estimated log *P* Values

compd type	log *P*	ratio (1-octanol/buffer)
hydrophilic	<0	30:140, 40:130
moderately lipophilic	0−1	70:110, 110:70
lipophilic	>1	3:180, 4:180

Equal amounts of phosphate buffer (0.1 M, pH 7.4) and 1-octanol were mixed and shaken vigorously for 5 min to saturate the phases. The mixture was left until separation of the two phases occurred, and the buffer was retrieved. Stock solutions of the test compounds were diluted with buffer to a concentration of 1 μM. For each compound, six determinations, that is, three determinations per 1-octanol/buffer ratio, were performed in different wells of a 96-well plate. The respective volumes of buffer containing analyte (1 μM) were pipetted to the wells and covered by saturated 1-octanol according to the chosen volume ratio. The plate was sealed with aluminum foil, shaken (1350 rpm, 25 °C, 2 h) on a Heidolph Titramax 1000 plate-shaker (Heidolph Instruments GmbH & Co. KG, Schwabach, Germany), and centrifuged (2000 rpm, 25 °C, 5 min, 5804 R Eppendorf centrifuge, Hamburg, Germany). The aqueous phase was transferred to a 96-well plate for analysis by LC−MS.

The log $D_{7.4}$ coefficient was calculated from the 1-octanol/buffer ratio (o/b), the initial concentration of the analyte in buffer (1 μM), and the concentration of the analyte in buffer (c_B) with eq 3:

$$\log D_{7.4} = \log\left(\frac{1\ \mu\text{M} - c_B}{c_B}\ \frac{1}{o/b}\right) \tag{3}$$

Aqueous Solubility. Solubility was determined in a 96-well format using the μSOL Explorer solubility analyzer (pIon, version 3.4.0.5). For each compound, measurements were performed at pH 3.0 and 7.4 in triplicate. For this purpose, six wells of a deep well plate, that is, three wells per pH value, were filled with 300 μL of PRISMA HT universal buffer, adjusted to pH 3.0 or 7.4 by adding the requested amount of NaOH (0.5 M). Aliquots (3 μL) of a compound stock solution (10−50 mM in DMSO) were added and thoroughly mixed. The final sample concentration was 0.1−0.5 mM, and the residual DMSO concentration was 1.0% (v/v) in the buffer solutions. After 15 h, the solutions were filtered (0.2 μm 96-well filter plates) using a vacuum to collect manifold (Whatman Ltd., Maidstone, U.K.) to remove the precipitates. Equal amounts of filtrate and 1-propanol were mixed and transferred to a 96-well plate for UV/vis detection (190−500 nm, SpectraMax 190). The amount of material dissolved was calculated by comparison with UV/vis spectra obtained from reference samples, which were prepared by dissolving compound stock solution in a 1:1 mixture of buffer and 1-propanol (final concentrations 0.017−0.083 mM).

Parallel Artificial Membrane Permeation Assay (PAMPA). Effective permeability (log P_e) was determined in a 96-well format with the PAMPA.[60] For each compound, measurements were performed at pH 5.0 and 7.4 in quadruplicates. Eight wells of a deep well plate, that is, four wells per pH value, were filled with 650 μL of PRISMA HT universal buffer adjusted to pH 5.0 or 7.4 by adding the requested amount of NaOH (0.5 M). Samples (150 μL) were withdrawn from each well to determine the blank spectra by UV/vis spectroscopy (190−500 nm, SpectraMax 190). Then analyte dissolved in DMSO was added to the remaining buffer to yield 50 μM solutions. To exclude precipitation, the optical density was measured at 650 nm, with 0.01 being the threshold value. Solutions exceeding this threshold were filtered. Afterward, samples (150 μL) were withdrawn to determine the reference spectra. Further 200 μL was transferred to each well of the donor plate of the PAMPA sandwich (pIon, P/N 110163). The filter membranes at the bottom of the acceptor plate were infused with 5 μL of GIT-0 lipid solution, and 200 μL of Acceptor Sink Buffer was filled into each acceptor well. The sandwich was assembled, placed in the GutBox, and left undisturbed for 16 h. Then it was disassembled and samples (150 μL) were transferred from each donor and acceptor well to UV plates for determination of the UV/vis spectra. Effective permeability (log P_e) was calculated from the compound flux deduced from the spectra, the filter area, and the initial sample concentration in the donor well with the aid of the PAMPA Explorer software (pIon, version 3.5).

Colorectal Adenocarcinoma (Caco-2) Cell Permeation Assay. Caco-2 cells were cultivated in tissue culture flasks (BD Biosciences) with DMEM high glucose medium, containing L-glutamine (2 mM), nonessential amino acids (0.1 mM), penicillin (100 U/mL), streptomycin (100 μg/mL), and fetal bovine serum (10%). The cells were kept at 37 °C in humidified air containing 5% CO_2, and the medium was changed every second day. When approximately 90% confluence was reached, the cells were split in a 1:10 ratio and distributed to new tissue culture flasks. At passage numbers between 60 and 65, they were seeded at a density of 5.3×10^5 cells per well to Transwell six-well plates (Corning Inc.) with 2.5 mL of culture medium in the basolateral and 1.8 mL in the apical compartment. The medium was renewed on alternate days. Permeation experiments were performed between days 19 and 21 after seeding. Prior to the experiment, the integrity of the Caco-2 monolayers was evaluated by measuring the transepithelial electrical resistance (TEER) with an Endohm tissue resistance instrument (World Precision Instruments Inc., Sarasota, FL, USA). Only wells with TEER values higher than 250 Ω cm^2 were used. Experiments were performed in the apical-to-basolateral (absorptive) and basolateral-to-apical (secretory) directions in triplicate. Transport medium (DMEM without sodium pyruvate and phenol red) was withdrawn from the donor compartments of three wells and replaced by the same volume of compound stock solution (10 mM in DMSO) to reach an initial sample concentration of 62.5 μM. The Transwell plate was then shaken (600 rpm, 37 °C) on a Heidolph

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

Titramax 1000 plate-shaker. Samples (40 μL) were withdrawn from the donor and acceptor compartments 30 min after initiation of the experiment, and the compound concentrations were determined by LC–MS (see below). Apparent permeability (P_{app}) was calculated according to eq 4:

$$P_{app} = \frac{dQ}{dt}\frac{1}{Ac_0} \quad (4)$$

where dQ/dt is the compound flux (mol s^{-1}), A is the surface area of the monolayer (cm^2), and c_0 is the initial concentration in the donor compartment (mol cm^{-3}).[60] After the experiment, TEER values were assessed again for each well and results from wells with values below 250 Ω cm^2 were discarded.

Plasma Protein Binding (PPB). PPB was determined in a 96-well format using a high throughput dialysis block (HTD96b; HTDialysis LCC, Gales Ferry, CT, USA). For each compound, measurements were performed in triplicate. Dialysis membranes (MWCO 12-14 K; HTDialysis LCC) were hydrated according to the instructions of the manufacturer and placed into the dialysis block. Human plasma was centrifuged (5800 rpm, 5 °C, 10 min), the pH of the supernatant (without floating plasma lipids) was adjusted to 7.4 by adding the requested amount of HCl (4 M), and analyte was added to yield a final concentration of 10 μM. Equal volumes (150 μL) of plasma containing the analyte or TRIS-HCl buffer (0.1 M, pH 7.4) were transferred to the compartments separated by the dialysis membrane. The block was covered with a sealing film and left undisturbed (5 h, 37 °C). Afterward, samples (90 μL) were withdrawn from the buffer compartments and diluted with plasma (10 μL). From the plasma compartments, samples (10 μL) were withdrawn and diluted with TRIS-HCl buffer (90 μL). The solutions were further diluted with ice-cooled MeCN (300 μL) to precipitate the proteins and centrifuged (3600 rpm, 4 °C, 10 min). The supernatants (50 μL) were retrieved, and the analyte concentrations were determined by LC–MS (see below). The fraction bound (f_b) was calculated as follows (eq 5):

$$f_b = 1 - \frac{c_b}{c_p} \quad (5)$$

where c_b is the concentration of the analyte withdrawn from the buffer compartment before dilution and c_p is the concentration in the plasma compartment. The values were accepted if the recovery of analyte was between 80% and 120% of the initial amount.

Cytochrome P450 Mediated Metabolism. Incubations consisted of pooled male rat liver microsomes (0.5 mg microsomal protein/mL), test compound (2 μM), $MgCl_2$ (2 mM), and NADPH (1 mM) in a total volume of 300 μL TRIS-HCl buffer (0.1 M, pH 7.4) and were performed in a 96-well plate on a Thermomixer Comfort (Eppendorf). Compounds and microsomes were preincubated (37 °C, 700 rpm, 10 min) before NADPH was added. Samples (50 μL) at $t = 0$ min and after an incubation time of 5, 10, 20, and 30 min were quenched with 150 μL of ice-cooled MeOH, centrifuged (3600 rpm, 4 °C, 10 min), and 80 μL of supernatant was transferred to a 96-well plate for LC–MS analysis (see below). The metabolic half-life ($t_{1/2}$) was calculated from the slope of the linear regression from the log percentage remaining compound versus incubation time relationship. Control experiments without NADPH were performed in parallel.

LC–MS Measurements. Analyses were performed using an 1100/1200 series HPLC system coupled to a 6410 triple quadrupole mass detector (Agilent Technologies, Inc., Santa Clara, CA, USA) equipped with electrospray ionization. The system was controlled with the Agilent MassHunter Workstation Data Acquisition software (version B.01.04). The column used was an Atlantis T3 C18 column (2.1 mm × 50 mm) with a 3 μm particle size (Waters Corp., Milford, MA, USA). The mobile phase consisted of eluent A (H_2O containing 0.1% formic acid (for **10a–f,h,i**), or 10 mM ammonium acetate, pH 5.0 in 95:5, H_2O/MeCN (for **10g,j**)) and eluent B (MeCN containing 0.1% formic acid). The flow rate was maintained at 0.6 mL/min. The gradient was ramped from 95% A/5% B to 5% A/95% B over 1 min and then held at 5% A/95% B for 0.1 min. The system was then brought back to 95% A/5% B, resulting in a total duration of 4 min. MS parameters such as fragmentor voltage, collision energy, polarity were optimized individually for each analyte, and the molecular ion was followed for each compound in the multiple reaction monitoring mode. The concentrations of the analytes were quantified by the Agilent Mass Hunter Quantitative Analysis software (version B.01.04).

In Vivo Studies. *Animals.* Female C3H/HeN mice weighing between 19 and 25 g were obtained from Charles River Laboratories (Sulzfeld, Germany) or Harlan (Venray, The Netherlands) and were housed three or four per cage. The mice were kept under specific pathogen-free conditions in the Animal House of the Department of Biomedicine, University Hospital of Basel, and animal experimentation guidelines according to the regulations of the Swiss veterinary law were followed. After 7 days of acclimatization, 9- to 10-week-old mice were used for the studies. Animals had free access to chow and water at any time and were kept in a 12 h/12 h light/dark cycle. For administration volumes and sampling the good practice guidelines were followed.[103]

Pharmacokinetic Studies. The single-dose studies for the first experiment set were performed by intravenous application of FimH antagonists at a dosage of 50 mg/kg body weight, followed by plasma and urine sampling. Antagonists were diluted in PBS (Sigma-Aldrich) for injection into the tail vein. Blood and urine samples (10 μL) were taken at 6 and 30 min and at 1, 2, 4, 6, and 8 h after injection. For the PK studies with **10j**, the antagonist was dissolved in PBS with 5% DMSO (Sigma-Aldrich) and injected into the tail vain (0.625 mg/kg) or given orally (1.25 mg/kg) using a gavage (syringes from BD Micro Fine, U-100 Insuline, 30 G with BD Microlance 3, 25 G needles, Becton Dickinson and Soft-Ject, 1 mL syringes from Henke Sass Wolf; gavage from Fine Science Tools). Blood and urine were sampled (10 μL) after 7, 13, 20, 30, 45 min and after 1, 1.5, 2, 2.5, 3, 4, 6, 8, and 24 h. Both blood and urine samples were directly diluted after sampling with MeOH (Acros Organics) to precipitate the proteins and centrifuged for 11 min at 13 000 rpm. The supernatants were transferred to a 96-well plate (Agilent Technologies, 0.5 mL, polypropylene), and the analyte concentrations were determined by LC–MS (see above).

Infection Study. For all infection studies, the drinking water of the mice was replaced by water containing 5% glucose (monohydrate from AppliChem, BioChemica), 3 days before the start of the experiment. **10j** was dosed at 1.25 mg/kg (in 5% DMSO and PBS) and 10 mg/kg (in 5% DMSO in PBS containing 1% Tween 80, all purchased from Sigma-Aldrich) and applied orally via gavage to six and four mice, respectively, as described in the section Pharmacokinetic Studies, 40 min prior to infection. Ciprofloxacin (Ciproxin solution, 2 mg/mL, Bayer) was dosed with 8 mg/kg, which would correspond to a human dose of 500 mg,[81] subcutaneously 10 min prior to infection with UTI89 to 4 mice. The values for the control group (PBS, po) resulted from the infection of 11 mice. Four mice were orally treated with the formulation vehicle for **10j** (5% DMSO in PBS containing 1% Tween 80) and termed controls formulation. Before infection, remaining urine in the bladder was expelled by gentle pressure on the abdomen. Mice were anesthetized in 2.5 vol % isoflurane/oxygen mixture (Attane, Minrad Inc., USA) and placed on their back. Infection was performed transurethrally using a polyethylene catheter (Intramedic polyethylene tubing, inner diameter 0.28 mm, outer diameter 0.61 mm, Becton Dickinson), on a syringe (Hamilton Gastight Syringe 50 μL, removable 30G needle, BGB Analytik AG, Switzerland). After gentle insertion of the catheter into the bladder, 50 μL of bacterial suspension of UTI89 (5.5×10^9 to 2.25×10^{10} CFU/mL) was slowly injected. This corresponded to approximately 10^7–10^8 CFU per mouse. Mice were killed by CO_2 3 h after inoculation, and bladder and kidneys were aseptically removed. Organs were homogenized in 1 mL of PBS using a tissue lyser (Retsch, Germany). Serial dilutions of bladder and kidneys were plated on Levine Eosin Methylene Blue Agar plates (Becton Dickinson), and CFUs were counted after overnight incubation at 37 °C.

■ ASSOCIATED CONTENT

Supporting Information

HPLC data and chromatograms of target compounds and ^{1}H NMR spectra of the synthetic compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221–2239

AUTHOR INFORMATION

Corresponding Author

*Phone: +41 61 267 15 51. Fax: +41 61 267 15 52. E-mail: beat.ernst@unibas.ch.

Author Contributions

[‖]S.K., L.P., K.M., D.E., and A.S. contributed equally to the project.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Prof. Dr. med. Radek Skoda, Department of Biomedicine, University Hospital Basel, Switzerland, for giving us access to the animal facility. The financial support by the Swiss National Science Foundation (SNF Interdisciplinary Grant K-32K1-120904) is gratefully acknowledged.

ABBREVIATIONS USED

ΔH, change in enthalpy; ΔS, change in entropy; AUC, area under the curve; BSA, bovine serum albumin; C_{max}, maximal concentration; Caco-2, colorectal adenocarcinoma; CFU, colony forming unit; CL_{tot}, total clearance; CRD, carbohydrate recognition domain; C_0, initial concentration; DL, detection limit; FITC, fluorescein isothiocyanate; FP, fluorescence polarization; ITC, isothermal titration calorimetry; iv, intravenous; K_D, dissociation constant; MAC_{90}, minimal antiadhesion concentration to inhibit 90% adhesion; PAMPA, parallel artificial membrane permeation assay; P_{app}, apparent permeability; PD, pharmacodynamics; P_e, effective permeability; PK, pharmacokinetics; po, per os; sc, subcutaneous; UPEC, uropathogenic *Escherichia coli*; UTI, urinary tract infection; V_z, volume of distribution in terminal phase

REFERENCES

(1) Foxman, B.; Barlow, R.; D'Arcy, H.; Gillespie, B.; Sobel, J. D. Urinary tract infection: self-reported incidence and associated costs. *Ann. Epidemiol.* **2000**, *10*, 509−515.

(2) Ronald, A. The etiology of urinary tract infection: traditional and emerging pathogens. *Am. J. Med.* **2002**, *113* (Suppl. 1A), 14S−19S.

(3) Fihn, S. D. Acute uncomplicated urinary tract infection in women. *N. Engl. J. Med.* **2003**, *349*, 259−266.

(4) Hooton, T. M.; Besser, R.; Foxman, B.; Fritsche, T. R.; Nicolle, L. E. Acute uncomplicated cystitis in an era of increasing antibiotic resistance: a proposed approach to empirical therapy. *Clin. Infect. Dis.* **2004**, *39*, 75−80.

(5) Sanchez, G. V.; Master, R. N.; Karlowsky, J. A.; Bordon, J. M. In vitro antimicrobial resistance of urinary *Escherichia coli* isolates among U.S. outpatients from 2000 to 2010. *Antimicrob. Agents Chemother.* **2012**, *56*, 2181−2183.

(6) Clatworthy, A. E.; Pierson, E.; Hung, D. T. Targeting virulence: a new paradigm for antimicrobial therapy. *Nat. Chem. Biol.* **2007**, *3*, 541−548.

(7) Mulvey, M. A.; Schilling, J. D.; Martinez, J. J.; Hultgren, S. J. Bad bugs and beleaguered bladders: interplay between uropathogenic *Escherichia coli* and innate host defenses. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8829−8835.

(8) Schilling, J. D.; Mulvey, M. A.; Hultgren, S. J. Structure and function of *Escherichia coli* type 1 pili: new insight into the pathogenesis of urinary tract infections. *J. Infect. Dis.* **2001**, *183* (Suppl. 1), S36−S40.

(9) Wiles, T. J.; Kulesus, R. R.; Mulvey, M. A. Origins and virulence mechanisms of uropathogenic *Escherichia coli*. *Exp. Mol. Pathol.* **2008**, *85*, 11−19.

(10) Capitani, G.; Eidam, O.; Glockshuber, R.; Grütter, M. G. Structural and functional insights into the assembly of type 1 pili from *Escherichia coli*. *Microbes Infect.* **2006**, *8*, 2284−2290.

(11) Le Trong, I.; Aprikian, P.; Kidd, B. A.; Forero-Shelton, M.; Tchesnokova, V.; Rajagopal, P.; Rodriguez, V.; Interlandi, G.; Klevit, R.; Vogel, V.; Stenkamp, R. E.; Sokurenko, E. V.; Thomas, W. E. Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like β sheet twisting. *Cell* **2010**, *141*, 645−655.

(12) Sharon, N. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochim. Biophys. Acta* **2006**, *1760*, 527−537.

(13) Firon, N.; Itzhak, O.; Sharon, N. Interaction of mannose-containing oligosaccharides with the fimbrial lectin of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **1982**, *105*, 1426−1432.

(14) Firon, N.; Ofek, I.; Sharon, N. Carbohydrate specificity of the surface lectins of *Escherichia coli*, *Klebsiella pneumoniae*, and *Salmonella typhimurium*. *Carbohydr. Res.* **1983**, *120*, 235−249.

(15) Bouckaert, J.; Berglund, J.; Schembri, M.; De Genst, E.; Cools, L.; Wuhrer, M.; Hung, C.-S.; Pinkner, J.; Slättegard, R.; Zavialov, A.; Choudhury, D.; Langermann, S.; Hultgren, S. J.; Wyns, L.; Klemm, P.; Oscarson, S.; Knight, S. D.; De Greve, H. Receptor binding studies disclose a novel class of high-affinity inhibitors of the *Escherichia coli* FimH adhesin. *Mol. Microbiol.* **2005**, *55*, 441−455.

(16) Firon, N.; Ashkenazi, S.; Mirelman, D.; Ofek, I.; Sharon, N. Aromatic alpha-glycosides of mannose are powerful inhibitors of the adherence of type 1 fimbriated *Escherichia coli* to yeast and intestinal epithelial cells. *Infect. Immun.* **1987**, *55*, 472−476.

(17) Sperling, O.; Fuchs, A.; Lindhorst, T. K. Evaluation of the carbohydrate recognition domain of the bacterial adhesin FimH. Design, synthesis and binding properties of mannoside ligands. *Org. Biomol. Chem.* **2006**, *4*, 3913−3922.

(18) Han, Z.; Pinkner, J. S.; Ford, B.; Obermann, R.; Nolan, W.; Wildman, S. A.; Hobbs, D.; Ellenberger, T.; Cusumano, C. K.; Hultgren, S. J.; Janetka, J. W. Structure-based drug design and optimization of mannoside bacterial FimH antagonists. *J. Med. Chem.* **2010**, *53*, 4779−4792.

(19) Klein, T.; Abgottspon, D.; Wittwer, M.; Rabbani, S.; Herold, J.; Jiang, X.; Kleeb, S.; Lüthi, C.; Scharenberg, M.; Bezençon, J.; Gubler, E.; Pang, L.; Smiesko, M.; Cutting, B.; Schwardt, O.; Ernst, B. FimH antagonists for the oral treatment of urinary tract infections: from design and synthesis to in vitro and in vivo evaluation. *J. Med. Chem.* **2010**, *53*, 8627−8641.

(20) Cusumano, C. K.; Pinkner, J. S.; Han, Z.; Greene, S. E.; Ford, B. A.; Crowley, J. R.; Henderson, J. P.; Janetka, J. W.; Hultgren, S. J. Treatment and prevention of urinary tract infection with orally active FimH inhibitors. *Sci. Transl. Med.* **2011**, *3*, 109ra115.

(21) Han, Z.; Pinkner, J. S.; Ford, B.; Chorell, E.; Crowley, J. M.; Cusumano, C. K.; Campbell, S.; Henderson, J. P.; Hultgren, S. J.; Janetka, J. W. Lead optimization studies on FimH antagonists: discovery of potent and orally bioavailable ortho-substituted biphenyl mannosides. *J. Med. Chem.* **2012**, *55*, 3945−3959.

(22) Pang, L.; Kleeb, S.; Lemme, K.; Rabbani, S.; Scharenberg, M.; Zalewski, A.; Schädler, F.; Schwardt, O.; Ernst, B. FimH antagonists: structure−activity and structure−property relationships for biphenyl α-D-mannopyranosides. *ChemMedChem* **2012**, *7*, 1404−1422.

(23) Jiang, X.; Abgottspon, D.; Kleeb, S.; Rabbani, S.; Scharenberg, M.; Wittwer, M.; Haug, M.; Schwardt, O.; Ernst, B. Anti-adhesion therapy for urinary tract infections—a balanced PK/PD profile proved to be key for success. *J. Med. Chem.* **2012**, *55*, 4700−4713.

(24) Schwardt, O.; Rabbani, S.; Hartmann, M.; Abgottspon, D.; Wittwer, M.; Kleeb, S.; Zalewski, A.; Smiesko, M.; Cutting, B.; Ernst, B. Design, synthesis and biological evaluation of mannosyl triazoles as FimH antagonists. *Bioorg. Med. Chem.* **2011**, *19*, 6454−6473.

(25) Brument, S.; Sivignon, A.; Dumych, T. I.; Moreau, N.; Roos, G.; Guérardel, Y.; Chalopin, T.; Deniaud, D.; Bilyy, R. O.; Darfeuille-Michaud, A.; Bouckaert, J.; Gouin, S. G. Thiazolylaminomannosides as potent antiadhesives of type 1 piliated *Escherichia coli* isolated from Crohn's disease patients. *J. Med. Chem.* **2013**, *56*, 5395−5406.

(26) Lindhorst, T. K.; Kieburg, C.; Krallmann-Wenzel, U. Inhibition of the type 1 fimbriae-mediated adhesion of *Escherichia coli* to erythrocytes by multiantennary D-mannosyl clusters: the effect of multivalency. *Glycoconjugate J.* **1998**, *15*, 605−613.

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

(27) Nagahori, N.; Lee, R. T.; Nishimura, S.-L.; Pagé, S.; Roy, R.; Lee, Y. C. Inhibition of adhesion of type 1 fimbriated *Escherichia coli* to highly mannosylated ligands. *ChemBioChem* **2002**, *3*, 836−844.
(28) Appeldoorn, C. C. M.; Joosten, J. A. F.; Maate, F. A.; Dobrindt, U.; Hacker, J.; Liskamp, R. M. J.; Khan, A. S.; Pieters, R. J. Novel multivalent mannose compounds and their inhibition of the adhesion of type 1 fimbriated uropathogenic *E. coli*. *Tetrahedron: Asymmetry* **2005**, *16*, 361−372.
(29) Patel, A.; Lindhorst, T. K. A modular approach for the synthesis of oligosaccharide mimetics. *Carbohydr. Res.* **2006**, *341*, 1657−1668.
(30) Touaibia, M.; Wellens, A.; Shiao, T. C.; Wang, Q.; Sirois, S.; Bouckaert, J.; Roy, R. Mannosylated G(0) dendrimers with nanomolar affinities to *Escherichia coli* FimH. *ChemMedChem* **2007**, *2*, 1190−1201.
(31) Durka, M.; Buffet, K.; Iehl, J.; Holler, M.; Nierengarten, J.-F.; Taganna, J.; Bouckaert, J.; Vincent, S. P. The functional valency of dodecamannosylated fullerenes with *Escherichia coli* FimH—towards novel bacterial antiadhesives. *Chem. Commun.* **2011**, *47*, 1321−1323.
(32) Bouckaert, J.; Li, Z.; Xavier, C.; Almant, M.; Caveliers, V.; Lahoutte, T.; Weeks, S. D.; Kovensky, J.; Gouin, S. G. Heptyl α-D-mannosides grafted on a β-cyclodextrin core to interfere with *Escherichia coli* adhesion: an in vivo multivalent effect. *Chem.—Eur. J.* **2013**, *19*, 7847−7855.
(33) Scharenberg, M.; Schwardt, O.; Rabbani, S.; Ernst, B. Target selectivity of FimH antagonists. *J. Med. Chem.* **2012**, *55*, 9810−9816.
(34) Choudhury, D.; Thompson, A.; Stojanoff, V.; Langermann, S.; Pinkner, J.; Hultgren, S. J.; Knight, S. D. X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* **1999**, *285*, 1061−1066.
(35) Hung, C.-S.; Bouckaert, J.; Hung, D.; Pinkner, J.; Widberg, C.; DeFusco, A.; Auguste, C. G.; Strouse, R.; Langermann, S.; Waksman, G.; Hultgren, S. J. Structural basis of tropism of *Escherichia coli* to the bladder drug in urinary tract infection. *Mol. Microbiol.* **2002**, *44*, 903−915.
(36) Wellens, A.; Garofalo, C.; Nguyen, H.; Van Gerven, N.; Slättegard, R.; Henalsteens, J.-P.; Wyns, L.; Oscarson, S.; De Greve, H.; Hultgren, S. J.; Bouckaert, J. Intervening with urinary tract infections using anti-adhesives based on the crystal structure of the FimH-oligomannose-3 complex. *PLoS One* **2008**, *3*, e2040.
(37) Wellens, A.; Lahmann, M.; Touaibia, M.; Vaucher, J.; Oscarson, S.; Roy, R.; Remaut, H.; Bouckaert, J. The tyrosine gate as a potential entropic lever in the receptor-binding site of the bacterial adhesin FimH. *Biochemistry* **2012**, *51*, 4790−4799.
(38) Totsika, M.; Kostakioti, M.; Hannan, T. J.; Upton, M.; Beatson, S. A.; Janetka, J. W.; Hultgren, S. J.; Schembri, M. A. A FimH inhibitor prevents acute bladder infection and treats chronic cystitis caused by multidrug-resistant uropathogenic *Escherichia coli* ST131. *J. Infect. Dis.* **2013**, *208*, 921−928.
(39) Meanwell, M. A. Synopsis of some recent tactical application of bioisosteres in drug design. *J. Med. Chem.* **2011**, *54*, 2529−2591.
(40) Prieto, M.; Zurita, E.; Rosa, E.; Luñoz, L.; Lloyd-Williams, P.; Giralt, E. Arylboronic acids and arylpinacolboronate esters in Suzuki coupling reactions involving indoles. Partner role swapping and heterocycle protection. *J. Org. Chem.* **2004**, *69*, 6812−6820.
(41) Schulz, M. J.; Coats, S. J.; Hlasta, D. J. Microwave-assisted preparation of aryltetrazoleboronate esters. *Org. Lett.* **2004**, *6*, 3265−3268.
(42) Devos, A.; Remion, J.; Frisque-Hesbain, A. M.; Colens, A.; Ghosez, L. Synthesis of acyl halides under very mild conditions. *J. Chem. Soc., Chem. Commun.* **1979**, 1180−1181.
(43) Rabbani, S.; Jiang, X.; Schwardt, O.; Ernst, B. Expression of the carbohydrate recognition domain of FimH and development of a competitive binding assay. *Anal. Biochem.* **2010**, *407*, 188−195.
(44) Waetherman, R. V.; Kiessling, L. L. Fluorescence anisotropy assay reveals affinities of C- and O-glycosides for concanavalin A. *J. Org. Chem.* **1996**, *61*, 534−538.
(45) Cer, R. Z.; Mudunuri, U.; Stephens, R.; Lebeda, F. J. IC50-to-Ki: a web-based tool for converting IC50 to Ki values for inhibitors of enzyme activity and ligand binding. *Nucleic Acids Res.* **2009**, *37*, W441−W445.
(46) Lynch, B. A.; Loiacono, K. A.; Tiong, C. L.; Adams, S. E.; MacNeil, I. A. A fluorescence polarization based Src-SH2 binding assay. *Anal. Biochem.* **1997**, *247*, 77−82.
(47) Wu, P.; Brasseur, M.; Schindler, U. A high-throughput STAT binding assay using fluorescence polarization. *Anal. Biochem.* **1997**, *249*, 29−36.
(48) Huang, X. Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. *J. Biomol. Screening* **2003**, *8*, 34−38.
(49) Cooper, A. *Biophysical Chemistry*, 2nd ed.; RSC Publishing: Cambridge, U.K., 2011; pp 122−123.
(50) Scharenberg, M.; Jiang, X.; Pang, L.; Navarra, G.; Rabbani, S.; Binder, F.; Schwardt, O.; Ernst, B. Kinetic properties of carbohydrate−lectin interactions: FimH antagonists. *ChemMedChem* **2014**, *9*, 78−83.
(51) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. Group contribution to the thermodynamic properties of non-ionic solutes in dilute aqueous solution. *J. Solution Chem.* **1981**, *10*, 563−595.
(52) Hansch, C.; Leo, A.; Taft, R. W. A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* **1991**, *91*, 165−195.
(53) Chen, A.; Wadso, I. Simultaneous determination of delta G, delta H and delta S by an automatic microcalorimetric titration technique: application to protein ligand binding. *J. Biochem Biophys Methods* **1982**, *6*, 307−316.
(54) Freire, E.; Mayorga, O. L.; Straume, M. Isothermal titration calorimetry. *Anal. Chem.* **1990**, *62*, 950A−959A.
(55) Wiseman, T.; Williston, S.; Brandts, J. F.; Lin, L.-N. Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **1989**, *179*, 131−137.
(56) Turnbull, W. B.; Daranas, A. H. On the value of c: can low affinity systems be studied by isothermal titration calorimetry? *J. Am. Chem. Soc.* **2003**, *125*, 14859−14866.
(57) Sigurskjold, B. W. Exact analysis of competition ligand binding by displacement isothermal titration calorimetry. *Anal. Biochem.* **2000**, *277*, 260−266.
(58) Velazquez-Campoy, A.; Freire, E. Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat. Protoc.* **2006**, *1*, 186−191.
(59) Dearden, J. C.; Bresnen, G. M. The measurement of partition coefficients. *QSAR Comb. Sci.* **1988**, *7*, 133−144.
(60) Kansy, M.; Senner, F.; Gubernator, K. Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. *J. Med. Chem.* **1998**, *41*, 1007−1010.
(61) Hubatsch, I.; Ragnarsson, E. G. E.; Artursson, P. Determination of drug permeability and predicition of drug absorption in Caco-2 monolayers. *Nat. Protoc.* **2007**, *2*, 2111−2119.
(62) Banker, M. J.; Clark, T. H.; Williams, J. A. Development and validation of a 96-well equilibrium dialysis apparatus for measuring plasma protein binding. *J. Pharm. Sci.* **2003**, *92*, 967−974.
(63) Obach, R. S. Prediction of human clearance of twenty-nine drugs from hepatic microsomal intrinsic clearance data: an examination of in vitro half-life approach and nonspecific binding to microsomes. *Drug Metab. Dispos.* **1999**, *27*, 1350−1359.
(64) Chaturvedi, P. R.; Decker, C. J.; Odinecs, A. Prediction of pharmacokinetic properties using experimental approaches during early drug discovery. *Curr. Opin. Chem. Biol.* **2001**, *5*, 452−463.
(65) Di, L.; Kerns, E. H. Profiling drug-like properties in discovery research. *Curr. Opin. Chem. Biol.* **2003**, *7*, 402−408.
(66) Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235−249.
(67) Curatolo, W. Physical chemical properties of oral drug candidates in the discovery and exploratory development settings. *Pharm. Sci. Technol. Today* **1998**, *1*, 387−393.
(68) Ishikawa, M.; Hashimoto, Y. Improvement in aqueous solubility in small molecule drug discovery programs by disruption of molecular planarity and symmetry. *J. Med. Chem.* **2011**, *54*, 1539−1554.

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

(69) Avdeef, A.; Bendels, S.; Di, L.; Faller, B.; Kansy, M.; Sugano, K.; Yamauchi, Y. PAMPA − critical factors for better predictions of absorption. *J. Pharm. Sci.* **2007**, *96*, 2893−2909.
(70) Artursson, P.; Karlsson, J. Correlation between oral drug absorption in humans and apparent drug permeability coefficients in human intestinal epithelial (Caco-2) cells. *Biochem. Biophys. Res. Commun.* **1991**, *175*, 880−885.
(71) Feng, B.; LaPerle, J. L.; Chang, G.; Varma, M. V. S. Renal clearance in drug discovery and development: molecular descriptors, drug transporters and disease state. *Expert Opin. Drug. Metab. Toxicol.* **2010**, *6*, 939−952.
(72) Schmidt, S.; Gonzalez, D.; Derendorf, H. Significance of protein binding in pharmacokinetics and pharmacodynamics. *J. Pharm. Sci.* **2010**, *99*, 1107−1122.
(73) Weisiger, R. A. Dissociation from albumin: A potentially rate-limiting step in the clearance of substances by the liver. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1563−1567.
(74) Smith, D. A.; Jones, B. C.; Walker, D. K. Design of drugs involving the concepts and theories of drug metabolism and pharmacokinetics. *Med. Res. Rev.* **1996**, *16*, 243−266.
(75) Van de Waterbeemd, H.; Smith, D. A.; Beaumont, K.; Walker, D. K. Property-based design: optimization of drug absorption and pharmacokinetics. *J. Med. Chem.* **2001**, *44*, 1313−1333.
(76) Varma, M. V. S.; Feng, B.; Obach, R. S.; Troutman, M. D.; Chupka, J.; Miller, H. R.; El-Kattan, A. Physicochemical determinants of human renal clearance. *J. Med. Chem.* **2009**, *52*, 4844−4852.
(77) Waring, M. J. Lipophilicity in drug discovery. *Expert Opin. Drug Discovery* **2010**, *5*, 235−248.
(78) Zhang, Y.; Huo, M.; Solver, P. K. An add-in program for pharmacokinetic and pharmacodynamic data analysis in Microsoft Excel. *Comput. Methods Programs Biomed.* **2010**, *99*, 306−314.
(79) Scharenberg, M.; Abgottspon, D.; Cicek, E.; Jiang, X.; Schwardt, O.; Rabbani, S.; Ernst, B. Cytometry-based assay for screening FimH antagonists. *Assay Drug Dev. Technol.* **2011**, *9*, 455−464.
(80) Hooton, T. M. Fluoroquinolones and resistance in the treatment of uncomplicated urinary tract infection. *Int. J. Antimicrob. Agents* **2003**, *22*, 65−72.
(81) Jakobsen, L.; Cattoir, V.; Jensen, K. S.; Hammerum, A. M.; Nordmann, P.; Frimodt-Møller, N. Impact of low-level fluoroquinolone resistance genes qnrA1, qnrB19, and qnrS1 on ciprofloxacin treatment of isogenic *Escherichia coli* strains in a murine urinary tract infection model. *J. Antimicrob. Chemother.* **2012**, *67*, 2438−2444.
(82) Mulvey, M. A. Adhesion and entry of uropathogenic *Escherichia coli*. *Cell. Microbiol.* **2002**, *4*, 257−271.
(83) Ballatore, C.; Huryn, D. M.; Smith, A. B. Carboxylic acid (bio)isosteres in drug design. *ChemMedChem* **2013**, *8*, 385−395.
(84) Justice, S. S.; Hung, C.; Theriot, J. A.; Fletcher, D. A.; Anderson, G. G.; Footer, M. J.; Hultgren, S. J. Differentiation and developmental pathways of uropathogenic *Escherichia coli* in urinary tract pathogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1333−1338.
(85) Mulvey, M. A.; Schilling, J. D.; Hultgren, S. J. Establishment of a persistent *Escherichia coli* reservoir during the acute phase of a bladder infection. *Infect. Immun.* **2001**, *69*, 4572−9.
(86) Houtman, J. C.; Brown, P. C.; Bowden, B.; Yamaguchi, H.; Appella, E.; Samelson, L. E.; Schuck, P. Studying multisite binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: application to adaptor protein complexes in cell signaling. *Protein Sci.* **2007**, *16*, 30−42.
(87) *Maestro*, version 9.3; Schrödinger, LLC: New York, NY, 2012.
(88) *MacroModel*, version 9.9; Schrödinger, LLC: New York, NY, 2012.
(89) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Thompson, J. D.; Winget, P.; Lynch, B. J.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMSOL*, version 7.1; University of Minnesota: Minneapolis, MN, 2003; based in part on the following: Liotard, D. A.; Healy, E. F.; Ruiz, J. M.; Dewar, M. J. S. *AMPAC*, version 2.1; Semichem, Inc.: Shawnee, KS.
(90) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. *J. Phys. Chem.* **1996**, *100*, 16385−16398.
(91) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1993**, *115*, 5348−5348 [Erratum to *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909].
(92) Mulvey, M. A.; Schilling, J. D.; Hultgren, S. J. Establishment of a persistent *Escherichia coli* reservoir during the acute phase of a bladder infection. *Infect. Immun.* **2001**, *69*, 4572−4579.
(93) Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **1993**, *26*, 795−800.
(94) Leslie, A. G. W. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **2006**, *62*, 48−57.
(95) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **2011**, *67*, 235−242.
(96) McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **2007**, *63*, 32−41.
(97) Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **2004**, *60*, 2126−2132.
(98) Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L.-W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58*, 1948−1954.
(99) van Aalten, D. M. F.; Bywater, R.; Findlay, J. B. C.; Hendlich, M.; Hooft, R. W. W.; Vriend, G. PRODRG, a program for generating molecular topologies and unique molecular descriptors from coordinates of small molecules. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 255−262.
(100) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **2010**, *66*, 12−21.
(101) Bezençon, J.; Wittwer, M. B.; Cutting, B.; Smiesko, M.; Wagner, B.; Kansy, M.; Ernst, B. pK_a determination by ^{1}H NMR spectroscopy—an old methodology revisited. *J. Pharm. Biomed. Anal.* **2014**, *93*, 147−155.
(102) (a) VCCLAB, Virtual Computational Chemistry Laboratory, 2005. http://www.vcclab.org (accessed August 14, 2012). (b) Tetko, I. V.; Gasteiger, J.; Todeschini, R.; Mauri, A.; Livingstone, D.; Ertl, P.; Palyulin, V. A.; Radchenko, E. V.; Zefirov, N. S.; Makarenko, A. S.; Tanchuk, V. Y.; Prokopenko, V. V. Virtual computational chemistry laboratory—design and description. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 453−463.
(103) Diehl, K.-H.; Hull, R. A. Good practice guide to the administration of substances and removal of blood, including routes and volumes. *J. Appl. Toxicol.* **2001**, *21*, 15−23.

DOI: 10.1021/jm501524q
J. Med. Chem. 2015, 58, 2221−2239

2.2.4 Paper 3

Catch-bond mechanism of the bacterial adhesin FimH

This paper describes the in-depth study of the molecular mechanism that regulates the catch-bond behavior of the lectin protein FimH. The full-length protein, complemented with a synthetic donor strand, proved to be an excellent model system. Crystallography, molecular dynamics, and kinetic studies revealed the details of the allosteric regulation, which enables FimH to react to shear stress. Using cell-tracking experiments, the importance of the dynamic switching between a low and high affinity state for cell motility was demonstrated. Moreover, a road-map for studying other catch-bond interactions was outlined.

Contribution to the project:
Giulio Navarra designed, synthesized, and characterized the fluorescent tracer GN-FP-4 and wrote the corresponding parts of the supporting information.

This paper was published in *Nature Communications*:

Maximilian M. Sauer, Roman P. Jakob, Jonathan Eras, Sefer Baday, Deniz Eriş, Giulio Navarra, Simon Bernèche, Beat Ernst, Timm Maier, Rudi Glockshuber

nature COMMUNICATIONS

ARTICLE

Received 21 Jul 2015 | Accepted 13 Jan 2016 | Published 7 Mar 2016

DOI: 10.1038/ncomms10738 OPEN

Catch-bond mechanism of the bacterial adhesin FimH

Maximilian M. Sauer[1,*], Roman P. Jakob[2,*], Jonathan Eras[1], Sefer Baday[2,3], Deniz Eriş[4], Giulio Navarra[4], Simon Bernèche[2,3], Beat Ernst[4], Timm Maier[2] & Rudi Glockshuber[1]

Ligand–receptor interactions that are reinforced by mechanical stress, so-called catch-bonds, play a major role in cell–cell adhesion. They critically contribute to widespread urinary tract infections by pathogenic *Escherichia coli* strains. These pathogens attach to host epithelia via the adhesin FimH, a two-domain protein at the tip of type I pili recognizing terminal mannoses on epithelial glycoproteins. Here we establish peptide-complemented FimH as a model system for fimbrial FimH function. We reveal a three-state mechanism of FimH catch-bond formation based on crystal structures of all states, kinetic analysis of ligand interaction and molecular dynamics simulations. In the absence of tensile force, the FimH pilin domain allosterically accelerates spontaneous ligand dissociation from the FimH lectin domain by 100,000-fold, resulting in weak affinity. Separation of the FimH domains under stress abolishes allosteric interplay and increases the affinity of the lectin domain. Cell tracking demonstrates that rapid ligand dissociation from FimH supports motility of piliated *E. coli* on mannosylated surfaces in the absence of shear force.

[1] Institute of Molecular Biology and Biophysics, Department of Biology, ETH, Zurich, Otto-Stern-Weg 5, 8093 Zurich, Switzerland. [2] Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland. [3] SIB Swiss Institute of Bioinformatics, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland. [4] Institute of Molecular Pharmacy, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to T.M. (email: timm.maier@unibas.ch).

Cell–cell adhesion often occurs under dynamically varying conditions and mechanical stress. In many cell–cell adhesion systems, the lifetime of adhesin–receptor complexes is increased under tensile mechanical force via 'catch-bonds', which permit capture or retention of cells under flow conditions while still allowing for release under reduced mechanical force. Catch-bond interactions are prominent in vascular systems and are formed, for example, by selectins for leukocyte recruitment[1,2], by cadherins controlling tissue integrity[3,4] in the epithelial adhesion of cancer cells[5] and by the interactions between T-cell receptors (TCRs) and peptide-bound major histocompatibility complexes (MHC) on antigen-presenting cells[6,7]. Catch-bonds also play a major role in bacterial adhesion and infection by uropathogenic *Escherichia coli* strains, which are responsible for the vast majority of urinary tract infections (UTIs) in humans[8]. A first critical step in the establishment of infection is bacterial adhesion to urothelial cells under flow conditions, which is mediated by $0.1-2\,\mu$m long, proteinaceous filaments on the bacterial surface termed type 1 pili[9,10]. Type 1 pili are composed of up to 3,000 copies of the subunit FimA building the pilus rod, as well as the subunits FimF, FimG and FimH forming the distal tip fibrillum[11]. The adhesin FimH at the fimbrial tip specifically binds in a catch-bond mode[12] to terminal α-D-linked mannoses of N-linked glycans of the receptor uroplakin 1a on urinary epithelial cells[13]. Owing to its important role in establishing infection, FimH is an attractive target for the development of anti-adhesive drugs for UTI treatment[14,15].

FimH is a two-domain protein, composed of an N-terminal, mannoside-binding lectin domain ($FimH_L$) and a C-terminal pilin domain ($FimH_P$). $FimH_P$ possesses an incomplete immunoglobulin-like fold that is completed by insertion of an N-terminal donor strand of FimG, the subsequent subunit in pilus assembly[11]. The two-domain architecture of FimH is a prerequisite for catch-bond formation because the interactions between $FimH_L$ and $FimH_P$ determine the conformational state and ligand-binding properties of $FimH_L$ (refs 12,16,17). A 'compressed' $FimH_L$ conformation was observed in the crystal structure of FimH in the context of the type 1 pilus tip fibrillum in the absence of ligands, with an open binding site and interactions to $FimH_P$ mediated via three loop segments: the swing (amino acids (aa.) 27–33), linker (aa. 154–160) and insertion loops (aa. 112–118)[17]. In contrast, an 'extended' $FimH_L$ conformation was observed in crystal structures of the isolated, ligand-bound $FimH_L$ domain[18–23] and in the complex between FimH and the pilus assembly chaperone FimC, where FimC prevents the interactions between $FimH_L$ and $FimH_P$ (ref. 24). This extended form of $FimH_L$ is characterized by a closed ligand-binding pocket and rearranged swing, linker and insertion loops.

Notably, isolated $FimH_L$ was reported to show a ligand-binding affinity about two orders of magnitude higher than that of full-length FimH in the tip fibrillum[17,25]. Together with mutagenesis experiments disrupting the interdomain interface[26], these data indicated that ligand-binding is linked to domain separation in FimH, and that mechanical force shifts the ligand-binding affinity towards that of the isolated $FimH_L$. However, fundamental aspects of the mechanism underlying the force-dependent binding of FimH remained unknown: (i) How is domain-associated, full-length FimH interacting with ligands? (ii) Does ligand-binding directly induce domain separation? (iii) How are interdomain interactions linked to the ligand-binding affinity of FimH and the kinetics of ligand-binding and dissociation?

To address these questions, we designed a stable, soluble variant of full-length FimH that is equivalent in its structural and functional properties to those of FimH in the assembled fimbrial tip. This variant allowed us to obtain high-resolution structural snapshots of all functional states of FimH and to obtain a complete characterization of ligand-binding kinetics in solution. Together with molecular dynamics simulations, these data reveal a three-state mechanism of FimH catch-bond formation. $FimH_P$ accelerates ligand release from $FimH_L$ via dynamic allostery by 100,000-fold. In addition, using single-cell tracking experiments, we show that the modulation of ligand affinity by $FimH_P$ is not only required for adhesion under mechanical stress, but also for efficient bacterial surface motility in the absence of shear force. Our results provide a first complete structural and kinetic description of a catch-bond system and establish a framework for the analysis of the distinct catch-bond mechanisms in other systems, which also commonly couple interdomain interactions to ligand affinity.

Results

Construction of a peptide-complemented FimH. Isolated FimH with its non-complemented pilin domain is only marginally stable and shows aggregation tendency under physiological conditions[27]. To establish a stable, isolated FimH molecule with all properties of FimH in the tip fibrillum, we complemented $FimH_P$ with the donor-strand peptide of FimG (FimG residues $1-14$; termed DsG). The FimH · DsG complex was obtained in good yields and purified after an *in vitro* reaction, mimicking the first donor-strand exchange (DSE) reaction during pilus assembly *in vivo*. In this reaction, the FimG donor strand displaces the pilus assembly chaperone FimC from FimH (Fig. 1a):

$$\begin{aligned}&\mathrm{FimC\cdot FimH + DsG \xrightarrow{k_{DSE}} FimH\cdot DsG + FimC}\\&\left(\mathrm{pH\ 7.0, 37\,^{\circ}C;}\ k_{\mathrm{DSE}} \approx 0.5\,\mathrm{M^{-1}s^{-1}}\right)\end{aligned} \tag{1}$$

The experiments described in the following were performed with FimH from the faecal *E. coli* strain F18 ($FimH^{F18}$), which is structurally identical to the most prevalent variants in uropathogenic infection[25], and FimH from the wild-type *E. coli* strain K12 ($FimH^{K12}$), which differ in three amino acids in $FimH_L$ (K12→F18: Val27Ala, Asn70Ser, Ser78Asn; Supplementary Fig. 1a). The isolated lectin domains (residues 1–159) of both FimH variants ($FimH_L^{K12}$ and $FimH_L^{F18}$) were produced by direct expression in the *E. coli* periplasm and were purified as described[27].

Ligand-free FimH · DsG resembles FimH in the fimbrial tip. The crystal structure of the binary complex $FimH^{F18}$ · DsG was determined at atomic resolution by molecular replacement (Table 1). $FimH^{F18}$ · DsG comprises the jellyroll fold $FimH_L$ and the immunoglobulin-like $FimH_P$ domain complemented with the FimG donor strand (Fig. 1b and Supplementary Fig. 1b). It closely resembles unliganded FimH in the fimbrial tip complex (Fig. 1c)[17], with a root-mean-square deviation of C_α positions (C_α r.m.s.d.) of 1.1 Å. The individual $FimH_P$ and $FimH_L$ domains are even more closely resembling unliganded, fimbrial FimH (r.m.s.d. 0.45 and 0.55 Å, respectively) and undergo only a minimal hinge-bending rotation of 4° (Fig. 1c).

The DsG peptide in $FimH^{F18}$ · DsG is in identical position as compared with the N-terminal FimG extension in the fimbrial tip structure; it interacts with β-strand 2 and 9 of $FimH_P$ (Fig. 1d). All contacts in the $FimH_L$–$FimH_P$ interdomain region (Supplementary Fig. 1c,d)[17] as well as the conformation of the empty ligand-binding pocket observed in FimH in the fimbrial tip are preserved in $FimH^{F18}$ · DsG. Thus, $FimH^{F18}$ · DsG represents the ligand-**free** state of fimbrial FimH, with **A**ssociated $FimH_L$ and $FimH_P$ ($\mathbf{A_{free}}$ state) and is an elegant minimal system to analyse the crosstalk between ligand-binding and interdomain interactions underlying the formation of catch-bonds by FimH.

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms10738

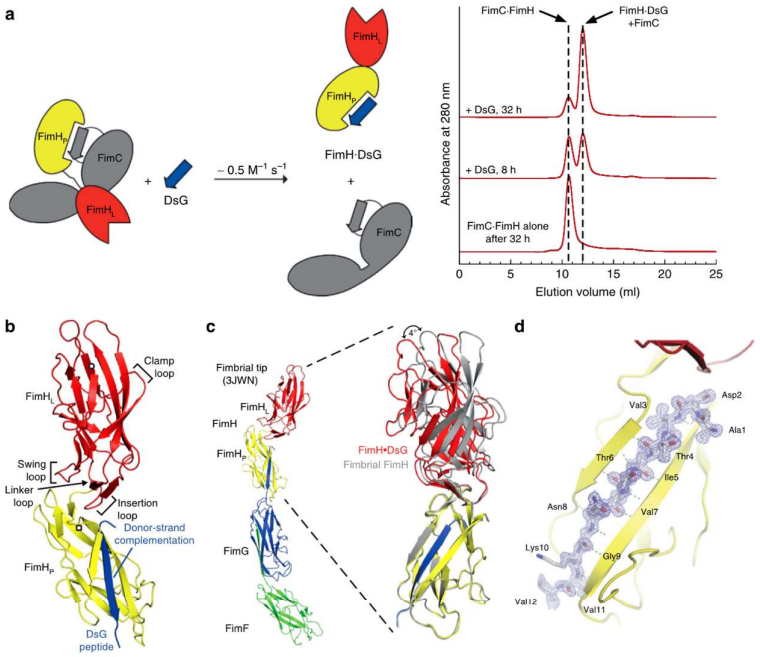


Figure 1 | FimH · DsG resembles fimbrial tip FimH. (**a**) Preparation of the FimH · DsG complex by DSE. Left: reaction scheme of the DSE reaction, in which DsG displaces the FimC chaperone from the FimH pilin domain. Right: kinetics of the FimH · DsG complex formation at 37 °C, monitored by analytical gel filtration. DSE was initiated by mixing the FimC · FimH complex (15 μM) with excess DsG peptide (50 μM). Samples were removed after different incubation times, rapidly cooled on ice and immediately subjected to gel filtration. The reaction can be followed by the decrease in the FimC · FimH complex concentration and the simultaneous increase in the concentrations of FimH · DsG and free FimC (FimC and FimH · DsG coelute as a single peak at ~12 ml). The chromatogram at the bottom shows that the FimC · FimH complex is stable against dissociation/aggregation under the chosen conditions. The rate constant of DSE estimated from these data is ~0.5 $M^{-1}s^{-1}$. (**b**) Structure of $FimH^{F18}$ · DsG (lectin domain $FimH_L$, red; pilin domain $FimH_P$, yellow; DsG, blue; circle and square indicate N- and C termini, respectively). (**c**) FimH from the fimbrial tip structure (left, PDB ID: 3JWN (ref. 17); FimG, blue; FimF, green) is superposed onto $FimH^{F18}$ · DsG based on their pilin domains (aa. 160–279), in the superposition (right) fimbrial FimH is shown in grey. (**d**) Close-up on the DsG peptide (stick representation) bound to $FimH^{F18}$ · DsG with $2F_o$–F_c electron density map. Backbone hydrogen bonds of the DsG peptide and β-strands 2 (β2) and 9 (β9) of $FimH_P$ are indicated.

Persistence of domain association in ligand-bound FimH · DsG. To test whether ligand-binding causes domain separation in FimH, we determined the co-crystal structure of the ternary complex of $FimH^{F18}$ · DsG with *n*-heptyl α-D-mannoside (HM), an established model ligand of FimH[20], as well as crystal structures of the isolated $FimH_L^{F18}$ and $FimH_L^{K12}$ lectin domains in complex with HM (Table 1). $FimH^{F18}$ · DsG · HM adopts the same closed conformation of the ligand-binding site as previously observed in other $FimH_L$–ligand complexes (Fig. 2a,b and Supplementary Fig. 2)[28]. The mannopyranose moiety of HM is coordinated by the side chains of Asp54, Gln133, Asn135 and Asp140, and the main chain of Phe1 and Asp47, and the *n*-heptyl aglycone of HM is sandwiched between Tyr48 and Tyr137. Compared with the $\mathbf{A_{free}}$ form, all loops surrounding the binding pocket close down onto the HM ligand. The most substantial conformational difference to $\mathbf{A_{free}}$ is observed for the clamp loop (aa. 8–16), whose tip moves almost 6 Å towards HM (Supplementary Fig. 1e,f).

Besides the closing of the ligand-binding pocket, the overall conformation of ligand-free $FimH_L$ in $\mathbf{A_{free}}$ and HM-bound $FimH^{F18}$ · DsG is closely similar (C_α r.m.s.d. 1.1 Å; Fig. 2a,b and Supplementary Fig. 2). Unexpectedly, the structural change in the ligand-binding site in $FimH^{F18}$ · DsG · HM was not transmitted to the domain interface, where the interdomain contacts and the conformations of the swing, linker and insertion loops remained intact. The lectin domain in the $FimH^{F18}$ · DsG · HM complex thus differs drastically from HM-bound isolated $FimH_L$ domains with respect to the swing, linker and insertion loop conformations (Fig. 2c). On the basis of the persistence of the domain **A**ssociation in the ligand-**bound** form, this state of FimH was termed $\mathbf{A_{bound}}$.

To test the stability of the $\mathbf{A_{bound}}$ state against domain separation and to exclude potential effects of selective crystallization, a molecular dynamics (MD) simulation of the $\mathbf{A_{bound}}$ state was conducted using the CHARMM36 force field of the

Table 1 | Data collection and refinement statistics

	$FimH^{F18}$ · DsG	$FimH^{K12}$ · DsG	$FimH^{K12}$ · DsG · HM	$FimH^{F18}$ · DsG · HM	$FimH^{K12}$ · DsF · HM	$FimH_L^{F18}$ · HM	$FimH_L^{K12}$ · HM
Data collection							
Space group	C 1 2 1	C 1 2 1	P 1	P 21 3	I 21 21 21	C 2 2 21	P 21 21 21
Cell dimensions							
a, *b*, *c* (Å)	99.3, 35.5, 72.8	99.5, 35.6, 72.9	56.5, 77.6, 78.1	128.4, 128.4, 128.4	94.5, 147.1, 250.8	140.1, 176.1, 28.3	63.0 68.4 95.9
α, β, γ (°)	90, 105, 90	90, 105, 90	101.5, 111.1, 96.3	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	70–1.14 (1.21–1.14)	48–1.14 (1.2–1.14)	52–2.54 (2.63–2.54)	128–2.4 (2.5–2.4)	48.1–3.0 (3.19–3.0)	54–1.42 (1.47–1.42)	52–1.7 (1.76–1.7)
R_{merge}	0.039 (0.760)	0.038 (0.691)	0.161 (0.859)	0.207 (2.424)	0.427 (1.741)	0.121 (1.242)	0.129 (1.278)
$CC_{1/2}$	100.0 (68.0)	100.0 (84.2)	99.2 (82.2)	99.8 (56.3)	95.8 (46.7)	100.0 (71.0)	99.9 (57.7)
$I/\sigma I$	16.6 (1.6)	17.8 (2.8)	8.79 (2.0)	18.3 (1.6)	14.4 (1.2)	17.9 (2.0)	15.1 (1.7)
Completeness (%)	97.6 (83.5)	94.7.0 (81.5)	90.7 (89.4)	99.9 (99.9)	99.9 (98.7)	98.9 (94.4)	96.9 (79.3)
Redundancy	3.1 (2.5)	3.2 (2.5)	3.4 (3.5)	20.4 (20.0)	6.6 (4.1)	16.9 (13.4)	11.5 (6.3)
Refinement							
Resolution (Å)	70.1–1.14	47.9–1.14	52.6–2.54	74.1–2.4	48.13–3.0	54–1.42	52.6–1.7
No. reflections	273774 (87283)	277806 (85217)	122641 (11994)	567915 (28396)	172553 (35275)	83287 (6222)	518524 (23173)
R_{work}/R_{free}	0.165/0.185	0.153/0.175	0.227/0.276	0.155/0.179	0.220/0.251	0.149/0.175	0.171/0.196
No. atoms	3102	3086	8855	2866	8751	3099	3232
Protein	2590	2590	8690	2558	8652	2427	2407
Ligand/ion	—	—	—	19	86	57	38
Water	512	496	165	284	13	596	471
B-factors							
Protein	13.8	13.3	43.7	33.5	53.2	13.2	21.3
Ligand/ion	—	—	39.7	21.3	32.3	23.0	20.7
Water	26.2	23.6	22.4	39.1	25.9	28.2	40.6
R.m.s deviations							
Bond length (Å)	0.009	0.01	0.019	0.013	0.004	0.015	0.004
Bond angles (°)	1.29	1.45	1.78	1.38	0.82	1.62	1.02

*Values in parenthesis are for the highest resolution shell.

NAMD package (Supplementary Fig. 3a,b). The domain association remained intact over 100 ns of simulation time without substantial changes in the domain interface; fluctuations were limited to the clamp loop region close to the ligand-binding site. On *in silico* removal of the HM ligand after initial equilibration, the $\mathbf{A_{bound}}$ state underwent a spontaneous transition to the $\mathbf{A_{free}}$ state after ~75 ns of simulation time via an opening of the clamp loop (Supplementary Fig. 3c,d), reproducing the experimentally observed dependence of the binding-site conformation on ligand-binding. Thus, the MD simulations indicate that $\mathbf{A_{bound}}$ is a stable conformational state of FimH induced by ligand-binding.

Trapping of a domain-separated state of full-length FimH. The increase in apparent affinity of FimH to its target glycans under tensile mechanical forces[12,29] has previously been linked to a separation of the $FimH_L$ and $FimH_P$ (ref. 17). To trap a potential domain-separated state of FimH for structural characterization in the absence of tensile force, we considered FimH variants with weakened interdomain interactions. We had shown previously that $FimH_P$ also accepts the donor strand of the non-cognate subunit FimF (DsF). However, $FimH_P$ is slightly less stabilized by complementation with DsF than with the natural donor-strand DsG[30]. We hypothesized that such complementation with DsF instead of DsG could also result in a mild destabilization of the interdomain interface in full-length FimH.

We then determined the co-crystal structure of $FimH^{K12}$ · DsF with HM ($FimH^{K12}$ · DsF · HM) at 3.0 Å resolution with four molecules in the asymmetric unit. Three $FimH^{K12}$ · DsF · HM molecules closely resembled the $\mathbf{A_{bound}}$ state (r.m.s.d. of 0.6 Å to $FimH^{F18}$ · DsG · HM) with a preserved interdomain interface. In the fourth molecule, however, the $FimH_L$ and $FimH_P$ domains were separated and they adopted a drastically different relative orientation with an angle between the domains of ~45° instead of ~150° in the other three molecules (Fig. 2a). $FimH_P$ is virtually identical in all four FimH molecules in the crystal (r.m.s.d. 0.4 Å). In contrast, the $FimH_L$ domain differs significantly between the fourth, domain-separated and the three full-length FimH molecules in the crystal. It shows closest similarity to the isolated $FimH_L$ · HM (r.m.s.d. 0.45 Å); in particular, all interdomain loops adopt identical conformations, which are incompatible with domain association (Figs 2c and 3). Remarkably, in the bent fourth molecule, no interactions between $FimH_L$ and $FimH_P$ other than the direct covalent linkage are detected, equivalent to a breakdown of the total 500 $Å^2$ interdomain interface of the $\mathbf{A_{bound}}$ state (Fig. 3). This molecule thus represents a third state, the domain-Separated, ligand-**bound** state of FimH, $\mathbf{S_{bound}}$. The complete absence of non-covalent interdomain interactions indicates that the $\mathbf{S_{bound}}$ state does not possess a defined relative domain orientation in solution, and that the observed, kinked conformation has been selected only by crystal packing.

To analyse the transition trajectory of the $\mathbf{A_{bound}}$ to the $\mathbf{S_{bound}}$ state, we removed the $FimH_P$ domain after equilibration from the $\mathbf{A_{bound}}$ state *in silico* for a 180-ns molecular dynamics simulation (Supplementary Fig. 3e,f). In contrast to the transition between the $\mathbf{A_{bound}}$ and $\mathbf{A_{free}}$ states on ligand removal, a sharp transition to the conformation of $FimH_L$ in the $\mathbf{S_{bound}}$ state was not observed. The conformation only slowly moved towards $\mathbf{S_{bound}}$; however, the $FimH_L$ loops that had formed in the former interdomain interaction kept fluctuating throughout the simulation, indicating lower cooperativity and potentially a higher activation energy for the $\mathbf{A_{bound}} \rightarrow \mathbf{S_{bound}}$ compared with the $\mathbf{A_{bound}} \rightarrow \mathbf{A_{free}}$ transition.

A comparison of the structural dynamics in the $\mathbf{A_{bound}}$ and $\mathbf{S_{bound}}$ states clearly reveals differences in the $FimH_L$–$FimH_P$ interface region. The root-mean-square fluctuations of atom positions (r.m.s.f.) increase in the swing and insertion loop from a background level of ~0.7 Å in $\mathbf{A_{bound}}$ to 1.5 and 2 Å in $\mathbf{S_{bound}}$, respectively. Surprisingly, despite the virtually identical conformations of the entire ligand-binding site depicted by X-ray crystallography (Fig. 2b), the clamp loop, which exhibits the most significant conformational changes between the open and closed

NATURE COMMUNICATIONS | DOI: 10.1038/ncomms10738

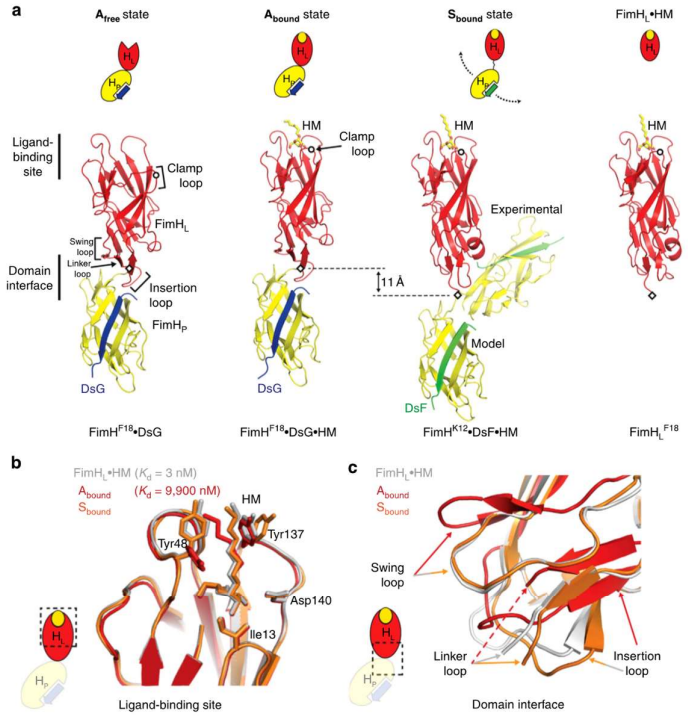


Figure 2 | Crystallographic analysis of FimH conformational states. (**a**) $FimH^{F18}$ · DsG in the $\mathbf{A_{free}}$ (left) and in the $\mathbf{A_{bound}}$ states ($FimH^{F18}$ · DsG · HM) in comparison with the $\mathbf{S_{bound}}$ state of $FimH^{K12}$ · DsF · HM and the isolated $FimH_L^{F18}$ · HM (right). The $FimH_L$, $FimH_P$, DsF and DsG are coloured in red, yellow, green and blue. The experimentally *in crystallo* trapped orientation of $FimH_P$ in $FimH^{K12}$ · DsF · HM and a modelled position based on a hinge motion stretching around Gly157 is indicated. A schematic representation for each crystal structure, similar to Fig. 1a, is given. The tip of the clamp loop and the C terminus of $FimH_L$ are indicated as a circle and diamond, respectively. (**b**) Comparison of the conformation of the ligand-binding site in the $\mathbf{A_{bound}}$ (red) and $\mathbf{S_{bound}}$ (orange) states with the isolated lectin domain $FimH_L$ (grey) and (**c**) comparison of the interdomain interface of the lectin domain.

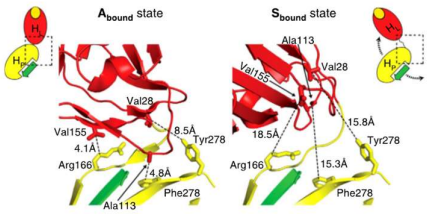


Figure 3 | The interdomain region in the $\mathbf{S_{bound}}$ state. Close-up of the interdomain region of $FimH^{K12}$ · DsF · HM in the $\mathbf{A_{bound}}$ form (left) and $FimH^{K12}$ · DsF · HM in the $\mathbf{S_{bound}}$ state (right). A cartoon representation for each crystal structure, similar to Fig. 1a, is given. Key residues in the interface are shown as sticks. $FimH_P$, $FimH_L$ and DsF are coloured in yellow, red and green, respectively.

conformations, exhibits strongly reduced fluctuations in $\mathbf{S_{bound}}$, with r.m.s.f. decreasing by up to 1.5 Å (Supplementary Fig. 3g,h). This change in clamp loop dynamics provides a mechanistic link between domain association and ligand-binding in full-length FimH.

Domain association alters FimH–ligand-binding kinetics. To analyse the ligand-binding properties of FimH · DsG, we exploited the increase in intrinsic tryptophan fluorescence in the FimH · DsG complexes of ~10% on HM binding (Fig. 4a). This difference was used to measure the dissociation constant of HM binding by equilibrium titration (Fig. 4b) and the rates of HM binding and dissociation by stopped-flow fluorescence kinetics (Fig. 4c,d). The FimH · DsG constructs showed uniform binding and dissociation kinetics, consistent with the view that domain-separated states of FimH are not significantly populated in the absence of shear force. The results revealed equilibrium dissociation constants (K_d) of 3.6 and 9.9 μM for $FimH^{K12}$ · DsG

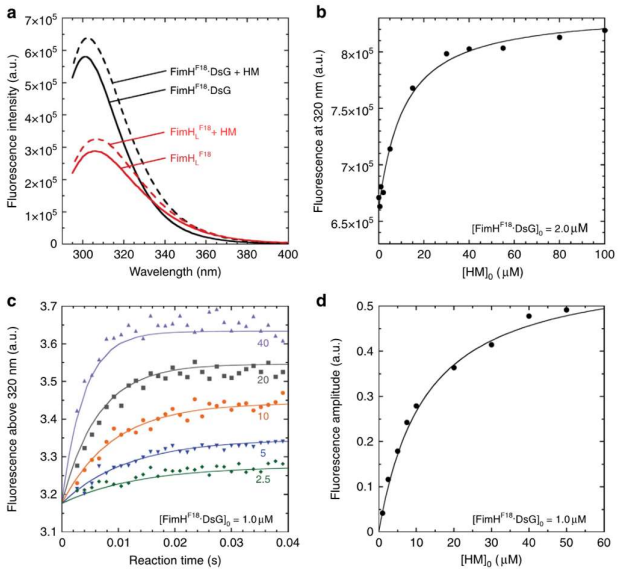


Figure 4 | Kinetics of HM binding and release by full-length FimH. (**a**) Fluorescence spectra (excitation at 280 nm) of FimH$_L^{F18}$ (2 μM; red lines) and FimH^{F18} · DsG (2 μM; black lines) in the absence (solid lines) or presence of 200 μM HM (dotted lines). (**b**) Equilibrium titration of FimH^{F18} · DsG (2 μM) with HM, recorded via the fluorescence increase at 320 nm. The total concentration of HM is plotted against the recorded fluorescence signal. Data were fitted (solid line) according to equation (2) (*cf.* experimental section) and yielded a K_d value of 9.9 ± 1.5 μM. (**c**) Stopped-flow fluorescence kinetics of HM binding to FimH^{F18} · DsG (1.0 μM), recorded via the fluorescence change above 320 nm. The HM concentration was varied between 0 and 50 μM. Five representative traces are shown (HM concentrations are given in μM). The fluorescence traces were globally fitted according to a second-order binding and first-order dissociation reaction (solid lines; Table 2). (**d**) Amplitudes of the reactions monitored in **c**, plotted against the total HM concentration. Data were fitted (solid line) according to equation (2), yielding a K_d value of 12 ± 1 μM.

and FimH^{F18} · DsG, respectively (Table 2). HM binding to FimH · DsG is extremely dynamic and was characterized by fast association rates (k_{on}) of 5.0×10^6 and $4.9 \times 10^6\,M^{-1}s^{-1}$, respectively, and rapid dissociation reactions (Supplementary Fig. 4). The rates of HM dissociation (k_{off}) of 22 and $58\,s^{-1}$ for FimH^{K12} · DsG and FimH^{F18} · DsG translate into dissociation half-lives of only 32 and 12 ms, respectively.

In contrast to full-length FimH, isolated FimH$_L^{K12}$ showed no change in tryptophan fluorescence on HM binding. We therefore determined the HM affinity of isolated FimH$_L$ indirectly by a competition experiment based on a newly designed fluorescent ligand, the fluorescein-labelled α-D-mannoside GN-FP-4 (Supplementary Fig. 5a–e and Supplementary Note 1). Displacement of GN-FP-4 from FimH$_L$ by increasing HM concentrations under equilibrium conditions showed that both FimH$_L^{K12}$ and FimH$_L^{F18}$ bind HM with 3,300-fold higher affinity compared with the respective FimH · DsG complexes (K_d values 1.1 and 3.0 nM, respectively; Fig. 5a and Table 2). In an inverse competition experiment (Supplementary Fig. 5f–j), in which HM in preformed FimH$_L$ · HM complexes was displaced by GN-FP-4, off-rates of 2.0×10^{-4} and $3.5 \times 10^{-4}\,s^{-1}$ were determined for FimH$_L^{K12}$ and FimH$_L^{F18}$, respectively, corresponding to dissociation half-lives of 58 and 33 min (Table 2). On the basis of these measured off-rates and equilibrium dissociation constants, k_{on} rates of 1.8×10^5 and $1.2 \times 10^5\,M^{-1}s^{-1}$ were calculated for FimH$_L^{K12}$ and FimH$_L^{F18}$, respectively. The on-rates for the isolated FimH$_L$ domains are thus 30-fold lower than those of the corresponding full-length FimH · DsG complexes.

Together, these results demonstrate that the 3,300-fold higher affinity of the isolated FimH$_L$ compared with full-length FimH results from a more than 100,000-fold lower ligand dissociation rate in isolated FimH$_L$, combined with a ligand-binding rate reduced by only 30-fold (Table 3). The 3,300-fold higher affinity for HM of FimH$_L$ relative to FimH · DsG translates into a free energy of $20\,kJ\,mol^{-1}$ for the interaction between FimH$_L$ and FimH$_P$ in full-length FimH. This corresponds very well with the mechanical work required for domain separation, as a displacement of FimH$_L$ from FimH$_P$ by 11 Å for complete domain separation (Fig. 2a)[17], and a force of 40 pN required to populate the domain-separated state of FimH[31] yields a value of $26.5\,kJ\,mol^{-1}$.

Domain association in FimH promotes bacterial motility. Uropathogenic *E. coli* require firm adhesion to the urinary epithelium under flow conditions to escape clearance by urine excretion. On the other hand, bacterial adhesion must be weak enough in the absence of external shear to allow flagellar motility as a prerequisite for the invasion of new tissue areas[32,33]. While the role of FimH catch-bond binding for adhesion under flow conditions had clearly been demonstrated[12,29,34], the relevance of

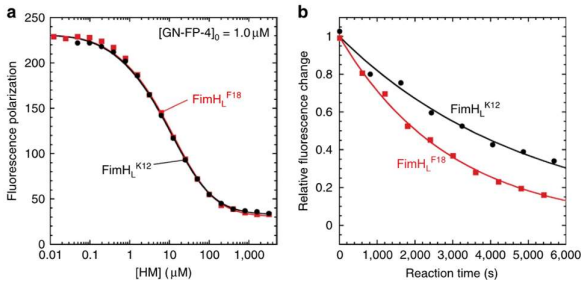


Figure 5 | HM binding and release by the isolated FimH lectin domain $FimH_L$. Analysis of $FimH_L \cdot HM$ interactions based on competition between HM and the synthetic fluorescent GN-FP-4 ligand. (**a**) HM binding to $FimH_L$ analysed by displacement of GN-FP-4 from $FimH_L$ variants as indicated. An equimolar mixture of $FimH_L$ and GN-FP-4 (1 μM each) was incubated with different HM concentrations (10 nM–3.2 mM) for >18 h. GN-FP-4 displacement is monitored by a decrease in fluorescence polarization at 528 ± 20 nm (excitation at 485 nm). Data were fitted (solid lines) according to a mechanism in which two ligands compete for the same binding site, with fixed K_d values for GN-FP-4 binding (*cf.* Table 2). (**b**) Kinetics of HM dissociation from $FimH_L$. A solution with equimolar concentrations of $FimH_L$ and HM (3 μM each, guaranteeing >95% occupancy with HM) was mixed with excess GN-FP-4 (10 μM), and the decrease in GN-FP-4 fluorescence at 520 nm as a consequence of HM dissociation and GN-FP-4 binding was recorded (Supplementary Fig. 5f–j). The obtained first-order kinetics are independent of the GN-FP-4 concentration and thus directly monitor HM dissociation.

Table 2 | Kinetics and thermodynamics of HM binding to $FimH_L$ or FimH · DsG at pH 7.4 and 25 °C.

Protein	k_{on} ($M^{-1}s^{-1}$)	k_{off} (s^{-1})	k_{off}/k_{on} (M)[a]	K_d (amplitude analysis; M)[b]	K_d (equilibrium titration; M)[c]
$FimH_L^{K12}$	$1.8 \pm 0.6 \times 10^{5}$[d]	$2.0 \pm 0.4 \times 10^{-4}$	n.a.	n.a.	$1.1 \pm 0.1 \times 10^{-9}$[e]
$FimH^{K12}$ · DsG	$5.0 \pm 0.1 \times 10^{6}$	$2.2 \pm 0.1 \times 10^{1}$	$4.3 \pm 0.1 \times 10^{-6}$	$4.2 \pm 0.3 \times 10^{-6}$	$3.6 \pm 0.3 \times 10^{-6}$
$FimH_L^{F18}$	$1.2 \pm 0.4 \times 10^{5}$[d]	$3.5 \pm 0.8 \times 10^{-4}$	n.a.	n.a.	$3.0 \pm 0.2 \times 10^{-9}$[e]
$FimH^{F18}$ · DsG	$4.9 \pm 0.1 \times 10^{6}$	$5.8 \pm 0.1 \times 10^{1}$	$1.2 \pm 0.04 \times 10^{-5}$	$1.2 \pm 0.1 \times 10^{-5}$	$9.9 \pm 1.5 \times 10^{-6}$

HM, *n*-heptyl α-D-mannoside.
The rate constants k_{on} and k_{off} were determined from experiments as shown in Figs 4 and 5. The K_D values for the complex formation between HM constructs and the different FimH constructs were determined (a) from the ratio of rate constants (k_{off}/k_{on}), (b) from the analysis of the amplitudes as in Fig. 4d and (c) from equilibrium titration as in Fig. 4b. (d) k_{on} values were calculated with k_{off} and K_d. (e) Values of K_d were obtained from competition equilibria with the fluorescent mannoside GN-FP-4 (Fig. 5a) and the following K_d values of GN-FP-4 binding determined in Supplementary Fig. 5: $FimH_L^{K12}$; $K_d = 7.0 \pm 0.1 \times 10^{-11}$ M; $FimH_L^{F18}$; $K_d = 1.8 \pm 0.2 \times 10^{-10}$ M.

Table 3 | Comparison of HM binding by variants of $FimH_L$ versus FimH · DsG.

FimH variant	k_{on} (FimH · DsG)/ k_{on} ($FimH_L$)	k_{off} (FimH · DsG)/ k_{off} ($FimH_L$)	K_d (FimH · DsG)/ K_d ($FimH_L$)
K12	28	110,000	3,300
F18	41	170,000	3,300

HM, *n*-heptyl α-D-mannoside.

rapid ligand dissociation under static conditions for flagellar motility remained unclear because of the complex interplay of flagellar swimming and the avidity of multivalent surface interactions by hundreds of *E. coli* pili. Here we employed single-cell tracking of piliated *E. coli* cells moving on surfaces coated with mono-mannosylated bovine serum albumin (1M-BSA), an established model system for analysing FimH-based adhesion[12,29], for a classification of cell motility into two states, attached or mobile (for details see Methods and Supplementary Fig. 6). To study the influence of FimH interdomain interactions, we compared isogenic *E. coli* strains producing either wild-type $FimH^{F18}$ or the $FimH^{F18}$-variant Ala188Asp, which is characterized by a destabilized interaction between $FimH_P$ and $FimH_L$ (ref. 26) and serves here to mimic the $\mathbf{S_{bound}}$ state in the absence of shear force[35]. The overall fraction of adherent $FimH^{F18}$-piliated bacteria on 1M-BSA-coated surfaces was identical to background levels on non-adhesive BSA-coated surfaces at 10–12% of tracked bacteria (Fig. 6a). In contrast, $FimH^{F18}$–Ala188Asp-piliated bacteria showed an increased fraction of adherent cells of 24% already at the beginning of cell tracking after a 1-min dead time, which further increased during the 5-min observation period to 48% (Fig. 6a and Supplementary Movies 1 and 2).

Cell tracking permitted quantitative analysis of the transition of individual cells between a mobile and an attached state (Supplementary Fig. 7). On non-adhesive control surfaces coated only with BSA, less than 2% of all $FimH^{F18}$- or $FimH^{F18}$–Ala188Asp bacteria showed shifts between the two states of motion (Fig. 6b and Supplementary Fig. 7c). However, on adhesive 1M-BSA surfaces, 13.9% of all $FimH^{F18}$ tracks (green in Fig. 6b) exhibited a single transient attachment event with a mean duration of 6.9 s (Supplementary Fig. 7d). For $FimH^{F18}$–Ala188Asp-piliated bacteria, only 7.2% of the cells showed attachment/detachment, but with fivefold longer adhesion (35.2 s; Supplementary Fig. 7d). Remarkably, the fraction of cells that permanently stayed attached after adhesion to 1M-BSA until

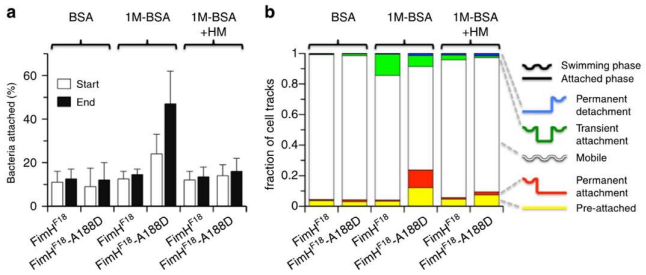


Figure 6 | Cell-tracking analysis of bacterial motility on mannosylated surfaces. *E. coli* cells piliated with FimH^{F18} or the FimH^{F18}-Ala188Asp variants were tracked under static conditions in the absence of shear force. (**a**) The fraction of bacteria attached to mannose-coated (1M-BSA) or BSA-coated surface (negative control) at the beginning of the time-lapse movies (white bars) and after 5 min (black bars) are given. Bacterial motility on 1M-BSA was analysed in the absence and presence of HM. The delay between application of bacteria and movie recording was ~1 min. (**b**) Fraction of tracked cells that were pre-attached (yellow; speed $<0.5\,\mu m\,s^{-1}$), permanently attach (red), were mobile (white), transiently attach (green) or permanently detach (blue) during the entire observation time (5 min). Right: schematic depiction of the observed cell behaviour. FimH^{F18}-piliated *E. coli* show almost exclusively transient attachment events on 1M-BSA. FimH^{F18}-Ala188Asp-piliated *E. coli* show less transient attachment but enhanced permanent attachment to 1M-BSA. Transient and permanent attachment to 1M-BSA is significantly reduced in the presence of HM. For each experiment five to seven independent replicates were analysed.

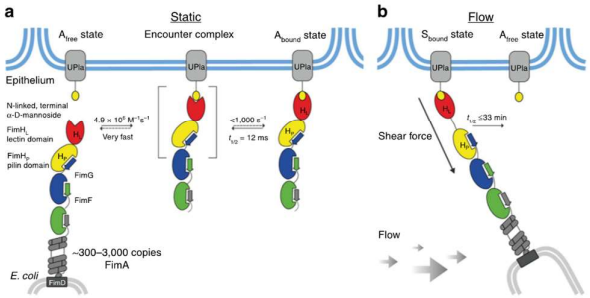


Figure 7 | Catch-bond mechanism of FimH-mediated cell adhesion. (**a**) In the absence of tensile mechanical force, formation of the FimH-Uroplakin 1a (UPIa) complex comprises the highly dynamic transition of the $\mathbf{A_{free}}$ to the $\mathbf{A_{bound}}$ state. The reaction likely proceeds via a transient encounter complex (indicated in square brackets). The reaction of the encounter complex to $\mathbf{A_{bound}}$ is not rate-limiting and must have a half-life of less than 1 ms. Dissociation of the receptor from the FimH lectin domain in the $\mathbf{A_{bound}}$ state is promoted via dynamic allostery by the pilin domain that acts as a negative allosteric regulator. The reaction from $\mathbf{A_{bound}}$ to the encounter complex corresponds to k_{Off}. Fast binding and release of UPIa by FimH enables bacterial motility on the bladder epithelium. (**b**) Shear force increases the population of the $\mathbf{S_{bound}}$ state of FimH, in which the pilin and lectin domains are separated. The dissociation of $\mathbf{S_{bound}}$ under shear force is slowed down up to 100,000-fold compared with $\mathbf{A_{bound}}$. The indicated rate constants and half-lives correspond to the interaction between FimH^{F18} and the model ligand HM. Rate limiting reactions are indicated by solid arrows, and fast, non-limiting reactions by dashed arrows.

the end of the observation period was much larger for FimH^{F18}–Ala188Asp (11.5%) than for FimH^{F18} (0.6%; red in Fig. 6b). Those permanently attached cells escape kinetic analysis; thus, the true average attachment time for FimH^{F18}–Ala188Asp must be considerably larger than 35.2 s. Permanent attachment is also the main cause of the increased fraction of attached cells for FimH^{F18}–Ala188Asp-piliated bacteria (Fig. 6a).

Altogether, cell-tracking analysis revealed that enforced domain separation in the FimH^{F18}–Ala188Asp variant resulted in reduced detachment rates and a larger proportion of permanently attached cells. These results directly demonstrate, at the cellular level, the importance of fast, spontaneous ligand dissociation catalysed by interdomain allostery in FimH–ligand complexes for bacterial motility in the absence of tensile mechanical forces.

Discussion

The characterization of full-length FimH had so far been restricted to the analysis of the adhesive properties of piliated *E. coli* cells and binding studies with the purified type 1 pilus tip fibrillum. With the FimH · DsG complex, we have now established a model system for quantitative studies of the interaction of FimH with carbohydrate ligands. Soluble FimH · DsG efficiently mimicks FimH in the context of the assembled tip fibrillum, is readily available in milligram quantities and permits the determination of ligand-binding and release kinetics in solution.

Using FimH · DsG, we obtained high-resolution snapshots of three functionally relevant states of FimH (Fig. 7). In the absence of ligands, FimH adopts the $\mathbf{A_{free}}$ state with associated $FimH_L$ and $FimH_P$ and an open conformation of the ligand-binding site, which is responsible for the 30-fold faster ligand-binding of full-length FimH as compared with the isolated $FimH_L$ domain. Ligand-binding in the absence of shear force induces the $\mathbf{A_{bound}}$ state with a closed binding site. In contrast to earlier hypotheses[17], the transition from $\mathbf{A_{free}}$ to $\mathbf{A_{bound}}$ is restricted to the ligand-binding site, while all interactions between $FimH_L$ and $FimH_P$ observed in the $\mathbf{A_{free}}$ state remain preserved in $\mathbf{A_{bound}}$. The $\mathbf{A_{free}} \rightarrow \mathbf{A_{bound}}$ transition most likely follows an induced fit mechanism, in which the formation of an encounter complex between FimH · DsG and HM is rate-limiting and followed by a fast, unimolecular rearrangement to the $\mathbf{A_{bound}}$ state, in agreement with the observation that binding of the model ligand HM remained rate-limiting for the formation of $\mathbf{A_{bound}}$ even at the highest HM concentrations used. Stopped-flow-binding kinetics indicate that the lifetime of the proposed encounter complex before $\mathbf{A_{bound}}$ formation is below 1 ms (Fig. 4c). Under tensile mechanical force applied to the FimH–ligand complex, mimicked here by the destabilized variant FimH · DsF and crystal packing forces, the domain-separated state of FimH, $\mathbf{S_{bound}}$, is formed. In the $\mathbf{S_{bound}}$ state, $FimH_L$ and $FimH_P$ no longer interact specifically and are only connected via the linker segment comprising FimH residues 154–160. In this $\mathbf{S_{bound}}$ state, $FimH_L$ adopts a conformation closely resembling isolated $FimH_L$ with bound ligand.

Notably, ligand dissociation from FimH · DsG is 100,000-fold faster than that from the isolated $FimH_L$ domain. This is striking because the respective crystal structures revealed indistinguishable ligand interactions and binding-site conformations in the FimH · DsG · HM and $FimH_L$ · HM complexes (Fig. 2b). MD simulations identified a considerable increase in the conformational dynamics of the $FimH_L$–ligand-enclosing clamp loop in the $\mathbf{A_{bound}}$ state as the most likely cause of the dramatic increase in k_{off} in the FimH · DsG · HM complex. The altered dynamics in $FimH_L$ in the $\mathbf{A_{bound}}$ state are the result of the presence of $FimH_P$, which can be described as a negative allosteric regulator[36–38]. The allosteric communication from the $FimH_P$–$FimH_L$ interface to the ligand-binding site reaches over 40 Å, and is mediated via changes in protein dynamics rather than in static structure, in line with a general model of dynamic allostery[39,40]. Our data demonstrate that the interdomain interactions in FimH (i) maintain the open conformation of the binding pocket and guarantee rapid ligand-binding and (ii) intramolecularly catalyse ligand dissociation by more than 100,000-fold. Rapid ligand-binding and short lifetimes of the FimH ligand complex allow for rapid dissociation of individual pili from their ligands in the absence of shear force. Our biophysical data demonstrate that this mechanism is conserved between the K12 and the F18 *E. coli* strains.

Different mechanistic models, such as the two-pathway[41], the deformation[42] and the sliding re-binding model[43], have been developed to describe catch-bond interactions, often based on powerful single-molecule atomic force measurements. These models included the principle of allosteric control of ligand-binding affinity[26,31], which was clearly fully confirmed in our present study. However, these conceptual models did not reveal the underlying atomic-scale mechanisms in different catch-bond systems. For most catch-bond systems, including the cadherin–catenin binding to actin filaments[3,44], integrin epithelial cell adhesion[45,46] and TCR–MHC interactions[6,7,47], structural information is, if at all, available only for one state or from computer simulations. One exception is the selectins, which employ catch-bond binding for leukocyte recruitment. Selectins are multidomain cell surface receptors, which consist of a lectin domain for complex carbohydrate binding, linked via an epidermal growth factor (EGF)-like domain to a variable number of short consensus repeat domains and a transmembrane-anchoring helix. Selectins exist in two conformations, a bent and an extended one, which differ in the angle between their lectin and EGF-like domain. Ligand-binding and conformational changes in the ligand-binding site are directly linked via a complex allosteric coupling mechanism to the adoption of the extended conformation[48,49]. Tensile mechanical force under flow conditions acts along the axis of the ligand-binding site and the Lec-EGF interface resulting in a stabilization of the extended conformation and thus increased ligand complex lifetimes[2,49]. Moreover, in FimH, catch-bond behaviour is mediated by the interplay of a lectin and an anchoring domain that does not interact with the ligand. Ligand-binding by FimH in the absence of shear force results in a closing of the ligand-binding site, but, in contrast to selectins, is not directly linked to altered interdomain interactions. Here mechanical force promotes domain separation and completely releases $FimH_L$ from $FimH_P$, which acts as an activator of ligand release via dynamic allostery. Remarkably, selectins and the fimbrial adhesin FimH thus employ entirely different mechanisms for establishing catch-bond behaviour by crosstalk between a lectin and an anchoring domain that provides tethering to a shaft. In both systems, the selectins and fimbrial adhesion, the shaft structures linking the terminal lectin/coupling domains to the cell surface, may contribute to the overall catch-bond behaviour, either via directly influencing coupling domain behaviour or via their general elastic properties[50,51].

The cell-tracking experiments indicate the importance of rapid ligand release from the high-mannose-type glycoprotein receptor uroplakin 1a in the lower urinary tract[52] for flagellar motility of piliated bacteria, and hence their ability to colonize new tissue areas under certain conditions during infection[12,29,53]. This provides a plausible explanation for the fact that low-affinity FimH variants were preserved in numerous uropathogenic *E. coli* strains. Binding of terminal mannoses with low affinity in the absence of shear force may also play a role in preventing the clearance of uropathogenic *E. coli* from the urinary tract by competitive binding to the Tamm–Horsfall protein in the urine[54]. In turn, populating the $\mathbf{S_{bound}}$ state with an extremely low dissociation rate ensures tight bacterial adhesion under the mechanical forces of urine excretion. FimH is a promising target for anti-adhesive therapy of UTI because FimH antagonists, in contrast to antibiotics, are not exerting selection pressure towards resistance formation[18,55,56]. Previous ligand-binding studies on the isolated $FimH_L$ domain mimic the domain-separated $\mathbf{S_{bound}}$ state of FimH. This state is characterized by extremely low off-rates and is promoted *in vivo* only after ligand-binding and the onset of flow conditions. Our kinetic data on ligand dissociation from full-length FimH demonstrate that rapid, competitive displacement of FimH from its carbohydrate ligands by FimH antagonists is well possible in the absence of shear force. Thus, full-length FimH (for example, in the form of the FimH · DsG complex established in this study) instead of the isolated $FimH_L$ domain is the relevant target for the development of anti-adhesive drugs. Importantly, the concept of the FimH · DsG model system can now be expanded to other related adhesive pilus adhesins. In combination with the novel fluorescent GN-FP-4 ligand, this model system paves the way for efficient screening for anti-adhesive drug candidates.

Methods

Materials. The synthetic DsG (sequence: ADVTITVNGKVVAKR) and DsF peptide (sequence: ADSTITIRGYVRDNG; >95% purity) were purchased from JPT (Germany). Guanidinium chloride ('AA-Grade' for spectroscopy) was

obtained from NIGU Chemie (Germany). Standard chemical of highest purity available was obtained Sigma, Merck or AppliChem. If not mentioned otherwise, chromatography media for protein purification were purchased from GE Healthcare (UK). Oligonucleotides were from Microsynth (Switzerland).

Construction of expression plasmids. Expression plasmids for the periplasmic production of the *E. coli* F18 FimH lectin domain ($FimH_L^{F18}$) and for the periplasmic co-expression of full-length $FimH^{F18}$ with FimC were based on the expression plasmids $pfimH_L$ and pfimH-fimC-ATG, respectively, for the analogous proteins from *E. coli* K12 (ref. 27). Six silent mutations replacing rare codons were introduced into the *E. coli* F18 *fimH* gene (*fimH*^{F18}) contained in the plasmid pGB2-24 (ref. 57) with the QuikChange mutagenesis kit (Agilent Technologies, Switzerland) to improve periplasmic expression. The coding sequence of the modified *fimH*^{F18} gene was amplified by PCR using the primers 5′-GATCCTCTA GAGGAGGGATGATTGTAATGAAACGAG-3′ and 5′-TTTCAAGCTTATTGA TAAACAAAAGTCACG-3′ and cloned into pfimH-fimC-ATG[27] via the XbaI and HindIII sites (thereby replacing the *fimH*^{K12} gene) and yielded $pfimH^{F18}$–fimC–ATG. The gene encoding $FimH_L^{F18}$ was amplified with the primers 5′-GATCCTCT AGAGGAGGGATGATTGTAATGAAACGAG-3′ and 5′-CAGCCAAGCTTAG CCAGTAGGCACCACCAC-3′ and ligated via the XbaI and HindIII sites into ptrc99a-f1-stopp[27]. Protein production in the resulting plasmid $pfimH_L^{F18}$ is under control of the trc–promotor/*lac* operator.

Protein production and purification. For purification of the complexes FimC · $FimH^{K12}$ and FimC · $FimH^{F18}$, *E. coli* HM125 harbouring the corresponding co-expression plasmid was grown at 30 °C in 2YT medium containing ampicillin (100 μg ml^{-1}). At an OD_{600} of 1.5, isopropyl-β-D-thiogalactoside (IPTG) was added to a final concentration of 1 mM. The cells were further grown for 12 − 18 h, harvested by centrifugation, suspended in cold 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 1 mg ml^{-1} polymyxin B sulfate (18 ml l^{-1} of culture) and stirred at 4 °C for 1.5 h. After centrifugation, the supernatant (periplasmic extract) was dialysed against 20 mM Tris-HCl pH 8.0 and applied to a QA52 (Whatman, Maidstone, UK) column equilibrated with the same buffer. The flow-through containing the respective FimC · FimH complex was dialysed against 20 mM MOPS–NaOH pH 7.0, loaded onto a Resource S column equilibrated with the same buffer and the complexes were eluted with a linear NaCl gradient (0 − 400 mM). Fractions containing FimC · FimH were pooled and loaded onto a Superdex 75 (HiLoad 26/60) column equilibrated with 20 mM NaH_2PO_4–NaOH pH 7.4, 50 mM NaCl. Fractions containing the pure complex were pooled and stored at 4 °C until further use. Typically, 3 − 5 mg of the purified complex were obtained per litre of bacterial culture.

For expression of the isolated *E. coli* $FimH_L^{K12}$ and $FimH_L^{F18}$, *E. coli* HM125 transformed with the respective expression plasmid was grown at 30 °C in M9 medium containing ampicillin (100 μg ml^{-1}) to an OD_{600} of 1.0, and expression was induced with 1 μM IPTG. After further growth for 12 h, cells were subjected to periplasmic extraction (see above). The extracts were mixed with 0.11 volumes of 1 M acetic acid–NaOH pH 4.5, dialysed against 10 mM acetic acid–NaOH pH 4.5 and then loaded onto a SP-Sepharose column equilibrated with the same buffer. The flow-through was collected and its pH was adjusted to 8.0 by addition of 1 M Tris-HCl pH 8.2. This solution was then applied to a Q-Sepharose column equilibrated with 20 mM Tris-HCl pH 8.0. The flow-through containing $FimH_L$ was loaded onto a Resource S column dialysed against 20 mM formic acid–NaOH pH 4.0. The protein was eluted with a linear NaCl gradient (0 − 1 M). Fractions containing pure $FimH_L$ were pooled, dialysed against water and stored at −20 °C. The identity of the purified proteins was confirmed by electrospray ionization (ESI)–mass spectrometry ($FimH_L^{F18}$; calculated: 16,934.9 Da; measured: 16,935.0 Da; $FimH_L^{K12}$; calculated: 16,963.0 Da; measured: 16,962.8 Da). About 11 mg of the pure $FimH_L$ was obtained per litre of bacterial culture.

Production of FimH · DsG and $FimH^{K12}$ · DsF complexes. The respective FimC · FimH complex (40 μM) was incubated with a threefold molar excess of the DsG peptide and incubated in 20 mM NaH_2PO_4–NaOH, pH 7.0, 50 mM NaCl for 48 h at 37 °C. The reaction mixture containing isolated FimC, the FimH · DsG complex and excess DsG was dialysed against 20 mM acetic acid–NaOH pH 4.5 and loaded onto a Resource S (6 ml) column equilibrated with the same buffer. The FimH · DsG complex was eluted with a linear NaCl gradient (0 − 400 mM). Fractions containing the pure complex were pooled, dialysed against water and stored at 4 °C. The FimH · DsG partially dissociated during ESI–mass spectrometry analysis, so that masses of the intact complexes and free FimH were obtained: $FimH^{K12}$ · DsG: calculated mass: 30,635.3 Da; measured mass: 30,636.0 Da; $FimH^{F18}$ · DsG: calculated mass: 30,607.3 Da; measured mass: 30,607.0 Da; $FimH^{K12}$; calculated mass: 29,064.5 Da; measured mass: 29,064.0 Da; $FimH^{F18}$; calculated mass: 29,036.4 Da; measured mass: 29,036.0 Da. The overall yields of the purified FimH · DsG complexes relative to the initial amount of FimC · FimH were in the range of 50–55%. The $FimH^{K12}$ · DsF complex was generated and purified as described for the FimH · DsG complexes ($FimH^{K12}$ · DsF: calculated mass: 30,702.2 Da; measured mass: 30,702.5 Da). The FimH · DsF complex was prepared from the FimC · FimH complex after mixing with DsF exactly according to the protocol described above for FimH · DsG and obtained in similar yields. Despite the non-natural interaction between $FimH_F$ and DsF, the $FimH^{K12}$ · DsF complex was formed four times faster than the $FimH^{K12}$ · DsG complex at pH 7.0 and 37 °C, with a rate constant of 2.2 ± 0.5 M^{-1} s^{-1}. The FimH · DsF complex was stable against dissociation and unspecific aggregation.

Determination of protein concentrations. Protein concentrations were measured via the specific absorbance at 280 nm, using the following extinction coefficients ($FimH^{K12}$ and $FimH^{F18}$ have identical extinction coefficients): FimC · FimH (59,090 M^{-1} cm^{-1}), $FimH_L$ (24,670 M^{-1} cm^{-1}), FimH · DsG (35,090 M^{-1} cm^{-1}) and FimH · DsF (36,580 M^{-1} cm^{-1}). The concentrations of DsG and DsF were determined via their absorbance at 205 nm (42,650 and 49,700 M^{-1} cm^{-1}, respectively).

Synthesis of the fluorescent-labelled FimH ligand GN-FP-4. To a stirred solution of mannoside **1** (25 mg, 0.061 mmol)[55] in dry dimethylformamide (DMF; 1 ml), N-hydroxysuccinimide (21 mg, 0.183 mmol) was added, followed by N,N′-dicyclohexylcarbodiimide (9.2 mg, 0.073 mmol). The mixture was stirred at room temperature for 2 h, then *N*-Boc-ethylendiamine (10.7 mg, 0.067 mmol) was added and the reaction was stirred for an additional 10 h. After cooling to 0 °C, the reaction mixture was diluted with water and concentrated. Chromatography on silica gel (CH_2Cl_2/MeOH) yielded 23 mg (0.042 mmol, 68%) of *tert*-butyl (3′-chloro-4′-(α-D-mannopyranosyloxy)-biphenyl-4-yl-carboxamido)ethyl)carbamate. This product was dissolved in CH_2Cl_2 (3 ml) and trifluoroacetic acid (TFA, 1 ml) was added. The solid dissolved during addition of TFA. After 10 min at room temperature the reaction was complete. The mixture was evaporated and excess TFA was removed in high vacuum. The intermediate *N*-(2-aminoethyl)-3′-chloro-4′-(α-D-mannopyranosyloxy)-biphenyl-4-carboxamide TFA salt (23 mg, 0.042 mmol, quant.) was used without purification in the next step. After dissolution in dry DMF (0.5 ml), triethylamine (12.8 mg, 0.127 mmol) was added. The mixture was cooled to 0 °C, then fluorescein isocyanate (14.8 mg, 0.038 mmol) was added and the mixture was stirred for 3 h in the dark. After the addition of water, DMF was removed azeotropically, the residue dissolved in MeOH/10% acetic acid and evaporated. Chromatography on silica gel (CH_2Cl_2/MeOH) yielded compound **2**, contaminated with triethylammonium acetate. Therefore, after dissolution in MeOH, 0.5 N HCl in MeOH was added, the mixture evaporated and chromatographed on silica gel to yield pure compound GN-FP-4 ((3′-Chloro-*N*-(2-(3-(3′,6′-dihydroxy-3-oxo-3*H*-spiro[isobenzofuran-1,9′-xanthen]-5-yl)-thioureido)ethyl)-4′-(α-D-mannopyranosyloxy)-biphenyl-4-carboxamide)[58] (15 mg, 47%). $[\alpha]_D^{20}$ + 12.1 (*c* 0.3, MeOH); ^{1}H NMR (500 MHz, CD_3OD): δ = 8.12 (s, 1H), 7.92 (d, *J* = 8.3 Hz, 2H, Ar–H), 7.70 (dd, *J* = 5.0, 13.1 Hz, 2H, Ar–H), 7.64 (d, *J* = 8.3 Hz, 2H, Ar–H), 7.54 (dd, *J* = 2.2, 8.6 Hz, 1H, Ar–H), 7.46 (d, *J* = 8.7 Hz, 1H, Ar–H), 7.09 (d, *J* = 8.2 Hz, 1H, Ar–H), 6.74 (s, 2H), 6.69 (d, *J* = 1.4 Hz, 2H, Ar–H), 6.55 (d, *J* = 8.4 Hz, 2H, Ar–H), 5.63 (d, *J* = 1.3 Hz, H-1), 4.15 (dd, *J* = 1.8, 3.1 Hz, H-2), 4.03 (dd, *J* = 3.4, 9.5 Hz, H-3), 3.94 (s, 2H, CH_2), 3.86-3.64 (m, 6H, H-4, H-5, H-6, CH_2); ^{13}C NMR (126 MHz, CD_3OD): δ = 153.21, 143.84, 136.41, 129.66, 129.18, 127.76, 127.70, 125.37, 118.64, 103.62 (Ar-C), 100.75 (C-1), 76.00 (C-5), 72.41 (C-3), 71.86 (C-2), 68.24 (C-4), 62.69 (C-6), 40.76 (CH_2); ESI–MS: *m*/*z*: Calculated for $C_{42}H_{37}ClN_3O_{12}S$ $[M + H]^+$: 842.2, found: 842.2.

Fluorescence spectroscopy. Fluorescence emission spectra of FimH variants were recorded between 300 and 450 nm (excitation at 280 nm) at 25 °C in 1.0 × 0.4-cm quartz cuvettes on a QM 7/2003 spectrofluorimeter (PTI, USA) equipped with a magnetic stirrer. Protein concentrations were 1 − 2 μM in 20 mM MOPS–NaOH pH 7.4. Fluorescence spectra of GN-FP-4 ($\varepsilon_{495\,nm}$ = 54,900 M^{-1} cm^{-1}) were recorded between 500 and 650 nm (excitation at 497 nm) in the same buffer.

Kinetics of HM binding to FimH · DsG. The rate constants of binding (k_{on}) and dissociation (k_{off}) for the complex between FimH · DsG and HM were measured at 25 °C in 20 mM MOPS–NaOH pH 7.4 in a SX20 stopped-flow instrument (Applied Photophysics, UK). A constant FimH · DsG concentration of 1 or 2 μM was used. FimH · DsG was mixed with different concentrations of HM (2 − 100 μM), and binding was monitored by the increase in fluorescence above 320 nm (excitation at 280 nm). The fluorescence traces were globally fitted with *Dynafit*[59] according to a second-order binding and first-order dissociation reaction. As an additional control, the fluorescence amplitudes of the individual reactions were plotted against the total HM concentration and fitted according to equation (2). The deduced dissociation constants reproduced the K_d values obtained with equilibrium titration within experimental error.

Equilibrium titration of FimH · DsG with HM. The binding equilibrium between FimH · DsG and HM was followed at 25 °C in 20 mM MOPS–NaOH pH 7.4 on a QM 7/2003 spectrofluorometer (PTI) by the increase in fluorescence at 320 nm on HM binding (excitation at 280 nm). Measurements were performed with a stirred 1 × 0.4-cm quartz cuvette. The concentration of FimH · DsG was kept constant at 2 μM and the concentration of HM was varied between 0 and 200 μM. The samples were equilibrated overnight, and their fluorescence intensities were recorded for 30 s and averaged. The fluorescence intensities were plotted against total HM

concentration and fitted according to equation (2)

$$F = (F_\infty - F_0)\cdot\frac{[P]_0 + [L]_0 + K_d - \sqrt{([P]_0 + [L]_0 + K_d)^2 - 4\cdot[P]_0\cdot[L]_0}}{2\cdot[P]_0} + F_0 \quad (2)$$

where F is the monitored fluorescence signal, F_0 is the fluorescence signal in absence of ligand, F_∞ is the fluorescence signal at full saturation with ligand, K_d is the dissociation constant, $[P]_0$ is the total concentration of FimH · DsG and $[L]_0$ is the total concentration of HM.

Equilibrium titration of $FimH_L$ with GN-FP-4. The binding equilibrium between $FimH_L$ and GN-FP-4 at 25 °C in 20 mM MOPS–NaOH pH 7.4, supplemented with 0.001% Tween 20 to prevent unspecific adsorption effects at nanomolar concentrations, was recorded by the decrease in GN-FP-4 fluorescence at 520 nm (excitation at 497 nm). Measurements were performed with a stirred 1 × 0.4-cm quartz cuvette. The concentration of GN-FP-4 was kept constant at 1.0 or 2.0 nM and the concentration of $FimH_L$ was varied between 0 and 10 nM. The samples were equilibrated overnight, and their fluorescence intensities at 520 nm were recorded for 30 s and averaged. The experimental data were fitted according to equation (2).

Displacement of HM from the $FimH_L$ by GN-FP-4. The rate constant of dissociation (k_{off}) for HM from $FimH_L$ at 25 °C in 20 mM MOPS–NaOH pH 7.4 was measured indirectly by binding of excess GN-FP-4 to $FimH_L$ after dissociation of HM, recorded with the decrease in GN-FP-4 fluorescence at 520 nm (excitation at 497 nm). A mixture of $FimH_L$ and HM (3 μM each), pre-incubated for at least 18 h, was mixed with different amounts of excess GN-FP-4 (final concentrations: 10 − 40 μM), and GN-FP-4 fluorescence was recorded every 10 min for 10 s, averaged and the data fitted according to first-order kinetics. The obtained rate constants were independent of GN-FP-4 concentration and thus identical to the dissociation rate of HM from $FimH_L$.

Determination of the $FimH_L$ · HM dissociation constant. The affinity of $FimH_L$ for HM at 25 °C in 20 mM MOPS–NaOH pH 7.4 was determined by the competition between HM and GN-FP-4 for binding to $FimH_L$. A mixture of $FimH_L$ and GN-FP-4 (1 μM each) was incubated with different concentrations of HM (10–3.2 mM) and incubated for at least 18 h. The displacement of GN-FP-4 by HM was recorded on by the decrease in the fluorescence polarization at 528 ± 20 nm (excitation at 485 ± 20 nm) on a microplate reader (Biotek, USA), using flat black-bottom 96-well microtitre plates (Greiner, Austria). The fluorescence polarization data were fitted with *Dynafit*[59] according to an equilibrium competition mechanism, with the total concentrations of $FimH_L$, GN-FP-4 and HM (variable) and the respective K_d of GN-FP-4 (Table 2) as input, and K_d of HM and the fluorescence polarization at zero and infinite HM concentration as open parameters.

Crystallization of FimH variants. All crystallization experiments were performed at 4 °C with the sitting drop vapour diffusion method. For crystallization, $FimH^{F18}$ · DsG and $FimH^{K12}$ · DsG (0.1 − 0.2 μl, 15 mg ml^{-1} in H_2O) was mixed with 0.1 − 0.2 μl of precipitant (25% (w/v) polyethylene glycol (PEG) 3350, 0.2 M magnesium chloride, 0.1 M BisTris-HCl pH 5.5 at 4 °C. Crystals of $FimH^{F18}$ · DsG and $FimH^{K12}$ · DsG grew within 4 − 6 weeks and are of the space group C2, with one molecule per asymmetric unit. $FimH^{K12}$ · DsG crystals in space group P1 grew at 0.2 M Na Malonate, 20% PEG3350 within 2 months at 4 °C. For crystallization of the $FimH^{F18}$ · DsG · HM complex, a threefold excess over FimH · DsG was used (protein concentration and protein/precipitant ratios were as described for FimH · DsG). Crystals of the space group $P2_13$ appeared after 1 month in 30% (v/v) 2-Methyl-2,4-pentanediol, 0.1 M sodium cacodylate, 0.2 M magnesium acetate pH 6.5 at 4 °C. $FimH^{K12}$ · DsF · HM crystals appeared after 2 months in 30% w/v PEG 5000, 0.1 M 2-(*N*-morpholino)ethanesulfonic acid (MES) monohydrate, 0.2 M ammonium sulfate pH 6.5 at 20 °C (2.5-fold excess of ligand over protein). $FimH_L^{F18}$ · HM crystallized in 17% PEG 2000 MME, 0.1 M HEPES–NaOH pH 7.5 at 4 °C. $FimH_L^{K12}$ · HM crystallized in 1.5 M $(NH4)_2SO_4$, 0.2 M Na acetate pH 5.5 at 20 °C.

Crystallographic data collection. All crystals, except for $FimH^{F18}$ · DsG · HM, were cryo-preserved by the addition of ethane-1,2-diol to a final concentration of 20% (v/v). The precipitant solution used for the crystallization of $FimH^{F18}$ · DsG · HM already contained 30% (v/v) methyl-2,4-pentanediol, which acts as cryoprotectant. Crystals were flash-cooled in liquid nitrogen. All measurements were carried out at the SLS beamline X06DA and X06SA (Swiss Light Source, Paul Scherrer Institute, Switzerland) at 100 K. All data were integrated, indexed and scaled using the XDS software package[60] (5% of the reflections were set aside as test set). Data collection statistics are summarized in Table 1.

Crystallographic structure determination. All structures were solved by molecular replacement using structures of isolated $FimH_L$ (AA1-158, PDB ID: 3MCY[18], and the pilin domain of FimC · FimH (AA160-297, PDB ID: 1QUN[24], as search models with the programme Phaser[61]). Model building and structure refinement were performed with Coot (ref. 62) and PHENIX (ref. 63). Twelve out of thirteen residues could be built for the FimG donor strands in the crystal structures, and only the C-terminal lysine residue had weak electron density. Refinement statistics are summarized in Table 1.

Molecular dynamics simulations. Four molecular systems were prepared for $FimH^{F18}$. The first system was constructed using the $\mathbf{A_{bound}}$ state of $FimH^{F18}$ · DsG · HM (Supplementary Fig. 3a,b) and the second system is equivalent but HM was removed (Supplementary Fig. 3c,d). The third system contains only the $FimH_L$ and HM from the $\mathbf{A_{bound}}$ X-ray structure (Supplementary Fig. 3e,f). The fourth system was prepared for $FimH_L$ and HM based on the $\mathbf{S_{bound}}$ state in the $FimH^{K12}$ · DsF · HM crystal structure (Supplementary Fig. 3g,h).

The CHARMM-GUI web server[64] was used to prepare the molecular systems, which were solvated with TIP3 water molecules and ionized with 50 mM NaCl. Each system contains between 50,000 and 60,000 atoms. All simulations were performed with the NAMD simulation package (version 2.9) (ref. 65). The CHARMM36 force field was used for the protein, and parameters for HM were generated using the CHARMM General Force Field programme (version 0.9.7 beta). Electrostatic interactions were calculated using the particle-mesh Ewald method[66] with a grid spacing of 1 Å. The cutoff for the van der Waals interactions was taken at 12 Å with a switching function used after 10 Å. Time step for the integration of dynamics was 2 fs. Simulations were performed in an isothermal–isobaric ensemble, with a pressure of 1 atm and a temperature of 300 K.

Cell tracking on mannose-BSA-coated surfaces. The *E. coli* KB18 strain[67] was kindly provided by Professor Evgeni Sokurenko and served as host for the generation of recombinant strains. KB18 contains the pPKL114 plasmid[57], which encodes the whole *fim* operon with a translational stop linker upstream of the *fimH* gene. KB18 was co-transformed with the pGB2-24 plasmid, which was isolated from the ELT115 strain and encodes *fimH*[196] (kindly provided by Professor Evgeni Sokurenko). Single-nucleotide point mutations were introduced in *fimH*[196] using overlap extension PCR following standard molecular techniques to obtain *fimH*^{F18} and *fimH*^{F18}-Ala188Asp. The PCR products were cloned into pGB2-24 by the ApaLI and SphI sites, and KB18 was transformed with the resulting plasmid.

E. coli strains were grown from frozen stocks in LB medium supplemented with antibiotics (100 μg ml^{-1} ampicillin and 25 μg ml^{-1} chloramphenicol) until late log phase (OD_{600} of 1.0–1.2) and diluted to an OD_{600} of 0.01 before movie acquisition.

Cell culture dishes (35 mm, Corning Inc., Corning, NY) were incubated with 50 μl of 50 μg ml^{-1} 1M-BSA in 0.02 M bicarbonate buffer for 75 min at 37 °C. The dishes were then washed three times and quenched with 0.1% PBS–BSA to remove unbound 1M-BSA and block remaining sites on the plastic surface to prevent nonspecific binding of bacteria. Controls were prepared by treating cell culture dishes with 0.1% PBS–BSA only. The bacterial suspension was added to the cell culture dishes for microscopy studies.

Cell tracking was carried out at room temperature under static conditions. A bacterial suspension of 50 μl in the late logarithmic growth phase was placed onto the cell culture dishes (diluted to OD_{600} of 0.01), and a cover slide was placed on top. The delay between sample placement and start of the movie acquisition was about 1 min. Time-lapse movies were recorded with a × 20 phase contrast objective using a CMOS digital camera (The Imaging Source Europe, Bremen, Germany) mounted on a Nikon Ti Eclipse inverted microscope and using the NIS Elements Basic Research software (Nikon, Zurich, Switzerland). Phase contrast images in an ∼5-μm-thick surface layer were taken at four to five frames per second over 5 min. The dead time of movie acquisition was ∼1 min. The resulting images were segmented by creating a projection of the average intensities over all frames to remove the background and by subsequent thresholding using the Maximum Entropy method in Fiji[68] to obtain binary images (examples shown in Supplementary Movies 1 and 2). The segmented movies were imported into Imaris (Bitplane, Zurich, Switzerland) and tracked through the autoregressive algorithm. A time filter was applied to exclude all tracks with a length below 15 s. Tracks longer than 15 s were reviewed individually and edited manually, if necessary. Five to seven independent movies were recorded for each experimental set-up: $FimH^{F18}$ or $FimH^{F18}$-Ala188Asp on 1M-BSA-coated dishes and $FimH^{F18}$ and $FimH^{F18}$-Ala188Asp on BSA-coated cell culture dishes. *E. coli* piliated with $FimH^{F18}$ or $FimH^{F18}$-Ala188Asp binding to 1M-BSA in the absence (1,815 and 1,283 individual tracks, respectively) and in the presence of 200 μM HM (1,175 and 1,071 individual tracks, respectively) were analysed respectively. For *E. coli* piliated with $FimH^{F18}$ or $FimH^{F18}$-Ala188Asp binding to BSA 1,314 and 1,065 individual tracks, respectively, were analysed. Bacteria with a speed of <0.5 μm s^{-1} were classified as attached, all other bacteria were classified as mobile. Owing to limitation in the spatial and temporal resolution of movie acquisition, we did not further subdivide bacterial swimming into motility behaviours as 'rolling'[29,53], 'roaming', 'orbiting' and so on. The individual cell tracks were classified into four classes: no motility change during observation (pre-attached or mobile), transient attachment, permanent attachment and permanent detachment. For $FimH^{F18}$ on 1M-BSA surfaces, 13.9% (251 out of 1,815 tracks) of all tracks showed a single transient attachment event (Supplementary Fig. 7d). In total, 67 out of the 251

bacteria that underwent a first transient adhesion attached and detached from the surface a second time. For these cells, the average time between detachment and re-attachment was only 13.5 s (Supplementary Fig. 7e), suggesting that re-binding may be favoured by proximity to the surface as compared with the initial attachment. The mean velocity on 1M-BSA, as compared with BSA-coated surfaces, was reduced for both $FimH^{F18}$-piliated (4.2 and 7.4 $\mu m\,s^{-1}$, respectively) and $FimH^{F18}$–Ala188Asp-piliated bacteria (3.5 and 8.1 $\mu m\,s^{-1}$, respectively; Supplementary Fig. 7a,b). This reduction of the mean velocity originates from two different phenomena: in $FimH^{F18}$-piliated cells it is caused by a change from fast swimming to a slower mode of motion (Supplementary Fig. 7b; Supplementary Movie 1), which is consistent with bacterial surface rolling due to weak, short-lived mannose-based interactions[29,53]. In contrast, for $FimH^{F18}$–Ala188Asp-piliated bacteria, the reduction of the mean velocity results from an increase in the fraction of adherent cells on 1M-BSA compared with BSA (see main text). In the presence of 200 μM HM, the mean velocity on 1M-BSA is increased for both $FimH^{F18}$-piliated (6.5 $\mu m\,s^{-1}$) and $FimH^{F18}$–Ala188Asp-piliated bacteria (5.9 $\mu m\,s^{-1}$; Supplementary Fig. 7a) and transient and permanent attachment is reduced by 75% and 85%, respectively (Fig. 6 and Supplementary Fig. 7c).

References

1. Marshall, B. T. *et al.* Direct observation of catch bonds involving cell-adhesion molecules. *Nature* **423,** 190–193 (2003).
2. Evans, E., Leung, A., Heinrich, V. & Zhu, C. Mechanical switching and coupling between two dissociation pathways in a P-selectin adhesion bond. *Proc. Natl Acad. Sci. USA* **101,** 11281–11286 (2004).
3. Buckley, C. D. *et al.* Cell adhesion. The minimal cadherin-catenin complex binds to actin filaments under force. *Science* **346,** 1254211 (2014).
4. Akiyoshi, B. *et al.* Tension directly stabilizes reconstituted kinetochore-microtubule attachments. *Nature* **468,** 576–579 (2010).
5. Fiore, V. F., Ju, L., Chen, Y., Zhu, C. & Barker, T. H. Dynamic catch of a Thy-1-alpha5beta1 + syndecan-4 trimolecular complex. *Nat. Commun.* **5,** 4886 (2014).
6. Liu, B., Chen, W., Evavold, B. D. & Zhu, C. Accumulation of dynamic catch bonds between TCR and agonist peptide-MHC triggers T cell signaling. *Cell* **157,** 357–368 (2014).
7. Das, D. K. *et al.* Force-dependent transition in the T-cell receptor beta-subunit allosterically regulates peptide discrimination and pMHC bond lifetime. *Proc. Natl Acad. Sci. USA* **112,** 1517–1522 (2015).
8. Ronald, A. The etiology of urinary tract infection: traditional and emerging pathogens. *Am. J. Med.* **113,** 14S–19S (2002).
9. Jones, C. H. *et al.* FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proc. Natl Acad. Sci. USA* **92,** 2081–2085 (1995).
10. Hahn, E. *et al.* Exploring the 3D molecular architecture of *Escherichia coli* type 1 pili. *J. Mol. Biol.* **323,** 845–857 (2002).
11. Waksman, G. & Hultgren, S. J. Structural biology of the chaperone-usher pathway of pilus biogenesis. *Nat. Rev. Microbiol.* **7,** 765–774 (2009).
12. Thomas, W. E., Trintchina, E., Forero, M., Vogel, V. & Sokurenko, E. V. Bacterial adhesion to target cells enhanced by shear force. *Cell* **109,** 913–923 (2002).
13. Zhou, G. *et al.* Uroplakin Ia is the urothelial receptor for uropathogenic *Escherichia coli*: evidence from *in vitro* FimH binding. *J. Cell Sci.* **114,** 4095–4103 (2001).
14. Sharon, N. Carbohydrates as future anti-adhesion drugs for infectious diseases. *Biochim. Biophys. Acta* **1760,** 527–537 (2006).
15. Ofek, I., Hasty, D. L. & Sharon, N. Anti-adhesion therapy of bacterial diseases: prospects and problems. *FEMS Immunol. Med. Microbiol.* **38,** 181–191 (2003).
16. Guo, B. & Guilford, W. H. Mechanics of actomyosin bonds in different nucleotide states are tuned to muscle contraction. *Proc. Natl Acad. Sci. USA* **103,** 9844–9849 (2006).
17. Le Trong, I. *et al.* Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like beta sheet twisting. *Cell* **141,** 645–655 (2010).
18. Han, Z. *et al.* Structure-based drug design and optimization of mannoside bacterial FimH antagonists. *J. Med. Chem.* **53,** 4779–4792 (2010).
19. Wellens, A. *et al.* The tyrosine gate as a potential entropic lever in the receptor-binding site of the bacterial adhesin FimH. *Biochemistry* **51,** 4790–4799 (2012).
20. Bouckaert, J. *et al.* Receptor binding studies disclose a novel class of high-affinity inhibitors of the *Escherichia coli* FimH adhesin. *Mol. Microbiol.* **55,** 441–455 (2005).
21. Wellens, A. *et al.* Intervening with urinary tract infections using anti-adhesives based on the crystal structure of the FimH-oligomannose-3 complex. *PLoS ONE* **3,** e2040 (2008).
22. Brument, S. *et al.* Thiazolylaminomannosides as potent antiadhesives of type 1 piliated *Escherichia coli* isolated from Crohn's disease patients. *J. Med. Chem.* (2013).
23. Vanwetswinkel, S. *et al.* Study of the structural and dynamic effects in the FimH adhesin upon alpha-d-heptyl mannose binding. *J. Med. Chem.* **57,** 1416–1427 (2014).
24. Choudhury, D. *et al.* X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic *Escherichia coli*. *Science* **285,** 1061–1066 (1999).
25. Schwartz, D. J. *et al.* Positively selected FimH residues enhance virulence during urinary tract infection by altering FimH conformation. *Proc. Natl Acad. Sci. USA* **110,** 15530–15537 (2013).
26. Yakovenko, O. *et al.* FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. *J. Biol. Chem.* **283,** 11596–11605 (2008).
27. Vetsch, M., Sebbel, P. & Glockshuber, R. Chaperone-independent folding of type 1 pilus domains. *J. Mol. Biol.* **322,** 827–840 (2002).
28. Gouin, S., Roos, G. & Bouckaert, J. *Discovery and Application of FimH Antagonists* 1–46 (Springer Berlin Heidelberg, 2014).
29. Thomas, W. E., Nilsson, L. M., Forero, M., Sokurenko, E. V. & Vogel, V. Shear-dependent 'stick-and-roll' adhesion of type 1 fimbriated *Escherichia coli*. *Mol. Microbiol.* **53,** 1545–1557 (2004).
30. Puorger, C. *et al.* Infinite kinetic stability against dissociation of supramolecular protein complexes through donor strand complementation. *Structure* **16,** 631–642 (2008).
31. Thomas, W. *et al.* Catch-bond model derived from allostery explains force-activated bacterial adhesion. *Biophys. J.* **90,** 753–764 (2006).
32. Lane, M. C. *et al.* Role of motility in the colonization of uropathogenic *Escherichia coli* in the urinary tract. *Infect. Immun.* **73,** 7644–7656 (2005).
33. Wright, K. J., Seed, P. C. & Hultgren, S. J. Uropathogenic *Escherichia coli* flagella aid in efficient urinary tract colonization. *Infect. Immun.* **73,** 7657–7668 (2005).
34. Nilsson, L. M., Thomas, W. E., Trintchina, E., Vogel, V. & Sokurenko, E. V. Catch bond-mediated adhesion without a shear threshold: trimannose versus monomannose interactions with the FimH adhesin of *Escherichia coli*. *J. Biol. Chem.* **281,** 16656–16663 (2006).
35. Aprikian, P. *et al.* Interdomain interaction in the FimH adhesin of *Escherichia coli* regulates the affinity to mannose. *J. Biol. Chem.* **282,** 23437–23446 (2007).
36. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J. & Lee, A. L. Hidden dynamic allostery in a PDZ domain. *Proc. Natl Acad. Sci. USA* **106,** 18249–18254 (2009).
37. Tzeng, S. R. & Kalodimos, C. G. Protein dynamics and allostery: an NMR view. *Curr. Opin. Struct. Biol.* **21,** 62–67 (2011).
38. Laskowski, R. A., Gerick, F. & Thornton, J. M. The structural basis of allosteric regulation in proteins. *FEBS Lett.* **583,** 1692–1698 (2009).
39. Cooper, A. & Dryden, D. T. Allostery without conformational change. A plausible model. *Eur. Biophys. J.* **11,** 103–109 (1984).
40. Tsai, C. J., del Sol, A. & Nussinov, R. Allostery: absence of a change in shape does not imply that allostery is not at play. *J. Mol. Biol.* **378,** 1–11 (2008).
41. Pereverzev, Y. V., Prezhdo, O. V., Forero, M., Sokurenko, E. V. & Thomas, W. E. The two-pathway model for the catch-slip transition in biological adhesion. *Biophys. J.* **89,** 1446–1454 (2005).
42. Pereverzev, Y. V. & Prezhdo, O. V. Force-induced deformations and stability of biological bonds. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **73,** 050902 (2006).
43. Lou, J. Z. & Zhu, C. A structure-based sliding-rebinding mechanism for catch bonds. *Biophys. J.* **92,** 1471–1485 (2007).
44. Manibog, K., Li, H., Rakshit, S. & Sivasankar, S. Resolving the molecular mechanism of cadherin catch bond formation. *Nat. Commun.* **5,** 3941 (2014).
45. Choi, Y. I. *et al.* Dynamic control of beta1 integrin adhesion by the plexinD1-sema3E axis. *Proc. Natl Acad. Sci. USA* **111,** 379–384 (2014).
46. Kong, F., Garcia, A. J., Mould, A. P., Humphries, M. J. & Zhu, C. Demonstration of catch bonds between an integrin and its ligand. *J. Cell Biol.* **185,** 1275–1284 (2009).
47. Depoil, D. & Dustin, M. L. Force and affinity in ligand discrimination by the TCR. *Trends Immunol.* **35,** 597–603 (2014).
48. Somers, W. S., Tang, J., Shaw, G. D. & Camphausen, R. T. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1. *Cell* **103,** 467–479 (2000).
49. Preston, R. C. *et al.* E-selectin ligand complexes adopt an extended high-affinity conformation. *J. Mol. Cell Biol.* **8,** 62–72 (2015).
50. Zakrisson, J., Wiklund, K., Axner, O. & Andersson, M. The shaft of the type 1 fimbriae regulates an external force to match the FimH catch bond. *Biophys. J.* **104,** 2137–2148 (2013).
51. Whitfield, M. J., Luo, J. P. & Thomas, W. E. Yielding elastic tethers stabilize robust cell adhesion. *PLoS Comput. Biol.* **10,** e1003971 (2014).
52. Xie, B. *et al.* Distinct glycan structures of uroplakins Ia and Ib: structural basis for the selective binding of FimH adhesin to uroplakin Ia. *J. Biol. Chem.* **281,** 14644–14653 (2006).
53. Anderson, B. N. *et al.* Weak rolling adhesion enhances bacterial surface colonization. *J. Bacteriol.* **189,** 1794–1802 (2007).
54. Pak, J., Pu, Y., Zhang, Z. T., Hasty, D. L. & Wu, X. R. Tamm-Horsfall protein binds to type 1 fimbriated *Escherichia coli* and prevents *E. coli* from binding to uroplakin Ia and Ib receptors. *J. Biol. Chem.* **276,** 9924–9930 (2001).
55. Klein, T. *et al.* FimH antagonists for the oral treatment of urinary tract infections: from design and synthesis to *in vitro* and *in vivo* evaluation. *J. Med. Chem.* **53,** 8627–8641 (2010).

56. Cusumano, C. K. *et al.* Treatment and prevention of urinary tract infection with orally active FimH inhibitors. *Sci. Transl. Med.* **3,** 109ra115 (2011).
57. Sokurenko, E. V., Courtney, H. S., Ohman, D. E., Klemm, P. & Hasty, D. L. FimH family of type 1 fimbrial adhesins: functional heterogeneity due to minor sequence variations among fimH genes. *J. Bacteriol.* **176,** 748–755 (1994).
58. Kleeb, S. *et al.* FimH antagonists: bioisosteres to improve the *in vitro* and *in vivo* PK/PD profile. *J. Med. Chem.* **58,** 2221–2239 (2015).
59. Kuzmic, P. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. *Anal. Biochem.* **237,** 260–273 (1996).
60. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66,** 133–144 (2010).
61. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
62. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60,** 2126–2132 (2004).
63. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* 58, 1948–1954, 2002).
64. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29,** 1859–1865 (2008).
65. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
66. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).
67. Sokurenko, E. V., Courtney, H. S., Maslow, J., Siitonen, A. & Hasty, D. L. Quantitative differences in adhesiveness of type 1 fimbriated *Escherichia coli* due to structural differences in fimH genes. *J. Bacteriol.* **177,** 3680–3686 (1995).
68. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9,** 676–682 (2012).

Acknowledgements

We thank the staff at the Swiss Light Source (Villigen, Switzerland) for outstanding support for crystallographic data collection and the Imaging Facility as well as Research IT/A, and Mazur at Biozentrum, Basel, for supporting cell-tracking experiments. Discussions with Dr Jochen Reinstein (Max-Planck-Institute for Medical Research) are gratefully acknowledged. This work was supported by the Swiss National Science Foundation (Projects R'Equip 145023 (T.M) and Project 144183 (B.E., T.M. and R.G.), 310030B_138657 (R.G.) and 31003A_156304 (R.G.)). Calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing core facility at University of Basel and the Swiss National Supercomputing Centre.

Author contributions

M.M.S., R.P.J., R.G. and T.M. conceived the study and experimental approach; M.M.S. and J.E. cloned, purified and crystallized proteins and carried out kinetic experiments; R.P.J. crystallized proteins, collected X-ray diffraction data and determined crystal structures; G.N. and B.E. synthesized GN-FP-4. Se.B. and Si.B. performed molecular dynamics simulations. D.E. and R.P.J. performed cell-tracking experiments: and M.M.S., R.P.J, J.E., B.E., R.G. and T.M. wrote the paper with input from all authors.

Additional information

Accession codes: The atomic coordinates and structure factors have been deposited in the Protein Data Bank under the accession codes 4XO8, 4XO9, 4XOA, 4XOB, 4XOC, 4XOD, 4XOE.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Sauer, M. M. *et al.* Catch-bond mechanism of the bacterial adhesin FimH. *Nat. Commun.* 7:10738 doi: 10.1038/ncomms10738 (2016).

2.2.5 Paper 4

Label-Free FimH Protein Interaction Analysis Using Silicon Nanoribbon BioFETs

In this paper the label-free detection of FimH protein with biosensors based on silicon nanoribbons (SiNRs) is described. SiNRs configured as ion-sensitive field-effect transistors (ISFETs) are of great interest for diagnosis. The performance of the system was compared to the well-established commercial instrument Biacore T-200. The results highlighted some crucial elements for achieving successful label-free detection of pathologically relevant proteins with ISFETs and provided the first proof of concept for this application.

Contribution to the project:
Giulio Navarra designed, synthesized, and characterized the mannose-based ligands that were used in the study. He established the surface functionalization methods and prepared protein samples for the measurements. He additionally performed the Biacore experiments.

This paperwas published in *ACS Nano*:

Mathias Wipf, Ralph L. Stoop, Giulio Navarra, Said Rabbani, Beat Ernst, Kristine Bedner, Christian Schönenberger, and Michel Calame

ACS SENSORS

Article

pubs.acs.org/acssensors

Label-Free FimH Protein Interaction Analysis Using Silicon Nanoribbon BioFETs

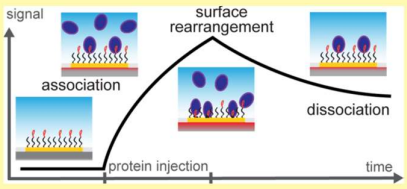
Mathias Wipf,*[,†] Ralph L. Stoop,[†] Giulio Navarra,[‡] Said Rabbani,[‡] Beat Ernst,[‡] Kristine Bedner,[§] Christian Schönenberger,[†] and Michel Calame[†]

[†]Department of Physics, and [‡]Institute of Molecular Pharmacy, Pharmacenter, University of Basel, 4003 Basel, Switzerland

[§]Laboratory for Micro- and Nanotechnology, Paul Scherrer Institut, 5232 Villigen, Switzerland

Supporting Information

ABSTRACT: The detection of biomarkers at very low concentration and low cost is increasingly important for clinical diagnosis. Moreover, monitoring affinities for receptor-antagonist interactions by time-resolved measurements is crucial for drug discovery and development. Biosensors based on ion-sensitive field-effect transistors (BioFETs) are promising candidates for being integrated into CMOS structures and cost-effective production. The detection of DNA and proteins with silicon nanowires has been successfully demonstrated using high affinity systems such as the biotin–streptavidin interaction. Here, we show the time-resolved label-free detection of the interaction of the bacterial FimH lectin with an immobilized mannose ligand on gold-coated silicon nanoribbon BioFETs. By comparing our results with a commercial state of the art surface plasmon resonance system, additional surface effects become visible when using this charge based detection method. Furthermore, we demonstrate the effect of sensor area on signal-to-noise ratio and estimate the theoretical limit of detection.



KEYWORDS: *silicon, nanoribbon, biosensor, BioFET, FimH, protein, gold, ion-sensitive field-effect transistor, signal-to-noise ratio*

Detection and quantification of biological and chemical species are central to many areas of research in life sciences and healthcare, ranging from diagnosing diseases to discovery and screening of new drug molecules. Monitoring the binding affinities and kinetics of protein–ligand interactions is crucial in drug research. A real-time measurement of molecular interactions by a sensing device reveals the valuable information on binding affinities[1] and offers a useful tool for disease diagnosis,[2] genetic screening[3] and drug discovery.[4] The search for new therapeutic candidates often requires screening of compound libraries. At present, the state of the art is surface plasmon resonance (SPR).[5] However, the high throughput screening application of this technique is rather limited and cost-intensive. An alternative method to measure protein–ligand interactions is silicon nanowire (SiNW) field-effect transistors (FETs).[6,7] The direct transduction of the analyte–surface interaction into an electrical signal allows real-time and high-throughput detection of biomolecules. Immobilizing the ligand directly on the sensor surface allows highly specific, label-free detection.[8,9] In the past, it has been demonstrated that FET based biosensors (BioFETs) allow the detection of biomolecular interactions down to picomolar concentrations.[9−11] However, most of this research has been focused on reducing the limit of detection (LOD). So far, studies on quantifying the signals, specifically binding affinities and kinetic data, have primarily focused on DNA interaction[12] and biotin–streptavidin interactions.[4] However, the biotin–streptavidin binding is one of the strongest noncovalent interactions known in nature (its dissociation constant K_D is on the order of $\approx 10^{-14}$ M),[13] and therefore, its significance for interaction studies and benchmark for minimum LOD is questionable.

In this work, we demonstrate the real-time detection of a therapeutically relevant protein with gold-coated silicon nanoribbons (SiNRs). Clear concentration dependent signals were obtained upon protein injection. The simultaneous measurement of several SiNRs in active and control arrays increased the amount of data and allowed the comparison of different sensor dimensions. Our results are a proof of concept for the use of BioFETs for kinetic studies of protein–ligand binding. As analyte we have chosen the therapeutically relevant FimH lectin. Lectins are highly specific carbohydrate-binding proteins, that are involved in numerous physiological and pathophysiological processes, including cell–cell recognition, inflammation, immune response, cancer, and pathogen tropism.[14,15] FimH is a bacterial lectin. Its expression is highly correlated with urinary tract infections (UTIs), for which *E. coli* expressing the FimH protein at the tip of their pili are the main causative agent. In the human urinary tract, FimH enables bacterial adhesion to the urothelium, which is the first step of

Received: February 5, 2016
Accepted: May 10, 2016
Published: May 10, 2016

DOI: 10.1021/acssensors.6b00089
ACS Sens. 2016, 1, 781−788

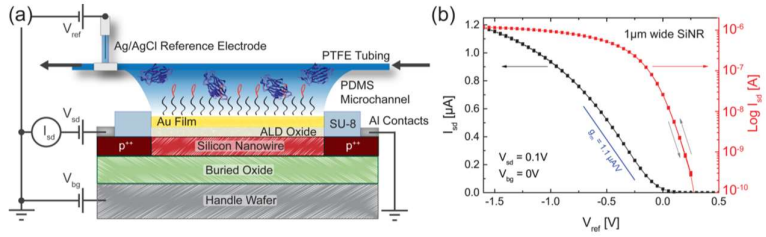


Figure 1. Measurement setup. (a) Cross section of the fabricated device and a schematic of the silicon nanoribbon biosensor setup. The gold film, deposited on top of the HfO_2 gate oxide, is covered by a SAM of MHDA to which the ligands are attached by amine coupling. PDMS microchannels and PTFE tubings are used as fluidic system. A peristaltic pump (indicated by arrows) and a flow-trough Ag/AgCl reference electrode are used to control the flow speed and the liquid potential, respectively. A constant voltage V_{sd} = 0.1 V is applied across source and drain. The back gate voltage V_{bg} is applied to the bulk silicon support (handle) wafer (generally set to 0 V) and the liquid gate voltage V_{ref} is applied to the reference electrode. FimH proteins in the solution bind to the ligands and thereby change the surface potential Ψ_0, which leads to a change in source-drain current I_{sd}. (b) Source-drain current (I_{sd}) vs liquid potential (V_{ref}) for a 1 μm wide gold-coated SiNR in pH 8 buffer solution. For the time-resolved measurements the SiNRs are operated in the linear region where the transconductance g_m is constant as indicated by the blue line.

the infection.[16,17] In previous work, we have synthesized and evaluated high affinity FimH antagonists, demonstrating their therapeutic potential for the treatment of UTIs.[18−21] Since a crucial factor for the efficacy of a therapeutic agent is the half-life of the drug−receptor complex, kinetics of the binding process and equilibrium dissociation constants are of special interest. We show that BioFETs are potential candidates to compete with SPR, the state of the art method to study these parameters. The possibility for high integration, up-scaling, and the low cost of the BioFET technology[22] are very attractive features from which diagnostics and drug discovery could benefit in the near future.

■ RESULTS AND DISCUSSION

Silicon Nanoribbon Biosensor. Sensor chips containing 48 SiNRs were fabricated on silicon on insulator wafers by a top-down microfabrication process according to previously published protocols.[2,3,24] The nanoribbon pattern was defined by electron-beam lithography and transferred by reactive ion etching of SiO_2 and anisotropic wet etching of Si. HfO_2 gate dielectric (8 nm) was deposited by atomic layer deposition (ALD) to ensure high gate oxide capacitance and low leakage currents. Contact regions were highly doped by ion implantation and sealed with SU-8 photoresist to prevent liquid contact. SiNRs used for biosensing were coated with a 20 nm thick gold film to reduce the pH response and enable thiol-based surface functionalization, as described in our previous work.[25] Further, it allows the direct comparison to SPR based biosensors. Polydimethylsiloxane (PDMS) microfluidic channels, with a flow through Ag/AgCl reference electrode embedded in the tubing, were used for well controlled liquid transport. However, potential fluctuations from air bubbles limit the signal-to-noise ratio (SNR). Therefore, a liquid cell with ≈15 μL volume and embedded Ag/AgCl reference electrode was used as an alternative to study the SNR.

Measurements were performed at constant liquid flow and at a fixed working point, i.e. source-drain voltage V_{sd} = 0.1 V, back-gate voltage V_{bg} = 0 V, and constant liquid potential (V_{ref}) to operate the SiNRs in the linear regime. Changes in surface potential (Ψ_0) upon analyte binding shift the threshold voltage (V_{th}) which changes I_{sd}. The relation for a p-type semiconductor is given by the transconductance ($g_m = \partial I_{sd}/\partial V_{ref}$) according to

$$\Delta\Psi_0 = -\Delta V_{th} = -\frac{\Delta I_{sd}}{g_m} = \frac{q_A}{C_0}[B]_0\frac{[A]}{K_D + [A]} \quad (1)$$

Here q_A is the electric charge given by an adsorbed analyte and C_0 is the capacitive coupling (in [F/m^2]) between the charge of the analyte molecule within the double layer and the bulk solution. It is influenced by the double layer capacitance and hence dependent on the ionic strength of the buffer solution.[26,27] $[B]_0$ is the total number of surface bound ligands per unit area. The last term describes the ratio of surface bound analytes at equilibrium, given by the site binding model.[24,28] $[A]$ is the analyte bulk concentration and K_D is the equilibrium dissociation constant, which describes the protein−ligand affinity. g_m can be determined by $I_{sd} - V_{ref}$ measurements of each SiNR or by applying gate steps in the time-resolved measurement. Using this conversion introduced by Duan et al.[4] the signal is no longer a function of the FET performance and only depends on $\Delta\Psi_0$ induced by the analyte.

In Figure 1, a schematic cross section of the SiNR biosensor setup is shown. Proteins injected to the liquid system adsorb to the functional layer and change Ψ_0. To ensure that the proteins are within the electrical double layer and hence their charges affect the surface potential, the measurements were performed in buffer with reduced ionc strength. 10 mM HEPES buffer pH 8 (with a Debye length $\lambda_D \geq 3$ nm) was used. A peristaltic pump (MCP, Ismatec) and a valve selector system (CHEM-INERT VICI, Valco Instruments Co. Inc.) were used to exchange the solutions. The liquid potential was applied to the reference electrode (V_{ref}). A Keithley 2636A source meter with two channels was used to apply the source-drain bias V_{sd} and to measure source-drain current I_{sd}. A switching box (Keithley 3706) was used to address all the 48 nanoribbons on the chip. The back gate-voltage V_{bg} was applied at the handle wafer. All devices were automatically controlled by a self-made LabView program.

Figure 1b shows the transfer curve $I_{sd}(V_{ref})$ of a 1 μm wide gold-coated SiNR in pH 8 buffer solution. The p-type transistor is operated in accumulation mode. The transconductance is extracted from the linear regime.

DOI: 10.1021/acssensors.6b00089
ACS Sens. 2016, 1, 781−788

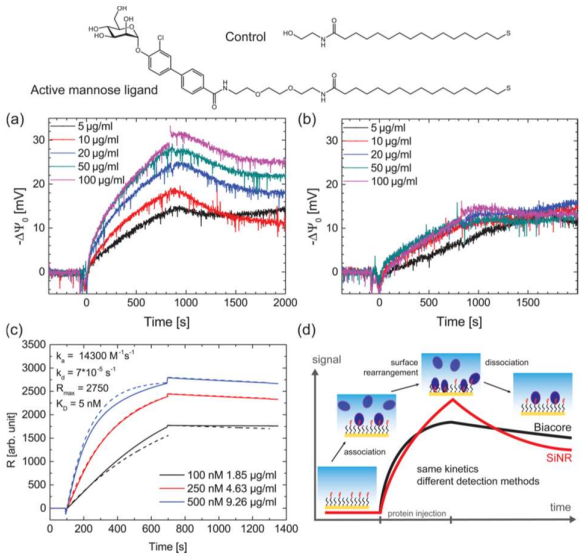


Figure 2. FimH binding kinetics. Real-time sensor response upon injection of FimH proteins at different concentrations. (a) Active SiNR shows pronounced concentration dependent protein adsorption and initial desorption upon rinsing with buffer after 900 s. (b) Control SiNR shows nonspecific adsorption of FimH proteins, which we associate with the lipophilic character of the MHDA monolayer. (c) Reference experiment measured in the SPR system (Biacore T200) on a Au chip functionalized with the active mannose ligand. The signal starts to saturate already at smaller FimH concentrations and dissociation is less pronounced. 1:1 Langmuir kinetic fits are indicated by the dashed lines. An equilibrium dissociation constant of $K_D \approx 5$ nM is obtained. (d) Schematic of a binding cycle comparing typical sensor responses of SiNRs and Biacore. Association of proteins to the surface ligands occurs upon FimH injection and dissociation upon switching to running buffer. Since a very similar surface on the SiNRs and the Biacore chip is expected, the binding kinetics should be similar. The difference in signal can be explained by the two different detection methods. Whereas the surface plasmon resonance detects larger molecules within ≥100 nm from the surface, the BioFET detects charges within ≤3 nm from the surface (at the used buffer concentration). We expect the adsorbed proteins to interact with the hydrophobic MHDA layer and move closer to the sensor surface. This surface rearrangement is a slow process and only affects the SiNR signal.

The ligands used for the sensor surface functionalization for specific (active) and unspecific (control) protein adsorption are shown in Figures 2 and 3. Two different methods were used. In a two-step method, the gold surface was first coated with a monolayer of 16-mercaptohexadecanoic acid (MHDA) and afterward a high affinity mannoside was attached by amine coupling. Ethanolamine, which is uncharged at pH 8 was used as control. Additionally a one-step method with disulfide bonds (Figure 3) for direct ligand immobilization on gold was used. We did not observe a difference in binding kinetics for the mannoside ligand using the two different functionalization methods. Functionalization scheme, control experiments as well as details for the FimH protein expression and purification are shown in the Supporting Information. To exclude signals from background salt concentration the protein was dialyzed against 10 mM HEPES buffer pH 8.

FimH Protein Detection. In Figure 2a, the real-time sensor response of a SiNR with active mannose ligand for five different FimH concentrations in 10 mM HEPES buffer pH 8 ranging from 5 μg/mL up to 100 μg/mL (1 μg/mL ≈ 54 nM) is shown. Since the aim of affinity interaction studies is not to detect the analyte at physiological concentration, but to obtain and compare the affinity of antagonists, the concentration range was chosen to obtain kinetic data within acceptable measurement times. After each cycle, the surface was regenerated by flushing the system with 6 M urea for 10 min. At pH 8, FimH is negatively charged, leading to an increase in I_{sd} upon protein adsorption. Using a p-type semiconductor, $-\Delta\Psi_0$ is plotted as a function of time. The straight line, obtained for the first 400 s prior to the binding event, was subtracted to correct drift and to set the baseline to zero. Time = 0 s is defined as the onset of FimH adsorption. The response to FimH is clearly concentration dependent, but does not follow 1:1 Langmuir kinetics perfectly. In particular because the slope of the association saturates at high protein concentration and no equilibrium is observed even after 15 min. The variation in dissociation for the 5 μg/mL signals (active and control) can be associated with a change in baseline drift.

Figure 2b shows the response of a control SiNR. A weaker signal is observed, which we attribute to nonspecific adsorption of FimH to the lipophilic layer of the MHDA functionalization.

DOI: 10.1021/acssensors.6b00089
ACS Sens. 2016, 1, 781−788

ACS Sensors Article

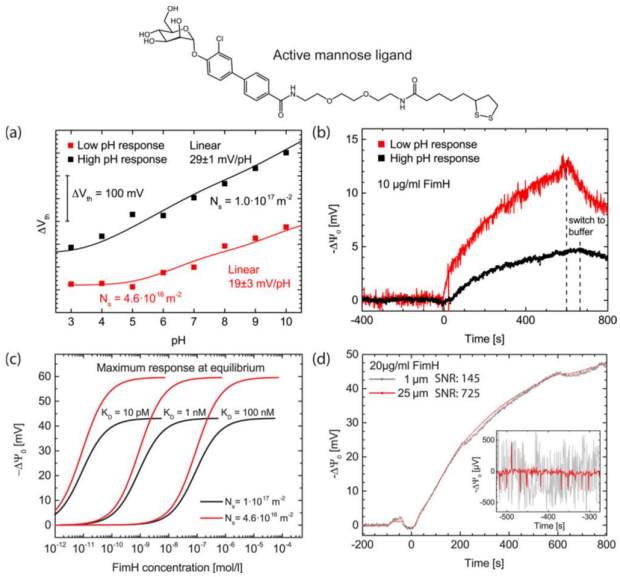


Figure 3. Signal-to-noise ratio. Competing surface reactions limit the signal. (a) pH response (ΔV_{th} vs pH) for gold-coated SiNRs functionalized with the active mannose ligand shown on top. V_{th} is extracted from $I_{sd} - V_{ref}$ sweeps. The two different data sets show the same sample measured after different FimH measurement series. The lines correspond to the site binding model (eq 3) at different hydroxyl group density N_s ($pK_a = 9$, $pK_b = 7$). Depending on N_s, the linear response around pH 8 varies from ≈19 mV/pH (low pH response) to 29 mV/pH (high pH response). (b) Real-time sensor response for 10 μg/mL FimH. The curves correspond to the same functionalized SiNRs as shown in (a). The response to FimH is clearly increased by roughly a factor of 2 when N_s is low. Increased noise is visible coming from voltage fluctuations induced by air bubbles. (c) Theoretical FimH response at equilibrium as a function of FimH concentration based on the site binding model (eq 3) at two different hydroxyl group densities (N_s) for different protein−ligand interaction affinities (K_D). Based on pH and FimH measurements, the following parameters were chosen: $[B]_0 = 3 \times 10^{16}$ m^{-2}, $N_s = 1 \times 10^{17}$ m^{-2}, 4.6×10^{16} m^{-2}, $C_{dl} = 0.1$ Fm^{2-}, pH = 8 and $pK_a = 9$, $pK_b = 7$, and the visible net charge per protein $q_A = 2e$. (d) Signal-to-noise ratio for different nanoribbon widths. Surface potential in a real-time measurement for two different nanoribbons (1 and 25 μm) with the active mannose ligand. Both nanoribbons show the same signal upon injection of 20 μg/mL FimH. The inset shows the RMS noise for the baseline which is equivalent to the standard deviation of the measurement points ($\sigma = \sqrt{\text{variance}}$). $\sigma_{1\mu m} = 325$ μV, $\sigma_{25\mu m} = 65$ μV. The SNR is clearly increased for the larger sensor area (scales with $\sqrt{\text{area}}$).

Control experiments were performed with a commercial SPR-based biosensor (Biacore T200, GE Healthcare, Uppsala, Sweden). The response of a functionalized Au chip (active mannose ligand) is shown in Figure 2c. Although the same functionalization scheme was used, the signal in the Biacore shows different kinetics as compared to the BioFET. In particular, saturation starts at lower concentration and dissociation is less pronounced. A K_D of ≈5 nM is extracted by 1:1 Langmuir kinetic fits, indicated by the dashed lines.

The surface of the two different sensors is expected to be identical since the same surface functionalization was applied. While differences in ligand density $[B]_0$ cannot be excluded, the dissociation constant K_D is expected to be the same. However, there is a clear difference in association and dissociation rates (k_a, k_d) using the two different systems. External factors such as flow speed can influence these rates. Mechanical force studies have shown that FimH-mediated bacterial adhesion is influenced by shear forces and therefore depends on the flow rate.[29,30] Although, in our work, FimH is dissolved in buffer and is not membrane bound, the flow speed at the sensor surface could be a cause of the difference in signal. Here we would like to mention that the outcome of affinity assays performed in commercial SPR systems vary for different users and strongly depend on equipment maintenance and operation.[31,32] However, at the same total flow rate (26 μL/min), which was adjusted to be comparable to the SPR measurement (20 μL/min), we did observe very similar binding kinetics using different flow geometries (microchannel Figure 2 and liquid cell Figure 3). On the contrary, at slow speed, the transport of the analyte to the reaction site is becoming a limiting factor which strongly affects the binding kinetics. We have tested commonly used kinetic models, such as the two-compartment model for transport limited kinetics[4,33,34] to fit the BioFET data. However, they cannot explain the signals satisfactorily. As we generally expect similar kinetics and affinity of the protein−ligand interaction for both detection systems, different effects

DOI: 10.1021/acssensors.6b00089
ACS Sens. 2016, 1, 781−788

which could be the origin of the discrepancy in kinetics are discussed in the following.

(I) The effective protein surface concentration is considerably lower as initially injected. Using a flow rate of 26 μL/min, it takes ≈50 s for the liquid to pass the liquid system and reach the SiNR surface. Proteins accumulate at the side walls and thereby the bulk concentration gets depleted. The materials in contact with the solution, PTFE, PDMS, SU-8, and HfO_2, are known to adsorb proteins.[35−37] With increasing side wall coverage this interaction diminishes and hence, bulk concentration increases with time. This would explain why no saturation is observed after 900 s. However, this effect can not explain the increased dissociation rate in the BioFET. Even if the concentration is taken as a free fitting parameter, an apparent affinity constant of $K_D \approx 300$ nM is found, which is 2 orders of magnitude higher than reported values of this particular protein−ligand interaction.[38]

(II) Different sensing mechanisms are used for the two systems. While the BioFETs sense charges localized within a few nanometers from the surface (characterized by the Debye length), the SPR system measures the change in plasmon resonance frequency upon mass adsorption to the surface (change in refractive index). The depth of the evanescent wave is roughly 2 orders of magnitude larger as the Debye length,[39] which results in a different sensitivity on analyte distance to the surface. Surface rearrangement[40] and surface induced conformational changes of adsorbed proteins[41] within a few Ångstroms affect the BioFET signal, whereas the influence on the SPR signal is marginal. Figure 2d shows a scheme of a protein binding cycle and a qualitative picture of the difference in signal. As proteins bind to the surface the signal increases for both sensors until surface coverage has reached equilibrium. While the total amount of bound proteins stays constant, the SPR signal saturates. However, the BioFET is extremely sensitive to surface rearrangements, i.e., proteins approaching the SiNR at high surface coverage by a conformation change or interaction with the MHDA monolayer. We expect this process to be much slower than the protein−ligand association, which is why the signal does not saturate even if the numbers of proteins bound to the surface does not change. In addition the slope of the BioFET response saturates at very high protein concentrations. This indicates that the available binding sites are already occupied and the change in Ψ_0 has to have a different origin than the binding of additional proteins. The difference in dissociation can also be explained by this qualitative model, when proteins again undergo a rearrangement at the surface upon flushing with buffer.

We expect that both proposed effects influence the BioFET signals. However, an established model including microscopic surface rearrangement effects, which only become visible by using BioFETs, is still lacking.

Signal-to-Noise Ratio. For biosensors, the limit of detection (LOD) is an important figure of merit. It is directly related to the SNR and ultimately limited by the protein−ligand affinity. As the electrical noise is intrinsic to the device quality and geometry,[42,43] the signal strongly depends on the surface properties. In previous work,[28] we have shown that competing surface reactions of other species than the analyte can limit the sensitivity of the sensor. The competing adsorption reactions of the individual species are coupled via the surface potential. In the case of gold-coated BioFETs, the response to pH variations affects the signal of the FimH proteins. Only due to the very low pH response of the gold film we were able to detect clear signals from FimH adsorption.

In Figure 3a, the pH response for gold-coated BioFETs functionalized with the active mannose ligand (one-step disulfide bond) is demonstrated. The threshold voltage is extracted from $I_{sd} - V_{ref}$ sweeps. Due to harsh surface treatments (cleaning and functionalization) between different measurements, the gold film on the SiNR surface was altered. We observed a gradual increase in pH response. We anticipate that using UV-ozone, organic solvents, and a wide range of pH buffers oxidizes the gold surface, leading to an increase in the number of surface hydroxyl groups.[25,28] Since the FimH measurements were performed at pH 8, the pH range from pH 5 to 9 was of interest. The pH response (linear fit from pH 5 to 10) varies from ≈19 to 29 mV/pH. Using the extended site binding model from reference[28] where the density of proton sensitive hydroxyl groups and FimH ligands are included (FimH concentration is set to ≈0 M) the pH response of the functionalized gold surface can be fitted to extract the density of hydroxyl groups (N_s). We find that N_s changed by roughly a factor of 2.

In Figure 3b, the FimH response of the respective measurements are compared. For the increased N_s the FimH response was clearly reduced. The data supports the model of pH as competing surface reactions, which is exemplified in Figure 3c. It shows the theoretical response to a protein at a ligand density of $[B]_0 = 3 \times 10^{16}$ m^{-2} for two different N_s as a function of protein concentration. The curves denote the change in surface potential at equilibrium, calculated with the site binding model including competing surface reactions as described by eq 3 in the Methods. The detectable concentration range predominantly depends on K_D, the affinity of the protein−ligand interaction (indicated by three example values). However, with increasing N_s, the response to the protein decreases. Simultaneously the sensitive concentration range becomes narrower. In summary the FimH signal increases for a low pH response, where Figure 3b and c agrees qualitatively. This holds for any ISFET system, where decreasing the number of surface sites of a competing reaction enhances the response to the targeted analyte. We assume both gold surfaces used for the SPR and SiNR measurements are comparable. However, the parameter N_s primarily affects the surface potential and only secondarily affects the binding kinetics. Though, as for SPR systems where the surface potential is not measured, the parameter N_s becomes negligible.

Using PDMS microchannels and remote liquid gating by placing the reference electrode in the tubing increases current fluctuations, caused by unstable gating due to moving air bubbles. To analyze the signal-to-noise ratio we reduced external noise, by using a larger liquid cell with the reference electrode included in the immediate vicinity of the nanoribbons; see the Supporting Information for details on the fluidic setup. Figure 3 d shows the response of two active SiNRs of two different areas (6×1 μm^2 and 6×25 μm^2) upon injection of 20 μg/mL FimH. The signal ($\Delta\Psi_0$) is the same for both sensor dimensions. However, the noise decreases with larger sensor area. The inset in Figure 3d shows the noise in the baseline of the two SiNRs. The root-mean-square (RMS) noise, which is equivalent to the standard deviation of the measurement points ($\sigma = \sqrt{\text{variance}}$), is 325 μV for the 1 μm SiNR and 65 μV for the 25 μm SiNR. As we have analyzed in our previous work where we studied the low frequency $1/f$ noise in silicon nanowires,[43] the gate-referred voltage noise S_{VG}

DOI: 10.1021/acssensors.6b00089
ACS Sens. 2016, 1, 781−788

scales with $1/(WL)$, where W and L represent the silicon channel width and length, respectively. Further we showed that the sensor width has no influence on pH response.[44] For the SiNR dimensions presented here, the change in surface potential is independent of the sensor width since the total charge from adsorbed proteins is proportional to the area. Hence, the signal-to-noise ratio ($\frac{\Delta\Psi_0}{\sqrt{S_{VG}}}$) scales with $\sqrt{\text{area}}$, which is shown here as it increases from 145 for the 1 μm wide SiNR to 725 for the 25 μm wide SiNR.

CONCLUSION

By the detection of FimH, a therapeutically relevant protein with an important role in UTI, we have successfully demonstrated the use of gold-coated SiNRs as biosensors. Real-time detection without labeling was achieved at a very high signal-to-noise ratio of ≥700. The SNR is shown to increase with $\sqrt{\text{area}}$, which is an important aspect for the design of a biosensor with high device density. The use of gold as surface material has two tremendous advantages as compared to oxides. First, the pH response is strongly reduced which enables the detection of other species than protons. Second, surface functionalization of gold has been extensively investigated which simplifies the development of protocols for ligand immobilization on the sensor and allows the direct comparison with SPR measurements. Being able to observe association and dissociation represents a first step toward the use of BioFETs as affinity sensors. However, the accurate determination of the protein binding affinity and kinetics remains challenging when comparing the data with SPR measurements from a commercial Biacore system. This might be due to the enhanced sensitivity of BioFETs to surface rearrangements which is potentially advantageous for very local measurements of biochemical species. For successful detection of proteins, the screening limitations of the ionic environment, the binding affinity of the targeted analyte, the intrinsic electrical noise, as well as competing surface reactions have to be considered and finally the sensor needs to be engineered accordingly. Our results propose that SiNR BioFETs have a great potential to be used in disease diagnosis and drug discovery. Because of the large scale integration of SiNR arrays at low cost biosensing based on silicon nanoribbons offers a promising alternative to the currently used methodologies.

METHODS

Device Fabrication. The SiNR samples were produced by a top-down approach, according to previously described protocols.[24,25,43] Silicon on insulator (SOI) wafers (Soitec, France) with a buried oxide (BOX) layer of 145 nm in thickness were used. The 85 nm thick p-Si(100) device layer with a resistivity of 8.5−11.5 Ωcm was oxidized thermally until a 15 nm thick SiO_2 layer was grown. The structures were defined by electron beam lithography (EBL) and carved out by reactive ion etching of SiO_2 and anisotropic wet etching of the Si device layer with tetramethylammonium hydroxide (TMAH) and isopropyl alcohol 9:1 at 45°. The resulting nanoribbons with Si (111) side faces were of 6 μm in length, 80 nm in height and 1 or 25 μm wide. Heavy doping of the source and drain contact areas was done by BF_2^+ ions (energy = 33 keV, dose 2.3×10^{15} cm^{-2}). The dopants were activated by thermal annealing in a forming gas (6 min at 950 °C). Eight nm HfO_2 as dielectric was deposited by atomic layer deposition (ALD) at 200 °C (Savannah S100, Cambridge NanoTech). The ohmic contacts were opened by local etching of the ALD oxide, metallized by Al−Si(1%) and annealed at 450 °C. The structures for the gold film were defined by EBL. Five nm chromium and 20 nm gold was deposited by electron beam evaporation. A 2 μm thick protection layer (SU-8 2002, MicroChem) with 6 μm wide openings, defined by UV-lithography was used as liquid protection for the contact areas. The samples were wire bonded into a chip carrier and the bonds were sealed with epoxy (Epotek 353ND).

Microchannels were produced by pouring polydimethylsiloxane (PDMS, SYLGARD 184 Silicone Elastomer) onto SU-8 (SU-8 100 MicroChem) patterned Si wafers, degassing, and heating at 60 °C for 2 h. Polytetrafluoroethylene (PTFE) tubes were used to connect the microchannels to a peristaltic pump and the electrolyte solutions.

Surface Functionalization. The sensor surface was rinsed with DI, cleaned in UV-ozone for 20 min, and enclosed by the PDMS microchannel, separating the chip in active and control channels. The channels were then rinsed with ethanol for ≈30 min.

One-Step Functionalization. The mannose ligand (synthesis is given in the Supporting Information) was dissolved in ethanol (2 mM). The control channel was treated with lipoic acid dissolved in ethanol (2 mM). The microchannels were flushed with 200 μL of the respective solution, and then 200 μL was slowly injected over ≈15 h using a syringe pump. After the functionalization, the channels were washed with ethanol before the PDMS microchannel was removed for the measurement.

Two-Step Functionalization. SAM formation of 16-mercaptohexadecanoic acid (MHDA) (2 mM in ethanol) for 16 h at 4 °C and afterward rinsed with ethanol. After surface activation with EDC and *N*-hydroxysuccinimide (NHS) for ≈30 min, the ligands were injected to the microchannels (details are given in the Supporting Information).

Protein and Buffer Solution. FimH carbohydrate recognition domain (FimH-CRD) with a thrombin cleavage site (Th) linked to a 6His tag (FimH-CRD-Th-6His, 18.6 kDa) was expressed in *E. coli* strain HM125 and purified by affinity chromatography as described previously.[38,45] The purified protein was dialyzed against 10 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) buffer pH 8. Protein concentrations ranging from 1 to 100 μg/mL (54 nM to 2.7 μM) were used. An intermediate ionic strength was chosen to have a well buffered solution and a Debye length of ≥3 nm. The theoretical isoelectric point of the FimH protein is at pH 6.7, so the protein is negatively charged in pH 8 buffer solution. For the pH measurements in Figure 3, standard pH buffer solutions (Titrisol, Merck) were used.

Surface Regeneration. Surface regeneration was accomplished by denaturing the structure of the analyte. Usually strong bases or acids as well as detergents are used to denature proteins. However, since pH also affects the surface potential of the gold-coated nanoribbons, we chose concentrated urea (6 M) as regeneration solution, since pH was similar to the running buffer.

Competing Surface Reactions. Besides the ligands immobilized on the surface, additional hydroxyl groups (MOH) have to be assumed due to the residual pH response of the gold surface.[25] Analyte ([*A*], FimH protein) adsorption as well as deprotonation and protonation of MOH change the surface charge and hence the surface potential. The system can be described by three equilibrations:[28]

$$\begin{aligned} &\mathrm{MOH} \rightleftharpoons \mathrm{MO^-} + \mathrm{H^+},\ K_a \\ &\mathrm{MOH_2^+} \rightleftharpoons \mathrm{MOH} + \mathrm{H^+},\ K_b \\ &[AB] \rightleftharpoons [A] + [B],\ K_D \end{aligned} \tag{2}$$

K_a, K_b, and K_D are the equilibrium dissociation constants. $[A]$ is the analyte concentration, $[B]$ is the number of free ligands per unit area. The surface potential is related to the surface charge by $\Psi_0 = \sigma_0/C_{dl}$ where σ_0 is the total number of surface charge and C_{dl} is the double layer capacitance. Including the Boltzmann distribution for the proton activity, $a_{H^+}^{surface} = a_{H^+}\exp(-e\Psi_0/kT)$, with e as elementary charge, k the Boltzmann constant, and T as absolute temperature, we get

$$\Psi_0 = \frac{q_A}{C_{dl}}[B]_0\frac{[A]}{[A]+K_D} + \frac{e}{C_{dl}}N_s \times \frac{a_{H^+}{}^2 - K_aK_b\,e^{-e\Psi_0/kT}}{a_{H^+}{}^2 + a_{H^+}K_b\,e^{e\Psi_0/kT} + K_aK_b\,e^{2e\Psi_0/kT}} \tag{3}$$

DOI: 10.1021/acssensors.6b00089
ACS Sens. 2016, 1, 781−788

where the first term is given by the protein adsorption with q_A being the charge per protein and $[B]_0$ being the total number of surface bound ligands per unit area. For simplicity, a uniform distribution of surface and bulk proteins can be assumed, since the protein size is larger as the Debye length. The second term describes the intrinsic proton sensitivity.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssensors.6b00089.

Details on the measurement setup (Figures S1, S2); gold surface functionalization (Figure S3); supporting control measurements showing the overview of a protein detection measurement (Figure S4); influence of switching solutions (Figure S5), SiNR array sensing (Figure S6) and surface regeneration (Figure S7); nonspecific response to bovine serum albumin (Figure S8); response of control surfaces to FimH (Figure S9); and completive pH measurements (Figure S10); Figure S11: Synthesis of ligands (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mathias.wipf@gmail.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors gratefully acknowledge the support by the Swiss Nanoscience Institute (SNI), the Swiss Nano-Tera program, the European Commission under the FP7-NMP project Hysens (263091), and the Swiss National Science Foundation as part of the NCCR Molecular Systems Engineering.

■ REFERENCES

(1) Wilson, G. S.; Gifford, R. Biosensors for real-time in vivo measurements. *Biosens. Bioelectron.* **2005**, *20*, 2388−2403.

(2) Cheng, M. M.-C.; Cuda, G.; Bunimovich, Y. L.; Gaspari, M.; Heath, J. R.; Hill, H. D.; Mirkin, C. A.; Nijdam, A. J.; Terracciano, R.; Thundat, T.; et al. Nanotechnologies for biomolecular detection and medical diagnostics. *Curr. Opin. Chem. Biol.* **2006**, *10*, 11−19.

(3) Rothberg, J. M.; Hinz, W.; Rearick, T. M.; Schultz, J.; Mileski, W.; Davey, M.; Leamon, J. H.; Johnson, K.; Milgrew, M. J.; Edwards, M.; et al. An Integrated Semiconductor Device Enabling Non-Optical Genome Sequencing. *Nature* **2011**, *475*, 348−352.

(4) Duan, X.; Li, Y.; Rajan, N. K.; Routenberg, D. A.; Modis, Y.; Reed, M. A. Quantification of the Affinities and Kinetics of Protein Interactions Using Silicon Nanowire Biosensors. *Nat. Nanotechnol.* **2012**, *7*, 401−407.

(5) Cooper, M. A. Optical biosensors in drug discovery. *Nat. Rev. Drug Discovery* **2002**, *1*, 515−528.

(6) Rajan, N. K.; Duan, X.; Reed, M. A. Performance limitations for nanowire/nanoribbon biosensors. *Wiley Interdiscip Rev. Nanomed Nanobiotechnol* **2013**, *5*, 629−645.

(7) Elfström, N.; Karlström, A. E.; Linnros, J. Silicon Nanoribbons for Electrical Detection of Biomolecules. *Nano Lett.* **2008**, *8*, 945−949.

(8) Curreli, M.; Zhang, R.; Ishikawa, F.; Chang, H.-K.; Cote, R.; Zhou, C.; Thompson, M. E. Real-Time, Label-Free Detection of Biological Entities Using Nanowire-Based FETs. *IEEE Trans. Nanotechnol.* **2008**, *7*, 651−667.

(9) Stern, E.; Vacic, A.; Reed, M. A. Semiconducting Nanowire Field-Effect Transistor Biomolecular Sensors. *IEEE Trans. Electron Devices* **2008**, *55*, 3119−3130.

(10) Stern, E.; Klemic, J. F.; Routenberg, D. A.; Wyrembak, P. N.; Turner-Evans, D. B.; Hamilton, A. D.; LaVan, D. A.; Fahmy, T. M.; Reed, M. A. Label-Free Immunodetection with CMOS-Compatible Semiconducting Nanowires. *Nature* **2007**, *445*, 519−522.

(11) Cui, Y.; Wei, Q.; Park, H.; Lieber, C. M. Nanowire Nanosensors for Highly Sensitive and Selective Detection of Biological and Chemical Species. *Science* **2001**, *293*, 1289−1292.

(12) Bunimovich, Y. L.; Shin, Y. S.; Yeo, W.-S.; Amori, M.; Kwong, G.; Heath, J. R. Quantitative Real-Time Measurements of DNA Hybridization with Alkylated Nonoxidized Silicon Nanowires in Electrolyte Solution. *J. Am. Chem. Soc.* **2006**, *128*, 16323−16331.

(13) DeChancie, J.; Houk, K. N. The Origins of Femtomolar Protein-Ligand Binding: Hydrogen-Bond Cooperativity and Desolvation Energetics in the Biotin-(Strept)Avidin Binding Site. *J. Am. Chem. Soc.* **2007**, *129*, 5419−5429.

(14) Sharon, N. Lectins: Carbohydrate-specific Reagents and Biological Recognition Molecules. *J. Biol. Chem.* **2007**, *282*, 2753−2764.

(15) Ghazarian, H.; Idoni, B.; Oppenheimer, S. B. A glycobiology review: carbohydrates, lectins, and implications in cancer therapeutics. *Acta Histochem.* **2011**, *113*, 236−247.

(16) Zhang, L.; Foxman, B. Molecular epidemiology of Escherichia coli mediated urinary tract infections. *Front. Biosci., Landmark Ed.* **2003**, *8*, E235−E244.

(17) Cegelski, L.; Marshall, G. R.; Eldridge, G. R.; Hultgren, S. J. The biology and future prospects of antivirulence therapies. *Nat. Rev. Microbiol.* **2008**, *6*, 17−27.

(18) Klein, T.; Abgottspon, D.; Wittwer, M.; Rabbani, S.; Herold, J.; Jiang, X.; Kleeb, S.; Luthi, C.; Scharenberg, M.; Bezençon, J.; et al. FimH Antagonists for the Oral Treatment of Urinary Tract Infections: From Design and Synthesis to in Vitro and in Vivo Evaluation. *J. Med. Chem.* **2010**, *53*, 8627−8641.

(19) Jiang, X.; Abgottspon, D.; Kleeb, S.; Rabbani, S.; Scharenberg, M.; Wittwer, M.; Haug, M.; Schwardt, O.; Ernst, B. Antiadhesion Therapy for Urinary Tract InfectionsA Balanced PK/PD Profile Proved To Be Key for Success. *J. Med. Chem.* **2012**, *55*, 4700−4713.

(20) Pang, L. J.; Kleeb, S.; Lemme, K.; Rabbani, S.; Scharenberg, M.; Zalewski, A.; Schadler, F.; Schwardt, O.; Ernst, B. FimH Antagonists: Structure-Activity and Structure-Property Relationships for Biphenyl a-D-Mannopyranosides. *ChemMedChem* **2012**, *7*, 1404−1422.

(21) Kleeb, S.; Pang, L.; Mayer, K.; Eris, D.; Sigl, A.; Preston, R. C.; Zihlmann, P.; Sharpe, T.; Jakob, R. P.; Abgottspon, D.; et al. FimH Antagonists: Bioisosteres To Improve the in Vitro and in Vivo PK/PD Profile. *J. Med. Chem.* **2015**, *58*, 2221−39.

(22) Livi, P.; Shadmani, A.; Wipf, M.; Stoop, R. L.; Rothe, J.; Chen, Y.; Calame, M.; Schönenberger, C.; Hierlemann, A. Sensor System including Silicon Nanowire Ion Sensitive FET Arrays and CMOS readout. *Sens. Actuators, B* **2014**, *204*, 568−577.

(23) Knopfmacher, O.; Tarasov, A.; Fu, W.; Wipf, M.; Niesen, B.; Calame, M.; Schönenberger, C. Nernst Limit in Dual-Gated Si-Nanowire FET Sensors. *Nano Lett.* **2010**, *10*, 2268−2274.

(24) Tarasov, A.; Wipf, M.; Bedner, K.; Kurz, J.; Fu, W.; Guzenko, V. A.; Knopfmacher, O.; Stoop, R. L.; Calame, M.; Schönenberger, C. True Reference Nanosensor Realized with Silicon Nanowires. *Langmuir* **2012**, *28*, 9899−9905.

(25) Wipf, M.; Stoop, R. L.; Tarasov, A.; Bedner, K.; Fu, W.; Wright, I. A.; Martin, C. J.; Constable, E. C.; Calame, M.; Schönenberger, C. Selective Sodium Sensing with Gold-Coated Silicon Nanowire Field-Effect Transistors in a Differential Setup. *ACS Nano* **2013**, *7*, 5978−5983.

(26) van Hal, R.; Eijkel, J.; Bergveld, P. A General Model to Describe the Electrostatic Potential at Electrolyte Oxide Interfaces. *Adv. Colloid Interface Sci.* **1996**, *69*, 31−62.

(27) Schasfoort, R.; Bergveld, P.; Kooyman, R.; Greve, J. Possibilities and Limitations of Direct Detection of Protein Charges by Means of an Immunological Field-Effect Transistor. *Anal. Chim. Acta* **1990**, *238*, 323−329.

(28) Stoop, R. L.; Wipf, M.; Müller, S.; Bedner, K.; Wright, I. A.; Martin, C. J.; Constable, E. C.; Fu, W.; Tarasov, A.; Calame, M.; et al.

DOI: 10.1021/acssensors.6b00089
ACS Sens. 2016, 1, 781−788

Competing surface reactions limiting the performance of ion-sensitive field-effect transistors. *Sens. Actuators, B* **2015**, *220*, 500−507.
(29) Anderson, B. N.; Ding, A. M.; Nilsson, L. M.; Kusuma, K.; Tchesnokova, V.; Vogel, V.; Sokurenko, E. V.; Thomas, W. E. Weak Rolling Adhesion Enhances Bacterial Surface Colonization. *J. Bacteriol.* **2007**, *189*, 1794−1802.
(30) Yakovenko, O.; Sharma, S.; Forero, M.; Tchesnokova, V.; Aprikian, P.; Kidd, B.; Mach, A.; Vogel, V.; Sokurenko, E.; Thomas, W. E. FimH Forms Catch Bonds That Are Enhanced by Mechanical Force Due to Allosteric Regulation. *J. Biol. Chem.* **2008**, *283*, 11596−11605.
(31) Cannon, M. J.; Papalia, G. A.; Navratilova, I.; Fisher, R. J.; Roberts, L. R.; Worthy, K. M.; Stephen, A. G.; Marchesini, G. R.; Collins, E. J.; Casper, D.; et al. Comparative analyses of a small molecule/enzyme interaction by multiple users of Biacore technology. *Anal. Biochem.* **2004**, *330*, 98−113.
(32) Katsamba, P. S.; Navratilova, I.; Calderon-Cacia, M.; Fan, L.; Thornton, K.; Zhu, M.; Bos, T. V.; Forte, C.; Friend, D.; Laird-Offringa, I.; et al. Kinetic analysis of a high-affinity antibody/antigen interaction performed by multiple Biacore users. *Anal. Biochem.* **2006**, *352*, 208−221.
(33) Myszka, D. G.; He, X.; Dembo, M.; Morton, T. A.; Goldstein, B. Extending the range of rate constants available from BIACORE: interpreting mass transport-influenced binding data. *Biophys. J.* **1998**, *75*, 583−594.
(34) Gaster, R. S.; Xu, L.; Han, S.-J.; Wilson, R. J.; Hall, D. A.; Osterfeld, S. J.; Yu, H.; Wang, S. X. Quantification of Protein Interactions and Solution Transport Using High-Density GMR Sensor Arrays. *Nat. Nanotechnol.* **2011**, *6*, 314−320.
(35) Zardeneta, G.; Mukai, H.; Marker, V.; Milam, S. B. Protein interactions with particulate Teflon: Implications for the foreign body response. *J. Oral Maxillofac. Surg.* **1996**, *54*, 873−878.
(36) Sikanen, T.; Wiedmer, S. K.; Heikkila, L.; Franssila, S.; Kostiainen, R.; Kotiaho, T. Dynamic coating of SU-8 microfluidic chips with phospholipid disks. *Electrophoresis* **2010**, *31*, 2566−2574.
(37) Mora, M. F.; Giacomelli, C. E.; Garcia, C. D. Electrophoretic Effects of the Adsorption of Anionic Surfactants to Poly-(dimethylsiloxane)-Coated Capillaries. *Anal. Chem.* **2007**, *79*, 6675−6681.
(38) Scharenberg, M.; Jiang, X.; Pang, L.; Navarra, G.; Rabbani, S.; Binder, F.; Schwardt, O.; Ernst, B. Kinetic Properties of CarbohydrateLectin Interactions: FimH Antagonists. *ChemMedChem* **2014**, *9*, 78−83.
(39) Habauzit, D.; Chopineau, J.; Roig, B. SPR-based biosensors: a tool for biodetection of hormonal compounds. *Anal. Bioanal. Chem.* **2007**, *387*, 1215−1223.
(40) Rabe, M.; Verdes, D.; Seeger, S. Understanding protein adsorption phenomena at solid surfaces. *Adv. Colloid Interface Sci.* **2011**, *162*, 87−106.
(41) Roach, P.; Farrar, D.; Perry, C. C. Interpretation of Protein Adsorption: Surface-Induced Conformational Changes. *J. Am. Chem. Soc.* **2005**, *127*, 8168−8173.
(42) Tarasov, A.; Fu, W.; Knopfmacher, O.; Brunner, J.; Calame, M.; Schönenberger, C. Signal-to-noise ratio in dual-gated silicon nanoribbon field-effect sensors. *Appl. Phys. Lett.* **2011**, *98*, 012114.
(43) Bedner, K.; Guzenko, V. A.; Tarasov, A.; Wipf, M.; Stoop, R. L.; Rigante, S.; Brunner, J.; Fu, W.; David, C.; Calame, M.; et al. Investigation of the dominant 1/f noise source in silicon nanowire sensors. *Sens. Actuators, B* **2014**, *191*, 270−275.
(44) Bedner, K.; Guzenko, V. A.; Tarasov, A.; Wipf, M.; Stoop, R. L.; Just, D.; Rigante, S.; Fu, W.; Minamisawa, R. A.; David, C.; et al. pH Response of Silicon Nanowire Sensors: Impact of Nanowire Width and Gate Oxide. *Sens. Mater.* **2013**, *25*, 567−576.
(45) Rabbani, S.; Jiang, X.; Schwardt, O.; Ernst, B. Expression of the carbohydrate recognition domain of FimH and development of a competitive binding assay. *Anal. Biochem.* **2010**, *407*, 188−195.

DOI: 10.1021/acssensors.6b00089
ACS Sens. 2016, 1, 781−788

3. The PapG Adhesins and Their Role in UTIs

3.1 Introduction

3.1.1 Pyelonephritis

Pyelonephritis is a UTI that involves the renal parenchyma and the renal pelvis [1]. In the USA approximately 250,000 cases, predominantly women, are registered every year, 100,000 of which require hospitalization [2,3]. The estimated cost due to pyelonephritis in the year 2000 was 2.14 billion USD [4]. In immunocompetent patients, in absence of urinary tract abnormalities, the infection is considered uncomplicated. In presence of host factors such as immunosuppression, diabetes mellitus, pregnancy, sickle cell anemia, obstruction, renal stones or any structural or functional abnormalities of the urinary tract, pyelonephritis is considered as complicated [4]. 1-2% of all pregnancies are complicated by pyelonephritis [5].

3.1.2 Pathogenesis

In most of the infections reaching the kidneys bacteria ascend from the lower urinary tract. Descending infections with hematic origin are rare and limited to chronically ill or immunosuppressed patients [1]. The main etiological agent is *E. coli*, which accounts for 70-80% of all uncomplicated infections [6,7]. Untreated pyelonephritis can lead to serious organ damage and even death [8].

3.1.3 UPEC with kidney tropism

Uropathogenic *E. coli* are highly specialized bacteria that evolved an arsenal of virulence factors (VFs), supporting colonization of the urinary tract. Among them, Type 1 and type P pili have been ascribed pivotal roles in the establishment of UTIs [9,10]. Type 1 pili bind in fact to glycoproteins bearing terminal mannose residues

[11], which are widely expressed in the bladder [12,13] and less abundant to completely absent in the upper urinary tract [14,15]. Type P pili are instead expressed in most UPEC isolates from pyelonephritis patients [16] and bind to glycolipids exposing Galα(1-4)Galβ (galabiose) moieties [17], which are abundant in the human kidney [18]. Although type-1 pili have been shown to clearly play a role in bladder colonization as well as in the establishment of recurrent infections, their function in the invasion of the upper urinary tract is not fully understood. The adhesive properties of type-P pili have been shown to facilitate bacterial invasion of the kidneys in some studies [19-21], but others could not find any differences among the infectivity of UPEC possessing functional or non-functional pili [22]. However, care should be used in extrapolating the results from mouse models to humans, due to large anatomic differences. Studies in primates and in humans [21] suggest the involvement of the type P pili in the infection, especially in the early stage [19].

3.1.4 Type P pilus

The type P-pilus is similar to the type-1 pilus already described in chapter 2. It consists of a right-handed helical rod [23,24] of nearly 10 nm diameter [25] composed of ≈ 1000 copies of the major subunit PapA and a flexible tip-fibrillum [25,26]. The latter is formed by one copy of PapK, followed by 5–10 copies of PapE, one copy of PapF and one copy of the adhesin PapG (figure 1) [27], and has a diameter of about 3 nm [26]. The total length of the pilus is about 1 μM [25]. The biogenesis follows the chaperone/usher pathway. The subunits are secreted from the cytosol through the secYEG translocon [27]. In the periplasmic space, the chaperone PapD [28,29] bind them by donor stand complementation (DSC) [30,31], and afterwards the complex is captured by the usher PapC [32]. The next subunit binds to the usher and undergoes donor stand exchange (DSE) [33]. The reiteration of this process allows the assembly of the whole pilus.

3.1.5 PapG Adhesins

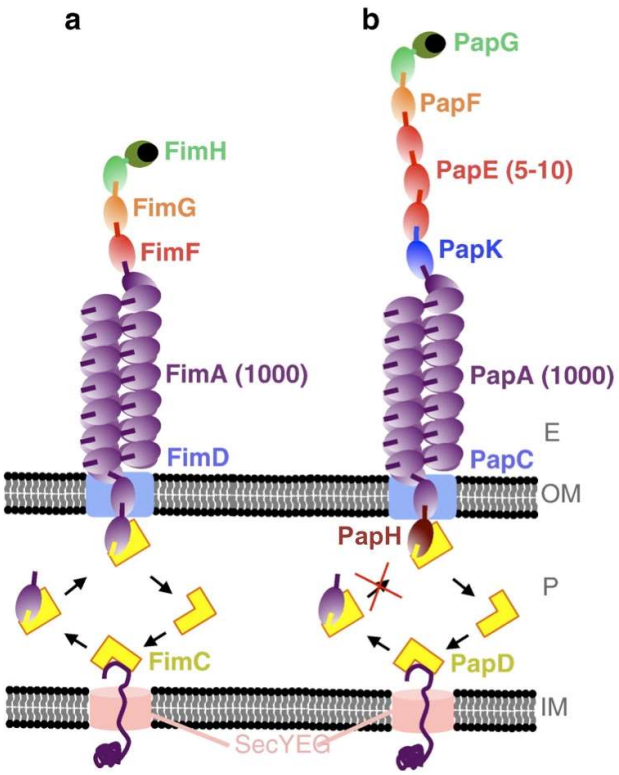


Figure 1. Comparison of type-1 (a) and type-P pili (b) (adapted from ref. [33]). Extracellular space (E), periplasm (P), outer membrane (OM), inner membrane (IM). The red cross over the last arrow of the assembly cycle of the type-P pilus symbolizes PapH function, i.e. cycle stop and pilus anchoring.

The PapG adhesins are located at the tip of the P pili [27] and are the only subunits of the whole pilus responsible for binding [17]. Four isoforms have been identified, which can be grouped into three binding subtypes (I-III). All adhesins recognize glycolipids containing the galabiose core. However, the residues flanking the galabiose have different influence on each of the three classes.

3.1.6 Natural Ligands of PapG Adhesins

The minimal binding epitope for all PapG adhesins is galabiose [17]. Nevertheless, the single adhesins' subtypes bind preferentially to specific ligand; class I adhesins prefer GbO3, while class II bind more strongly to GbO4, and class III to GbO5 [16], although these sharp differences have been questioned, due to the high variability of the phenotypic assays with which they were determined [34]. Moreover, all PapG adhesins bind well to the isolated GbO5 pentasaccharide, but only class III adhesins do so also when it is bound to the cellular membrane [18]. In 1998, Stapleton *et al.* demonstrated that all P adhesins bind avidly to the isolated hexasaccharide of sialosyl galactosyl globoside (SGG), which is expressed on kidney and vaginal epithelial cells

of nonsecretors. They correlated this behavior with the increased risk of contracting UTIs in these patients' population [35].

The different affinity for Galα(1-4)Gal isoreceptors exhibited by class I-III adhesins (table 1) has been suggested to play a role in host specificity [18].

Table 1. Binding of ^{35}S-labelled recombinant *E.coli* strains expressing G adhesins I-III, to globosides immobilized on a thin plate [18,36]. +++*, "very strong" binding; + +, "strong" binding; +, "moderate" binding; (+), "weak" binding; -, no binding at all. *SGG was measured in a separate experiment and compared to GbO3 and GbO4, where it showed stronger binding than all other glycolipids to all adhesins' classes [35].

Name	**Structure**	**Binding**		
		I	**II**	**III**
GbO3	Galα1-4Galβ1-4Glcβ1-1Cer	++	++	-
GbO4	Gal*N*Acβ1-3Galα1-4Galβ1-4Glcβ1-1Cer	++	++	+
GbO5	Gal*N*Acα1-3Gal*N*Acβ1-3Galα1-4Galβ1-4Glcβ1-1Cer	++	++	++
SGG*	NeuAcα2-3Galβ1-3Gal*N*Acβ1-3Galα1-4Galβ1-4Glcβ1-1Cer	+++	+++	+++

In fact, dog UTI isolates express mainly class-III adhesins and bind cells expressing GbO5 (abundant in dog urinary tract) [18], whereas human isolates from patients with a normal urinary tract produce mainly class-II adhesins [18,36], recognizing preferentially GbO4, one of the major globosides expressed in humans [18].

3.1.7 PapG class-II Adhesin

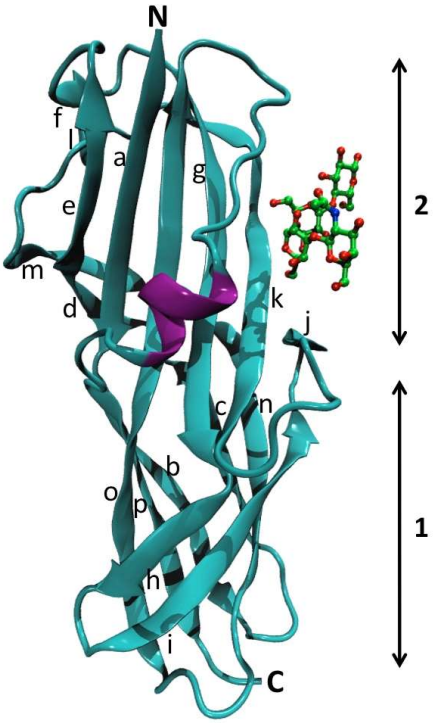


Figure 2. PapG-II lectin domain (PDB code 1J8R) [39]. The arrows delimit-tate regions 1 and 2. β-sheets are marked in alphabetical order, starting from the N-terminus. The protein is represented in cartoon style (cyan β-sheets, purple α-helix). The ligand (GbO4 tetra-saccharide) is represented in ball and sticks with green carbon, red oxygen, and blue nitrogen atoms).

PapG-II is of particular medical interest, because of its strong association with pyelonephritis in humans [17,18,37,38]. PapG-II presents two domains, a receptor-binding N-terminal domain (lectin domain) and a C-terminal domain that links it to the rest of the pilus (pilin domain) [30]. The lectin domain is formed mainly by β-sheet structures, organized in two regions; region 1 consists of a large β-barrel, formed by seven strands. Region 2 is formed by a large β-sheet, flanked on one side by two double-stranded small β-sheets and on the other by an α-helix, and it comprises the binding pocket (figure 2) [39]. The pilin domain shows an incomplete Ig-like fold, with a missing strand that is complemented by the chaperone PapD (during pilus biogenesis) or by the subunit PapF (in the assembled pilus) [40].

The binding specificity of PapG-II has been studied and its natural target in the human urinary tract identified. It binds preferentially to GbO4 [41], which contains the minimal galabiose core, flanked by a Gal*N*Ac residue at the non-reducing end and a glucose unit at the reducing terminus, and is abundant in the human upper urinary tract [35]. PapG adhesins have been suggested to play a role also in bacteremia [42].

3.1.7.1 Binding Pocket of PapG-II

The binding site of PapG-II is located on one side with respect to the main axis of the protein (figure 2). In the first published crystal structure, the tetrasaccharide of GbO4 is in a pocket formed by strands j, g, and k, helix A, and the loop connecting strand o to helix A (figure 3) [41]. The ligand is bound in a V shape, with one arm comprising Gal*N*Ac and the adjacent Gal (rings A and B respectively), and the other arm Gal and Glc (rings C and D). Ring D forms a water-mediated H-bond between the anomeric oxygen and $N_{\varepsilon 1}$ of Trp107 and two charge-assisted H-bonds between 3-OH and Arg170. The α-face [43] of ring D is partially involved in apolar contacts with Trp107. The β-face [44] of ring C stacks on Trp107 (strand k). The 3-OH of ring C is involved in an H-bond with Gly104, whilst 6-OH is H-bonded to Glu59 and water-bridged to the backbone carbonyl of Lys106. Ring B forms a dense network of H-bonds: 2-OH forms a water-mediated H-bond to Glu59, the oxygen in position 3 accepts a H-bond from Lys172, 4-OH binds to Glu91, and 6-OH binds to the NH of Gly104 and is water-bridged to Lys103.

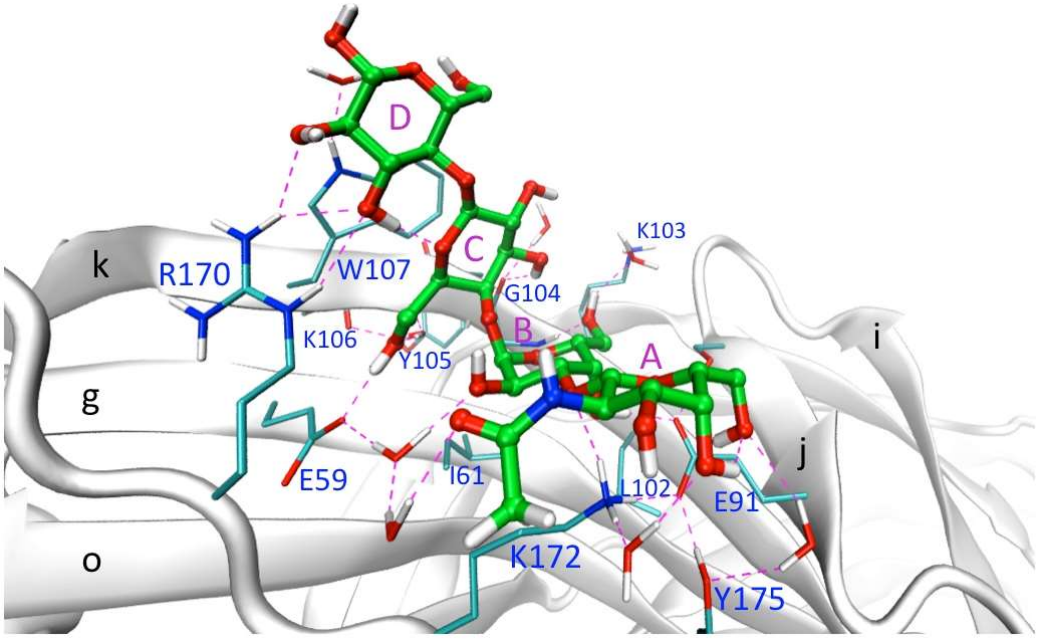


Figure 3. Binding pocket of PapG-II with the natural tetrasaccharide epitope of GbO4. The protein is represented in cartoon style in white color. The β-sheets are labeled with low case according to ref. [41]. The ligand is in ball and sticks, with carbon atoms in green, hydrogen in white, oxygen in red, and nitrogen in blue. Water molecules are shown as wires. Amino acids are labeled in one-letter code and numbered according to the PDB file 1J8R. Each monosaccharide unit is labeled with purple capital letter. Hydrogen bonds are highlighted by magenta dashed lines.

The apolar part of C1 and C2 of residue B lies on top of a hydrophobic patch formed by the hydrophobic portions of Lys172, Ile61 (g strand), and Leu102 (k strand).
The Gal*N*Ac (ring A) is involved in more loose interactions; the polar groups are involved in water-mediated contacts: the acetyl group with Glu59, 4-OH with Lys172, and 6-OH with the hydroxyl group of Tyr175 and with the backbone NH of Arg92.
The location of the binding site on one side of PapG-II has some implications on target recognition. As the GbO4 tetrasaccharide is thought to assume a close-to-90° orientation relative to the lipid chains [45], the adhesin has to orient parallel to the epithelium for reaching its target. The tip fibrillum could provide enough flexibility, as suggested by electron microscope (EM) analysis [41]. The positively charged patch on the surface of PapG-II, in close proximity to the binding site, might help the docking, thanks to the interaction with the polar head groups of the lipid bilayer of the cellular membranes [41].

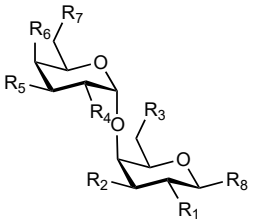
3.1.7.2 Catch-bond or slip-bond?

Nilsson *et al.* investigated the binding behavior of PapG-II by means of stick-and-roll experiments. The results suggested catch-bond behavior at shear-stress ≤4.3 pN/μm^2 [44]. However, further investigations with force-measuring optical tweezers, in which the single pilus can be studied, support a slip-bond character for the interaction between PapG-II and galabiose [46]. This is in agreement with the rigid body-type interaction already described by Dodson et al. [41] Interestingly, despite the rather low affinity of this interaction [41], the mechanical strength of the bond is enough to allow the unfolding of the pilus subunits under tensile force [47], probably an evolutionary optimization of the adhesion organelle, which allows better adaptation to the colonized niche [41,48].

3.1.7.3 PapG Antagonists

The correlation of PapG adhesins and UTI inspired researchers to develop molecules able to inhibit its binding to the host's ligands [41,48-56]. Starting from the natural targets, an impressive number of mimetics was tested. As the core motif is the Galα(1-4)Gal disaccharide, its modification was initially investigated.

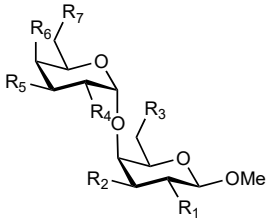
Table 2. Data from ref. [50]. Inhibition of hemagglutination by *E. coli* bearing class I adhesins. [a]Compound **25** was assayed with a different batch of bacteria. [b]Values were calculated using compound **1** as reference.



Cpd	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	IC_{50} (mM)	rIC_{50}[b]
1	OH	OH	OH	OH	OH	OH	OH	OMe	0.18	1.0
2	**H**	OH	OH	OH	OH	OH	OH	OMe	0.3	1.7
3	OH	**H**	OH	OH	OH	OH	OH	OMe	0.98	5.4
4	OH	**Me**	OH	OH	OH	OH	OH	OMe	0.37	2.1
5	OH	**Et**	OH	OH	OH	OH	OH	OMe	3.4	18.9
6	OH	**OMe**	OH	OH	OH	OH	OH	OMe	1.6	8.9
7	OH	OH	**H**	OH	OH	OH	OH	OMe	4.2	23.3
8	OH	OH	**F**	OH	OH	OH	OH	OMe	9.2	51.1
9	OH	OH	**OMe**	OH	OH	OH	OH	OMe	>25	–
10	OH	OH	OH	**H**	OH	OH	OH	OMe	2.3	12.8
11	OH	OH	OH	OH	**H**	OH	OH	OMe	6.4	35.6
12	OH	OH	OH	OH	**OMe**	OH	OH	OMe	0.082	0.5
13	OH	OH	OH	OH	OH	**H**	OH	OMe	10	55.6
14	OH	OH	OH	OH	OH	**F**	OH	OMe	3.3	18.3
15	OH	OH	OH	OH	OH	***epi***	OH	OMe	5.0	27.8
16	OH	OH	OH	OH	OH	OH	**H**	OMe	3.6	20.0
17	OH	OH	OH	OH	OH	OH	**F**	OMe	0.33	1.8
18	OH	OH	OH	OH	OH	OH	OH	**OEt**	0.13	0.7
19	OH	OH	OH	OH	OH	OH	OH	**O*i*Bu**	0.082	0.5
20	OH	OH	OH	OH	OH	OH	OH	**$OCH_2CH_2SiMe_3$**	0.046	0.3
21	OH	OH	OH	OH	OH	OH	OH	**$OCH_2CH(OH)SiMe_3$**	0.12	0.7
22	OH	OH	OH	OH	OH	OH	OH	**O-4GlcβOCH$_2$CH$_2$SiMe$_3$**	0.26	1.4
23	OH	OH	OH	OH	OH	OH	OH	**αOMe**	2.1	11.7
24	OH	OH	OH	OH	OH	OH	OH	**αβOH**	1.3	7.2
25	OH	OH	OH	OH	OH	OH	OH	**SEt**	–[a]	0.5[a]

Kihlberg *et al.* systematically replaced each hydroxy group of the disaccharide by H, F, or OMe and tested the modified disaccharides with PapG-I (table 2). Additional modifications, e.g. at the reducing end were tested as well (table 2, compounds 5, 15, and 20-25). The results furnished the first mapping of the binding site. Whereas most modifications reduced or even abolished the binding, the introduction of lipophilic aglycones and the methylation of the 3'-OH were beneficial (table 2) [50]. However, affinities for PapG-II were tested only at a leater point in time [51]. When Striker *et al.* examined the binding epitope of PapG-II [51], they reported that the substitution of 3-OH by either hydrogen or a methyl group slightly improved the affinity as compared to methyl galabiose **1** (table 3).

Table 3. Data from ref. [51]. Inhibition of hemagglutination by *E. coli* bearing class II adhesins. [a]Statistically significantly better inhibitor. [b]Statistically significantly worse inhibitor.



Cpd	R_1	R_2	R_3	R_4	R_5	R_6	R_7	MIC (mM)	Inhib. Power (%)
1	OH	OH	OH	OH	OH	OH	OH	7.4 ± 2.3	100
2	**H**	OH	OH	OH	OH	OH	OH	17.1 ± 11.9	51
3	OH	**H**	OH	OH	OH	OH	OH	4.8 ± 2.2	143
4	OH	**Me**	OH	OH	OH	OH	OH	3.8 ± 2.8	191[a]
5	OH	**Et**	OH	OH	OH	OH	OH	≥ 25.1	≤ 25
7	OH	OH	**H**	OH	OH	OH	OH	≥ 33.5[b]	≤ 22
10	OH	OH	OH	**H**	OH	OH	OH	≥ 25.9	≤ 30
11	OH	OH	OH	OH	**H**	OH	OH	≥ 34.4	≤ 21
13	OH	OH	OH	OH	OH	**H**	OH	≥ 33.8	≤ 21
16	OH	OH	OH	OH	OH	OH	**H**	≥ 35.4	≤ 20

Moreover, the presence of a glucose unit at the reducing end of galabiose strongly increased affinity. When Nilsson *et al.* investigated the isosteric replacement of the ring oxygen atom of the galactose moiety with sulfur, they observed a loss of affinity. As possible reasons small conformational changes or the disruption of a cooperative H-bond were discussed [57]. In 1998, Hansen *et al.* published the synthesis and evaluation of amino and carboxy analogs of galabiose and showed that 2'-OH of the galabiose can be efficiently replaced by an amino group, increasing the affinity for PapG-I, whilst the other modifications reduced the affinity [53]. However, no data for the binding to PapG-II were provided. The first study with both PapG-I and PapG-II

was published by Ohlsson *et al* in 2002. The most potent inhibitors to date are presented in figure 4 [55].

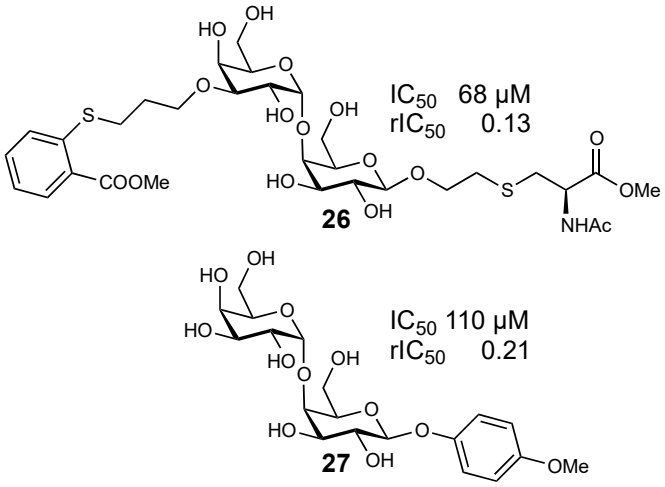


Figure 4. Strongest inhibitors of hemagglutination by *E. coli* bearing PapG-II adhesins, which are described in literature.

An SPR investigation of PapG-II binding to a small set of natural and synthetic inhibitors was published shortly after by Larsson *et al*. [54] Hydroxy groups in the 2'- and 3'-position were suggested as possible anchor points for improving the affinity. The docking results of the reported compound **27** in the PapG-II binding pocket was also disclosed, suggesting that the phenyl ring could be involved not only in hydrophobic interactions, but also in a π-cation interaction with Arg170. It is interesting to observe the drop in affinity of the thioglycoside **28** (figure 5, K_D = 488 µM) as compared to the isosteric compound **27** (figure 4, K_D = 170 µM); proposed to result on conformational changes [52]. Ohlsson *et al.* tested a large number of synthetic inhibitors for binding to the PapG-II adhesin, and developed a small set of QSAR models [56]. However, no improved binders as compared to compound **26** were discovered. In one SPR study, multivalent compounds were synthesized and their inhibitory power against bacterial adhesion to galabiose-coated BSA was measured. Surprisingly, only PapG-I was inhibited. The most potent compound showed 2 µM IC_{50} [57].

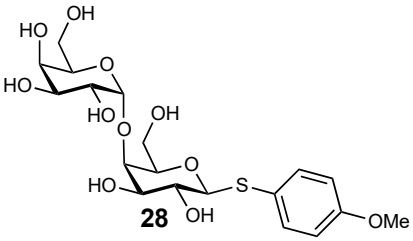


Figure 5.

3.1.8 Aims of this Project

This project is aiming at a deeper understanding of PapG-II binding to its cellular target and at the development of improved antagonists. To reach these goals structure-based as well as fragment-based approaches were used, with support from

crystallography, isothermal titration calorimetry, and NMR techniques. The collected data will help to develop further drug-like PapG-II inhibitors.

References

[1] K. Ramakrishnan, D.C. Scheid. Diagnosis and Management of Acute Pyelonephritis in Adults. *Am. Fam. Phys.* **2005**, *71*(5), 933-942.

[2] W. E. Stamm, T. M. Hooton, J. R. Johnson. Urinary tract infections: from pathogenesis to treatment. *Infect. Dis.* **1989**, *159*(3) 400-406.

[3] K. L. Klemstine, P. D. Brown, B. Foxman. Acute pyelonephritis in US hospitals in 1997: hospitalization and in-hospital mortality. *Ann Epidemiol.* **2003,** *13*(2), 144-150.

[4] P. Brown, M. Ki, B. Foxman. Acute Pyelonephritis Among Adults Cost of Illness and Considerations for the Economic Evaluation of Therapy. *Pharmacoeconomics* **2005,** *23*(11), 1123-1142.

[5] L. C. Gilstrap, S. M. Ramin. Urinary tract infections during pregnancy. *Obstet. Gyn. Clin. N. Am.* **2001,** *28*(3), 581-91.

[6] W. E. Stamm, T. M. Hooton. Management of urinary tract infections in adults. *N. Engl. J. Med.* **1993**, *329*, 1328-34.

[7] D. J. Farrell, I. Morrissey, D. DeRubeis, M. Robbins, D. Felmingham. A UK multicenter study of the antimicrobial susceptibility of bacterial pathogens causing urinary tract infection. *J. Infect.* **2003**, *46*(2), 94-100

[8] G. B. Piccoli, V. Consiglio, M. C. Deagostini, M. Serra, M. Biolcati, F. Ragni, A. Biglino, A. De Pascale, M. F. Frascisco, A. Veltri, and F. Porpiglia. The clinical and imaging presentation of acute "non complicated" pyelonephritis, A new profile for an ancient disease. *Nephrology* **2011**, *12*, 68

[9] M. A. Mulvey. Adhesion and entry of uropathogenic Escherichia coli. *Cell Microbiol.* **2002**, *4*(5), 257–271.

[10] M. A. Mulvey, J. D. Schilling, J. J. Martinez, S. J. Hultgren. Bad bugs and beleaguered bladders: interplay between uropathogenic Escherichia coli and innate host defenses. *Proc. Natl. Acad. Sci. USA* **2000**, *97*(16), 8829–8835.

[11] K. A. Krogfelt, H. Bergmans, P. Klemm. Direct evidence that the FimH protein is the mannose-specific adhesin of Escherichia coli type 1 fimbriae. *Infect. Immun.* **1990**, 58, *6*, 1995-1998.

[12] X. R. Wu, T. T. Sun, J. J. Medina. In vitro binding of type 1-fimbriated Escherichia coli to uroplakins Ia and Ib: relation to urinary tract infections. *Proc. Natl. Acad. Sci. USA* **1996**, *93*(18), 9630-9635.

[13] B. Xie, G. C. Zhou, C. Shiu-Yung, E. Shapiro, X. P. Kong, X. R. Wu, T. T. Sun, C. E. Costello. Distinct glycan structures of uroplakins Ia and Ib:

structural basis for the selective binding of FimH adhesin to uroplakin Ia. *J. Biol. Chem.* **2006**, *281*(21), 14644-14653.

[14] R. Romih, P. Koro, W. de Mello Jr, K. Jezernik. Differentiation of epithelial cells in the urinary tract. *Cell Tissue Res.* **2005**, *320*(2), 259-268.

[15] F.-X. Liang, M. C. Bosland, H. Huang, R. Romih, S. Baptiste, F.-M. Deng, X.-R. Wu, E. Shapiro, T.-T. Sun. Cellular basis of urothelial squamous metaplasia: roles of lineage heterogeneity and cell replacement. *J. Cell. Biol.* **2005**, *171*(5), 835-844.

[16] H. Leffler, C. Svanborg-Eden. Chemical identification of a glycosphingolipid receptor for Escherichia coli attaching to human urinary tract epithelial cells and agglutinating human erythrocytes. *FEMS Microbiol. Lett.* **1980**, *8*(3), 127–134.

[17] B. Lund, F. Lindberg, B. I. Marklund, S. Normark. The PapG protein is the alpha-D-galactopyranosyl-(1–4)-beta-D-galactopyranose-binding adhesin of uropathogenic Escherichia coli. *Proc Natl Acad Sci USA* 1987, *84*(16), 5898–5902.

[18] N. Stromberg, B.I. Marklund, B. Lund, D. Iliver, A. Hamers, W. Gaastra, K.A. Karlsson, S. Normark. Host-specificity of uropathogenic Escherichia coli depends on differences in binding specificity to Gal alpha 1-4Gal-containing isoreceptors. *EMBO J.* **1990**, *9*(6), 2001-2010.

[19] J. A. Roberts, B. I. Marklund, D. Ilver, D. Haslam, M. B. Kaack, G. Baskin, M. Louis, R. Möllby, J. Winberg, S. Normark. The Gal(alpha 1-4)Gal-specific tip adhesin of Escherichia coli P-fimbriae is needed for pyelonephritis to occur in the normal urinary tract. *Proc Natl Acad Sci USA* **1994**, *91*(25), 11889-11893

[20] C. C. Tseng, J. J. Huang, M. C. Wang, A. B. Wu, W. C. Ko, W. C. Chen, J. J. Wu. PapG II adhesin in the establishment and persistence of Escherichia coli infection in mouse kidneys. *Kidn. Int.* **2007**, *71*(8), 764-770.

[21] B. Wullt, G. Bergsten, H. Connell, P. Röllano, N. Gebretsadik, R. Hull, C, Svanborg. P fimbriae enhance the early establishment of Escherichia coli in the human urinary tract. *Mol. Biol.* **2000**, *38*(3), 456-464.

[22] H. L. T. Mobley, K. G. Jarvis, J. P. Elwood, D. I. Whittle, C. V. Lockatell, R. G. Russell, D. E. Johnson, M. S. Donnenberg, J. W. Warren. Isogenic P-fimbrial deletion mutants of pyelonephritogenic Escherichia coli: the role of α Gal(1–4)β Gal binding in virulence of a wild-type strain. *Mol. Biol.* **1993**, *10*(1), 143-155.

[23] M. Gong, L. Makowski. Helical structure of P pili from Escherichia coli. Evidence from X-ray fiber diffraction and scanning transmission electron microscopy. *J. Mol. Biol.* **1992**, *228*(3), 735–742.

[24] E. Bullitt, L. Makowski. Structural polymorphism of bacterial adhesion pili. *Nature* **1995**, *373*(6510), 164–167.

[25] M. J. Kuehn, J. Heuser, S. Normark, S. J. Hultgren. P pili in uropathogenic E. coli are composite fibres with distinct fibrillar adhesive tips. *Nature* **1992**, *356*(6366), 252-255.

[26] C. H. Jones, J. S. Pinkner, R. Roth, J. Heuser, A. V. Nicholes, S. N. Abraham, S. J. Hultgren. FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the Enterobacteriaceae. *Proc. Natl. Acad. Sci. USA* **1995**, *92*(6), 2081–2085.

[27] C. H. Jones, P.N. Danese, J.S. Pinkner, T.J. Silhavy, S.J.Hultgren. The chaperone-assisted membrane release and folding pathway is sensed by two signal transduction systems, *EMBO J.* **1997**, *16*(21), 6394–6406.

[28] S. J. Hultgren, F. Lindberg, G. Magnusson, J. Kihlberg, J. M. Tennent, S. Normark. The PapG adhesin of uropathogenic Escherichia coli contains separate regions for receptor binding and for the incorporation into the pilus. *Proc. Natl. Acad. Sci. USA* **1989**, *86*(12), 4357-4361.

[29] F. Lindberg, J. M. Tennent, S. J. Hultgren, B. Lund, S. Normark. PapD, a periplasmic transport protein in P-pilus biogenesis. *J. Bact.* **1989**, *171*(11), 6052-6058.

[30] F. G. Sauer, K. Futterer, J. S. Pinkner, K. W. Dodson, S. J. Hultgren, G. Waksman. Structural basis of chaperone function and pilus biogenesis. *Science* **1999**, *285*(5430), 1058–1061.

[31] D. Choudhury, A. Thompson, V. Stojanoff, S. Langermann, J. Pinkner, S. J. Hultgren, S. D. Knight. X-ray structure of the FimC–FimH chaperone–adhesin complex from uropathogenic Escherichia coli. *Science* **1999**, *285*(5430), 1061–1066.

[32] H. Remaut, C. Tang, N. S. Henderson, J. S. Pinkner, T. Wang, S. J. Hultgren, D. G. Thanassi, G. Waksman, H. Li. Fiber formation across the bacterial outer membrane by the chaperone/usher pathway. *Cell* **2008**, *133*(4), 640-652.

[33] M. Båga, M. Norgren, S. Normark. Biogenesis of E. coli Pap pili: PapH, a minor pilin subunit involved in cell anchoring and length modulation. *Cell* **1987**, *49*(2), 241-251.

[33] S. Geibel, G. Waksman. The molecular dissection of the chaperone-usher pathway. *Bioch. Biophys. Acta* **2014**, *1843*(8), 1559-1567.

[34] J. R. Johnson, J. L. Swanson, T. J. Barela, J. J. J. Brown. Receptor Specificities of Variant Gal(1-4)Gal-Binding PapG Adhesins of Uropathogenic Escherichia coli as Assessed by Hemagglutination Phenotypes. *Inf. Dis.* **1997**, *175*(2), 373-381.

[35] A. E. Stapleton, M. R. Stroud, S. I. Hakomori, W. E. Stamm. The globoseries glycosphingolipid sialosyl galactosyl globoside is found in urinary tract tissues and is a preferred binding receptor In vitro for uropathogenic Escherichia coli expressing pap-encoded adhesins. *Inf. Immun.* **1998**, *66*(8), 3856-3861.

[36] C. C. Tseng, J. J. Huang, W. C. Ko, J. J. Yan, J. J. Wu. Decreased Predominance Of Papg Class II Allele In Escherichia Coli Strains Isolated From Adults With Acute Pyelonephritis And Urinary Tract Abnormalities. *J. Urol.* **2001**, *166*(5), 1643-1646.

[37] I.M. Johanson, K. Plos, B. I. Marklund, C. Svanborg. *Pap, papG* and *prsG* DNA sequences in *Escherichia coli* from the fecal flora and the urinary tract. *Microb. Pathogenesis* **1993**, *15*(2), 121-129.

[38] J. R. Johnson. papG Alleles among Escherichia coli Strains Causing Urosepsis: Associations with Other Bacterial Characteristics and Host Compromise. Infect. Immun. 1998, 66(9), 4568-4571.

[39] K. W. Dodson, J. S. Pinkner, T. Rose, G. Magnusson, S. J. Hultgren, G. Waksman. Structural basis of the interaction of the pyelonephritic E. coli adhesin to its human kidney receptor. *Cell* **2001**, *105*(6), 133-143.

[40] B. Ford, D. Verger, K. Dodson, E. Volkan, M. Kostakioti, J. Elam, J. Pinkner, G. Waksman, S. J. Hultgren. The structure of the PapD-PapGII pilin complex reveals an open and flexible P5 pocket. *J. Bacteriol.* **2012**, *194*(23), 6390-6397.

[41] A. Larsson, J. Ohlsson, K. W. Dodson, S. J. Hultgren, U. Nilsson, J. Kihlberg, Quantitative studies of the binding of the class II PapG adhesin from uropathogenic *Escherichia coli* to oligosaccharides *Bioorg. Med. Chem.* **2003**, *11*(10), 2255-2261.

[42] G. Otto, G. Magnusson, M. Svensson, J. H. Braconier, C. Svanborg. *pap* Genotype and P Fimbrial Expression in *Escherichia coli* Causing Bacteremic and Nonbacteremic Febrile Urinary Tract Infection. Clin. Infect. Dis. 2001, 32(11), 1523-1531.

[43] I. A. Rose, K. R. Hanson, K.D. Wilkinson, M. J. Wimmer. A suggestion for naming faces of ring compounds. *Proc. Natl. Acad. Sci. USA* **1980**, *77*(5), 2439-2441.

[44] L. M. Nilsson, W. E. Thomas, E. Trintchina, V. Vogel, E. V. Sokurenko. Catch bond-mediated adhesion without a shear threshold: trimannose versus monomannose interactions with the FimH adhesin of Escherichia coli. *J. Biol. Chem.* **2006**, *281*(24), 16656-16663.

[45] I. Pascher, M. Lundmark, P. G. Nyholm, S. Sundell. Crystal structures of membrane lipids. *Bioch. Bioph. Acta* **1992**, *1113*(3-4), 339-373.

[46] O. Björnham, H. Nilsson, M. Andersson, S. Schedin. Physical properties of the specific PapG-galabiose binding in E. coli P pili-mediated adhesion. *Eur. J. Bioph.* **2009**, *38*(2), 245-254.

[47] R. Lugmaier, S. Schedin, F. Kühner, M. Benoit. Dynamic restacking of Escherichia coli P-pili. *Eur. J. Bioph.* **2008**, *37*(2), 111-120.

[48] R. Striker, U. Nilsson, A. Stonecipher, G. Magnusson, S. J. Hultgren. Structural requirements for the glycolipid receptor of human uropathogenic Escherichia coli. *Molec. Biol.* **1995**, *16*(5), 1021-1029.

[49] J. Kihlberg, S. J. Hultgren, S. Normark, G. Magnusson. Probing of the combining site of the PapG adhesin of uropathogenic Escherichia coli bacteria by synthetic analogs of galabiose. *J. Am. Chem. Soc.* **1989**, *111*(16), 6364-6368.

[50] J. Ohlsson, A. Sundin, U. J. Nilsson. Conformational studies on phenyl thioglycosides: a remote effect on disaccharide linkage by phenyl aglycons attenuates recognition of galabiosides by a bacterial adhesin. *Chem. Comm.* **2003**, 3, 384-5.

[51] U. Nilsson, R. T. Striker, S. J. Hultgren, G. Magnusson. PapG adhesin from E. coli J96 recognizes the same saccharide epitope when present on whole bacteria and as isolated protein. *Bioorg. Med. Chem.* **1996**, *4*(11), 1809-1817.

[52] H. C. Hansen, G. Magnusson. Synthesis of some amino and carboxy analogs of galabiose; evaluation as inhibitors of the pilus protein PapGJ96 from Escherichia coli. *Carbohyd. Res.* **1998**, *307*(3-4), 233-242.

[53] J. Ohlsson, J. Jass, B. E. Uhlin, J. Kihlberg, U. J. Nilsson. Discovery of potent inhibitors of PapG adhesins from uropathogenic Escherichia coli through synthesis and evaluation of galabiose derivatives. *ChemBioChem* **2002**, *3*(8), 772-779.

[54] J. Ohlsson, A. Larsson, S. Haataja, J. Alajääski, P. Stenlund, J. S. Pinkner, S. J. Hultgren, J. Finne, J. Kihlberg, U. J. Nilsson. Structure-activity relationships of galabioside derivatives as inhibitors of E. coli and S. suis adhesins: nanomolar inhibitors of S. suis adhesins. *Org. Biomol. Chem.* **2005**, *3*(5), 886-900.

[55] U. Nilsson, R. Johansson, G. Magnusson. Synthesis, Conformational Analysis and Comparative Protein Binding of a Galabioside and Its Thioglycoside Analogues. *Chem. Eur. J.* **1996**, *2*(3), 295-302.

[56] H. C. Hansen, G. Magnusson. Synthesis of selected aminodeoxy analogues of galabiose and globotriose. *Carbohyd. Res.* **1999**, *322*(3-4), 166-180.

[57] A. Salminen, V. Loimaranta, A. F. J. Joosten, S. A. Khan, J. Hacker, R. J. Pieters, J. Finne. Inhibition of P-fimbriated Escherichia coli adhesion by

multivalent galabiose derivatives studied by a live-bacteria application of surface plasmon resonance. *J. Antimicrob. Chemother.* **2007**, *60*(3), 495-501.

3.2 Results

3.2.1 Outline

The PapG-II adhesin has been the target of extensive medicinal chemistry efforts for over 30 years. Several attempts to generate glycomimetics with improved affinity as compared to the natural ligands have failed. The best antagonist in term of affinity and structural simplicity is compound **1**, discovered by Ohlsson *et al*., exhibiting an IC_{50} of 120 µM in hemagglutination experiments.

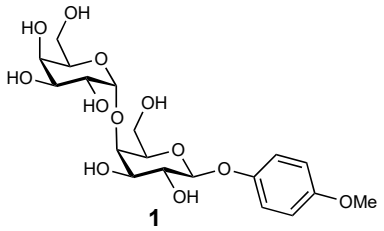
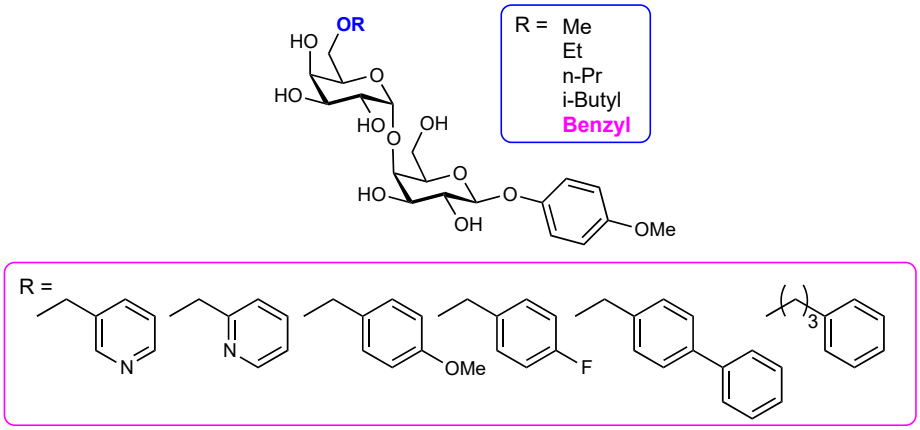


Figure 1. The structure of the most binding efficient PapG-II antagonist known to date.

By examining all reported small molecule antagonists of PapG-II and by molecular modeling, we identified a new promising exit vector, which had never been thoroughly exploited before.

Figure 2. Modifications of compound **1**. A small set of initial probe compounds (encircled in blue)



highlighted the interesting thermodynamic signature of the benzyl-substituted analog. The benzyl series was later expanded (encircled in magenta).

Disappointingly, despite intensive synthetic effort, no improved binders were discovered, confirming PapG-II as a hard-to-drug protein (*chapter 3.2.2*).

Furthermore, a fragment-based approach, which had been successfully applied to challenging targets, was explored. Thus, in our group, other lectins have been effectively targeted by linking a second-site fragment-sized ligand to a central glycomimetic core. However, the application of this strategy to PapG-II yielded only slightly improved binders (Manuscript 1).

Finally, literature data indicate that the hexasaccharidic epitope of sialyl galactosyl globoside **2** has stronger affinity for PapG-II than the tetrasaccharidic epitope of GbO4 (figure 3), which is the major natural target, and from which it differs by the two saccharidic units at the non-reducing end. The determination of the exact K_D for this interaction showed the surprisingly low value of 21.9 µM. Crystal structure and thermodynamic analyses suggested that the increased binding affinity originates from an entropy-driven contribution of the nonbinding saccharide moieties flanking the galabiose core (*Paper 5*).

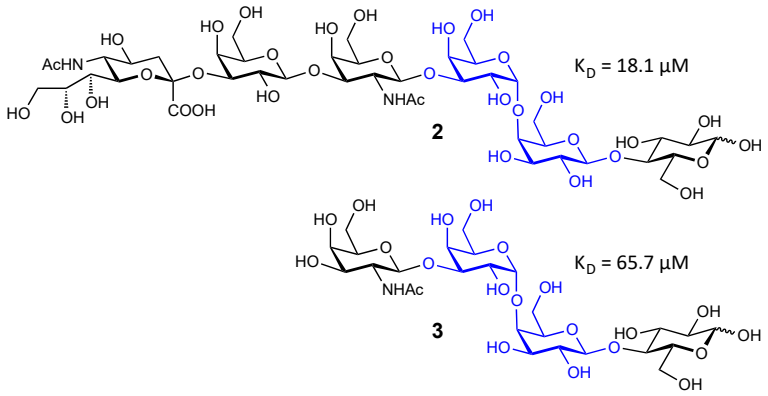


Figure 3. The structures and ITC-determined K_Ds of the hexasaccharidic epitope of sialyl galactosyl globoside **2** and of the tetrasaccharidic epitope of GbO4 **3**.

3.2.2 New Glycomimetics as Antagonists of the PapG-II Adhesin of *E. coli*

Contributions to the project:

Giulio Navarra designed and characterized all compounds used in this study except for compound **1** (commercial). He interpreted ITC, crystallography, and polymer-assay results and wrote the chapter. For the synthesis of the compounds, Lisa Beretta Piccoli importantly collaborated in the frame of her master thesis (cited in the text, were appropriate). Crystallization experiments were conducted by Dr. Roland Preston and Dr. Roman P. Jakob. The protein PapG-II was produced by Dr. Katja Stangier, Dr. Roland Preston, and Dr. Said Rabbani. The polymer assay was run by Dr. Said Rabbani.

3.2.2 New Glycomimetics as Antagonists of the PapG-II Adhesin of E. coli

Background

Despite extensive medicinal chemistry efforts devoted to the targeting of PapG-II, no affinity improvement as compared to the natural ligand has been achieved. The minimal binding epitope for all PapG adhesins is the galabiose core (Gal1-4αGal) [1]. Magnusson's and later Nilsson's groups explored the chemical modification of each hydroxyl group on the Gal1-4αGal core [2-7]. The most obvious exit vectors, namely anomeric position 1 and the 3'-position of galabiose, were thoroughly explored. However, the best binder **2** [8], exhibits only mid-micromolar affinity [4].

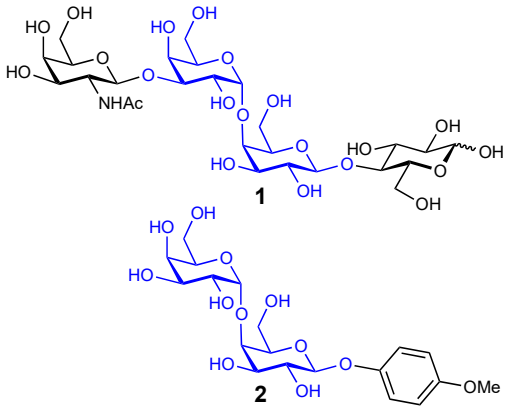


Figure 1. Structures of the tetrasaccharide epitope of GbO4 (natural ligand of PapG-II, **1**) and of the best mimetic **2**. The galabiose core is highlighted in blue.

Furthermore, 6'-position has been only poorly investigated. Its dehydroxylation reduces the affinity for PapG-II by at least 79% [9], but no other modifications were explored. Previous studies on PapG-I ($PapG_{J96}$) had evidenced a comparable effect; however, the isosteric substitution of the 6'-OH with a fluorine atom reduced the affinity only by 50% [2]. As the binding pocket of both adhesins recognizes the same pattern of OH groups on the galabiose [9], we speculated that for PapG-II a similar effect is to be expected and could be related to the H-bond acceptor character of this moiety, as suggested by Kihlberg *et al.* for PapG-I [2]. In the crystal structure of the natural tetrasaccharide receptor in complex with the lectin domain of PapG-II (PapG-

II_{LD}, PDB: 18JR), the corresponding OH group of the β-galactosyl unit accepts an H-bond from Gly104 and is involved in a water bridge to Lys103 [10]. For a deeper insight into the binding pocket, compound **2** [4] was synthesized and co-crystallized with PapG-II_{LD}, thus providing a clear picture of its binding mode. A series of new candidates was then synthesized and tested for binding to PapG-II_{LD}. Disappointingly, no improvement in affinity could be achieved. However, the results provided a better understanding of the binding event and will guide future medicinal chemistry efforts.

Results and Discussion

In the crystal structure of the complex PapG-II_{LD} : **2**, the galabiose core of compound **2** forms the same interactions as described for the natural tetrasaccharide Gal*N*Ac1-3βGal1-4αGal1-4βGlc (epitope of GbO4, compound **1**) [10]. However, in each asymmetric unit cell two molecules of PapG-II_{LD} are found (= chains A and B, respectively). The major difference among chains A and B involves the orientation of Lys103, which in chain A is forming a water bridge with the 6'-OH of **2**, and in chain B is pointing away from it, as highlighted in Figure 2. The fact that this interaction is not always formed suggested that its contribution to the binding energy is limited, and that targeting the lysine with appropriate manipulations of group 6'-OH could be beneficial. Moreover, if Lys103 is pointing away from the ligand (thus relaxing to an all-anti conformation), it would uncover a lipophilic patch formed by the backbone of residues 102-104 and the β-ε carbons of Lys103.

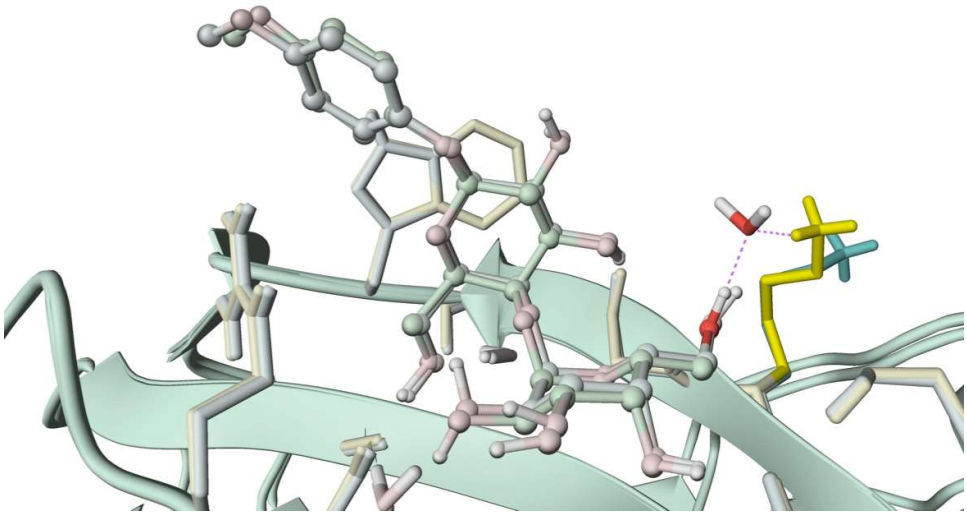
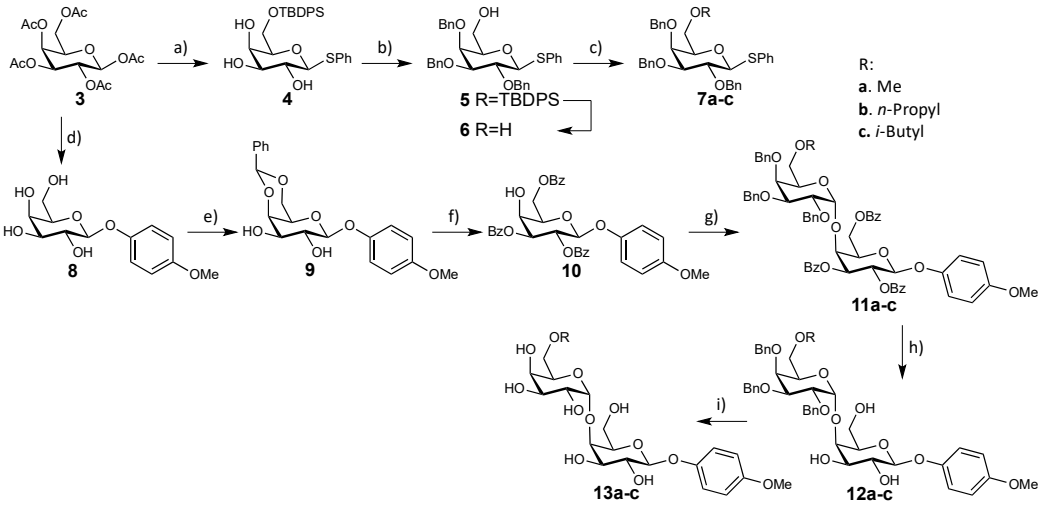


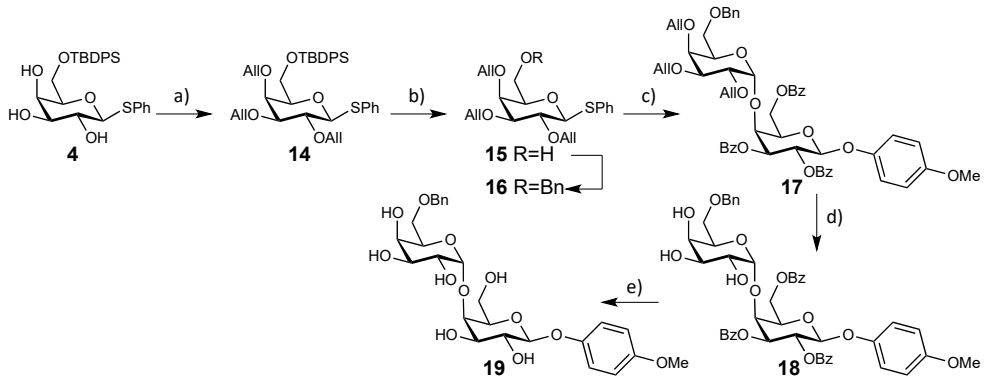
Figure 2. Highlight of Lys103 and the bridged water molecule, as found in the co-crystal **2** : PapG-II_{LD}. Overlap of chains A (yellow) and B (turquoise) from the same asymmetric unit. In chain A, the lysine is involved in a water bridge to 6'-OH. In chain B, Lys103 points away from the ligand.

To explore this hypothesis, "probe" compounds **13a-c**, **19**, and **25a** were synthesized as described in schemes 1-3. Charged moieties were not explored to avoid the introduction of additional polarity. The glycosyl donors **7a-c** were prepared starting from peracetylated α-D-galactose **3**, which was reacted with thiophenol in presence of $BF_3 \cdot Et_2O$, followed by removal of the acetyl groups with catalytic sodium methoxide in methanol. After regioselective protection of the 6-OH with *tert*-butyl(chloro)diphenylsilane (TBDPSCl), compound **4** was obtained. Subsequent benzylation of the remaining hydroxyl groups and selective removal of the silyl protective group gave access to **6**. Alkylation under basic conditions with the appropriate bromo- or iodoalkane provided the glycosyl donors **7a-c** [11]. The synthesis of compound **10** from **3** was adopted from Ohlsson, *et al*. [12]. Finally, the disaccharides **13a-c** were accessed through glycosylation of **10** with **7a-c**, followed by removal of the benzyl and benzoyl groups.



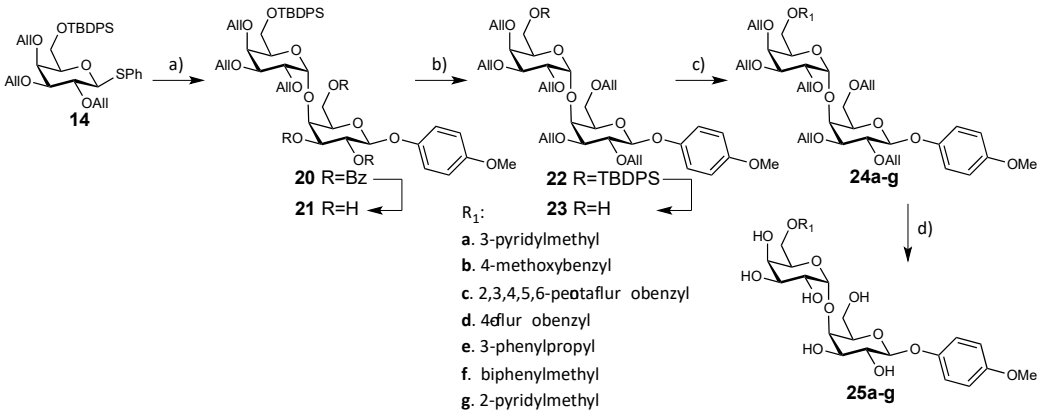
Scheme 1. a) i. PhSH, $BF_3 \cdot Et_2O$, DCM, rt, 2 h, ii. MeONa, MeOH, rt, overnight, 74%; iii. TBDPSCl, imidazole, DMF, 0 °C → rt, 8 h, 92%; b) i. BnBr, NaH, DMF, rt, 3 h, 60%, ii. TBAF 1 M, AcOH, 0 °C → rt, 1h, 75%; c) Base, RX, 0 °C → rt, 44-96%; d) i. 4-methoxyphenol, $BF_3 \cdot Et_2O$, DCM, rt, 4 h, ii. MeONa, MeOH, rt, overnight, 72%; e) $PhCH(OMe)_2$, *p*-TsOH, ACN, rt → 0 °C, 30 min, 85%; f) i. BzCl, DMAP, pyridine, 0 °C → rt, overnight, ii. 2 N HCl/THF 1:7, 55 °C, 24 h, iii. BzCl, pyridine, -10 °C, 75 min, 53%; g) i. **7a-c**, NIS, TMSOTf, Et_2O/DCM 2:1, -55 °C, 1-5 h, 78-27%; h) MeONa, MeOH, rt, 20-24 h, 96-84%; i) H_2, Pd-C, AcOH, rt, 85-81%.

The synthesis of 6'-*O*-arylmethyl compounds **19** (scheme 2) and **25a** (scheme 3) required a different approach to avoid the cleavage of the 6'-O-benzyl group during the final hydrogenation step. The key glycosyl donor **14** was prepared from **4**, through reaction with allyl bromide (scheme 2). Compound **16** was obtained by replacing the protective group in the 6-position and used for glycosylation of compound **10** (→ **17**). Full deprotection finally yielded compound **19**.



Scheme 2. a) i. PhSH, $BF_3 \cdot Et_2O$, DCM, rt, 2 h, ii. MeONa, MeOH, rt, overnight, 74%, iii. TBDPSCl, imidazole, DMF, 0 °C → rt, 8 h, 92%, iv. AllBr, NaH, DMF, rt, 30 min, 73%; b) i. TBAF 1 M, AcOH, 0 °C → rt, overnight, 78%, ii. NaH, BnBr, DMF, 0 °C → rt, 1 h, 77%; c) **10**, NIS, TMSOTf, Et_2O/DCM 2:1, -55 → 30 °C, 4 h, 39%; d) DMBA, $Pd(OAc)_2$, PPh_3, MeOH/DCM 3:1, rt → 40 °C, 5 h, 78%; e) MeONa, MeOH, rt, 18 h, 67%.

A similar approach proved ineffective for synthesizing compound **25a** and congeners. Therefore, the synthetic pathway in scheme 3 was designed. Compound **14** was glycosylated with **10**, to obtain **20**. The benzoyl groups of compound **20** were exchanged for allyl groups, and then the TBDPS group was removed, to yield **23**. Alkylation of **23** with the appropriate halogenide and removal of the allyl groups provided access to compound **25a-g**.



Scheme 3. a) **10**, NIS, TMSOTf, DCM, -45 → -15 °C, 5 h, 67%; b) i. MeONa, MeOH, rt, 40 h, 75%, ii. AllBr, NaH, DMF, 0 °C → rt, 50 min, 62%; iii. TBAF 1M, AcOH, 0 °C → rt, overnight, 90%; c) $ArCH_2X$, NaH, DMF, 0 °C → rt, 2-33 h, 74-50%; d) DMBA, $Pd(OAc)_2$, PPh_3, MeOH/DCM 3:1, rt → 40 °C, 7-24 h, 65-23%.

The binding to PapG-II_{LD} of the small set of probe compounds was investigated by isothermal titration calorimetry (ITC). As expected, the affinity of compound **13a**, which bears the 6'-OMe group, drops to about 50% of the value for **2**, due to a $\Delta\Delta G$ of 1.7 kJ/mol. This is in good agreement with previous studies on PapG-I [2].

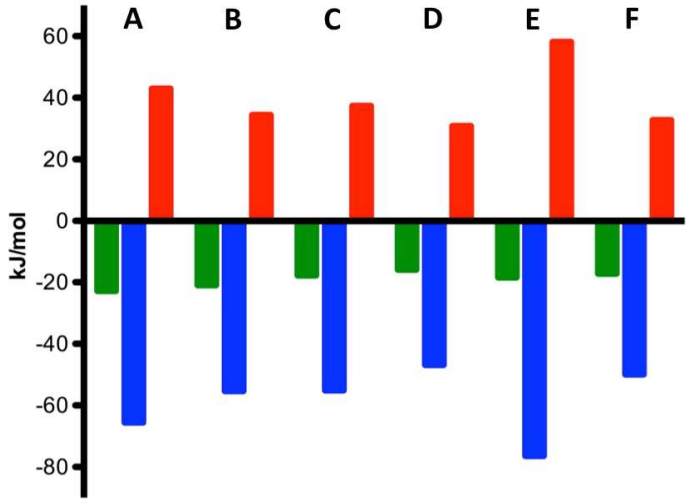
The difference in Gibbs free energy between compound **2** and **13a** is due to 8.0 kJ/mol for $\Delta\Delta H$ and -6.2 kJ/mol for $T\Delta\Delta S$, which can be related to the loss of a hydrogen bond (improved entropy and reduced enthalpic contribution), only partially compensated by the lower desolvation penalty (improved enthalpy). Disappointingly, increasing the size of the 6'-O alkyl substituent (compounds **13b** and **13c**) proved detrimental. For compound **13b** the reduction of affinity, compared to **13a**, is due almost exclusively to entropy, with $T\Delta\Delta S$ = 3.4 kJ/mol. This strongly suggests that the interaction with the protein is identical, but forces compound **13b**, or the binding pocket' side chains or both in an unfavorable conformation. Compound **13c**, however, behaves differently, with an entropy term being more favorable for **13a** ($T\Delta\Delta S$= -2.8 kJ/mol), but an enthalpy importantly reduced ($\Delta\Delta H$ = 6.4 kJ/mol). A possible

explanation is a steric clash of the bulky 6'-substituent with the protein. Because of the loose interaction, residual movements – and therefore lower entropic cost – can be expected.

Table 1. Thermodynamic signatures of PapG-II_{LD} ligands. C values are not reported as in all cases they were < 0.01, thus requiring *N* to be fixed to 1. K_D and ΔH values are average values of duplicate experiments at 95% c.i.

Entry	Cpd	K_D (95% c.i.)	*N*	$\Delta G°_{Obs}$ [kJ/mol]	$\Delta H°$ [kJ/mol] (95% c.i.)	$-T\Delta S°$ [kJ/mol]
A	**2**	102 µM (98 – 107)	1 (fix)	-22.8	-51.4 (-50.1 - -52.7)	28.6
B	**13a**	212 µM (203 – 223)	1 (fix)	-21.0	-43.4 (-41.8 - -45.1)	22.4
C	**13b**	677 µM (617 – 742)	1 (fix)	-18.1	-43.0 (-39.2 - -47.2)	25.8
D	**13c**	1.6 mM (970 – 3332)	1 (fix)	-15.9	-36.6 (-20.3 - -89.7)	20.7
E	**19**	598 µM (566 – 632)	1 (fix)	-18.4	-59.8 (-56.7 - -63.2)	41.4
F	**25a**	945 µM (846 – 1060)	1 (fix)	-17.3	-39.0 (-34.5 - -52.7)	21.7

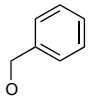
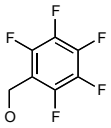
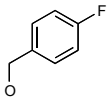
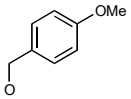
Figure 3. Bar graph of the data from table 1. Δ*G* values are in green, Δ*H* in blue, and –*T*Δ*S* in red.

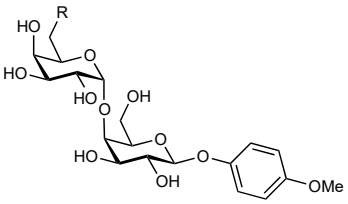


Compound **19** exhibits the most interesting thermodynamic profile. When compared to the reference compound **2**, $\Delta\Delta H$ = -8.4 kJ/mol and $T\Delta\Delta S$ = 12.8 kJ/mol were obtained. Desolvation alone does not explain this result, as according to the hydrophobic effect a more favorable entropy contribution is expected. A possible explanation is the formation of a π-cation interaction of the phenyl ring and Lys103.

The entropy penalty would then result from freezing of Lys103 side chain and at least partially from conformational restriction of the disaccharide. If this holds true, it should be possible to optimize the interaction with Lys103 and thus compensate for the large entropy cost. Maintaining the aromatic moiety, compounds **25b-g** were synthesized and tested for binding to PapG-II_{LD}. Due to the high protein demand for ITC experiments, this new set of molecules was evaluated in a cell-free plate-based competitive assay [13]. Results are summarized in table 2. Disappointingly, none of compounds **25b-g** showed an improved affinity compared to **2**. Compound **25f** was not evaluated due to low solubility. Interestingly, functionalization of the aromatic ring with one methoxy group (**25b**) or one fluorine atom (**25d**) in *para*-position, or the extension of the linker between the disaccharide and the aromatic ring (**25e**) all resulted in superior affinity, compared to compound **19**. Especially the relatively high affinity of compound **25d** was unexpected, suggesting that interactions other than π - cation are involved. To explain the observations and guide further studies, three representative molecules, namely **13a**, **19**, and **25e**, were selected for co-crystallization with PapG-II_{LD}.

Table 2. IC_{50} and rIC_{50} values for the series of benzylated compounds. *For compound **19** the relative K_D is reported.

Cpd	R	IC_{50}	rIC_{50}
2	OH	326.8	**1**
19		—	5.9*
25c		3500	10.7
25d		839.6	2.5
25g		6500	19.9
25b		715.2	2.2
25e		622.4	1.9



In Figure 4, crystal structures of various disaccharides bound to the protein are shown. The asymmetric crystal units contain two protein molecules; a close-up of the binding pocket of each is presented, in order to highlight small differences. Compound **13a** binds as predicted, but compounds **19** and **25e** show a surprising arrangement of the phenyl ring, which is stacking over the backbone amide bond connecting Ser89 and Trp88. Interestingly, Lys103 is pushed away from the disaccharide core because of the larger 6'-O substituent; although in the complex with compound **25e** it still presents two conformations, in presence of compound **19** it is frozen in one single conformation.

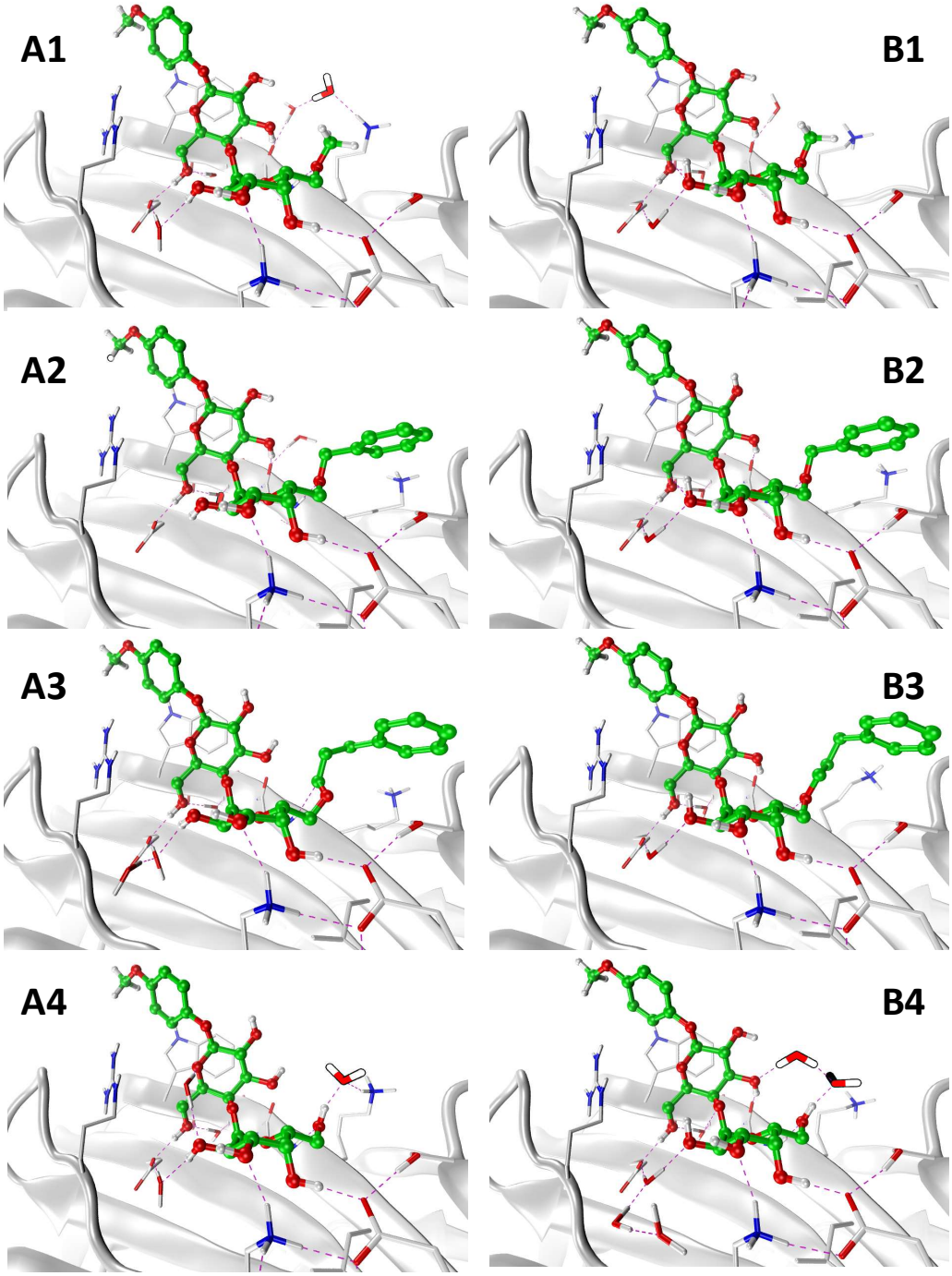


Figure 4. Binding pocket of PapG-II$_{LD}$ in complex with compounds **13a** (A1, B1), **19** (A2, B2), **25e** (A3, B3), and **2** (A4, B4). The two slightly different crystal structures (A and B) identified in each unit cell are compared side-by-side for each complex. The most important residues are depicted as white sticks. The protein structure is in cartoon representation. H-bonds are highlighted with dashed purple lines. Conserved water molecules are shown as sticks. Water molecules close to position 6'-O of the galabiose are highlighted with ticker sticks and Goodwell coloring style. Ligand carbon atoms are in green. Oxygen and nitrogen atoms are in red and blue, respectively.

The analysis of the scaled B-factors can provide insights into the local dynamics of proteins [14]. The direct comparison of B factors extracted from different structures is

not possible, as they are scaled differently, according to the refinement procedure [14]. For the calculation of the scaled B-factors the unity-based scaling and the *z*-score normalization are often applied. As both were shown to provide comparable results [14], in our study only the unity-based scaling according to equation 1 was applied.

$$B_{x_{scaled(i)}} = \frac{[B_{x(i)} - B_{\min(i)}]}{[B_{\max(i)} - B_{\min(i)}]} + 1 \qquad \text{eq. 1}$$

An important parameter when extracting information on residue's dynamics is the resolution. At low resolution (< 2.5Å) the B-factors become less informative [14]. In our case, all complexes have very high and comparable resolution, reinforcing our interpretation. However, the crystal structure of the apo form has only 2.5Å resolution and should be considered with care. In Figure 5 the scaled B-factors for the atoms of Lys103 are summarized in a bar graph. The highest values are found for the apo form, whilst the lowest in the complex with **2**. The disruption of the water-mediated bond with Lys103 (complex with **13a**) improves its movement.

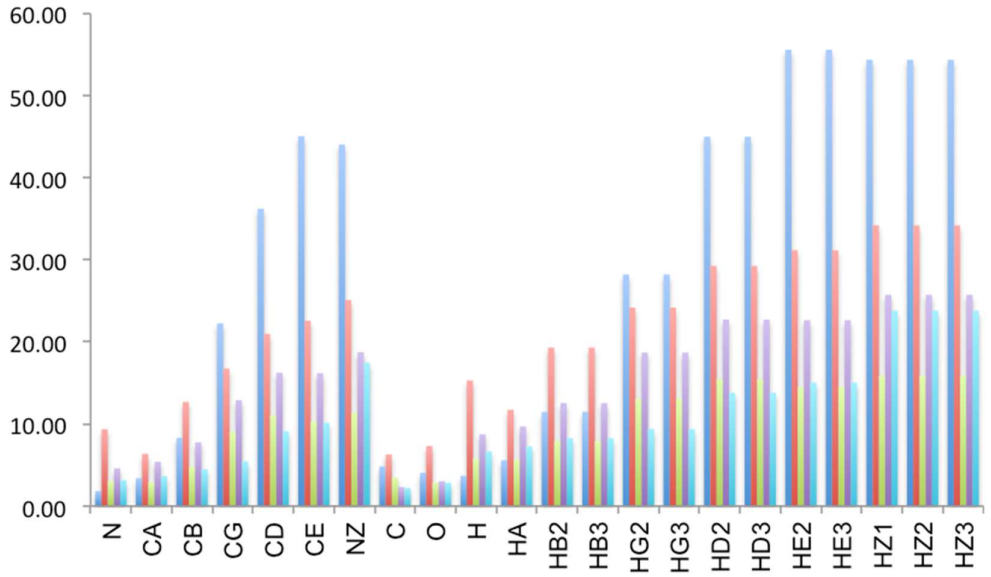


Figure 5. Scaled B-factors for the atoms of Lys103 from the crystals of PapG-II$_{LD}$. Apo form (blue) and in complex with **2** (cyan), **13a** (red), **19** (green), and **25e** (violet). N, C, O, H: nitrogen, carbon, oxygen, and hydrogen atoms involved in the amide bond in the backbone of the protein; CA to CE: carbon atoms α to ε of the lysine; NZ: nitrogen atom of the ε amine group; HA to HE: hydrogen atoms bound to carbons α to ε; HZ: hydrogen atoms of the ε amine group.

However, the introduction of the larger side chains of **19** and **25e** reduces the B-factors. The smallest values are recorder in presence of compound **19**. This structural

information explains the observed thermodynamic data for compound **19**. The lower desolvation cost for the benzyl ether substituent as compared to the OH group in position 6' reduces the enthalpic cost. The stacking on the backbone amide bond has suboptimal geometry and likely contributes poorly to the thermodynamic signature, as it shows a displacement value *x* of about 2.5Å (Figure 6). This value is almost 2-fold larger than in the minimum-energy arrangement of model systems [15].

From the comparison of the complexes of PapG-II_{LD} with **2** and **13a**, the broken water-mediated bond is estimated to contribute -1.7 kJ/mol $\Delta\Delta G$. The limitation of Lys103 motion contributes negative entropy (Figure 4, A4-B4 and A1-B1).

The increased affinity of compound **25e** compared to **19** is likely due to the increased motion of Lys103 (Figure 5) and to better stacking on the amide backbone. This compensates well for the perturbation of the H-bond between Gly104 and 6'-O (Figure 4).

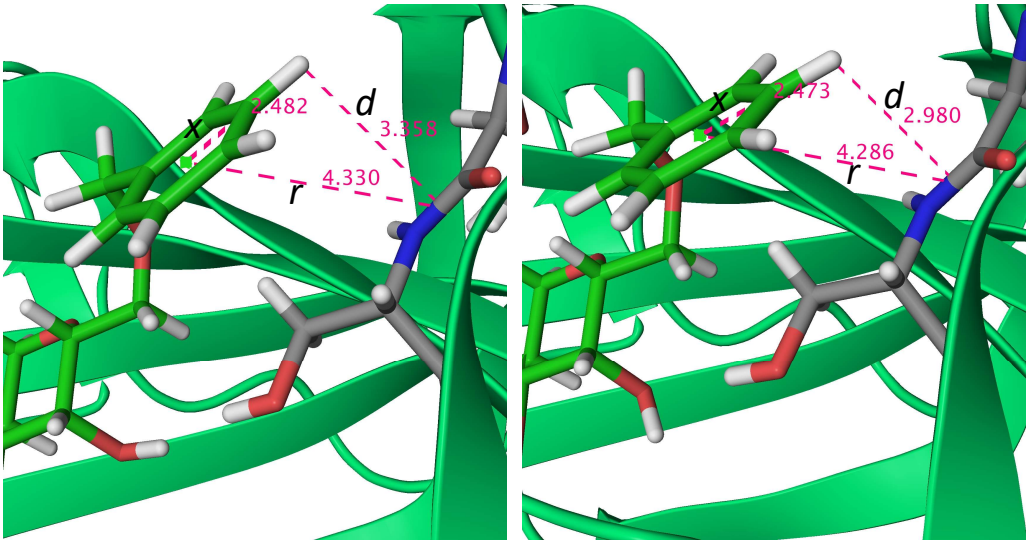


Figure 6. Stacking of the phenyl ring of the 6'-benzyl group in compound **19** on the backbone of the protein. Chain A (left) and chain B (right). Ribbons are in green, residue Ser88 and the backbone of Trp89 are shown as sticks (carbon in grey, oxygen in red, nitrogen in blue, hydrogen in white). The ligand is represented in sticks (carbon in green, oxygen in red, nitrogen in blue, hydrogen in white). Distances are shown in strawberry color and highlighted by dashed lines. Numbers are in angstroms. Distance between planes = *d*; distance between centroids = *r*; displacement = *x*.

The improvement in affinity of compounds **25b** and **25d** is difficult to rationalize. When related to the different dipole moments [15], this effect would be of opposite sign for the *p*-fluoro and *p*-methoxybenzyl substituents. Polarizability has been used to explain the improved interaction energy among substituted benzene dimers, due to the increased dispersion component [16]. However, values calculated on model

systems are not consistently correlated with the observed affinities (not shown). A further possible explanation involves a different binding mode, in which the fluorine atom of **25d** and the methoxy group of **25b** are interacting with the amino group of Lys103.

Conclusions

Co-crystallization of compound **2** with the lectin domain of PapG-II finally disclosed the binding mode of this structurally simplified mimetic of the natural epitope of GbO4. Based on these results and the examination of all previously published antagonists, a new, promising exit vector on the galabiose core was suggested. A series of probe compounds was evaluated for binding to PapG-II_{LD}. Compound **19** was identified as a new lead structure. However, none of the analogs **25a-g** showed better affinity, when compared to the reference compound **2**. Co-crystallization of **19** and **25g** with PapG-II_{LD} revealed a strained conformation of the side chain in position 6'-O of the galabiose core, that can explain the lack of improvements in affinity. These observations will help in rationally designing new antagonists.

Experimental

General methods:

NMR spectra were recorded on a Bruker Avance DMX-500 (500.1 MHz) spectrometer. Assignment of ^{1}H and ^{13}C NMR spectra was achieved using 2D methods (COSY, HSQC, HMBC). Chemical shifts are expressed in ppm using residual $CHCl_3$, CHD_2OD, or HDO as references. Optical rotations were measured using PerkinElmer polarimeter 341. Electron spray ionization mass spectra were obtained on a Waters micromass ZQ. Reactions were monitored by TLC using glass plates coated with silica gel 60 F_{254} (Merck) and visualized under UV light and/or by charring with a molybdate solution (a 0.02 M solution of ammonium cerium sulfate dihydrate and ammonium molybdate tetrahydrate in 10% aqueous H_2SO_4), or by oxidation with a 5% aqueous $KMnO_4$ solution. MPLC separations were carried out on

a CombiFlash Rf (Teledyne Isco) equipped with RediSep normal-phase or RP-C18 reversed-phase flash columns. LC−MS separations were done on a Waters system equipped with sample manager 2767, pump 2525, PDA 2525, and Micromass ZQ, using a SUNFIRE Prep C18 OBD 5 µM, 19x150 mm column, and eluting with the appropriate water/ACN gradient, in presence of 0.2% HCOOH. Commercially available reagents were purchased from Fluka, Aldrich, Alfa Aesar, or Abcr GmbH & Co. KG (Germany). Solvents were purchased from Sigma-Aldrich or Acros and were dried prior to use where indicated. Methanol (MeOH) was dried on 3Å heat-activated molecular sieves for at least 48 h before use. Figures were generated with Maestro (Schrödinger Release 2012: Maestro, version 9.3.5, Schrödinger, LLC, New York, NY, 2012), VMD [17], and Prism 5 (GraphPad Software Inc., San Diego, U.S.A.)

Synthesis

Phenyl 6-*O*-*tert*-butyldiphenylsilyl-1-thio-β-D-galactopyranoside (4):

From peracetylated galactose **3** (1.0 g, 2.6 mmol), phenyl 1-thio-β-D-galactopyranoside was synthesized as described in Ref. [12]. The obtained compound (591 mg, 2.17 mmol) and imidazole (222 mg, 3.26 mmol) were dissolved in 2 mL of DMF and cooled to 0 °C. TBDPSCl (657 mg, 2.39 mmol) was added dropwise with stirring. The mixture was allowed to warm to rt, and the solution was stirred for 7 h. To drive the reaction to completion, additional 0.1 eq of TBDPSCl were added, and stirring was prolonged for additional 8 h. Methanol was added (1 mL) to quench the reaction, then the solvents were removed under vacuum and by co-evaporation with xylenes. Chromatographic separation (DCM/MeOH gradient) yielded **4** (1.016 g, 92%).

$[\alpha]_D^{20}$ -13.8 (*c* 1.00, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{28}H_{34}NaO_5SSi$ $[M+Na]^+$: 533.2, found 533.1.

^1H and ^{13}C NMR data were in accordance with literature values [18].

Phenyl 6-*O*-*tert*-butyldiphenylsilyl-2,3,4-tris-*O*-benzyl-1-thio-β-D-galactopyranoside (5):

Compound **4** (1016 mg, 1.99 mmol) was dissolved in 1 mL of dry DMF and added dropwise to a suspension of NaH (191 mg referred to dry NaH, 7.96 mmol; total amount of the 60% dispersion in mineral oil added: 318 mg) in 8 mL of dry DMF at 0

°C. A solution of benzyl bromide (1.285 g, 7.560 mmol) in 7 mL of DMF was added dropwise, then the mixture was allowed to warm to rt, and the mixture was stirred until disappearance of the starting material on TLC. On TLC, some decomposition was evident. The reaction was quenched with MeOH (1 mL), then diluted with Et_2O and washed with water and satd aq. NH_4Cl. The organic layer was evaporated and chromatographed on silica (PE/EtOAc gradient) to yield **5** (564 mg, 36%), **6** (243 mg 23%), and 96 mg (7.6%) of perbenzylated phenylthiogalactoside.

Phenyl 6-O-tert-butyldiphenylsilyl-2,3,4-tris-O-benzyl-1-thio-β-D-galactopyranoside (**5**):

ESI-MS: *m/z*: Calcd for $C_{49}H_{52}NaO_5SSi$ $[M+Na]^+$: 803.3, found 803.4

^1H and ^{13}C NMR data were in accordance with literature values [19].

Phenyl 2,3,4,6-tetra-O-benzyl-1-thio-β-D-galactopyranoside:

^1H and ^{13}C NMR data were in accordance with literature values [12].

Phenyl 2,3,4-tris-*O*-benzyl-1-thio-β-D-galactopyranoside (6):

Compound **5** (564 mg, 0.722 mmol) was dissolved in TBAF (1M in THF, 2.15 mL, 7.44 mmol) at 0 °C, then glacial acetic acid was added (0.14 mL, 2.38 mmol) under argon. The reaction was allowed to proceed at rt, while monitoring by TLC (PE/EtOAc, 2:1). After 2 h the mixture was diluted with EtOAc and washed three times with brine, dried over Na_2SO_4, filtered and concentrated. The residue was chromatographed on silica (PE/EtOAc gradient) to yield **6** (292 mg, 74.5%).

$[\alpha]_D^{20}$ -9.4 (*c* 1.00, $CHCl_3$) [20].

ESI-MS: *m/z*: Calcd for $C_{33}H_{34}NaO_5S$ $[M+Na]^+$: 565.2, found 565.2

^1H and ^{13}C NMR data were in accordance with literature values [20].

Phenyl 6-*O*-methyl-2,3,4-tris-*O*-benzyl-1-thio-β-D-galactopyranoside (7a):

Compound **6** (243 mg, 0.432 mmol) was dissolved in 2 mL of dry DMF and added dropwise to a suspension of NaH (60% dispersion in mineral oil, 35 mg, 0.864 mmol) in 3 mL of dry DMF at 0 °C. MeI (122.6 mg, 0.864 mmol) was added, then the mixture was allowed to warm to rt, and the mixture was stirred until disappearance of the starting material on TLC (EtOAc/PE, 2:3). The reaction was then quenched with MeOH, evaporated, and taken up in DCM/water. The water phase was extracted 3

times with DCM, dried on Na_2SO_4, filtered, and evaporated. Chromatography on silica (PE/EtOAc gradient) yielded **7a** (240 mg, 96%).

ESI-MS: *m/z*: Calcd for $C_{34}H_{36}NaO_5S$ $[M+Na]^+$: 579.2, found 579.2.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.56 (d, *J* = 2.6 Hz, 2H, Ar), 7.42 – 7.22 (m, 15H, Ar), 7.17 (s, 3H, Ar), 4.97 (d, *J* = 11.5 Hz, 1H, PhC*H*H), 4.78 (d, *J* = 10.2 Hz, 1H, PhC*H*H), 4.70 (dd, *J* = 19.5, 12.7 Hz, 3H, 3 PhC*H*H), 4.66 – 4.59 (m, 2H, PhC*H*H, H-1), 3.98 - 3.90 (m, 2H, H-4, H-2), 3.62 - 3.50 (m, 4H, H-6a, H6b, H-3, H-5), 3.27 (s, 3H, OCH_3).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 138.89 (Ar), 138.49 (Ar), 138.36 (Ar), 134.24 (Ar), 131.67 (Ar), 128.89 (Ar), 128.55 (Ar), 128.43 (Ar), 128.41 (Ar), 128.33 (Ar), 127.98 (Ar), 127.83 (Ar), 127.81 (Ar), 127.68, (Ar) 127.61 (Ar), 127.17 (Ar), 87.78 (C-1), 84.25, 77.49, 77.37, 75.71, 74.57, 73.56, 72.75, 71.04, 59.21 (OCH_3).

Phenyl 6-*O*-propyl-2,3,4-tris-*O*-benzyl-1-thio-β-D-galactopyranoside (7b):

A dispersion of 60% NaH in mineral oil (9.2 mg, 0.23 mmol) was suspended in dry DMF (2 mL) at 0 °C, then compound **6** (62.2 mg, 0.115 mmol) in dry DMF (1.25 mL) was added dropwise, followed by 1-iodopropane (22 µL, 0.23 mmol). The mixture was then allowed to warm to rt. The reaction was monitored by TLC (PE/EtOAc, 3:2). After 2 h, 0.5 eq of NaH were added (2.3 mg, 0.06 mmol), followed by 0.5 eq of 1-iodopropane (5.5 µL, 0.06 mmol). After 2.5 h the reaction was quenched with MeOH. The solvents were evaporated and the residue was taken up in DCM. The organic solution was washed with water. The water phase was extracted three times with DCM. The collected organic phases were dried on Na_2SO_4, filtered, and concentrated. The residue was chromatographed on silica (PE/EtOAc gradient) to yield compound **7b** (34 mg, 51%).

$[\alpha]_D^{20}$ -0.5 (*c* 1.0, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{36}H_{40}NaO_5S$ $[M+Na]^+$: 607.3, found 607.1.

^{1}H-NMR (500 MHz, $CDCl_3$): δ = 7.52 – 7.43 (m, 2H, Ar), 7.34 – 7-12 (m, 15H, Ar), 7.15 – 7.10 (m, 3H, Ar), 4.90 (d, *J* = 11.5 Hz, 1H, PhC*H*H), 4.71 (d, *J* = 10.2 Hz, 1H, PhC*H*H), 4.67 – 4.65 (m, 3H, 3 PhC*H*H), 4.58 (d, 2H, *J* = 2.1 Hz, H-1), 3.89 (d, *J* = 2.8 Hz, 1H, H-4), 3.86 (t, *J* = 9.5 Hz, 1H, H-2), 3.54 (dd, *J* = 9.2, 2.8 Hz, 1H, H-3), 3.51 – 3.49 (m, 3H, H-5, H-6a, H-6b), 3.33 – 3.27 (m, 1H, $OCH\mathit{H}CH_2CH_3$), 3.25 – 3.12 (m, 1H, $OCH\mathit{H}CH_2CH_3$), 1.51 – 1.40 (m, 2H, $OCH_2C\mathit{H}_2CH_3$), 0.81 (t, *J* = 7.4 Hz, 3H, $OCH_2CH_2C\mathit{H}_3$).

^{13}C-NMR (126 MHz, $CDCl_3$): δ = 131.52 (Ar), 128.76 (Ar), 128.43 (Ar), 128.19 (Ar), 127.83 (Ar), 127.68 (Ar), 127.57 (Ar), 87.75 (C-1), 84.21, 77.27, 75.66, 74.47, 73.61, 73.21, 72.73, 68.98($CH_2CH_2CH_3$), 22.88 (CH_2CH_3), 10.58 (CH_3).

Phenyl 6-*O*-(2-methylpropyl)-2,3,4-tris-*O*-benzyl-1-thio-β-D-galactopyranoside (7c):

Powdered KOH (94.7 mg, 1.57 mmol) was suspended in dry DMSO (1 mL) and stirred for 10 min at 0 °C. Compound **6** (91.7 mg, 0.169 mmol) was added dropwise, followed by 1-bromo-2-methylpropane (0.185 mL, 1.69 mmol). The mixture was allowed to warm to rt. The reaction was monitored by TLC (PE/EtOAc, 3:2). After 6 h, an additional aliquot of KOH was added (94.7 mmol, 1.69 mmol). After 20 h, the reaction was complete. The mixture was diluted with water and the water phase was extracted three times with DCM. The collected organic phases were dried on Na_2SO_4, filtered, and concentrated. The residue was chromatographed on silica (PE/EtOAc gradient) to yield compound **5c** (44 mg, ≈ 44%), slightly contaminated with an impurity. The compound was directly used for the next step without further purification.

ESI-MS: *m/z*: Calcd for $C_{37}H_{42}NaO_5S$ $[M+Na]^+$: 621.3, found 621.2.

General procedure for the synthesis of compounds **11a-c**:

Glycosylation:

Compound **10** was synthesized from **3** as described in ref. [12]. It was dissolved in dry DCM, together with 1.2 eq of NIS and 1.2 eq of compounds **7a-c**. To the solution was added dry Et_2O to a final composition of 1:2 DCM/Et_2O, and final concentration of compound **10** of 0.03 M. The mixture was cooled to -55 °C, and then 0.17 eq of TMSOTf were added. The mixture was stirred at -55 °C for 1 – 5 h. To quench the reaction TEA was added (27 – 200 µL) at -55 °C. The temperature was then increased to rt, the mixture was diluted with DCM and washed subsequently with aq. $Na_2S_2O_3$, satd aq. $NaHCO_3$, and water. The organic phase was dried on Na_2SO_4, filtered and

concentrated. The residue was chromatographed on silica (PE/EtOAc gradient) to yield the fully protected disaccharides **10a-c**.

4-Methoxyphenyl (2,3,4-tri-*O*-benzyl-6-*O*-methyl-α-D-galactopyranosyl-(1→4)-2,3,6 -tri-*O*-benzoyl-β-D-galactopyranoside (11a):

7a (240 mg, 0.431 mmol). Reaction time: 1 h. Yield: 283 mg, 75%.

ESI-MS: *m/z*: Calcd for $C_{62}H_{60}NaO_{15}$ $[M+Na]^+$ 1067.4, found 1067.6.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.04 (d, *J* = 7.7 Hz, 2H, Ar), 7.96 (t, *J* = 8.1 Hz, 4H, Ar), 7.59 (t, *J* = 7.3 Hz, 1H, Ar), 7.45 (dd, *J* = 19.8, 7.4 Hz, 6H, Ar), 7.38 – 7.21 (m, 15H, Ar), 7.20 – 7.08 (m, 3H, Ar), 6.96 (d, *J* = 8.5 Hz, 2H, Ar), 6.65 (d, *J* = 8.5 Hz, 2H, Ar), 5.98 (t, *J* = 9.1 Hz, 1H, H-2), 5.31 (d, *J* = 10.5 Hz, 1H, H-3), 5.14 (d, *J* = 7.7 Hz, 1H, H-1), 5.01 – 4.76 (m, 7H, H-6a, H-6b, PhCH*H*, 2 $PhCH_2$), 4.72 (d, *J* = 11.7 Hz, 1H), 4.57 (d, *J* = 11.2 Hz, 1H, PhCH*H*), 4.45 (s, 1H, H-4), 4.36 – 4.27 (m, 1H, H-5'), 4.26 – 4.15 (m, 2H, H-3', H-5), 4.15 – 3.99 (m, 2H, H-2', H-4'), 3.69 (s, 3H, Ar-OCH_3), 3.23 (t, *J* = 8.7 Hz, 1H, H-6a'), 2.96 (s, 3H, OCH_3), 2.91 – 2.76 (m, 1H, H-6b').

^{13}C NMR (126 MHz, $CDCl_3$): δ = 166.43 (C=O), 166.10 (C=O), 165.36 (C=O), 155.56 (Ar), 151.31 (Ar), 138.94 (Ar), 138.75 (Ar), 138.40 (Ar), 133.34 (Ar), 133.25 (Ar), 133.20 (Ar), 130.08 (Ar), 129.89 (Ar), 129.81 (Ar), 129.69 (Ar), 129.58 (Ar), 129.11 (Ar), 128.53 (Ar), 128.48 (Ar), 128.41 (Ar), 128.34 (Ar), 128.30 (Ar), 128.20 (Ar), 128.18 (Ar), 127.59 (Ar), 127.48 (Ar), 118.83 (Ar), 114.44 (Ar), 101.09 (C-1), 100.00 (C-1'), 78.98, 77.38, 77.12, 76.87, 76.10, 75.47 (C-4), 75.04 (Ph*C*H_2), 74.66, 74.11, 74.05, 73.16, 72.58, 69.79 (C-6'), 69.62 (C-5'), 69.57 (C-2), 62.87 (C-6'), 60.42, 58.59 (OCH_3), 55.59 (Ar-OCH_3).

4-Methoxyphenyl (2,3,4-tri-*O*-benzyl-6-*O*-propyl-6-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (11b):

7b (90.0 mg, 0.154 mmol). Reaction time: 1 h. Yield: 73.8 mg, 54%.

$[\alpha]_D^{20}$ +26.0 (*c* 1.00, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{64}H_{64}NaO_{15}$ $[M+Na]^+$ 1095.4, found 1095.4.

^{1}H-NMR (500 MHz, $CDCl_3$): δ = 7.98 – 7.94 (m, 2H, Ar), 7.91 – 7.85 (m, 4H, Ar), 7.55 – 7.49 (m, 1H, Ar), 7.44 – 7.03 (m, 25H, Ar), 6.91 – 6.82 (m, 2H, Ar), 6.61 – 6.52 (m, 2H, Ar), 5.89 (dd, *J* = 10.5, 7.8 Hz, 1H, H-2), 5.20 (dd, *J* = 10.5, 3.0 Hz, 1H, H-3), 5.04 (d, *J* = 7.8 Hz, 1H, H-1), 4.87 – 4.66 (m, 7H, H-1', H-6a, H-6b, PhCH*H*, 2

$PhCH_2$), 4.62 (d, *J* = 11.8 Hz, 1H), 4.48 (d, *J* = 11.1 Hz, 1H, H-2'), 4.36 (d, *J* = 2.6 Hz, 1H, H-4), 4.23 (dd, *J* = 9.2, 5.5 Hz, 1H, H-5'), 4.16 – 3.96 (m, 4H, H-2', H-3', H-5, H-4'), 3.63 (s, 3H, OCH_3), 3.21 (t, *J* = 8.9 Hz, 1H, H-6'a), 2.92 (dt, *J* = 8.9, 7.0 Hz, 1H, $OCHCH_2CH_3$), 2.87 – 2.77 (m, 2H, H-6'b, $OCHCH_2CH_3$), 1.39 – 1.36 (m, 2H, $OCH_2CH_2CH_3$), 0.78 (t, *J* = 7.4 Hz, 3H, $OCH_2CH_2CH_3$).

^{13}C-NMR (126 MHz, $CDCl_3$): δ = 166.41 (C=O), 166.08 (C=O), 165.30 (C=O), 155.52 (Ar), 151.30 (Ar), 139.03 (Ar), 138.79 (Ar), 138.36 (Ar), 129.78 (Ar), 128.49 (Ar), 128.43 (Ar), 128.39 (Ar), 128.34 (Ar), 128.29 (Ar), 128.13 (Ar), 128.03 (Ar), 127.51 (Ar), 118.80 (Ar), 114.40 (Ar), 101.17 (C-1), 101.07 (C-1'), 78.89 (C-3'), 77.28, 77.03, 76.78, 76.09 (C-2'), 75.51 (C-4), 75.02 ($PhCH_2$), 74.80 ($PhCH_2$), 74.09, 73.13, 72.72, 72.51, 69.84 (C-5'), 69.58 (C-2), 67.66 (C-6'), 62.84, 55.58 (Ar-CH_3), 22.91 ($OCH_2CH_2CH_3$), 10.65 ($OCH_2CH_2CH_3$).

4-Methoxyphenyl [2,3,4-tri-*O*-benzyl-6-*O*-(2-methylpropyl)-α-D-galactopyranosyl]-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (11c):

7c (36.7 mg, 0.061 mmol). Reaction time: 3.5 h, at -55 °C; 1 h at -45 °C, then additional TMSOTf (1 µL, 0.050 mmol) at -55 °C and 1 h at -55 °C. Yield: 17.8 mg, 27%.

$[\alpha]_D^{20}$ +24.2 (*c* 1.05, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{33}H_{34}NaO_5S$ $[M+Na]^+$ 1109.4, found 1109.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.99 – 7.95 (m, 2H, Ar), 7.89 (d, *J* = 7.9 Hz, 4H, Ar), 7.53 (t, *J* = 7.4 Hz, 1H, Ar), 7.45 – 7.04 (m, 25H, Ar), 6.89 – 6.87 (m, 2H, Ar), 6.59 – 6.55 (m, 2H, Ar), 5.90 (dd, *J* = 10.5, 7.8 Hz, 1H, H-2), 5.19 (dd, *J* = 10.5, 2.9 Hz, 1H, H-3), 5.05 (d, *J* = 7.8 Hz, 1H, H-1), 4.85 - 4.81 (m, 3H, H-1', 2 PhC*H*H), 4.80 – 4.67 (m, 5H, H-6a, H-6b, 3 PhC*H*H), 4.63 (d, *J* = 11.7 Hz, 1H, PhC*H*H), 4.48 (d, *J* = 11.1 Hz, 1H, H-2'), 4.37 (d, *J* = 2.7 Hz, 1H, H-4), 4.24 (dd, *J* = 9.3, 5.4 Hz, 1H, H-5'), 4.17 – 4.06 (m, 2H, H-3', H-5), 4.05 – 3.98 (m, 2H, H-2', H-4'), 3.63 (s, 3H, OCH_3), 3.23 (t, *J* = 9.0 Hz, 1H, H-6a'), 2.82 (dd, *J* = 8.4, 5.0 Hz, 1H, H-6b'), 2.78 – 2.66 (m, 2H, isobutyl CH_2), 1.71 – 1.60 (m, 1H, isobutyl CH), 0.79 (dd, *J* = 19.1, 6.7 Hz, 6H, isobutyl $(CH_3)_2$).

^{13}C-NMR (126 MHz, $CDCl_3$): δ = 166.42 (C=O), 166.08 (C=O), 165.20 (C=O), 155.51 (Ar), 139.09 (Ar), 138.81 (Ar), 138.35 (Ar), 133.26 (Ar), 133.19 (Ar), 129.87 (Ar), 128.48 (Ar), 128.46 (Ar), 128.33 (Ar), 128.29 (Ar), 128.12 (Ar), 127.93 (Ar),

127.49 (Ar), 118.78 (Ar), 114.40 (Ar), 101.24 (C-1), 101.22 (C-1'), 78.99 (C-2'), 77.96 (O*C*H_2CH(CH_3)$_2$), 77.28, 77.02, 76.77, 76.11 (C-3'), 75.61 (C-4), 75.08, 75.01, 74.08, 73.15, 72.51, 69.80 (C-5'), 69.58, 67.76 (C-6'), 62.88 (C-6), 55.58 (Ar-OCH_3), 28.51 (isobutyl CH), 19.66 (isobutyl CH_3), 19.36 (isobutyl CH_3).

General procedure for the synthesis of compounds **12a-c**:

Cleavage of the benzoyl groups:

The fully protected disaccharides were dissolved in dry MeOH, at a final concentration of approximately 0.01 M. Freshly prepared 1 M MeONa in dry methanol was added in one portion (approx. 1% of the volume of the solution). After stirring at rt under argon for 20 – 40 h, Amberlite IR-120 was added to neutralize the solution. After filtration, the solution was concentrated. The residue was chromatographed on silica (PE/EtOAc gradient) to yield the tri-*O*-benzyl protected disaccharides **12a-c**.

4-Methoxyphenyl (2,3,4-tri-*O*-benzyl-6-*O*-methyl-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (12a):

11a (283 mg, 0.386 mmol). Reaction time 20 h. Yield: 167 mg, 84%.

$[\alpha]_D^{20}$ - 4.0 (*c* 1.0, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{41}H_{48}NaO_{12}$ $[M+Na]^+$ 755.3, found 755.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.45 – 7.27 (m, 15H, Ar), 6.99 (d, *J* = 8.2 Hz, 2H, Ar), 6.79 (d, *J* = 8.3 Hz, 2H, Ar), 4.92 (m, 2H, 2 PhC*H*H), 4.85 (s, 1H, H-1'), 4.77 (s, 2H, PhCH_2), 4.74 (d, *J* = 7.3 Hz, 1H, H-1), 4.70 (d, *J* = 11.8, 1H, PhC*H*H), 4.58 (d, *J* = 11.5 Hz, 1H, PhC*H*H), 4.18 – 4.08 (m, 2H, H-5', H-2'), 4.07 – 3.96 (m, 2H, H-3', H-4), 3.91 (s, 1H, H-4'), 3.83 – 3.69 (m, 7H, Ar-OCH_3, H-2, H-6a, H-6b), 3.62 (s, 1H), 3.56 – 3.41 (m, 2H, H-3, H-6'a), 3.30 (s, 3H, OCH_3), 3.25 (d, *J* = 9.4 Hz, 1H, H-6'b).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 155.39 (Ar), 151.31 (Ar), 138.16 (Ar), 138.10 (Ar), 137.14 (Ar), 128.77 (Ar), 128.60 (Ar), 128.59 (Ar), 128.49 (Ar), 128.40 (Ar), 128.19 (Ar), 127.89 (Ar), 127.84 (Ar), 127.43 (Ar), 118.51 (Ar), 114.49 (Ar), 102.57 (C-1), 101.02 (C-1'), 80.47, 79.12 (C-3'), 77.34, 77.09, 76.83, 75.75, 74.96, 74.68, 74.48 (C-

4'), 73.97, 73.84, 72.74, 71.99, 71.90 (C-6'), 71.38 ($PhCH_2$), 60.37 (C-6), 59.36 (OCH_3), 55.64 ($ArOCH_3$).

4-Methoxyphenyl (2,3,4-tri-*O*-benzyl-6-*O*-propyl-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (12b):

11b (70.7 mg, 0.066 mmol). Reaction time 40 h. Yield: 47.4 mg, 95%.

$[\alpha]_D^{20}$ -2.5 (*c* 1.0, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{43}H_{52}NaO_{12}$ $[M+Na]^+$ 783.4, found 783.3.

^{1}H-NMR (500 MHz, $CDCl_3$): δ = 7.43 – 7.33 (m, 15H, Ar), 7.08 – 7.01 (m, 2H, Ar), 6.93 – 6.83 (m, 2H, Ar), 4.97 (t, *J* = 11.6 Hz, 2H), 4.88 (d, *J* = 3.6 Hz, 1H, H-1'), 4.83 - 4.70 (m, 4H, H-1, H-1), 4.62 (d, *J* = 11.5 Hz, 1H), 4.24 – 4.11 (m, 2H, H-2', H-5'), 4.10 – 4.03 (m, 2H, H-4, H-3'), 3.96 (d, 1H, *J* = 1.8 Hz, H-4'), 3.88 – 3.72 (m, 7H, H-2, Ar-OCH_3, H-6a, H-6b, H-5), 3.60 – 3.50 (m, 2H, H-6'a, H-3), 3.45 – 3.27 (m, 3H, H-6'b, $CH_2CH_2CH_3$), 1.68 – 1.56 (m, 2H, $CH_2CH_2CH_3$), 0.97 - 0.91 (m, 3H, $CH_2CH_2CH_3$).

^{13}C-NMR (126 MHz, $CDCl_3$): δ = 155.38 (Ar), 151.30 (Ar), 138.22 (Ar), 138.14 (Ar), 137.15 (Ar), 128.77 (Ar), 128.62 (Ar), 128.58 (Ar), 128.49 (Ar), 128.39 (Ar), 128.19 (Ar), 127.87 (Ar), 127.84 (Ar), 127.44 (Ar), 118.53 (Ar), 114.49 (Ar), 102.54 (C-1) 101.12 (C-1'), 80.50 (C-4), 79.11 (C-3'), 77.83, 77.12, 76.87, 75.78 (C-2'), 74.95 ($PhCH_2$), 74.70 (C-4'), 74.58 ($PhCH_2$), 74.05 (C-3), 73.85, 73.43 ($CH_2CH_2CH_3$), 72.74, 71.99, 71.51 (C-5'), 69.95 (C-6'), 60.34, 55.64 ($ArOCH_3$), 22.85 ($CH_2CH_2CH_3$), 10.48 ($CH_2CH_2CH_3$).

4-Methoxyphenyl [2,3,4-tri-*O*-benzyl-6-*O*-(2-methylpropyl)-α-D-galactopyranosyl]-(1→4)-β-D-galactopyranoside (12c):

11c (17.8 mg, 0.016 mmol). Reaction time 24 h. Yield: 11.9 mg, 96%.

$[\alpha]_D^{20}$ -2.8 (*c* 1.0, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{26}H_{34}NaO_{12}$ $[M+Na]^+$ 797.4, found 797.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.46 – 7.30 (m, 15H, Ar), 7.03 – 7.01 (m, 2H, Ar), 6.84 – 6.82 (m, 2H, Ar), 4.96 (t, *J* = 11.4 Hz, 2H), 4.88 (d, *J* = 3.7 Hz, 1H, H-1'), 4.81 (s, 2H, $PhCH_2$), 4.78 (d, *J* = 7.5 Hz, 1H, H-1), 4.78 (d, *J* = 11.5 Hz, 1H, PhC*H*H), 4.61 (d, *J* = 11.4 Hz, 1H, PhC*H*H), 4.18 (dd, *J* = 6.8, 3.9 Hz, 2H, H-5', H-2'), 4.11 – 4.04 (m, 2H, H-3', H-4), 3.96 (d, *J* = 1.7 Hz, 1H, H-4'), 3.89 – 3.71 (m, 7H, $ArOCH_3$,

H-2, H-6a, H-6b, H-5), 3.59 – 3.50 (m, 2H, H-6'a, H-3), 3.31 (dd, J = 9.6, 4.2 Hz, 1H, H-6'b), 3.16 (m, 2H, isobutyl CH_2), 1.92 – 1.80 (m, 1H, isobutyl CH), 0.96 – 0.88 (m, 6H, isobutyl $(CH_3)_2$).

^{13}C-NMR (126 MHz, $CDCl_3$): δ = 155.38 (Ar), 151.26 (Ar), 138.19 (Ar), 138.09 (Ar), 137.07 (Ar), 128.76 (Ar), 128.63 (Ar), 128.57 (Ar), 128.51 (Ar), 128.37 (Ar), 128.15 (Ar), 127.83 (Ar), 127.41 (Ar), 118.48 (Ar), 114.47 (Ar), 102.48 (C-1), 101.23 (C-1'), 80.41 (C-4), 79.04, 78.69 (isobutyl CH_2), 77.28, 77.03, 76.77, 75.50 (C-2'), 74.99 ($PhCH_2$), 74.74 ($PhCH_2$), 74.65 (C-4'), 74.01 (C-3'), 73.82, 72.72 ($PhCH_2$), 72.08, 71.56 (C-5'), 70.15 (C-6'), 60.24, 55.64 ($ArOCH_3$), 29.71, 28.43 (isobutyl CH), 19.41 (isobutyl CH_3), 19.32 (isobutyl CH_3).

General procedure for the synthesis of compounds **13a-c**:

Cleavage of the benzyl groups:

The benzyl-protected disaccharides were dissolved in AcOH under argon, Pd on charcoal (5-10% Pd) was added (2:1 disaccharide/Pd-C) and the mixture was hydrogenated (1 atm H_2, balloon) with stirring. The reaction was monitored by TLC (DCM/MeOH, 9:1). After 2-13 h the mixture was diluted with MeOH and filtered through celite. The filtrate was concentrated and the residue chromatographed on silica (DCM/MeOH gradient) to yield compounds **13a-c**.

4-Methoxyphenyl (6-*O*-methyl-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (13a):

12a (160 mg, 0.218 mmol). Reaction time: 13 h. Yield: 86 mg, 85%.

$[\alpha]_D^{20}$ +8.0 (c 1.0, $CHCl_3$).

ESI-MS: m/z: Calcd for $C_{20}H_{30}NaO_{12}$ $[M+Na]^+$ 485.2, found 485.2.

^{1}H NMR (500 MHz, CD_3OD): δ = 7.08 (d, J = 8.5 Hz, 2H, Ar), 6.86 (d, J = 8.5 Hz, 2H, Ar), 5.01 (s, 1H, H-1'), 4.83 (d, J = 7.6 Hz, 1H, H-1), 4.44 (t, J = 6.0 Hz, 1H, H-5'), 4.07 (s, 1H, H-4), 3.92 (s, 1H, H-4'), 3.91 – 3.72 (m, 9H, H-2, H-2', $ArOCH_3$, H-6a, H-6b, H-5, H-3'), 3.65 (d, J = 10.1 Hz, 1H, H-3), 3.60 (d, J = 5.9 Hz, 2H, H-6'a, H-6'b), 3.42 (s, 3H, OCH_3).

^{13}C NMR (126 MHz, CD_3OD): δ = 156.79 (Ar), 153.00 (Ar), 119.31 (Ar), 115.55 (Ar), 104.08 (C-1), 102.71 (C-1'), 79.60 (C-4), 76.22, 74.60 (C-3), 72.98 (C-6'),

72.75 (C-2'), 71.28 (C-5'), 71.13 (C-4'), 71.03, 70.68, 61.13 (C-6), 59.57 (OCH_3), 56.11 ($ArOCH_3$).

4-Methoxyphenyl (6-*O*-propyl-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (13b):

12b (47.4 mg, 0.062 mmol). Reaction time: 2 h. Yield: 24.4 mg, 81%.

$[\alpha]_D^{20}$ +5.8 (*c* 1.0, MeOH).

ESI-MS: *m/z*: Calcd for $C_{22}H_{34}NaO_{12}$ $[M+Na]^+$ 513.2, found 513.1.

^{1}H-NMR (500 MHz, CD_3OD): δ = 7.08 (m, 2H, Ar), 6.87 (m, 2H, Ar), 5.00 (d, *J* = 1.9 Hz, 1H, H-1'), 4.83 (d, *J* = 7.6 Hz, 1H, H-1), 4.44 (t, *J* = 6.3 Hz, 1H, H-5'), 4.07 (d, *J* = 3.0 Hz, 1H, H-4), 3.95 (s, 1H, H-4'), 3.92 – 3.70 (m, 9H, H-2, $ArOCH_3$, H-2', H-6a, H-6b, H-5, H-3'), 3.68 – 3.59 (m, 3H, H-3, H-6'a, H-6'b), 3.49 (t, *J* = 6.4 Hz, 2H, $OCH_2CH_2CH_3$), 1.70 – 1.59 (m, 2H, $OCH_2CH_2CH_3$), 0.96 (t, *J* = 7.4 Hz, 3H, $OCH_2CH_2CH_3$).

^{13}C-NMR (126 MHz, CD_3OD): δ = 156.79 (Ar), 153.00 (Ar), 119.31 (Ar), 115.53 (Ar), 104.11 (C-1), 102.73 (C-1'), 79.50 (C-4), 76.24, 74.59 (C-3), 74.42, 72.78 (C-2), 71.36 (C-5'), 71.14 (C-2'), 71.03, 70.83, 70.72 (C-6'), 61.07, 56.09, 49.55, 49.38, 49.21, 49.04, 48.87, 48.70, 48.53, 23.85 ($OCH_2CH_2CH_3$), 10.91 ($OCH_2CH_2CH_3$).

4-Methoxyphenyl [6-*O*-(2-methylpropyl)-α-D-galactopyranosyl]-(1→4)-β-D-galactopyranoside (13c):

12c (11.9 mg, 0.015 mmol). Reaction time: 2 h. Yield: 6 mg, 79%.

$[\alpha]_D^{20}$ +2.6 (*c* 1.0, MeOH).

ESI-MS: *m/z*: Calcd for $C_{23}H_{36}NaO_{12}$ $[M+Na]^+$ 527.2, found 527.1.

^{1}H NMR (500 MHz, CD_3OD): δ = 7.11 – 7.05 (m, 2H, Ar), 6.89 – 6.82 (m, 2H, Ar), 5.01 (s, 1H, H-1'), 4.83 (d, *J* = 7.5 Hz, 1H, H-1), 4.46 (t, *J* = 6.5 Hz, 1H, H-5'), 4.07 (d, *J* = 3.0 Hz, 1H, H-4), 3.97 (d, *J* = 1.0 Hz, 1H, H-4'), 3.91 – 3.71 (m, 9H, H-2, H-2', H-3', H-6a, H-6b, H-5, $ArOCH_3$), 3.69 – 3.58 (m, 3H, H-6'a, H-6'b), 3.30 (d, *J* = 6.7 Hz, 2H, isobutyl CH_2), 1.91 (dp, *J* = 13.4, 6.7 Hz, 1H, isobutyl CH), 0.95 (d, *J* = 0.6 Hz, 3H, isobutyl CH_3), 0.94 (d, *J* = 0.6 Hz, 3H, isobutyl CH_3).

^{13}C-NMR (126 MHz, CD_3OD): δ = 156.80 (Ar), 153.00 (Ar), 119.30 (Ar), 115.51 (Ar), 104.10 (C-1), 102.70 (C-1'), 79.68 (isobutyl CH_2), 79.24 (C-4), 76.29, 74.57 (C-3), 72.79 (C-2), 71.37, 71.02, 70.95 (C-6'), 70.74 (C-5'), 61.00 (C-6), 56.07

($ArOCH_3$), 49.53, 49.36, 49.10, 49.02, 48.85, 48.68, 48.51, 29.54 (isobutyl CH), 19.76 (isobutyl CH_3).

Phenyl 2,3,4-tri-*O*-allyl-6-*O*-*tert*-butyldiphenylsilyl-1-thio-β-D-galactopyranoside (14):

To a cooled solution (0 °C) of AllBr (0.206 mL, 2.38 mmol) in dry DMF (3 mL) under argon, a 60% NaH dispersion in oil (95.2 mg, 2.38 mmol) was added, followed by a solution of **4** (320 mg, 0.627 mmol) in dry DMF (3 mL). The temperature was increased to rt. The reaction was monitored by TLC (PE/EtOAc, 1:1). After 1 h, the reaction was quenched with MeOH. The mixture was neutralized with a solution of 10% AcOH in MeOH, diluted with DCM, and washed with water. The water phase was extracted three times with DCM. The organic phases were collected and extracted twice with a 1:1 mixture of water and brine, dried over Na_2SO_4, filtered and concentrated. The residue was purified by chromatography on silica gel (PE/EtOAc gradient) to yield **14** (288.3 mg, 73%).

$[\alpha]_D^{20}$ +1.8 (*c* 1.0, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{37}H_{46}NaO_5SSi$ $[M+Na]^+$ 653.3, found 653.2.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.67 – 7.62 (m, 4H, Ar), 7.51 – 7.47 (m, 2H, Ar), 7.45 – 7.33 (m, 6H, Ar), 7.23 – 7.15 (m, 3H, Ar), 6.00 – 5.80 (m, 3H, 3 CH_2CHCH_2O), 5.33 – 5.06 (m, 6H, 3 CH_2CHCH_2O), 4.48 (d, *J* = 9.7 Hz, 1H, H-1), 4.38 – 4.36 (m, 1H, CH_2CHCHO), 4.27 – 4.16 (m, 4H, 2 CH_2CHCHO), 4.12 – 4.07 (m, 1H, CH_2CHCHO), 3.89 (d, *J* = 1.8 Hz, 1H, H-4), 3.86 (d, *J* = 7.7 Hz, 1H, H-6a), 3.81 – 3.78 (m, 1H, H-6b), 3.65 (t, *J* = 9.4 Hz, 1H, H-2), 3.42 (dd, *J* = 7.1, 6.2 Hz, 1H, H-5), 3.36 (dd, *J* = 9.2, 2.8 Hz, 1H, H-3), 1.07 (d, *J* = 5.2 Hz, 9H, $C(CH_3)_3$).

^{13}C-NMR (126 MHz, $CDCl_3$): δ = 135.58 (CH_2CHCH_2O), 135.56 (Ar), 135.07 (CH_2CHCH_2O), 134.93 (CH_2CHCH_2O), 134.51 (Ar), 133.29 (Ar), 131.17 (Ar), 129.77 (Ar), 128.70 (Ar), 127.75 (Ar), 127.74 (Ar), 126.83 (Ar), 116.89 (CH_2CHCH_2O), 116.68 (CH_2CHCH_2O), 116.17 (CH_2CHCH_2O), 87.76 (C-1), 83.77 (C-3), 78.68 (C-5), 74.44 (CH_2CHCH_2O), 73.55 (CH_2CHCH_2O), 73.13 (C-4), 72.74 (C-2), 71.64 (CH_2CHCH_2O), 62.24 (C-6), 26.90 ($C(CH_3)_3$).

Phenyl 2,3,4-tri-*O*-allyl-1-thio-β-D-galactopyranoside (15):

A solution of TBAF (1M in THF, 1.363 mL, 4.709 mmol) was neutralized with glacial acetic acid, cooled down to 0 °C and added to **14** (288 mg, 0.457 mmol). The

reaction was monitored with TLC (PE/EtOAc, 1:1). After 20 h, the reaction was complete. The product was diluted with EtOAc and washed twice with a mixture of brine and water (1:1). The organic phases were collected, dried over Na_2SO_4, filtered and concentrated. The residue was purified by chromatography on silica gel (PE/EtOAc gradient) to yield **15** (139.5 mg, 78%).

$[\alpha]_D^{20}$ +2.3 (*c* 0.80, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{21}H_{28}NaO_5S$ $[M+Na]^+$ 415.2, found 414.8.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.57 – 7.51 (m, 2H, Ar), 7.30 – 7.18 (m, 3H, Ar), 6.02 – 5.82 (m, 3H, 3 CH_2CHCH_2O), 5.34 – 5.12 (m, 6H, 3 CH_2CHCH_2O), 4.57 (d, *J* = 9.7 Hz, 1H, H-1), 4.37 (ddt, *J* = 12.8, 5.1, 1.4 Hz, 1H, CH_2CHCHO), 4.31 – 4.20 (m, 2H, 2 CH_2CHCHO), 4.19 – 4.16 (m, 2H, CH_2CHCH_2O), 4.11 – 4.04 (m, 1H, CH_2CHCHO), 3.90 (dd, *J* = 11.4, 7.3 Hz, 1H, H-6a), 3.76 (d, *J* = 2.3 Hz, 1H, H-4), 3.68 (m, 2H, H-2, H-6b), 3.49 – 3.44 (m, 1H, H-5), 3.40 (dd, *J* = 9.2, 2.9 Hz 1H, H-3).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 135.13 (CH_2CHCH_2O), 134.93 (CH_2CHCH_2O), 134.70 (CH_2CHCH_2O), 134.07 (Ar), 131.51 (Ar), 128.84 (Ar), 127.18 (Ar), 117.27 (CH_2CHCH_2O), 117.07 (CH_2CHCH_2O), 116.91 (CH_2CHCH_2O), 87.67 (C-1), 83.61 (C-3), 78.85 (C-5), 74.50 (CH_2CHCH_2O), 73.75(C-4), 73.51 (CH_2CHCH_2O), 71.77 (CH_2CHCH_2O), 62.54 (C-6).

Phenyl 2,3,4-tri-*O*-allyl-6-*O*-benzyl-1-thio-β-D-galactopyranoside (16):

A dispersion of 60% NaH in oil (25.7 mg, 0.642 mmol) was suspended in dry DMF (2 mL) at 0 °C was diluted with EtOAc and washed twice with a mixture of brine and water (1:1)**15** (126 mg, 0.321 mmol) in dry DMF (1.5 mL). The mixture was allowed to warm to rt. The reaction was monitored by TLC (PE/EtOAc, 1:1). After 1 h, the reaction was quenched with MeOH. The solvents were evaporated and the residue taken up in DCM. The organic phase was washed with water. The water phase was extracted three times with DCM. The organic phases were collected, dried over Na_2SO_4, filtered and evaporated. The residue was purified by chromatography on silica gel (PE/EtOAc gradient) to yield **16** (118.8 mg, 77%).

$[\alpha]_D^{20}$ -5.9 (*c* 0.9, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{28}H_{34}NaO_5S$ $[M+Na]^+$ 505.2, found 505.1.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.56 – 7.52 (m, 2H, Ar), 7.37 – 7.27 (m, 6H, Ar), 7.27 – 7.18 (m, 3H, Ar), 6.02 – 5.82 (m, 3H, 3 CH_2CHCH_2O), 5.38 – 5.07 (m, 6H, 3 CH_2CHCH_2O), 4.60 – 4.48 (m, 3H, H-1, $PhCH_2$), 4.34 (dd, *J* = 12.8, 5.3 Hz, 1H,

$CH_2CHCHHO$), 4.29 – 4.03 (m, 5H, 5 $CH_2CHCHHO$), 3.83 (d, *J* = 2.7 Hz, 1H, H-4), 3.75 – 3.62 (m, 3H, H-6a and 6b, H-2), 3.60 – 3.54 (m, 1H, H-5), 3.38 (dd, *J* = 9.2, 2.9 Hz, 1H, H-3).

^{13}C-NMR (126 MHz, $CDCl_3$): δ = 137.92 (Ar), 135-46 (CH_2CHCH_2O), 135.02 (CH_2CHCH_2O), 134.81 (CH_2CHCH_2O), 134.41 (Ar), 131.45 (Ar), 128.74 (Ar), 128.45 (Ar), 127.94 (Ar), 127.82 (Ar), 127.00 (Ar), 116.99 (CH_2CHCH_2O), 116.74 (CH_2CHCH_2O), 116.52 (CH_2CHCH_2O), 87.89 (C-1), 83.52 ($PhCH_2$), 77.37 (C-5), 74.51 (CH_2CHCH_2O), 73.68 (CH_2CHCH_2O), 73.64 (CH_2CHCH_2O), 73.31 (C-4), 71.57, 68.75 (C-6).

4-Methoxyphenyl (2,3,4-tri-*O*-allyl-6-*O*-benzyl-α-D-galactopyranosyl)-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (17):

Compound **10** (122 mg, 0.203 mmol), **16** (118 mg, 0.244 mmol) and NIS (54.9 mg, 0.244 mmol) were dissolved in dry DCM (2.2 mL), then dry Et_2O (4.3 mL) was added, and the mixture was cooled to -55 °C. TMSOTf (6.0 µL, 0.035 mmol) was then added, and the mixture stirred at the same temperature. The reaction was monitored by TLC (PE/EtOAc, 4:1). After 1 h the temperature was raised to -45 °C and after 2 h the temperature was increased to -30 °C. After 4 h TEA (0.120 mL) was added, and the reaction was stirred for 1 h at -30 °C. The temperature was increased to rt. The mixture was diluted with DCM, subsequently washed with aq. $Na_2S_2O_3$, aq. $NaHCO_3$ and water, dried over Na_2SO_4, filtered and concentrated. The residue was purified by chromatography on silica gel (PE/EtOAc gradient) to yield **17** (77.5 mg, 39%).

$[\alpha]_D^{20}$ +28.5 (*c* 1.0, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{56}H_{58}NaO_{15}$ $[M+Na]^+$ 993.4, found 993.5.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.03 – 7.95 (m, 2H, Ar), 7.89 (dt, *J* = 8.4, 4.2 Hz, 4H, Ar), 7.53 (dd, *J* = 10.5, 4.3 Hz, 1H, Ar), 7.47 – 7.38 (m, 3H, Ar), 7.35 (t, *J* = 7.4 Hz, 1H, Ar), 7.32 – 7.14 (m, 12H, Ar), 6.96 – 6.85 (m, 2H, Ar), 6.65 – 6.53 (m, 2H, Ar), 5.95 – 5.70 (m, 4H, 3 CH_2CHCH_2O, H-2), 5.31 – 4.92 (m, 8H, 3 CH_2CHCH_2O, H-1, H-3), 4.88 (d, *J* = 3.6 Hz, 1H, H-1'), 4.78 (d, *J* = 6.6 Hz, 2H, H-6a, H-6b), 4.42 (d, *J* = 2.7 Hz, 1H, H-4), 4.29 – 4.17 (m, 3H, H-5', 2 $CH_2CHCHHO$), 4.17 – 4.01 (m, 6H, 3 $CH_2CHCHHO$, H-5, $PhCH_2$), 3.95 (dd, *J* = 12.4, 6.0 Hz, 1H, $CH_2CHCHHO$), 3.89 – 3.81 (m, 2H, H-3', H-4'), 3.76 (dd, *J* = 10.0, 3.5 Hz, 1H, H-2'), 3.64 (s, 3H, $ArOCH_3$), 3.41 – 3.29 (m, 1H, H-6'a), 2.97 (dd, *J* = 8.3, 4.9 Hz, 1H, H-6'b).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 166.67 (C=O), 166.27 (C=O), 165.46 (C=O), 155.66 (Ar), 151.45 (Ar), 138.54 (Ar), 135.72 (CH_2*C*HCH_2O), 135.23 (CH_2*C*HCH_2O), 135.10 (CH_2*C*HCH_2O), 133.36 (Ar), 130.16 (Ar), 129.88 (Ar), 129.80 (Ar), 128.65 (Ar), 128.59 (Ar), 128.53 (Ar), 128.41 (Ar), 127.70 (Ar), 127.56 (Ar), 118.94 (Ar), 117.39 (*C*H_2CHCH_2O), 116.72 (*C*H_2CHCH_2O), 116.27 (*C*H_2CHCH_2O), 114.55 (Ar), 101.47 (C-1'), 101.23 (C-1), 78.27, 77.89 (C-3'), 76.48 (C-2'), 76.04, 75.84 (C-4), 74.46 (CH_2CH*C*H_2O), 74.35 (C-4'), 74.27 (C-3), 73.34, 71.41, 69.83 (CH_2CH*C*H_2O), 69.68, 67.51 (C-6'), 63.06 ($PhCH_2$), 55.72 ($ArOCH_3$).

4-Methoxyphenyl 6-*O*-benzyl-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (18):

DMBA (38.4 mg, 0.246 mmol), $Pd(OAc)_2$ (1.8 mg, 0.008 mmol) and PPh_3 (6.6 mg, 0.025 mmol) were mixed in a reaction tube. A solution of **17** (39.8 mg, 0.041 mmol) in dry MeOH (0.3 mL) and dry DCM (0.1 mL) was added. The reaction was stirred at rt, monitoring by TLC (PE/EtOAc, 1:1). After 2 h the temperature was increased to 40 °C. After 3 h the mixture was diluted with EtOAc and washed once with satd aq. $NaHCO_3$. The water phase was extracted twice with EtOAc. The organic phases were collected and washed once with brine, dried over Na_2SO_4, filtered and concentrated. The residue was purified by chromatography on silica gel (PE/EtOAc gradient) to yield **18** (27.1 mg, 78%).

$[\alpha]_D^{20}$ +21.8 (*c* 1.00, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{47}H_{46}NaO_{15}$ $[M+Na]^+$ 873.3, found 873.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.10 – 8.05 (m, 2H, Ar), 7.99 – 7.90 (m, 4H, Ar), 7.71 – 7.65 (m, 1H, Ar), 7.64 – 7.54 (m, 2H, Ar), 7.53 – 7.43 (m, 5H, Ar), 7.41 – 7.24 (m, 7H), 7.15 – 7.09 (m, 2H, Ar), 6.98 – 6.92 (m, 2H, Ar), 6.72 – 6.65 (m, 2H, Ar), 5.93 (dd, *J* = 10.5, 7.8 Hz, 1H, H-2), 5.33 (dd, *J* = 10.5, 3.0 Hz, 1H, H-3), 5.16 (d, *J* = 7.8 Hz, 1H, H-1), 5.10 (d, *J* = 3.1 Hz, 1H, H-1'), 4.89 (dd, *J* = 11.4, 7.4 Hz, 1H, H-6a), 4.75 (dd, *J* = 11.4, 6.4 Hz, 1H, H-6b), 4.56 (d, *J* = 2.9 Hz, 1H, H-4), 4.26 – 4.16 (m, 3H, H-5', PhCH*H*, H-5), 4.15 – 4.13 (m, 1H, H-4'), 4.03 (d, *J* = 12.0 Hz, 1H, PhC*H*H), 4.00 – 3.94 (m, 2H, H-2', H-3'), 3.72 (s, 3H, $ArOCH_3$), 3.07 (dd, *J* = 10.2, 5.0 Hz, 1H, H-6'a), 3.00 (dd, *J* = 10.2, 3.7 Hz, 1H, H-6'b).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 166.21 (C=O), 166.11 (C=O), 165.61 (C=O), 155.87 (Ar), 151.18 (Ar), 137.40 (Ar), 133.71 (Ar), 133.57 (Ar), 132.29 (Ar), 129.99

(Ar), 129.83 (Ar), 128.77 (Ar), 128.71 (Ar), 128.61 (Ar), 128.55 (Ar), 127.96 (Ar), 127.68 (Ar), 127.18 (Ar), 119.06 (Ar), 114.62 (Ar), 101.23 (C-1), 100.87 (C-1'), 74.09 (C-4), 73.92, 73.65 ($PhCH_2$), 72.99 (C-5), 70.89 (C-4'), 70.60, 69.98 (C-6'), 69.54, 69.12 (C-5'), 62.05 (C-6), 55.73 ($ArOCH_3$).

4-Methoxyphenyl 6-*O*-benzyl-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (19):

Compound **18** (53.6 mg, 0.063 mmol) was diluted in MeOH (7.5 mL) and freshly prepared 1 M MeONa/MeOH (59 µL) was added. The reaction was stirred at rt, monitoring by TLC (DCM/MeOH, 9:1). After 18 h, an additional aliquot of MeONa (30 µL) was added. After 24 h 10% AcOH in MeOH was added to neutralize the mixture. The mixture was evaporated and the residue chromatographed on silica gel (DCM/MeOH gradient) to yield **19** (22.7 mg, 67%).

$[\alpha]_D^{20}$ +3.4 (*c* 1.0, MeOH).

ESI-MS: *m/z*: Calcd for $C_{26}H_{34}NaO_{12}$ $[M+Na]^+$ 561.2, found 561.2.

^{1}H NMR (500 MHz, CD_3OD): δ = 7.41 (d, *J* = 7.3 Hz, 2H, Ar), 7.36 (t, *J* = 7.5 Hz, 2H, Ar), 7.29 (t, *J* = 7.3 Hz, 1H, Ar), 7.10 – 7.06 (m, 2H, Ar), 6.88 – 6.84 (m, 2H, Ar), 5.02 (d, *J* = 1.8 Hz, 1H, H-1'), 4.83 (d, *J* = 7.6 Hz, 1H, H-1), 4.62 (s, 2H, $PhCH_2$), 4.52 (t, *J* = 6.3 Hz, 1H, H-5'), 4.08 (d, *J* = 3.0 Hz, 1H, H-4), 3.95 (s, 1H, H-4'), 3.89 (dd, *J* = 10.1, 6.3 Hz, 1H, H-6a), 3.85 – 3.75 (m, 8H, H-2', H-2, H-3', H-5, H-6b, $ArOCH_3$), 3.70 – 3.63 (m, 3H, H-6'a, H-6'b, H-3).

^{13}C-NMR (126 MHz, CD_3OD): δ = 156.80 (Ar), 152.96 (Ar), 139.67 (Ar), 129.39 (Ar), 128.95 (Ar), 128.66 (Ar), 119.32 (Ar), 115.50 (Ar), 104.07 (C-1), 102.67 (C-1'), 79.37 (C-4), 76.25 (C-3'), 74.62 (C-3), 74.43 ($PhCH_2$), 72.74 (C-2), 71.29 (C-5'), 71.20 (C-4'), 71.01 (C-2'), 70.69 (C-5), 70.34 (C-6'), 61.05 (C-6), 56.05 ($ArOCH_3$).

4-Methoxyphenyl (2,3,4-tri-*O*-allyl-6-*O*-*tert*-butyldiphenylsilyl-α-D-galactopyranosyl)-(1→4)- 2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (20):

Compound **14** (518.0 mg, 0.821 mmol), **10** (590 mg, 0.985 mmol), NIS (222 mg, 0.985 mmol) and activated MS-4Å (20 mg) were mixed in dry DCM (18.6 mL) and stirred at rt for 30 min. The mixture was then cooled to -42 °C and TMSOTf (25 µL, 0.14 mmol) was added. The reaction was monitored by TLC (PE/EtOAc, 2:8). After 1 h the temperature was increased to -30 °C. After 3 h the reaction was quenched with

TEA at -40 °C, then warmed to rt, diluted with DCM, washed with 10% aq. $Na_2S_2O_3$, aq. $NaHCO_3$ and water. The organic phases were collected, dried over Na_2SO_4, filtered and concentrated. The residue was flash chromatographed on silica gel (PE/EtOAc gradient) to yield **20** (36 mg, 74%).

$[\alpha]_D^{20}$ +22.4 (*c* 1.00, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{65}H_{70}NaO_{15}Si$ $[M+Na]^+$ 1141.5, found 1141.2.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.10 (d, *J* = 7.4 Hz, 2H, Ar), 7.96 (d, *J* = 7.4 Hz, 2H, Ar), 7.83 (d, *J* = 7.5 Hz, 2H, Ar), 7.72 (d, *J* = 6.7 Hz, 2H, Ar), 7.67 (d, *J* = 6.5 Hz, 2H, Ar), 7.60 (t, *J* = 7.4 Hz, 1H, Ar), 7.50 (dt, *J* = 21.2, 7.6 Hz, 4H, Ar), 7.44 – 7.26 (m, 9H, Ar), 6.97 (d, *J* = 9.0 Hz, 2H, Ar), 6.92 (t, *J* = 7.8 Hz, 2H, Ar), 6.63 (d, *J* = 9.0 Hz, 2H, Ar), 6.03 – 5.91 (m, 2H, $CH_2C\mathit{H}CH_2O$, H-2), 5.91 – 5.71 (m, 2H, 2 $CH_2C\mathit{H}CH_2O$), 5.37 (dd, *J* = 17.2, 1.6 Hz, 1H, $CH\mathit{H}CHCH_2O$), 5.26 (dd, *J* = 10.3, 2.8 Hz, 1H, H-3), 5.20 – 5.18 (m, 1H), 5.18 – 5.15 (m, 2H, 2 $C\mathit{H}HCHCH_2O$), 5.11 – 5.03 (m, 2H, 2 $C\mathit{H}HCHCH_2O$), 5.00 (d, *J* = 10.3 Hz, 1H, $C\mathit{H}HCHCH_2O$), 4.90 – 4.82 (m, 3H, H-1', H-6a, H-6b), 4.52 (d, *J* = 2.0 Hz, 1H, H-4), 4.39 – 4.32 (m, 2H, $CH_2CHCH\mathit{H}O$, H-5'), 4.30 – 4.18 (m, 4H, 3 $CH_2CHCH\mathit{H}O$, H-5), 4.11 – 3.99 (m, 3H, 2 $CH_2CHCH\mathit{H}O$, H-4'), 3.94 (dd, *J* = 10.2, 2.5 Hz, 1H, H-3'), 3.81 – 3.74 (m, 2H, H-2', H-6'a), 3.70 (s, 3H, $ArOCH_3$), 3.64 (m, 1H, H-6'b), 1.10 (s, 9H, $C(CH_3)_3$).

^{13}C NMR (126 MHz, CDCl3): δ = 166.60 (C=O), 166.21 (C=O), 164.84 (C=O), 155.45 (Ar), 151.42 (Ar), 135.80, 135.77 ($CH_2\mathit{C}HCH_2O$), 135.64 ($CH_2\mathit{C}HCH_2O$), 135.23 ($CH_2\mathit{C}HCH_2O$), 134.96 (Ar), 133.47 (Ar), 133.42 (Ar), 133.11 (Ar), 130.14 (Ar), 130.02 (Ar), 129.75 (Ar), 129.67 (Ar), 129.59 (Ar), 129.55 (Ar), 128.90 (Ar), 128.44 (Ar), 128.36 (Ar), 128.17 (Ar), 127.77 (Ar), 127.66 (Ar), 118.53 (Ar), 117.29 ($\mathit{C}H_2CHCH_2O$), 116.35 ($\mathit{C}H_2CHCH_2O$), 116.03 ($\mathit{C}H_2CHCH_2O$), 114.41 (Ar), 101.87 (C-1), 100.99 (C-1'), 78.49 (C-3'), 77.10 (C-2'), 76.83 (C-4), 74.57 ($CH_2CH\mathit{C}H_2O$), 74.20 (C-3), 73.56 (C-4'), 73.02 ($CH_2CH\mathit{C}H_2O$), 71.79 (C-5), 71.50 ($CH_2CH\mathit{C}H_2O$), 69.72 (C-2), 64.39 (C-6), 60.95 (C-6'), 55.56 ($ArOCH_3$), 27.00 ($C(\mathit{C}H_3)_3$), 19.14 ($\mathit{C}(CH_3)_3$).

4-Methoxyphenyl (2,3,4-tri-*O*-allyl-6-*O*-*tert*-butyldiphenylsilyl-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (21):

Compound **20** (465 mg, 0.415 mmol) was dissolved in dry MeOH (3 mL) and freshly-prepared 1 M MeONa/MeOH (30 µL) was added. The reaction was stirred at rt and

monitored by TLC (PE/EtOAc, 1:1). After 4 h additional MeONa (30 µL) was added. After 24 h the mixture was diluted with MeOH and neutralized with 10% AcOH in MeOH. The mixture was concentrated and the residue was chromatographed on silica (PE/EtOAc gradient) to yield **21** (242.7 mg, 73%).

$[\alpha]_D^{20}$ +3.3 (*c* 1.0, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{44}H_{58}NaO_{12}$ Si $[M+Na]^+$ 829.4, found 829.5.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.74 – 7.64 (m, 4H, Ar), 7.50 – 7.39 (m, 6H, Ar), 7.03 (d, *J* = 9.0 Hz, 2H Ar), 6.83 (d, *J* = 9.0 Hz, 2H, Ar), 6.04 – 5.70 (m, 3H, 3 CH_2CHCH_2O), 5.37 – 5.32 (m, 1H, $CHHCHCH_2O$), 5.31 – 5.26 (m, 1H, $CHHCHCH_2O$), 5.23 (d, *J* = 10.3 Hz, 1H, $CHHCHCH_2O$), 5.21 – 5.13 (m, 2H, 2 $CHHCHCH_2O$), 5.11 – 5.06 (m, 1H, $CHHCHCH_2O$), 5.04 (d, *J* = 3.6 Hz, 1H, H-1'), 4.73 (d, *J* = 7.3 Hz, 1H, H-1), 4.45 – 4.32 (m, 2H, CH_2CHCH_2O), 4.23 – 4.11 (m, 4H, CH_2CHCH_2O, $CH_2CHCHHO$, H-5'), 4.10 – 4.03 (m, 2H, $CH_2CHCHHO$, H-4), 4.00 – 3.90 (m, 2H, H-2', H-4), 3.89 – 3.73 (m, 9H, H-3', $ArOCH_3$, H-6'a, H-6'b, H-6a, H-6b, H-5), 3.59 (dd, *J* = 9.9, 7.5 Hz, 1H, H-2), 3.55 – 3.48 (m, 1H, H-3), 1.09 (s, 9H, $C(CH_3)_3$).

^{13}C NMR (126 MHz, CDCl3): δ = 135.67 (Ar), 135.63 (CH_2CHCH_2O), 135.14 (CH_2CHCH_2O), 134.71 (CH_2CHCH_2O), 133.84 (Ar), 133.68 (Ar), 130.06 (Ar), 130.06 (Ar), 128.13 (Ar), 128.07 (Ar), 119.23 (CH_2CHCH_2O), 118.44 (Ar), 117.49 (Ar), 117.38 (CH_2CHCH_2O), 116.76 (Ar), 116.60 (CH_2CHCH_2O), 114.69 (Ar), 101.76 (C-1), 101.46 (C-1'), 80.21, 78.23 (C-3'), 76.37 (C-2'), 74.28, 73.89, 72.94 (C-3), 72.54 (C-2), 71.41 (CH_2CHCH_2O), 74.12 (C-4'), 62.82 (C-6'), 60.13 (C-6), (C-6'), 55.80 ($ArOCH_3$), 27.08($C(CH_3)_3$), 19.36 ($C(CH_3)_3$).

4-Methoxyphenyl (2,3,4-tri-*O*-allyl-6-*O*-*tert*-butyldiphenylsilyl-α-D-galacto-pyranosyl)-(1→4)-(2,3,6-tri-*O*-allyl-β-D-galactopyranoside) (22):

To a cooled suspension (0 °C) of a 60% NaH dispersion in oil (42.1 mg, 1.05 mmol) in dry DMF (2 mL) under argon, a solution of **21** (243 mg, 0.301 mmol) in dry DMF was added, followed by AllBr (0.156 mL, 1.81 mmol). The mixture was allowed to warm to rt. The reaction was monitored by TLC (PE/EtOAc, 1:1). After 45 min the reaction was quenched with MeOH. The mixture was neutralized with a solution of 10% AcOH in MeOH, diluted with DCM, and washed with water. The water phase was extracted three times with DCM, the organic phases were collected and washed

four times with a mixture of water and brine (1:1), dried over Na_2SO_4, filtered and concentrated. The residue was purified by chromatography on silica (PE/EtOAc gradient) to yield **22** (187 mg, 67%).

$[\alpha]_D^{20}$ +5.8 (*c* 0.9, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{53}H_{70}NaO_{12}$ Si $[M+Na]^+$ 949.5, found 949.6.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.75 – 7.61 (m, 4H, Ar), 7.48 – 7.33 (m, 6H, Ar), 7.10 – 6.96 (m, 2H, Ar), 6.85 – 6.74 (m, 2H, Ar), 6.03 – 5.81 (m, 5H, 5 CH_2CHCH_2O), 5.69 (ddd, *J* = 22.9, 10.8, 5.6 Hz, 1H, CH_2CHCH_2O), 5.36 (dq, *J* = 17.2, 1.7 Hz, 1H, $CHHCHCH_2O$), 5.28 – 5.26 (m, 1H, $CHHCHCH_2O$), 5.26 – 5.21 (m, 2H, 2 $CHHCHCH_2O$), 5.21 – 5.17 (m, 1H, $CHHCHCH_2O$), 5.17 – 5.15 (m, 2H, 2 $CHHCHCH_2O$), 5.15 – 5.09 (m, 3H, 3 $CHHCHCH_2O$), 5.08 – 5.04 (m, 1H, $CHHCHCH_2O$), 5.04 – 4.98 (m, 1H, $CHHCHCH_2O$) 4.93 (d, *J* = 1.9 Hz, 1H, H-1'), 4.74 (d, *J* = 7.7 Hz, 1H, H-1), 4.50 – 4.41 (m, 1H, $CH_2CHCHHO$), 4.37 – 4.26 (m, 4H, 3 $CH_2CHCHHO$, H-5'), 4.26 – 4.20 (m, 2H, CH_2CHCH_2O), 4.19 – 4.03 (m, 3H, H-4, 2 $CH_2CHCHHO$), 4.01 – 3.84 (m, 9H, H-2', $CH_2CHCHHO$, H-4, $CH_2CHCHHO$, H-6'a, H-3', CH_2CHCH_2O, H-6b), 3.76 (s, 3H, $ArOCH_3$), 3.72 – 3.66 (m, 2H, H-2, H-6'b), 3.62 (dd, *J* = 9.2, 6.0 Hz, 1H, H-6a), 3.57 (t, *J* = 6.5 Hz, 1H,), 3.32 (dd, *J* = 9.9, 3.0 Hz, 1H, H-3), 1.07 (s, 9H, $C(CH_3)_3$).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 155.27 (Ar), 151.93 (Ar), 135.96 (Ar), 135.82 (Ar), 135.73 (Ar), 135.49 (CH_2CHCH_2O), 135.43 (CH_2CHCH_2O), 135.32 (CH_2CHCH_2O), 134.74 (Ar), 133.53 (Ar), 133.46 (Ar), 129.82 (Ar), 127.88 (Ar), 127.84 (Ar), 118.72 (Ar), 117.32 (CH_2CHCH_2O), 117.05 (CH_2CHCH_2O), 116.79 (CH_2CHCH_2O), 116.58 (CH_2CHCH_2O), 116.15 (CH_2CHCH_2O), 115.94 (CH_2CHCH_2O), 114.57 (Ar), 103.31 (C-1), 100.55 (C-1), 79.45 (C-3), 79.05 (C-4), 78.16 (C-2), 76.80 (CH_2CHCH_2O), 74.77 (C-3'), 74.62 (C-4'), 74.14 (CH_2CHCH_2O), 74.11(CH_2CHCH_2O), 74.05 (C-5), 73.12 (CH_2CHCH_2O), 72.23 (CH_2CHCH_2O), 71.33 (CH_2CHCH_2O), 71.20 (C-2'), 70.91 (C-5'), 68.19 (C-6'), 61.55 (C-6), 55.78 ($ArOCH_3$), 27.10 $C(CH_3)_3$, 19.37 $C(CH_3)_3$.

4-Methoxyphenyl (2,3,4-tri-*O*-allyl-α-D-galactopyranosyl)-(1→4)-(2,3,6-tri-*O*-allyl-β-D-galactopyranoside) (23):

A solution of TBAF (1 M in THF, 0.585 mL, 2.02 mmol) was neutralized with glacial acetic acid, cooled down to 0 °C and added to compound **22** (182 mg, 0.196 mmol).

The mixture was stirred overnight at rt. The reaction was monitored by TLC (PE/EtOAc, 1:1). After complete consumption of the starting, the mixture was diluted with EtOAc and extracted twice with a mixture of brine and water (1:1). The organic phase was dried over Na_2SO_4, filtered and concentrated. The residue was chromatographed on silica (PE/EtOAc gradient) to yield **23** (12.6 mg, 91%).

$[\alpha]_D^{20}$ +14.8 (*c* 1.00, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{37}H_{52}NaO_{12}$ $[M+Na]^+$ 711.8, found 711.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.09 – 6.93 (m, 2H), 6.87 – 6.59 (m, 2H), 6.05 – 5.77 (m, 6H), 5.37 – 5.30 (m, 2H), 5.30 – 5.27 (m, 2H), 5.27 – 5.23 (m, 2H), 5.22 – 5.10 (m, 6H), 5.01 (d, *J* = 3.5 Hz, 1H, H-1'), 4.75 (d, *J* = 7.6 Hz, 1H, H-1), 4.54 – 4.36 (m, 2H, 2 CH_2CHCH*H*O), 4.32 – 4.21 (m, 3H, 3 CH_2CHCH*H*O), 4.21 – 4.07 (m, 6H, H-5', 5 CH_2CHCH*H*O), 4.05 – 3.90 (m, 4H, C-4, H-6a, 2 CH_2CHCH*H*O), 3.88 – 3.84 (m, 2H, H-2', H-4'), 3.84 – 3.79 (m, 2H, H-6'a, H-3'), 3.77 (s, 3H, $ArOCH_3$), 3.73 – 3.62 (m, 4H, H-2, H-6'b, H-6b, H-5), 3.38 (dd, *J* = 10.0, 3.0 Hz, 1H, H-3), 2.69 (dd, *J* = 9.3, 1.8 Hz, 1H, 6'-OH).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 155.34 (Ar), 151.86 (Ar), 135.42 (CH_2*C*HCH_2O), 135.31 (CH_2*C*HCH_2O), 135.18 (CH_2*C*HCH_2O), 134.96(CH_2*C*HCH_2O), 134.76 (CH_2*C*HCH_2O), 118.64 (Ar), 117.75 (CH_2*C*HCH_2O), 117.51 (*C*H_2CHCH_2O), 117.15 (*C*H_2CHCH_2O), 116.85 (*C*H_2CHCH_2O), 116.77 (*C*H_2CHCH_2O), 116.28 (*C*H_2CHCH_2O), 114.61 (Ar), 103.41 (C-1), 100.64 (C-1'), 80.84 (C-3), 78.79 (C-2), 78.55 (C-3'), 76.77 (C-4'), 76.05 (C-2'), 75.97 (C-4), 74.27 (C-5), 74.18 (CH_2CH*C*H_2O), 73.88 (CH_2CH*C*H_2O), 73.07 (CH_2CH*C*H_2O), 72.35 (CH_2CH*C*H_2O), 72.18 (CH_2CH*C*H_2O), 71.60 (CH_2CH*C*H_2O), 71.41 (C-5), 68.46 (C-6), 63.23 (C-6'), 55.80 ($ArOCH_3$).

General procedure for the synthesis of compounds **24a-g**:

Alkylation

A 60% dispersion of NaH in oil (1.5 eq.) was suspended in dry DMF (1 mL) at 0 °C, then the halogenide (2.5 eq.) was added, followed by a solution of the carbohydrate (1 eq.) in dry DMF (0.5 mL). The mixture was allowed to warm to rt. The reaction was monitored by TLC (PE/EtOAc, 8:2). When necessary, additional aliquots of NaH were added until completion of the reaction. After completion of the reaction the

mixture was poured on ice and diluted with DCM. The water phase was extracted three times with DCM. The organic phases were collected and dried over Na_2SO_4, filtered and concentrated. The residue was chromatographed on silica (PE/EtOAc gradient).

4-Methoxyphenyl [2,3,4-tri-*O*-allyl-6-*O*-(3-pyridylmethyl)]-α-D-galactopyrano-syl-(1→4)-2,3,4-tri-*O*-allyl-β-D-galactopyranoside (24a):

Halogenide: 3-Chloromethyl pyridine•HCl. In this case, a catalytic amount of TBAI was also added. After 6.5 h a byproduct started to be evident on the TLC and the reaction was quenched by pouring on water at 0 °C. Yield: 34.4 mg, 49%.

$[\alpha]_D^{20}$ +10.7 (*c* 1.00, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{43}H_{58}NO_{12}$ $[M+H]^+$ 780.9, found 780.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.58 (s, 1H, Py), 8.55 (d, *J* = 3.7 Hz, 1H, Py), 7.66 (dt, *J* = 7.8, 1.8 Hz, 1H, Py), 7.28 (dd, *J* = 7.8, 5.0 Hz, 1H, Py), 7.09 – 6.94 (m, 2H, Ar), 6.83 – 6.75 (m, 2H, Ar), 5.99 – 5.82 (m, 6H, 6 CH_2CHCH_2O), 5.37 – 5.29 (m, 2H, 2 $CHHCHCH_2O$), 5.29 – 5.25 (m, 2H, 2 $CHHCHCH_2O$), 5.25 – 5.22 (m, 2H, 2 $CHHCHCH_2O$), 5.21 – 5.14 (m, 2H, 2 $CHHCHCH_2O$), 5.14 – 5.06 (m, 4H, 4 $CHHCHCH_2O$), 4.99 (d, *J* = 3.1 Hz, 1H, H-1'), 4.74 (d, *J* = 7.7 Hz, 1H, H-1), 4.59 – 4.48 (m, 2H, $PyCH_2O$), 4.43 (dd, *J* = 8.6, 5.3 Hz, 1H, H-5'), 4.40 – 4.29 (m, 3H, 3 $CH_2CHCHHO$), 4.24 (m, 2H, 2 $CH_2CHCHHO$), 4.18 (dt, *J* = 5.2, 1.5 Hz, 2H, 2 $CH_2CHCHHO$), 4.13 (ddt, *J* = 12.8, 6.0, 1.3 Hz, 1H, $CH_2CHCHHO$), 4.11 – 4.04 (m, 2H, 2 $CH_2CHCHHO$), 4.02 (d, *J* = 3.1 Hz, 1H, H-4), 4.01 – 3.92 (m, 3H, H-6a, 2 $CH_2CHCHHO$), 3.91 (s, 1H, H-4'), 3.88 – 3.80 (m, 2H, H-2', H-3'), 3.76 (s, 3H, $ArOCH_3$), 3.73 – 3.65 (m, 2H, H-2, H-6'a), 3.64 – 3.57 (m, 2H, H-6b, H-5), 3.53 (dd, *J* = 8.3, 4.9 Hz, 1H, H-6'b), 3.34 (dd, *J* = 9.9, 3.1 Hz, 1H, H-3).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 155.30 (Py), 151.80 (Py), 149.48 (Py), 149.36 (Py), 135.74 (CH_2CHCH_2O), 135.60 (CH_2CHCH_2O), 135.44 (Ar and CH_2CHCH_2O), 135.33 (CH_2CHCH_2O), 135.24 (CH_2CHCH_2O), 134.65 (CH_2CHCH_2O), 133.67 (Ar), 123.51 (Py), 118.78 (Ar), 117.24 (CH_2CHCH_2O), 116.77 (CH_2CHCH_2O), 116.71 (CH_2CHCH_2O), 116.69 (CH_2CHCH_2O), 116.15 (CH_2CHCH_2O), 114.57 (Ar), 103.32 (C-1), 100.15 (C-1'), 80.52 (C-3), 78.58 (C-2), 78.36 (C-5), 76.47 (C-2'), 74.48 (C-4'), 74.28 (C-4), 74.17 (CH_2CHCH_2O), 74.05 (CH_2CHCH_2O), 73.97 (CH_2CHCH_2O),

73.14 (CH_2CHCH_2O), 72.28, 72.16 (CH_2CHCH_2O), 71.24 (CH_2CHCH_2O), 71.08 ($PyCH_2O$), 69.02 (C-5'), 68.63 (C-6'), 67.83 (C-6), 55.78 ($ArOCH_3$).

4-Methoxyphenyl [2,3,4-tri-*O*-allyl-6-*O*-(4-methoxyphenyl)]-α-D-galactopyranosyl)-(1→4)-2,3,6-tri-*O*-allyl-β-D-galactopyranoside (24b):

Halogenide: 4-Methoxybenzyl chloride. After 2 h, additional 0.2 eq. of NaH (1 mg, 0.027 mmol) and 2 eq. of 4-methoxybenzyl chloride (3 µL, 0.027 mmol) were added. After further 3.5 h, the reaction was complete. Yield: 65.4 mg, 60%.

$[\alpha]_D^{20}$ +9.2 (*c* 1.0, $CHCl_3$).

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.32 – 7.23 (m, 2H, Ar), 7.11 – 6.98 (m, 2H, Ar), 6.95 – 6.84 (m, 2H, Ar), 6.82 – 6.71 (m, 2H, Ar), 6.16 – 5.80 (m, 6H, 6 CH_2CHCH_2O), 5.35 – 5.16 (m, 7H, 7 $CHHCHCH_2O$), 5.16 – 5.07 (m, 5H, 5 $CHHCHCH_2O$), 4.96 (d, *J* = 3.4 Hz, 1H, H-1'), 4.73 (d, *J* = 7.7 Hz, 1H, H-1), 4.47 – 4.39 (m, 3H, H-5', $ArCH_2O$), 4.39 – 4.30 (m, 3H, 3 $CH_2CHCHHO$), 4.30 – 4.24 (m, 1H, $CH_2CHCHHO$), 4.23 – 4.10 (m, 4H, 4 $CH_2CHCHHO$), 4.10 – 4.02 (m, 2H, 2 $CH_2CHCHHO$), 4.01 (d, *J* = 3.1 Hz, 1H, H-4), 3.99 – 3.93 (m, 3H, 2 $CH_2CHCHHO$, H-6a), 3.91 (d, *J* = 7.0 Hz, 1H, H-4'), 3.85 (dd, *J* = 10.3, 3.4 Hz, 1H, H-3'), 3.83 – 3.80 (m, 2H, H-2'), 3.80 (s, 3H, $ArOCH_3$), 3.76 (s, 3H, $ArOCH_3$), 3.68 (dd, *J* = 9.9, 7.7 Hz, 1H, H-2), 3.66 – 3.55 (m, 3H, H-6'a, H-6b, H-5), 3.47 (dd, *J* = 8.2, 4.7 Hz, 1H, H-6'b), 3.33 (dd, *J* = 9.9, 3.1 Hz, 1H, H-3).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 159.45 (Ar), 155.28 (Ar), 151.81 (Ar), 135.92 (CH_2CHCH_2O), 135.49 (CH_2CHCH_2O), 135.48 (CH_2CHCH_2O), 135.40 (CH_2CHCH_2O), 135.29 (CH_2CHCH_2O), 134.64 (CH_2CHCH_2O), 130.35 (Ar), 129.77 (Ar), 128.80 (Ar), 118.81 (Ar), 117.24 (CH_2CHCH_2O), 116.82 (CH_2CHCH_2O), 116.73 (CH_2CHCH_2O), 116.66 (CH_2CHCH_2O), 116.62 (CH_2CHCH_2O), 116.09 (CH_2CHCH_2O), 114.56 (Ar), 114.12, 113.96 (Ar), 103.30 (C-1), 100.16 (C-1'), 80.30, 78.54, 78.36, 76.43, 74.40, 74.19 (CH_2CHCH_2O), 74.12 (CH_2CHCH_2O), 74.09 (CH_2CHCH_2O), 73.96 (CH_2CHCH_2O), 73.28 (CH_2CHCH_2O), 73.14, 72.27, 72.05, 71.09, 69.06 (C-6'), 67.81 (C-6), 55.78 ($ArOCH_3$), 55.43 ($ArOCH_3$).

4-Methoxyphenyl (2,3,4-tri-*O*-allyl-6-*O*-pentafluorophenylmethyl)-α-D-galactopyranosyl)-(1→4)-2,3,6- tri-*O*-allyl-β-D-galactopyranoside (24c):

Halogenide: 2,3,4,5,6-Pentafluorobenzyl bromide. After 1.5 h, additional 0.3 eq. of

NaH were added, then stirring was continued for further 2 h. Yield: 54.5 mg, 56%.

$[\alpha]_D^{20}$ +7.7 (*c* 1.0, $CHCl_3$).

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.09 – 6.96 (m, 2H, Ar), 6.85 – 6.75 (m, 2H, Ar), 6.00 – 5.80 (m, 6H, 6 CH_2CHCH_2O), 5.35 – 5.24 (m, 4H, 4 $CHHCHCH_2O$), 5.25 – 5.21 (m, 2H, 2 $CHHCHCH_2O$), 5.20 – 5.14 (m, 3H, 3 $CHHCHCH_2O$), 5.14 – 5.10 (m, 2H, 2 $CHHCHCH_2O$), 5.10 – 5.07 (m, 1H, $CHHCHCH_2O$), 4.96 (d, *J* = 3.3 Hz, 1H, H-1'), 4.74 (d, *J* = 7.7 Hz, 1H, H-1), 4.59 (dd, *J* = 24.9, 11.1 Hz, 2H, $ArCH_2O$), 4.43 – 4.32 (m, 3H, H-5', 2 $CH_2CHCHHO$), 4.32 – 4.19 (m, 3H, 3 $CH_2CHCHHO$), 4.20 – 4.15 (m, 2H, 2 $CH_2CHCHHO$), 4.14 – 3.99 (m, 5H, H-4, 4 $CH_2CHCHHO$), 3.99 – 3.86 (m, 4H, 2 $CH_2CHCHHO$, H-6a, H-4'), 3.87 – 3.78 (m, 2H, H-2', H-3'), 3.76 (s, 3H, $ArOCH_3$), 3.71 – 3.62 (m, 2H, H-2, H-6'a), 3.62 – 3.55 (m, 2H, H-6b, H-5), 3.51 (dd, *J* = 7.9, 4.7 Hz, 1H, H-6'b), 3.34 (dd, *J* = 9.9, 3.1 Hz, 1H, H-3).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 155.31 (Ar), 151.78 (CH_2CHCH_2O), 135.60 (CH_2CHCH_2O), 135.41 (CH_2CHCH_2O), 135.40 (CH_2CHCH_2O), 135.29 (CH_2CHCH_2O), 135.20 (CH_2CHCH_2O), 134.61 (CH_2CHCH_2O), 118.81 (Ar), 117.26 (CH_2CHCH_2O), 116.81 (CH_2CHCH_2O), 116.67 (CH_2CHCH_2O), 116.38 (CH_2CHCH_2O), 116.16 (CH_2CHCH_2O) (C-1'), 100.16 (C-1'), 80.55 (C-3), 78.51 (C-2), 78.38, 76.33 (C-2'), 74.30 (C-4), 74.26, 74.08, 73.88, 73.18, 72.27, 72.21, 71.18, 68.75 (C-3'), 68.53 (C-6), 67.67 (C-6'), 59.89 ($ArCH_2O$), 55.78 ($ArOCH_3$).

4-Methoxyphenyl [2,3,4-tri-*O*-allyl-6-*O*-(4-fluorophenyl)]-α-D-galactopyranosyl)-(1→4)-2,3,6-tri-*O*-allyl-β-D-galactopyranoside (24d):

Halogenide: 4-Fluorobenzyl bromide. After 1.5 h, additional 0.3 eq. of NaH were added, then stirring was continued for further 2 h. Yield: 79 mg, 70%.

$[\alpha]_D^{20}$ +9.7 (*c* 1.0, $CHCl_3$).

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.06 – 6.98 (m, 2H, Ar), 6.84 – 6.75 (m, 4H, Ar), 6.84 – 6.75 (m, 2H, Ar), 6.05 – 5.77 (m, 6H, 6xCH_2CHCH_2O), 5.37 – 5.19 (m, 6H, 3 CH_2CHCH_2O), 5.19 – 5.07 (m, 6H, 3 CH_2CHCH_2O), 4.98 (d, *J* = 3.2 Hz, 1H, H-1'), 4.74 (d, *J* = 7.7 Hz, 1H, H-1), 4.47 (s, 2H, $ArCH_2O$), 4.42 (dd, *J* = 8.9, 5.2 Hz, 1H, H-5'), 4.39 – 4.30 (m, 3H, 3 $CH_2CHCHHO$), 4.30 – 4.24 (m, 1H, $CH_2CHCHHO$), 4.24 – 4.16 (m, 3H, 3 $CH_2CHCHHO$), 4.16 – 4.11 (m, 1H, $CH_2CHCHHO$), 4.11 – 4.03 (m, 2H, 2 $CH_2CHCHHO$), 4.02 (d, *J* = 3.0 Hz, 1H, H-4), 4.00 – 3.93 (m, 3H, H-6a, 2 $CH_2CHCHHO$), 3.91 (s, 1H, H-4'), 3.88 – 3.80 (m, 2H, H-2', H-3'), 3.76 (s, 3H,

OC*H*$_3$), 3.73 – 3.67 (m, 1H, H-2), 3.67 – 3.62 (m, 1H, H-6'a), 3.62 – 3.56 (m, 2H, H-6b, H-5), 3.49 (dd, *J* = 8.3, 4.8 Hz, 1H, H-6'b), 3.34 (dd, *J* = 9.9, 3.1 Hz, 1H, H-3).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 163.50 (Ar), 161.55 (Ar), 155.29 (Ar), 151.78 (Ar), 135.81 (CH_2*C*HCH_2O), 135.44 (CH_2*C*HCH_2O), 135.42 (CH_2*C*HCH_2O), 135.33 (CH_2*C*HCH_2O), 135.25 (CH_2*C*HCH_2O), 134.63 (CH_2*C*HCH_2O), 134.03 (Ar), 134.01 (Ar), 129.83 and 129.77 (Ar), 118.78 (Ar), 117.21 (*C*H_2CHCH_2O), 116.78 (*C*H_2CHCH_2O), 116.76 (*C*H_2CHCH_2O), 116.63 (*C*H_2CHCH_2O), 116.62 (*C*H_2CHCH_2O), 116.10 (*C*H_2CHCH_2O), 115.47 (Ar), 115.30 (Ar), 114.55 (Ar), 103.30 (C-1), 100.13 (C-1'), 80.38 (C-3), 78.54 (C-2), 78.35 (C-2'), 76.46 (C-3'), 74.46 (C-4), 74.20 (CH_2CH*C*H_2O), 74.15 (C-4'), 74.05, 73.95 (C-5), 73.14 ($ArCH_2O$), 72.87 (CH_2CH*C*H_2O), 72.26 (CH_2CH*C*H_2O), 72.06(CH_2CH*C*H_2O), 71.17 (CH_2CH*C*H_2O), 69.04 (C-5'), 68.15 (C-6'), 67.80 (C-6), 55.76 ($ArOCH_3$).

4-Methoxyphenyl (2,3,4-tri-*O*-allyl-6-*O*-propylphenyl)-α-D-galactopyranosyl)-(1→4)-2,3,6-tri-*O*-allyl-β-D-galactopyranoside (24e):

Halogenide: 1-Bromo-3-phenylpropane. After 1.5 h, additional 0.3 eq. of NaH and 0.2 eq. of bromide were added. Reaction time: 2 h. Yield: 74.4 mg, 77%.

$[\alpha]_D^{20}$ + 11.2 (*c* 1.0, $CHCl_3$).

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.34 – 7.26 (m, 2H, Ar), 7.21 – 7.14 (m, 3H, Ar), 7.07 – 6.94 (m, 2H, Ar), 6.90 – 6.62 (m, 2H), 6.04 – 5.82 (m, 6H, 6 CH_2C*H*CH_2O), 5.37 – 5.20 (m, 6H, 6 CH*H*$CHCH_2O$), 5.18 – 5.08 (m, 6H, 6 CH*H*$CHCH_2O$), 4.98 (d, *J* = 2.9 Hz, 1H, H-1'), 4.75 (d, *J* = 7.7 Hz, 1H, H-1), 4.41 – 4.31 (m, 4H, 3 CH_2CHCH*H*O, H-5'), 4.31 – 4.25 (m, 2H, 2 CH_2CHCH*H*O), 4.24 – 4.17 (m, 2H, 2 CH_2CHCH*H*O), 4.17 – 4.09 (m, 3H, 3 CH_2CHCH*H*O), 4.03 (d, *J* = 3.0 Hz, 1H, H-4), 4.00 – 3.94 (m, 3H, 2 CH_2CHCH*H*O, H-6a), 3.93 – 3.90 (m, 1H, H-4'), 3.88 – 3.81 (m, 2H, H-2', H-3'), 3.76 (s, 3H, $ArOCH_3$), 3.71 (dd, *J* = 9.9, 7.7 Hz, 1H, H-2), 3.64 – 3.55 (m, 3H, H-6'a, H-6b, H-5), 3.53 – 3.41 (m, 3H, H-6'b, OC*H*$_2CH_2CH_2Ph$), 3.35 (dd, *J* = 9.9, 3.0 Hz, 1H, H-3), 2.70 (dt, *J* = 8.4, 6.4 Hz, 2H, OCH_2CH_2C*H*$_2Ph$), 2.05 – 1.68 (m, 2H, OCH_2C*H*$_2CH_2Ph$).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 155.27 (Ar), 151.79 (Ar), 142.01 (Ar), 135.88 (CH_2*C*HCH_2O), 135.46 (CH_2*C*HCH_2O), 135.35 (CH_2*C*HCH_2O), 135.30 (CH_2*C*HCH_2O), 134.64 (CH_2*C*HCH_2O), 128.50 (Ar), 128.48 (Ar), 125.96 (Ar),

118.79 (Ar), 117.21 (*C*H_2CHCH_2O), 116.83 (*C*H_2CHCH_2O), 116.72 (*C*H_2CHCH_2O), 116.66 (*C*H_2CHCH_2O), 116.59 (*C*H_2CHCH_2O), 116.06 (*C*H_2CHCH_2O), 114.55 (Ar), 103.29 (C-1), 100.17 (C-1'), 80.45 (C-3), 78.60 (C-2), 78.31 (C-3'), 77.32 (CH_2CH*C*H_2O), 76.49 (CH_2CH*C*H_2O), 74.41 (C-4'), 74.16 (C-4), 74.06 (C-5), 73.99 (CH_2CH*C*H_2O), 73.12 (CH_2CH*C*H_2O), 72.26 (C-2'), 72.12 (CH_2CH*C*H_2O), 71.16, 70.75 (CH_2CH*C*H_2O), 68.98 (C-5'), 68.46 (C-6'), 67.82 (C-6), 55.76 (ArOCH_3), 32.56 (O$CH_2$$CH_2$*C*$H_2$Ph), 31.47 (O$CH_2$*C*$H_2$$CH_2$Ph).

4-Methoxyphenyl [2,3,4-tri-*O*-allyl-6-*O*-(biphenyl-4yl-methyl)]-α-D-galactopyranosyl)-(1→4)-2,3, 6-tri-*O*-allyl-β-D-galactopyranoside (24f):

Halogenide: 4-(Bromomethyl)biphenyl. Reaction time: 1.5 h. Yield: 57.1 mg, 55%.

$[\alpha]_D^{20}$ +9.2 (*c* 1.0, $CHCl_3$).

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.62 – 7.54 (m, 4H, Ar), 7.47 – 7.39 (m, 4H, Ar), 7.35 (t, *J* = 7.4 Hz, 1H, Ar), 7.06 – 6.99 (m, 2H, Ar), 6.85 – 6.75 (m, 2H, Ar), 6.05 – 5.77 (m, 6H, 6 CH_2C*H*CH_2O), 5.38 – 5.20 (m, 6H, 6 CH*H*CHCH_2O), 5.20 – 5.09 (m, 6H, 6 CH*H*CHCH_2O), 4.99 (d, *J* = 3.1 Hz, 1H, H-1'), 4.75 (d, *J* = 7.6 Hz, 1H, H-1), 4.60 – 4.52 (m, 2H, ArCH_2O), 4.46 (dd, *J* = 9.0, 4.7 Hz, 1H, H-5'), 4.42 – 4.32 (m, 3H, 3 CH_2CHCH*H*O), 4.32 – 4.17 (m, 4H, 4 CH_2CHCH*H*O), 4.17 – 4-05 (m, 3H, 3 CH_2CHCH*H*O), 4.03 (d, J = 2.9 Hz, 1H, H-4), 4.01 – 3.92 (m, 4H, H-4', H-6a, 2 CH_2CHCH*H*O), 3.92 – 3.80 (m, 2H, H-3', H-2'), 3.76 (s, 3H, ArOCH_3), 3.73 – 3.67 (m, 2H, H-6'a, H-2), 3.65 – 3.57 (m, 2H, H-6b, H-5), 3.54 (dd, *J* = 8.2, 4.7 Hz, 1H, H-6'b), 3.35 (dd, *J* = 9.9, 3.0 Hz, 1H, H-3).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 155.29 (Ar), 151.79 (Ar), 141.03 (Ar), 140.90 (Ar), 137.23 (Ar), 135.86 (CH_2*C*HCH_2O), 135.46 (CH_2*C*HCH_2O), 135.43 (CH_2*C*HCH_2O), 135.36 (CH_2*C*HCH_2O), 135.27 (CH_2*C*HCH_2O), 134.63 (CH_2*C*HCH_2O), 128.90 (Ar), 128.59 (Ar), 127.44 (Ar), 127.33 (Ar), 127.24 (Ar), 118.81 (Ar), 117.23 (*C*H_2CHCH_2O), 116.84 (*C*H_2CHCH_2O), 116.79 (*C*H_2CHCH_2O), 116.67 (*C*H_2CHCH_2O), 116.10 (*C*H_2CHCH_2O), 114.55 (Ar), 103.31 (C-1), 100.15 (C-1'), 80.34 (C-3), 78.56 (C-3'), 78.38 (C-2), 76.45 (C-2'), 74.46, 74.21 (C-4), 74.17, 74.09 (CH_2CH*C*H_2O), 73.95 (C-5), 73.35 (ArCH_2O), 73.17, 72.27 (CH_2CH*C*H_2O), 72.10, 71.14 (CH_2CH*C*H_2O), 69.07 (C-5'), 68.16 (C-6'), 67.77 (C-6), 55.77 (ArOCH_3).

4-Methoxyphenyl [2,3,4-tri-*O*-allyl-6-*O*-(2-pyridylmethyl)]-α-D-galactopyrano-

syl)-(1→4)-2,3,6-tri-*O*-allyl-β-D-galactopyranoside (24g):

Halogenide: 2-(Bromomethy)pyridine□HBr. Reaction time: 1 h. Yield: 48 mg, 71%.

$[\alpha]_D^{20}$ -11.2, (*c* 1.60, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{43}H_{57}NNaO_{12}$ 802.4, found 802.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.58 – 8.50 (m, 1H, H-1 Py), 7.68 (td, *J* = 7.7, 1.8 Hz, 1H, H-3 Py), 7.44 (d, *J* = 7.8 Hz, 1H, H-4 Py), 7.19 (dd, *J* = 6.5, 5.0 Hz, 1H, H-2 Py), 7.08 – 6.97 (m, 2H, Ar-H), 6.87 – 6.75 (m, 2H, Ar-H), 6.02 – 5.80 (m, 6H, CH_2*C*HCH_2O), 5.34 (dd, *J* = 17.3, 1.8 Hz, 1H, CH*H*CHCH_2O), 5.30 – 5.19 (m, 5H, 5 CH*H*CHCH_2O), 5.18 – 5.16 (m, 1H, CH*H*CHCH_2O), 5.16 – 5.14 (m, 1H CH*H*CHCH_2O), 5.14 – 5.12 (m, 1H CH*H*CHCH_2O), 5.11 (d, *J* = 1.2 Hz, 2H, 2 CH*H*CHCH_2O), 5.08 (dd, *J* = 2.7, 1.6 Hz, 1H, CH*H*CHCH_2O), 4.99 (d, *J* = 2.9 Hz, 1H, H-1'), 4.75 (d, *J* = 7.7 Hz, 1H, H-1), 4.65 (s, 2H, -OC*H*$_2$Py), 4.47 (dd, *J* = 8.8, 5.4 Hz, 1H, H-5'), 4.41 – 4.32 (m, 3H, 3 CH_2CHCH*H*O), 4.31 – 4.26 (m, 1H, CH_2CHCH*H*O), 4.26 – 4.18 (m, 3H, 3 CH_2CHCH*H*O), 4.17 – 4.06 (m, 3H, 3 CH_2CHCH*H*O), 4.03 (d, *J* = 3.0 Hz, 1H, H-4), 4.01 – 3.92 (m, 4H, 2 CH_2CHCH*H*O, H-4', H-6a), 3.91 – 3.82 (m, 2H, H-2', H-3'), 3.76 (s, 3H, -OCH_3) 3.75 – 3.69 (m, 2H, H-2 and H-6'a), 3.63 – 3.56 (m, 3H, H-6'b, H-5, H-6b), 3.35 (dd, *J* = 9.9, 3.1 Hz, 1H, H-3).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 158.57 (Ar), 155.29 (Ar), 151.79 (Ar), 149.26, 136.67 (Ar), 135.79 (CH_2*C*HCH_2O), 135.46 (CH_2*C*HCH_2O), 135.40 (CH_2*C*HCH_2O), 135.28 (CH_2*C*HCH_2O), 134.62 (CH_2*C*HCH_2O), 122.47 (Ar), 121.44 (Ar), 118.80 (Ar), 117.22 (*C*H_2CHCH_2O), 116.89 (*C*H_2CHCH_2O), 116.82 (*C*H_2CHCH_2O), 116.71 (*C*H_2CHCH_2O), 116.64 (*C*H_2CHCH_2O), 116.09 (*C*H_2CHCH_2O), 114.55 (Ar), 103.32 (C-1), 100.13 (C-1'), 80.34 (C-3), 78.59 (C-3'), 78.35 (C-2), 76.50 (C-2'), 74.49, 74.35 (ArCH_2O), 74.24, 74.15 (C-4), 74.05 (CH_2CH*C*H_2O), 73.97 (C-5), 73.17 (CH_2CH*C*H_2O), 72.26, 72.09, 71.23, 69.07 (C-5'), 68.82 (C-6'), 67.82 (C-6), 55.76 (ArOCH_3).

General procedures for the synthesis of compounds **25a-g**:

Cleavage of the allyl groups [21]

DMBA (12 eq), $Pd(OAc)_2$ (0.3 eq) and PPh_3 (0.9 eq) were mixed in a reaction tube under argon, then a solution of the carbohydrate (1 eq.) in dry MeOH (0.5 mL) and

dry DCM (0.1 mL) was added. The reaction was stirred at 40 °C and monitored by TLC (9:1 DCM/MeOH). After completion of the reaction the mixture was diluted with DCM and washed once with aq. Na_2CO_3. The water phase was extracted four times with DCM. The organic phases were collected, dried over Na_2SO_4, filtered and concentrated. The residue was purified by chromatography on silica gel (DCM/MeOH gradient).

4-Methoxyphenyl [6-*O*-(3-pyridylmethyl)-α-D-galactopyranosyl]-(1→4)-β-D-galactopyranoside (25a):

Reaction time: 17 h. Purification: Chromatography on silica (DCM/MeOH gradient + 0.2% NH_4OH). Yield: 10.5 mg, 46%.

$[\alpha]_D^{20}$ -13.0 (*c* 0.10, MeOH).

ESI-MS: *m/z*: Calcd for $C_{25}H_{33}NNaO_{12}[M+Na]^+$ 562.5, found 562.1.

^{1}H NMR (500 MHz, CD_3OD): δ = 8.60 (s, 1H, Py), 8.49 (d, *J* = 3.8 Hz, 1H, Py), 7.95 (d, *J* = 7.9 Hz, 1H, Py), 7.47 (dd, *J* = 7.8, 4.9 Hz, 1H, Py), 7.13 – 7.02 (m, 2H, Ar), 6.92 – 6.80 (m, 2H, Ar), 5.03 (d, *J* = 2.9 Hz, 1H, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.70 – 4.65 (m, 2H, $PyCH_2O$), 4.55 (t, *J* = 6.3 Hz, 1H, H-5'), 4.09 (d, *J* = 3.1 Hz, 1H, H-4), 3.97 (s, 1H, H-4'), 3.90 (dd, *J* = 9.7, 6.0 Hz, 1H, H-6a), 3.88 – 3.79 (m, 4H, H-2', H-3', H-6'a, H-6b, H-5), 3.77 (s, 3H, $ArOCH_3$), 3.76 – 3.70 (m, 2H, H-2, H-6'b), 3.67 (dd, *J* = 10.0, 2.9 Hz, 1H, H-3).

^{13}C NMR (126 MHz, CD_3OD): δ = 156.80 (Py), 152.97 (Ar), 149.34 (Py), 149.12 (Py), 137.91 (Py), 136.31 (Ar), 125.23 (Py), 119.31 (Ar), 115.51 (Ar), 104.10 (C-1), 102.70 (C-1'), 79.37 (C-4), 76.26 (C-3'), 74.57 (C-3), 72.79 (C-5), 71.52 ($PyCH_2O$), 71.22 (C-5'), 71.11 (C-2'), 70.91 (C-4'), 70.67 (C-6'), 61.02 (C-6), 56.06 ($ArOCH_3$).

4-Methoxyphenyl [6-*O*-(4-methoxyphenyl)]-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (25b):

Reaction time: 24 h. Purification: Chromatography on silica (DCM/MeOH gradient), followed by reversed phase chromatography (RP, C18, water/ACN gradient). Yield: 21 mg, 51%.

$[\alpha]_D^{20}$ +2.3 (*c* 1.0, MeOH).

^{1}H NMR (500 MHz, CD_3OD): δ = 7.37 – 7.31 (m, 2H, Ar), 7.12 – 7.05 (m, 2H, Ar), 6.96 – 6.91 (m, 2H, Ar), 6.89 – 6.83 (m, 2H, Ar), 5.02 (s, 1H, H-1'), 4.84 (d, *J* = 7.6

Hz, 1H, H-1), 4.60 – 4.51 (m, 2H, $ArCH_2O$), 4.49 (t, *J* = 6.2 Hz, 1H, H-5'), 4.08 (d, *J* = 3.0 Hz, 1H, H-4), 3.94 (s, 1H, H-4'), 3.89 (dd, *J* = 10.5, 6.7 Hz, 1H, H-6a), 3.84 – 3.83 (m, 2H, H-2', H-3'), 3.82 – 3.79 (m, 5H, $ArOCH_3$, H-6b, H-5), 3.79 – 3.75 (m, 4H, H-2, $ArOCH_3$), 3.68 – 3.63 (m, 3H, H-3, H-6'a, H-6'b).

^{13}C NMR (126 MHz, CD_3OD): δ = 160.82 (Ar), 156.77 (Ar), 152.94 (Ar), 131.58 (Ar), 130.63 (Ar), 119.32 (Ar), 115.49 (Ar), 114.77 (Ar), 104.03 (C-1), 102.65 (C-1'), 79.43 (C-4), 76.21 (C-2), 74.62 (C-3), 74.09 ($ArCH_2O$), 72.72 (C-5), 71.28 (C-5'), 71.22 (C-4'), 71.03 (C-2'), 70.66 (C-3'), 70.08 (C-6'), 61.07 (C-6), 56.05 ($ArOCH_3$), 55.68 ($ArOCH_3$).

4-Methoxyphenyl (6-*O*-pentafluorophenyl)-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (25c):

Reaction time: 24 h. Purification: Chromatography on silica (DCM/MeOH gradient), followed by reversed phase chromatography (RP, C18, water/ACN gradient). Yield: 8.4 mg, 23%.

$[\alpha]_D^{20}$ +1.3 (*c* 1.0, MeOH).

ESI-MS: *m/z*: Calcd for $C_{26}H_{29}F_5NaO_{12}$ $[M+Na]^+$ 651.2, found 651.3.

^1H NMR (500 MHz, CD_3OD): δ = 7.11 – 7.04 (m, 2H, Ar), 6.89 – 6.83 (m, 2H, Ar), 4.99 (d, *J* = 3.0 Hz, 1H, H-1'), 4.82 (d, *J* = 7.5 Hz, 1H, H-1), 4.78 – 4.70 (m, 2H, $ArOCH_2$), 4.51 (t, *J* = 6.4 Hz, 1H, H-5'), 4.06 (d, *J* = 3.0 Hz, 1H, H-4), 3.94 – 3.91 (m, 1H, H-4'), 3.88 (dd, *J* = 10.4, 6.8 Hz, 1H, H-6a), 3.85 – 3.76 (m, 7H, H-2', $ArOCH_3$, H-6b, H-5, H-3'), 3.76 – 3.65 (m, 3H, H-2, H-6'a, H-6'b), 3.64 (dd, *J* = 10.1, 3.0 Hz, 1H, H-3).

^{13}C NMR (126 MHz, CD_3OD): δ = 156.79 (Ar), 152.96 (Ar), 119.28 (Ar), 115.49 (Ar), 104.06 (C-1), 102.52 (C-1'), 78.88 (C-4), 76.33 (C-5), 74.48 (C-3), 72.68 (C-2), 71.27 (C-3'), 70.69 (C-2', C-4', and C-6'), 70.50 (C-5'), 60.91 ($ArCH_2O$ and C-6), 56.05 ($ArOCH_3$).

4-Methoxyphenyl [6-*O*-(4-fluorophenyl)]-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (25d):

Reaction time: 7 h. Purification: Chromatography on silica (DCM/MeOH gradient), followed by reversed phase chromatography (RP, C18, water/ACN gradient). Yield: 22.5 mg, 44%.

$[\alpha]_D^{20}$ +4.6 (*c* 1.1, MeOH).

ESI-MS: *m/z*: Calcd for $C_{26}H_{33}FNaO_{12}$ $[M+Na]^+$ 579.2, found 579.3.

^{1}H NMR (500 MHz, CD_3OD): δ = 7.41 (dd, *J* = 8.2, 5.7 Hz, 2H, Ar), 7.20 – 7.01 (m, 4H, Ar), 6.87 – 6.80 (m, 2H, Ar), 4.99 (d, *J* = 1.62 Hz, 1H, H-1'), 4.81 (d, *J* = 7.6 Hz, 1H, H-1), 4.61 – 4.53 (m, 2H, $ArCH_2O$), 4.48 (t, *J* = 6.2 Hz, 1H, H-5'), 4.06 (d, *J* = 3.0 Hz, 1H, H-4), 3.92 (s, 1H, H-4'), 3.87 (dd, *J* = 10.2, 6.4 Hz, 1H, H-6a), 3.81 (s, 2H, H-3', H-2'), 3.80 – 3.72 (m, 6H, $ArOCH_3$, H-6b, H-2, H-5), 3.69 – 3.59 (m, 3H, H-3, H-6'a, H-6'b).

^{13}C NMR (126 MHz, CD_3OD): δ = 164.74 (Ar), 162.80 (Ar), 156.79 (Ar), 152.96 (Ar), 135.73 and 135.71 (Ar(F)), 130.92 and 130.86 (Ar(F)), 119.31 (Ar), 116.09 and 115.92 (Ar(F)), 115.51 (Ar), 104.07 (C-1), 102.69 (C-1'), 79.46 (C-4), 76.23, 74.60 (C-3), 73.58 ($ArOCH_2$), 72.76 (C-2), 71.26 (C-4'), 71.20 (C-5'), 70.98 (C-3'), 70.67 (C-2'), 70.30 (C-6'), 61.06 (C-6), 56.06 ($ArOCH_3$).

4-Methoxyphenyl [6-*O*-(3-phenylpropyl)]-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (25e):

Reaction time: 14h. Purification: Chromatography on silica (DCM/MeOH gradient), followed by reversed phase chromatography (RP, C18, water/ACN gradient). Yield: 27.2 mg, 55%.

$[\alpha]_D^{20}$ +3.9 (*c* 1.0, MeOH).

ESI-MS: *m/z*: Calcd for $C_{26}H_{38}NaO_{12}$ $[M+Na]^+$ 589.2, found 589.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 7.35 – 7.23 (m, 4H, Ar), 7.20 – 7.11 (m, 1H, Ar), 7.12 – 6.95 (m, 2H, Ar), 6.92 – 6.78 (m, 2H, Ar), 5.02 (d, *J* = 1.5 Hz, 1H, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.46 (s, 1H, H-5'), 4.08 (d, *J* = 3.0 Hz, 1H, H-4), 3.97 (s, 1H, H-4'), 3.90 (dd, *J* = 10.5, 6.6 Hz, 1H, H-6a), 3.86 – 3.74 (m, 8H, H-2', $ArOCH_3$, H-2, H-6b, H-3', H-5), 3.71 – 3.61 (m, 3H, H-6'a, H-6'b, H-3), 3.57 – 3.51 (m, 2H, $OCH_2CH_2CH_2Ph$), 2.73 (t, *J* = 7.7 Hz, 1H, $OCH_2CH_2CH_2Ph$), 1.99 – 1.89 (m, 2H, $OCH_2CH_2CH_2Ph$).

^{13}C NMR (126 MHz, CD_3OD): δ = 156.78 (Ar), 152.98 (Ar), 143.41 (Ar), 129.58 (Ar), 129.32 (Ar), 126.74 (Ar), 119.30 (Ar), 115.50 (Ar), 104.10 (C-1), 102.74 (C-1'), 79.56 (C-4), 76.22 (C-3'), 74.60 (C-3), 72.78 (C-2), 71.82 ($OCH_2CH_2CH_2Ph$), 71.33 (C-5'), 71.16 (C-5), 71.03 (C-4'), 70.90 (C-6'), 70.70 (C-2'), 61.06 (C-6), 56.06 ($ArOCH_3$), 33.28 ($OCH_2CH_2CH_2Ph$), 32.55 ($OCH_2CH_2CH_2Ph$).

4-Methoxyphenyl (6-*O*-biphenyl-1-yl-methyl)-α-D-galactopyranosyl)-(1→4)-β-D-galactopyranoside (25f):

Reaction time: 7 h. Purification: Chromatography on silica (DCM/MeOH gradient), followed by reversed phase chromatography (RP, C18, water/ACN gradient). Yield: 19.2 mg, 48%.

$[\alpha]_D^{20}$ +1.2 (*c* 1.0, MeOH).

^{1}H NMR (500 MHz, CD_3OD): δ =7.64 (d, *J* = 8.0 Hz, 4H, Ar), 7.51 (d, *J* = 8.2 Hz, 2H, Ar), 7.47 – 7.42 (m, 2H, Ar), 7.38 – 7.32 (m, 1H, Ar), 7.11 – 7.07 (m, 2H, Ar), 6.89 – 6.84 (m, 2H, Ar, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.71 – 4.62 (m, 2H, $ArCH_2O$), 4.53 (t, *J* = 6.3 Hz, 1H, H-5'), 4.09 (d, *J* = 3.1 Hz, 1H, H-4), 3.98 (s, 1H, H-4'), 3.90 (dd, *J* = 9.7, 5.9 Hz, 1H, H-6a), 3.86 – 3.83 (m, 2H, H-3', H-2'), 3.83 – 3.78 (m, 3H, H-2, H-6b, H-5), 3.77 (s, 3H, $ArOCH_3$), 3.75 – 3.71 (m, 2H, H-6'a, H-6'b), 3.67 (dd, *J* = 10.1, 3.1 Hz, 1H, H-3).

^{13}C NMR (126 MHz, CD_3OD): δ = 156.80 (Ar), 152.97 (Ar), 129.85 (Ar), 129.50 (Ar), 128.33 (Ar), 128.00 (Ar), 127.94 (Ar), 119.34 (Ar), 115.50 (Ar), 104.09 (C-1), 102.71 (C-1'), 79.45 (C-4), 76.24 (C-5), 74.65 (C-3), 74.11 ($ArCH_2O$), 72.77 (C-2), 71.30 (C-5'), 71.04 (C-4'), 70.70 (3'), 70.38 (C-6'), 61.06 (C-6), 56.05 ($ArOCH_3$).

4-Methoxyphenyl [6-*O*-(2-pyridylmethyl)-α-D-galactopyranosyl]-(1→4)-β-D-galactopyranoside (25g):

Reaction time: 7 h. No aqueous work-up. Purification: Chromatography on silica (DCM/MeOH gradient, + 0.2% NH_4OH), followed by reversed phase chromatography (RP, C18, water/ACN gradient). Yield: 22 mg, 65%.

$[\alpha]_D^{20}$ -1.9 (*c* 1.1, MeOH).

ESI-MS: m/z: Calcd for $C_{25}H_{34}NO_{12}$ $[M]^+$ 540.2, found 540.2.

^{1}H NMR (500 MHz, CD_3OD): δ = 8.80 (d, *J* = 5.6 Hz, 1H, Py-H), 8.53 (td, *J* = 8.0, 1.4 Hz, 1H, Py-H), 8.00 (d, *J* = 8.09 Hz, 1H, Py-H), 7.95 (t, J = 6.72, 1H, Py-H), 7.11 - 6.99 (m, 2H, Ar-H), 6.89 - 6.79 (m, 2H, Ar-H), 5.09 – 4.99 (m, 3H, $PyCH_2O$, H-1'), 4.83 (d, *J* = 7.0 Hz, 1H, H-1), 4.52 (t, *J* = 5.7 Hz, 1H, H-5'), 4.12 (d, *J* = 2.3 Hz, 1H, H-4), 3.97 (m, 1H, H-4'), 3.96 – 3.92 (m, 2H, H-6'a, H-6'b), 3.91 – 3.87 (m, 3H, H-2', H-3', H-6a), 3.86 – 3.83 (m, 1H, H-6b), 3.82 – 3.78 (m, 3H, H-2, H-3, H-5), 3.77 – 3.75 (m, 1H, H-3), 3.74 (s, 3H, $-OCH_3$).

^{13}C NMR (126 MHz, CD_3OD): δ = 156.82 (Py), 154.87 (Ar), 152.94 (Ar), 146.99 (Py), 142.93 (Py), 126.76 (Py), 125.90 (Py), 119.27 (Ar), 115.51 (Ar), 104.12 (C-1), 103.37 (C-1'), 81.64 (C-4), 75.91 (C-5), 74.72 (C-3), 72.91 (C-6'), 72.87 (C-2), 72.25 (C-5'), 71.15 (C-4'), 70.91 (C-2'), 70.22 (C-3'), 69.73 ($PyCH_2O$), 61.46 (C-6), 56.05 ($ArOCH_3$).

Isothermal Titration Calorimetry. ITC experiments were performed using a VP-ITC instrument (GE Healthcare, Uppsala, Sweden) at a temperature of 25° C, an injection volume between 6 and 15 µl, a reference power of 10 µcal/s, a stirring speed of 307 rpm, high feedback, a spacing time of 300 - 600 s and a filter period of 2 s. Preceding the measurements, $PapGII_{LD}$-6His was dialyzed against assay buffer [2-(*N*-morpholino)ethanesulfonic acid, 20 mM, pH 5.8] using Slide-A-Lyzer G2 dialysis cassettes (10K MWCO) (Thermo Fisher Scientific, Rockford, IL, USA). Protein concentration was determined by HPLC-UV against a BSA standard [22,23]. All experiments were performed with *c*-values below 1. Therefore, stoichiometry was fixed to 1 to allow reliable determination of K_D and ΔH° [24,25]. Two independent experiments evaluated the consistency of the measurements. Experiments with *c*-values below 0.01 were performed only once, due to high material consumption. Baseline adjustment and peak integration were carried out using Origin 7.0 as described by the manufacturer (OriginLab, Northampton, MA, USA). The first injection was always excluded from data analysis. The three-parameter nonlinear least-square fitting and calculation of 95% confidence intervals were calculated by performing a global analysis of multiple ITC experiments by public domain multimethod analysis software SEDPHAT to determine N (stoichiometry), K_D (dissociation constant) and ΔH° (change in enthalpy) [26].

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ = RT \ln K_D = -RT \ln K_A \qquad \text{eq. 2}$$

Thermodynamic parameters ΔG° (change in Gibb's free energy) and ΔS° (change in Entropy) are calculated using equation 2, where T is the absolute temperature, and R is the universal gas constant (8.314 J/mol K).

Cloning of PapGII: genomic DNA was prepared from the *uropathogenic E. coli* clinical isolate BI47 (University Hospital, Basel, Switzerland) using the extraction kit from Wizard (Promega, Wallisellen, Switzerland). The PapG-II lectin domain (PapG-$II_{LD,}$ amino acids: 1-196) was amplified by PCR using the iproof high fidelity *pfu* polymerase (BioRad, Basel, Switzerland). The forward and reverse primers were designed on the basis of the published *PapG-II* sequences and were synthesized at Microsynth (Balgach, Switzerland). The PCR-generated fragment was treated with restriction enzymes XhoI and NcoI and ligated into the appropriate cloning site of the expression vector pET-22b(+) (Novagen, Switzerland). The ligation products, with or without 6His-Tag, were amplified in chemocompetent *E. coli* DH5α . After single clone selection and plasmid DNA-minipreparation, the correctness of the constructs was confirmed by restriction control and DNA sequencing (Microsynth, Balgach, Switzerland). For protein expression the constructs containing $PapGII_{LD}$ without tag or $PapGII_{LD}$-6His were transformed into the *E. coli* strain AD494(DE3).

Protein expression: *E. coli* AD494(DE3) clones containing the constructs $PapGII_{LD}$-6His or PapG-II_{LD} were cultivated in LB-medium at 30 °C and 180 rpm. The PapG-II expression was induced with 1 mM final concentration of isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD_{600} of 0.8. The culture was allowed to grow for further 14 h at 30 °C and 180 rpm. The cells were then cooled on ice for 5 min and harvested by centrifugation at 5,000 rpm for 20 min at 4 °C. The pellet was suspended in a cold solution of 50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 5 mM EDTA and 1 mg/mL polymyxin B sulfate and stirred for 2 h at 4 °C. The periplasmic extract was collected by centrifugation at 11,000 rpm for 20 min at 4 °C. His-tagged protein was dialyzed overnight against 50 mM NaH_2PO_4, 300 mM NaCl, and 10 mM imidazole, pH 8 (Binding buffer), and applied to a Ni-NTA column attached to a fast protein liquid chromatography system and pre-equilibrated with binding buffer. The column wash step was performed with binding buffer and His-tagged protein was eluted with 50 mM NaH_2PO_4, 300 mM NaCl, and 250 mM imidazole, pH 8.0 (elution buffer). The eluted fraction was dialyzed overnight against a buffer containing 20 mM Hepes, pH 7.4, 150 mM NaCl and 1 mM $CaCl_2$ (assay buffer).

PapG-II_{LD} without tag was extracted from the periplasm similar to the 6His-tagged protein, dialyzed overnight against assay buffer and applied to a GbO3 functionalized

fractogel. The protein was eluted using a 1 M lactose solution. Protein purity was confirmed by SDS-PAGE analysis followed by Coomassie Brillant Bleu G-250 staining. The protein concentration was determined by HPLC as reported [23]. At 4 °C, the protein was stable up to 4 weeks. For long-term storage, the protein was frozen at -80 °C.

Competitive binding assay: To determine the affinity of the various PapG-II antagonists, the competitive binding assay described previously was applied [13]. Microtiter plates (F96 MaxiSorp, Nunc) were coated with 100 µL/well of a 10 µg/mL solution of PapG-II_{LD}-6His in 20 mM Hepes, 150 mM NaCl and 1 mM $CaCl_2$, pH 7.4 (assay buffer) overnight at 4 °C. The coating solution was discarded and the wells were blocked with 150 µL/well of 3% BSA in assay buffer for 2 h at 4 °C. After three washing steps with assay buffer (150 µL/well), a four-fold serial dilution of the test compound (50 µL/well) in assay buffer containing 5% DMSO and streptavidin-peroxidase coupled GbO3-PAA polymer (50 µL/well of a 0.5 µg/mL solution) were added. On each individual microtiter plate the reference compound **2** was tested in parallel. The plates were incubated for 3 h at 25 °C and 350 rpm and then carefully washed four times with 150 µL/well assay buffer. After the addition of 100 µL/well of the horseradish peroxidase substrate 2,2'-azino-di-(3-ethylbenzothiazoline-6-sulfonic acid) (ABTS), the colorimetric reaction was allowed to develop for 4 min, then stopped by the addition of 2% aqueous oxalic acid before the optical density (OD) was measured at 415 nm on a microplate-reader (Spectramax 190, Molecular Devices, California, USA). The IC_{50} values of the compounds tested in duplicates were calculated with prism software (GraphPad Software, Inc., La Jolla, USA). The IC_{50} defines the molar concentration of the test compound that reduces the maximal specific binding of GbO3-PAA polymer to PapG-II_{LD} by 50%.

Crystallization of PapG-II_{LD}

All crystallization experiments were carried out with 10 mg/mL in sitting-drop vapor diffusion experiments in a 1:1 ratio of protein and precipitant. Initial crystals were optimized by streak seeding. PapG-II_{LD} in spacegroup P1 was crystallized at 4 °C with 30% PEG2000 MME, 0.15 M KBr and 40 mM SiaGal disaccharide. PapG-II_{LD} crystals in $P2_12_12$ were grown at 20 °C with 10% PEG10000 and 30 mM SiaαGal. The complex of PapG-II_{LD}/**2** crystallized with **0.15 M** Zn acetate**, 0.05 M $ZnCl_2$, 0.1**

M Tris pH 7.5, 13% PEG6000 at 20 °C. The complexes with compounds **13a**, **19**, or **25e** were crystallized at 20 °C with 10-14% PEG10000 (fivefold excess of ligand).

Data collection and structure determination

PapG-II_{LD} crystals were cryo-preserved by addition of glycerol to a final concentration of 20% (v/v) and flash cooled in liquid nitrogen. All measurements were done at the SLS beamlines X06DA and X06SA (Swiss Light Source, Paul Scherrer Institute, Switzerland) at 100 K. All data were integrated, indexed and scaled using the XDS software [27,28], 5% of the reflections were set aside as test set, respectively. Data collection statistics are summarized in Table S1. All structures were solved by molecular replacement using the crystal structure of PapG-II_{LD} (PDB ID: 1J8S) [10] as search models with the program Phaser [29]. Model building and structure refinement were performed with Coot [30], PHENIX [31], Buster-TNT [32], and Refmac [33]. Geometric restraints for the ligands were generated with grade [32] and PRODRG [34]. Refinement statistics are summarized in Table S1.

Table S1. Statistics on diffraction data and refinement of PapG-II$_{LD}$ and its complexes

	PapG-II$_{LD}$ (Apo-P21212)	**PapG-II$_{LD}$ (Apo-P1)**	**PapG-II$_{LD}$: cpd 2 (P212121)**	**PapG-II$_{LD}$: cpd 2 (P21)**
PDB Identifier	4Z3I	4Z3J	4Z3G	4Z3H
Wavelength (Å)	1.00000	0.99986	0.97618	1.00000
Resolution range (Å)	46.01 - 1.74 (1.80 - 1.74)*	52.07 - 2.5 (2.59 - 2.5)	57.66 - 1.45 (1.502 - 1.45)	20.07 - 1.50 (1.554 - 1.5)
Space group	P 21 21 2	P 1	P 21 21 21	P 1 21 1
Unit cell	54.99 83.98 45.36	50.76 56.56 70.94	56.27 72.61 94.86	47.35 53.77 48.32
□, β, γ (°)	90 90 90	112.8 102.8 88.1	90 90 90	90 117.58 90
Total reflections	201325 (9104)	42079 (4352)	790755 (35901)	224884 (19076)
Unique reflections	21920 (710)	23346 (2342)	69025 (6338)	33319 (3003)
Multiplicity	12.6 (12.8)	1.8 (1.9)	11.5 (5.7)	6.7 (6.4)
Completeness (%)	98.4 (84.2)	95.2 (95.5)	99.2 (92.6)	96.5 (87.4)
Mean I/sigma(I)	20.5 (2.6)	8.5 (2.5)	21.2 (1.3)	15.2 (2.9)
Wilson B-factor	22.0	22.9	20.0	14.9
R-merge	0.090 (0.221)	0.080 (0.308)	0.0577 (0.909)	0.075 (0.614)
R-meas	0.09426	0.1137	0.06035	0.08173
CC1/2	0.999 (0.994)	0.991 (0.849)	0.999 (0.554)	0.998 (0.876)
CC*	1 (0.999)	0.998 (0.958)	1 (0.844)	1 (0.966)
R-work	0.159 (0.302)	0.219 (0.302)	0.166 (0.295)	0.147 (0.222)
R-free	0.200 (0.282)	0.245 (0.347)	0.201 (0.296)	0.173 (0.267)
Number of atoms	3523	6613	7291	2018
macromolecules	1671	6395	3333	1678
ligands	---	---	86	31
water	230	218	561	304
Protein residues	197	783	399	197
RMS(bonds)	0.008	0.003	0.008	0.013
RMS(angles)	1.17	0.85	1.26	1.64
Ramachandran favored (%)	99	98	99	99
Ramachandran outliers (%)	0	0	0	0
Clashscore	2.43	2.08	3.29	0.90
Average B-factor	25.9	28.7	25.7	20.8

Table S1 (*Contd.*). Statistics on diffraction data and refinement of PapG-II_{LD} and its complexes

	PapG-II_{LD} : cpd 13a (P212121)	**PapG-II_{LD} : cpd 19 (P212121)**	**PapG-II_{LD} : cpd 25e (P212121)**
PDB Identifier	-	-	-
Wavelength (Å)	0.97618	0.97618	0.97618
Resolution range (Å)	57.17 - 1.8 (1.864 - 1.8)*	58.08 - 1.66 (1.719 - 1.66)	58.12 - 1.74 (1.802 - 1.74)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	55.77 71.82 94.44	56.62 73.43 94.91	56.59 73.59 94.74
a, β, γ (°)	90 90 90	90 90 90	90 90 90
Total reflections	454206 (42384)	618436 (62710)	492976 (25896)
Unique reflections	35852 (3501)	47485 (4693)	37834 (2260)
Multiplicity	12.7 (12.1)	13.0 (13.4)	13.0 (11.5)
Completeness (%)	99.91 (99.40)	99.97 (99.96)	91.52 (55.53)
Mean I/sigma(I)	10.93 (1.26)	14.50 (1.21)	14.83 (1.32)
Wilson B-factor	24.1	21.45	23.21
R-merge	0.1741 (1.988)	0.1237 (1.943)	0.1196 (1.744)
R-meas	0.1816	0.1289	0.1246
CC1/2	0.998 (0.531)	0.999 (0.561)	0.999 (0.413)
CC*	0.999 (0.833)	1 (0.848)	1 (0.764)
R-work	0.1876 (0.2890)	0.1662 (0.2915)	0.1576 (0.2924)
R-free	0.2333 (0.3646)	0.1985 (0.3094)	0.1926 (0.3198)
Number of atoms	7087	7237	7148
macromolecules	3339	3339	3333
ligands	88	100	104
water	340	460	382
Protein residues	400	400	400
RMS(bonds)	0.008	0.008	0.006
RMS(angles)	1.18	1.25	1.13
Ramachandran favored (%)	99	99	99
Ramachandran outliers (%)	0	0	0
Clashscore	2.97	2.97	1.78
Average B-factor	27.8	27.9	29

Generation of figures

The figures were generated with Maestro (**Schrödinger Release 2012:** Maestro, version 9.3.5, Schrödinger, LLC, New York, NY, 2012), VMD [35], Prism 5 (GraphPad Software Inc., San Diego, U.S.A.), Microsoft Excel 2013, or ChemBioDraw Ultra (12.0.3.1216).

References

[1] B. Lund, F. Lindberg, B. I Marklund, S. Normark. The PapG protein is the alpha-D-galactopyranosyl-(1–4) – beta-D-galactopyranose-binding adhesin of uropa-thogenic Escherichia coli. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 5898–5902.

[2] J. Kihlberg, S. J. Hultgren, S. Normark, G. Magnusson. Probing of the combining site of the PapG adhesin of uropathogenic Escherichia coli bacteria by synthetic analogs of galabiose. *J. Am. Chem. Soc.* **1989**, *111*(16), 6364-6368.

[3] H. C. Hansen, G. Magnusson. Synthesis of some amino and carboxy analogs of galabiose; evaluation as inhibitors of the pilus protein PapGJ96 from Escherichia coli. *Carbohyd. Res.* **1998**, *307*(3-4), 233-242.

[4] J. Ohlsson, J. Jass, B. E. Uhlin, J. Kihlberg, U. J. Nilsson. Discovery of potent inhibitors of PapG adhesins from uropathogenic Escherichia coli through synthesis and evaluation of galabiose derivatives. *ChemBioChem* **2002**, *3*(8), 772-779.

[5] J. Ohlsson, A. Larsson, S. Haataja, J. Alajääski, P. Stenlund, J. S. Pinkner, S. J. Hultgren, J. Finne, J. Kihlberg, U. J. Nilsson. Structure-activity relationships of galabioside derivatives as inhibitors of E. coli and S. suis adhesins: nanomolar inhibitors of S. suis adhesins. *Org. Biomol. Chem.* **2005**, *3*(5), 886-900.

[6] U. Nilsson, R. Johansson, G. Magnusson. Synthesis, Conformational Analysis and Comparative Protein Binding of a Galabioside and Its Thioglycoside Analogues. *Chem. Eur. J.* **1996**, *2*(3), 295-302.

[7] H. C. Hansen, G. Magnusson. Synthesis of selected aminodeoxy analogues of galabiose and globotriose. *Carbohyd. Res.* **1999**, *322*(3-4), 166-180.

[8] C. Abad-Zapatero, J. T. Metz. Ligand efficiency indices as guideposts for drug discovery. *Drug Discov. Today* **2005**, *10*(7), 464-469.

[9] R. Striker, U. Nilsson, A. Stonecipher, G. Magnusson, S. J. Hultgren. Structural requirements for the glycolipid receptor of human uropathogenic Escherichia coli. *Molec. Biol.* **1995**, *16*(5), 1021-1029.

[10] K. W. Dodson, J. S. Pinkner, T. Rose, G. Magnusson, S. J. Hultgren, G. Waksman. Structural basis of the interaction of the pyelonephritic E. coli adhesin to its human kidney receptor. *Cell* **2001**, *105*(6), 133-143.

[11] L. Beretta Piccoli. Synthesis Of Novel Papg-II Adhesis Antagonists, Based On 1-*p*-Methoxyphenyl-α-D-Galactopyranosyl-(1→4)-β-D-Galactopyranoside, And Their Evaluation As Potential Candidates For Treatment Of Pyelonephritis PhD thesis. University of Basel, **2013**.

[12] J. Ohlsson, G. Magnusson. Galabiosyl donors; efficient synthesis from 1,2,3,4,6-penta-*O*-acetyl-β-D-galactopyranose. *Carbohyd. Res.* **2000**, *329*, 49-55.

[13] S. Rabbani, X. Jiang, O. Schwardt, B. Ernst. Expression of the carbohydrate recognition domain of FimH and development of a competitive binding assay. *Anal. Biochem.* **2010**, *407*, 188-195.

[14] B. Schneider, J.-C. Gelly, A. G. de Brevern, J. Černý. Local dynamics of proteins and DNA evaluated from crystallographic B factors. *Acta Crystallogr.* **2014,** *D70*, 2413-2419.

[15] M. Harder, B. Kuhn, F. Diederich. Efficient Stacking on Protein Amide Fragments. *ChemMedChem* **2013**, *8*, 397-404

[16] M. O. Sinnokrot and C. D. Sherrill. Unexpected Substituent Effects in Face-to-Face π-Stacking Interactions. *J. Phys. Chem. A* **2003**, *107*(41), 8377-8379.

[17] W. Humphrey, A., Dalke, K. Schulten. VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **1996**, *14*, 33-38.

[18] S. Roy, N. Roy. Synthesis of a Blocked Tetrasaccharide Related to the Repeating Unit of the Antigen from Shigella dysenteriae Type 9 in the Form of Its Methyl (R)-Pyruvate Ester and 2-(Trimethylsilyl)Ethyl Glycoside. *J. Carbohyd. Chem.* **2003**, *22*(7), 521-535.

[19] J. López-Prados, F. Cuevas, N.-C. Reichardt, J.-L. de Paz, E- Q. Morales, M. Martín-Lomas. Design and Synthesis of Inositolphosphoglycan Putative Insulin Mediators. *Org. Biomol. Chem.*, **2005**, *3*, 764–786

[20] V. Kumar, N. Yadav, K. P. R. Kartha. In(III) triflate-catalyzed detritylation and glycosylation by solvent-free ball milling. *Carbohydr. Res.* **2014**, *397*, 18-26.

[21] H. Tsukamoto, T. Suzuki, Y. Kondo. Remarkable Solvent Effect on Pd(0)-Catalyzed Deprotection of Allyl Ethers Using Barbituric Acid Derivatives: Application to Selective and Successive Removal of Allyl, Methallyl, and Prenyl Ethers. *Synlett* **2007**, *20*, 3131–3136.

[22] S. Mesch, K. Lemme, H. Koliwer-Brandl, D. S. Strasser, O. Schwardt, S. Kelm, B. Ernst. Kinetic and thermodynamic properties of MAG antagonists. *Carbohydr. Res.* **2010**, *345*, 1348-1359;

[23] F. Bitsch, R. Aichholz, J. Kallen, S. Geisse, B. Fournier, J. M. Schlaeppi. Identification of natural ligands of retinoic acid receptor-related orphan receptor α ligand-binding domain expressed in Sf9 cells—a mass spectrometry approach. *Anal. Biochem.* **2003**, *323*, 139-149.

[24] J. Tellinghuisen Isothermal titration calorimetry at very low *c. Anal. Biochem.* **2008**, *373*, 395-397.

[25] W. B. Turnbull, A. H. Daranas. On the Value of *c*: Can Low Affinity Systems Be Studied by Isothermal Titration Calorimetry? J. *Am. Chem. Soc.* **2003**, *125*, 14859-14866.

[26] J. C. D. Houtman, P. H. Brown, B. Bowden, H. Yamagushi, E. Appella, L. E. Samelson, P. Schuck,. Studying multi-site binary and ternary protein interactions by global analysis of isothermal titration calorimetry data in SEDPHAT: Application to adaptor protein complexes in cell signaling. *Protein Sci.* **2007**, *16*, 30-42.

[27] W. Kabsch. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*(Pt 2), 133-144.

[28] W. Kabsch. Xds *Acta Crystallogr D Biol. Crystallogr.* **2010**, *66*(Pt 2), 125-132.

[29] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read. Phaser crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*(Pt 4): 658-674.

[30] P. Emsley, K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*(Pt 12 Pt 1), 2126-2132.

[31] P. D. Adams, R. W. Grosse-Kunstleve, L. W. Hung, T. R. Ioerger, A. J. McCoy, N. W. Moriarty, R. J. Read, J. C. Sacchettini, N. K. Sauter and T. C. Terwilliger. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*(Pt 11): 1948-1954.

[32] E. Blanc, P. Roversi, C. Vonrhein, C. Flensburg, S. M. Lea and G. Bricogne. "Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT". *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*(Pt 12 Pt 1), 2210-2221.

[33] G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*(Pt 4), 355-367.

[34] A. W. Schuttelkopf, D. M. van Aalten. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*(Pt 8), 1355-1363.

[35] W. Humphrey, A. Dalke, K. Schulten. "VMD - Visual Molecular Dynamics". *J. Molec. Graphics*, **1996**, *14*, 33-38.

3.2.3 New PapG-II Antagonists by a Fragment-Based Approach

Giulio Navarra, Pascal Zihlmann, Brigitte Fiege, Said Rabbani, Rachel Hevey, Beat Ernst

[a] Institute of Molecular Pharmacy, Pharmacenter, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland

[b] Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Keywords:

Fragment-based drug design, PapG, Antiadhesive, Medicinal chemistry, Glycosides, Proton relaxation enhancement

Abstract

The PapG-II adhesion, a surface proteins of *E. coli,* is involved in the bacterial adhesion to epithelial cells in the human urinary tract. Their expression correlates with an increased risk of developing pyelonephritis, a potentially life-threatening disease. Therefore, antagonizing PapG-II offers a meaningful alternative to an antibiotic therapy. Despite intensive medicinal chemistry efforts, the most promising reported antagonist – carbohydrate mimic **4** – exhibits only mid-micromolar affinity. Based on previous studies, we predicted that a substantial affinity increase would be possible by linking antagonist **4** with fragments binding outside the carbohydrate recognition domain of PapG-II. Thus, a library of fragment-sized compounds was screened for binding to PapG-II by NMR. The best binder **6**, recognizing a second-site proximal to the main binding pocket, was connected to antagonist **4** via linkers of variable length. With the best representative, compound **43**, a clear increase in affinity was reached, albeit at the cost of higher complexity. Further thermodynamic and structural studies are ongoing to reveal the structural background for the improved affinity.

Introduction

Carbohydrate-lectin interactions play major biological roles in as different areas as inflammation, cell-cell recognition and bacterial adhesion [1]. Among bacteria, uropathogenic *E. coli* (UPEC) is a classic example of a pathogen that utilizes a carbohydrate epitope as a target to adhere to the host cell and start colonization. UPEC strains causing infections in the upper urinary tract use PapG adhesins, which exist in three molecular variants (classes I-III) [2-6]. The PapG-II adhesin is of particular medical interest, as it is strongly associated with pyelonephritis in humans [7-10], a potentially life-threatening disease and a frequent complication during pregnancy [11]. PapG-II recognizes preferentially the globoside GbO4, present in the upper urinary tract of humans. Its binding epitope is oligosaccharide **1** (Figure 1) consisting of the D-Galα(1-4)-D-Gal core, flanked by β(1-3)-linked D-Glc*N*Ac on the non-reducing end, and β(1-4)-linked D-Glc on the reducing end [12].

Since for the current treatment of pyelonephritis antibiotic are frequently used [13], bacterial resistance is rapidly increasing [13-15]. Therefore, the development of new therapeutic options is an urgent need. Preventing bacterial adhesion by blocking PapG-II interaction with its human target GbO4 is a promising new strategy not only to combat the infection but also to reduce the risk of resistance [16].

Despite intensive synthetic efforts and detailed structural information on PapG-II (apo structure, PDB code: 18JS; co-crystal with **1,** PDB code: 18JR [17]), the design of potent small-molecule inhibitors has been only partially successful. Major achievements are summarized in Figure 1. Starting from the natural epitope **1**, the minimal binding epitope **2** was identified [18]. The corresponding β-methyl galabioside (**3**) exhibited a 4-fold higher affinity [12]. With an aromatic aglycone (→ **4**), an even 9-fold increase of affinity was observed [19]. However, further small improvements of affinity could only be realized at the cost of much higher complexity (→ **5**) [19]. Finally, for the therapy of pyelonephritis, the affinity of the to-date available PapG-II antagonists is still insufficient and has to be improved by almost a factor 100.

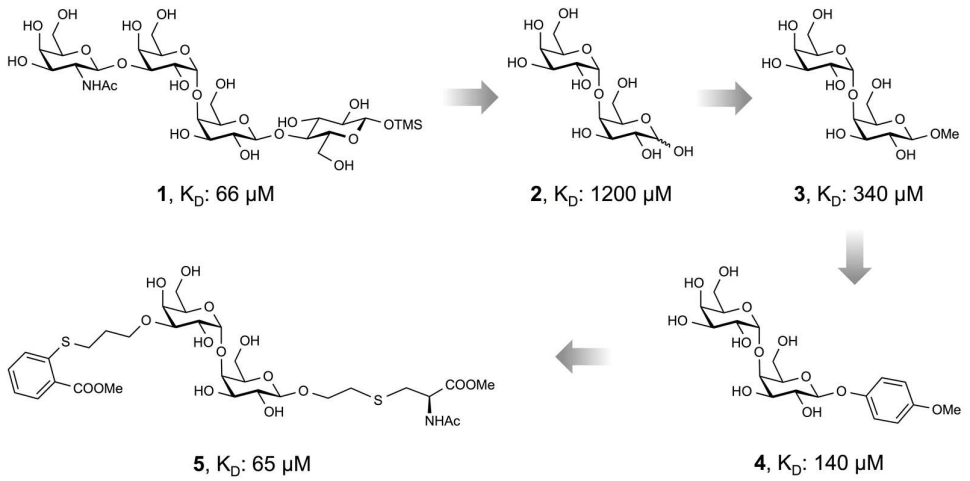


Figure 1. Relevant PapG-II ligands [18]: tetrasaccharide epitope of GbO4 (**1**, natural ligand); D-galabiose (**2**) (minimal binding epitope); methyl β-D-galabioside (**3**); *p*-methylphenyl β-D-galabioside (**4**); the yet most potent published antagonist **5**.

Recently, we published the results of two fragment-based campaigns that provided high affinity antagonists for MAG [20] and E-selectin [21]. As in those cases, the development of highly potent PapG-II antagonists is hampered by a shallow, water-accessible and highly polar binding site [17]. Therefore, a similar strategy was applied to discover fragments binding in close proximity of the carbohydrate recognition domain (CRD), which later on should be linked to the carbohydrate core structure.

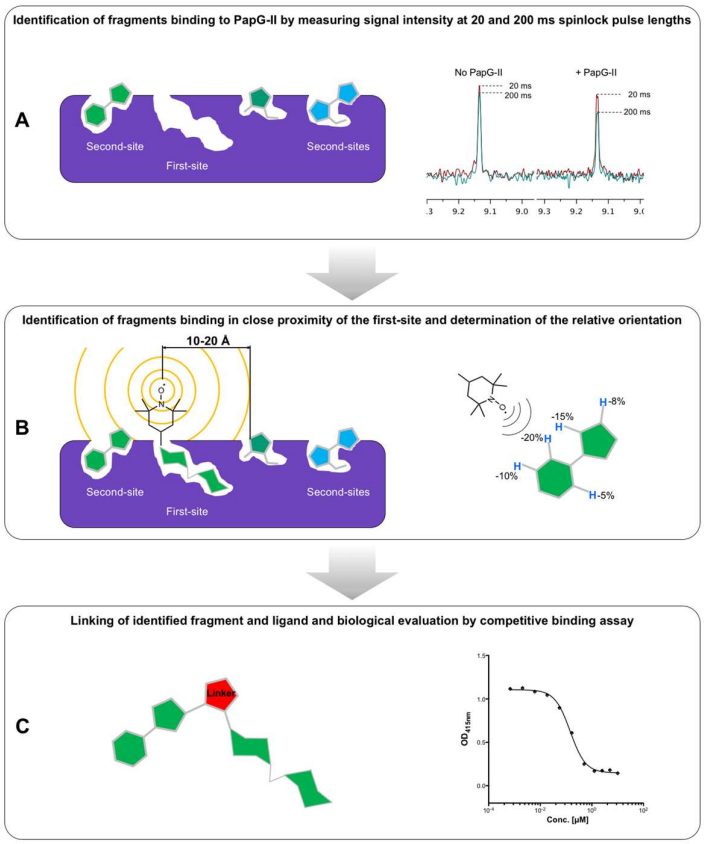


Figure 2. Fragment-based second-site ligand search: A) Fragments binding to PapG-II are identified by spin-locked filtered transverse magnetization (T2) decay experiments; B) Fragments binding in close proximity of the first-site ligand are identified by proton relaxation enhancement (PRE), which permits to deduce also their orientation relative to the first-site ligand; C) Different linkers are used to connect identified fragments and the first-site ligand, resulting in "linked compounds" which are evaluated for binding to PapG-II.

Results and Discussion

For the first step, namely the identification of fragments binding to PapG-II, a library of sixty Rule-of-Three [22] compliant fragments was screened by NMR. Detailed experimental procedures are reported in the supporting information. By this approach, nine hits (**6** – **14**) were identified (Figure 3-A).

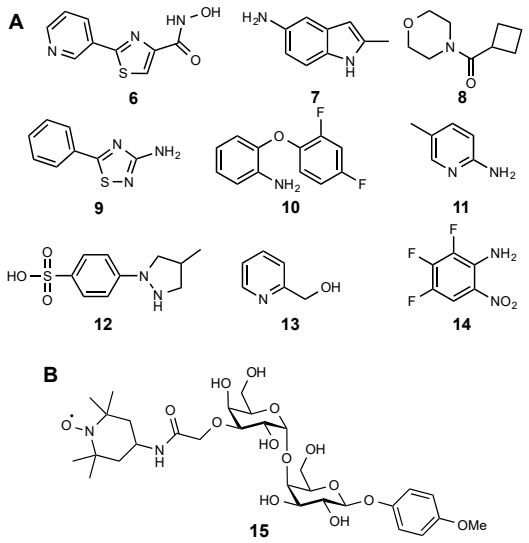


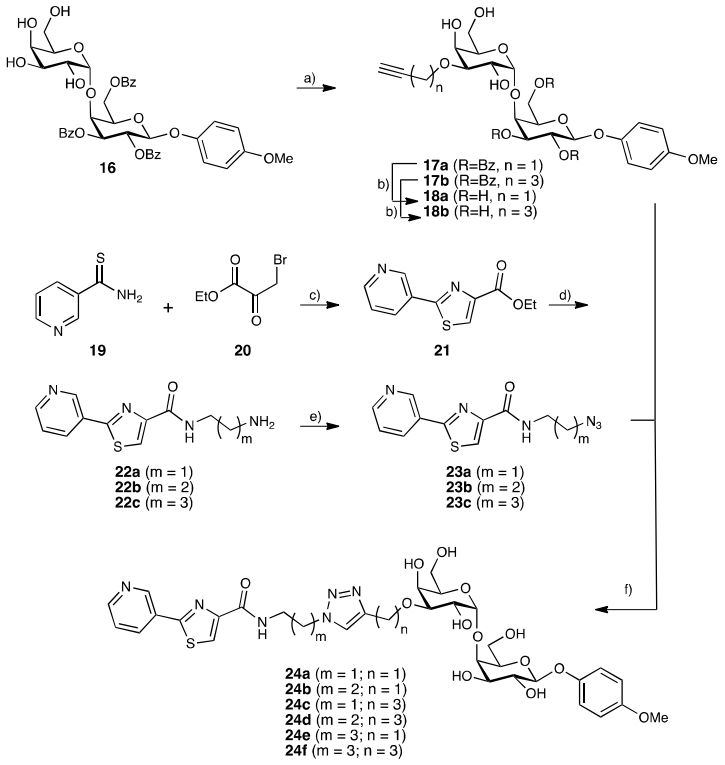
Figure 3. A) Fragments binding to PapG-II identified by NMR screening; B) TEMPO-labeled first site ligand **15** used to identify those fragments binding to PapG-II in close proximity to the first-site ligand.

To identify the fragments binding in close proximity of **4**, paramagnetic relaxation enhancement (PRE) experiments were performed (Figure 2-B) [23]. For this approach, the first-site ligand **4** (Figure 1) was labeled with a paramagnetic moiety (in this case TEMPO) to yield **15** (Figure 3-B, for the synthesis see supporting information). The TEMPO label was introduced at the 3'-position, because this is a tolerated exit vector [24].

With the spin-labeled antagonist **15** those fragments binding simultaneously to and in close proximity of the first-site ligand were identified. The unpaired electron on the TEMPO label accelerates substantially the transverse relaxation of protons (PRE effect) in its vicinity (~10 Å distance) [25], thus reducing the intensity of the proton signals. This effect is distance dependent and allows not only to identify proximal binders, but also to uncover their orientation relative to the paramagnetic center (Figure 2-B and supporting information). Although compounds **6** and **12** were identified, the later had to be excluded due to chemical stability issues. The signal reduction of all protons of fragment **6** revealed the hydroxamic acid moiety is pointing towards the first-site ligand and can therefore be used as a handle for linking.

For linking fragment **6** with the first-site ligand, several linking strategies are available. An ideal linker should allow the linked compound to adopt the proper

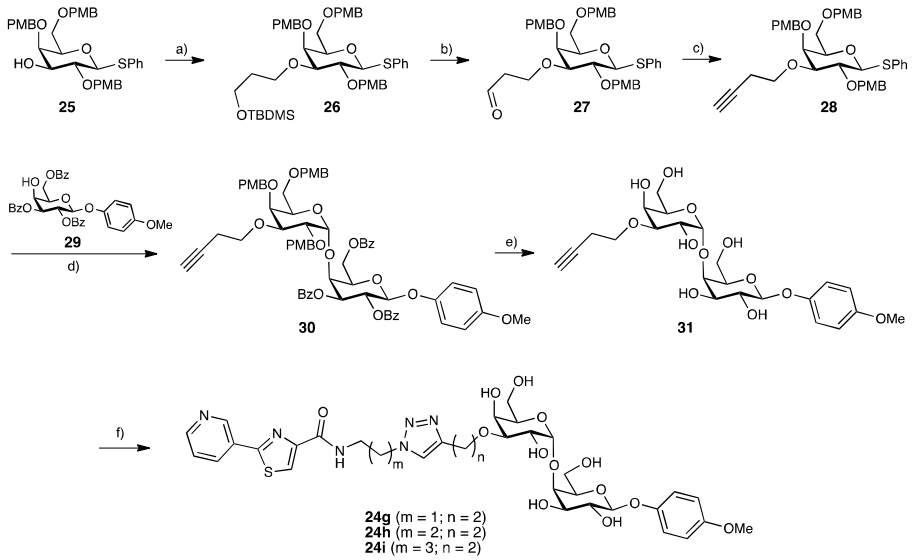
conformation and orientation for binding with minimal strain [26]. On the other hand, excessive flexibility should be avoided, causing an entropic penalty and thus reducing the affinity. And finally, ideally, the linker would also provide additional interactions with the protein. The first library was synthesized by copper-catalyzed azide-alkyne coupling (Schemes 1 & 2) [27]. In order to screen for different linker lengths, the homologous azides **23a-c** and a series of terminal alkynes were generated (Scheme 1). The former was built from compound **21**, which was obtained by a slightly modified known procedure from inexpensive **19** and **20** [28]. When **21** was heated up in the corresponding neat diamine, amines **22a-c** were obtained in good yields. Diazo-transfer with triflyl azide [29] finally afforded the azides **23a-c**. For the alkynes series, galabioside **16** was regioselectively alkylated in the 3'-position with the appropriate bromoalkyne. Whereas the alkylation with 3-bromopropyne and 5-bromo-1-pentyne worked well, 4-bromobutyne did not react with compound **16**.



Scheme 1. a) i. Bu_2SnO, toluene, reflux, 2-4h; ii. bromoalkyne, TBABr, toluene, 70-80 °C, overnight, 50-60%; b) i. NaOMe, MeONa, rt, 66-90%; c) EtOH, reflux, 70%; d) diamine, solvent-less, 70 °C, 45 min, 60%-quant; e) $CuSO_4$, water, MeOH, TfN_3, rt, 10-30 min, 89%-quant.; f) **23a-c**, $CuSO_4$, sodium ascorbate, *t*-BuOH:H_2O = 2:1, rt, 1-2h, 66-80%.

Therefore, a different route was devised for introducing the butynyl moiety (Scheme 2). Starting from the partially deprotected galactoside **25**, the *tert*-butyldimethylsilyl

ether of 3-bromo-1-propanol was introduced by nucleophilic substitution, to afford **26**. Desilylation, followed by oxidation with Dess-Martin periodinane [30] gave access to aldehyde **27**, which was converted into the terminal alkyne **28** under Corey-Fuchs conditions [31]. Glycosylation of galactoside **29** [32] with **28** provided galabioside **30**. The *p*-methoxybenzyl (PMB) protecting groups were removed with 10% trifluoroacetic acid (TFA), and the benzoyl esters were cleaved under Zemplén conditions [33], affording **31**. Click chemistry with the azides **23a-c** finally afforded the test compounds **24g-i**.



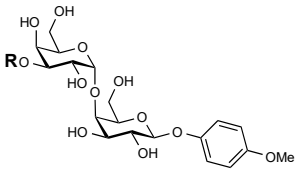
Scheme 2. a) NaH, 3-bromo-1-propanol TBDMS ether, DMF, 0 °C to rt, 71%; b) i. TBAF 1M, THF, rt, 1h, 88%; ii. Dess-Martin periodinane, DCM, rt, 40 min, 86%; c) i. CBr_4, PPh_3, DCM, 0 °C; ii. BuLi, THF, 90 min, 48%; d) TMSOTf, NIS, DCM/Et_2O, -55 °C, 57%; e) i. TFA 10% in DCM, rt, 5 min, quant.; ii. MeONa, MeOH, rt, 2h, 70%; f) **23a-c**, $CuSO_4$, sodium ascorbate, *t*-BuOH:H_2O = 2:1, rt, 1-2h, 48-72%.

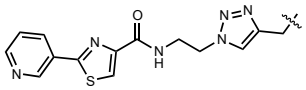
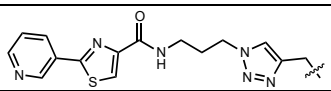
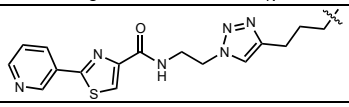
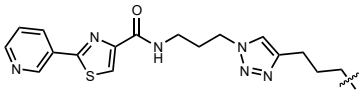
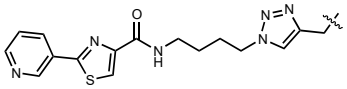
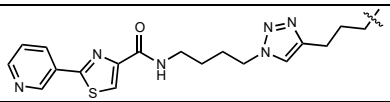
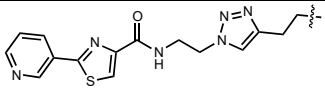
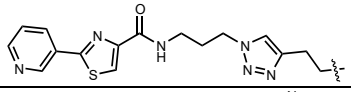
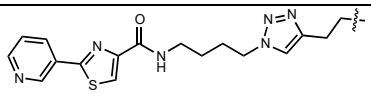
The affinity of the test compounds **24a-i** for PapG-II was determined in a competitive binding assay [34]. Although the majority of the triazole-linked compounds exhibited slightly better affinity then the reference compound **4**, only a minor improvement could be reached.

The limited affinity increase – the best compound **24i** exhibits only a two-fold stronger binding as compared to **4** – could be due to the rigidity of the linker preventing the fragment to reach the proximal second binding site in an optimal spatial orientation. In order to verify this assumption, we designed a second series compounds, with a more flexible linker.

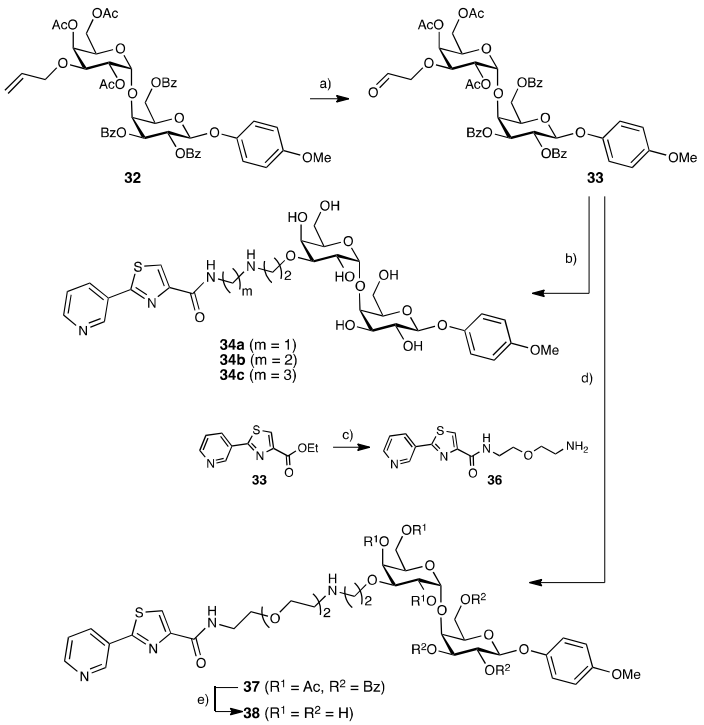
The linker for the second library was established by reductive amination (Scheme 3), leading to a flexible, sterically less demanding linker. Galabioside **32** was obtained from **16** by regioselective allylation of the 3'-position, followed by peracetylation [35]. By treatment with OsO_4 and $NaIO_4$ the terminal alkene function was converted into aldehyde **33**.

Table 1. Affinity values from the competitive binding assay for the triazole library



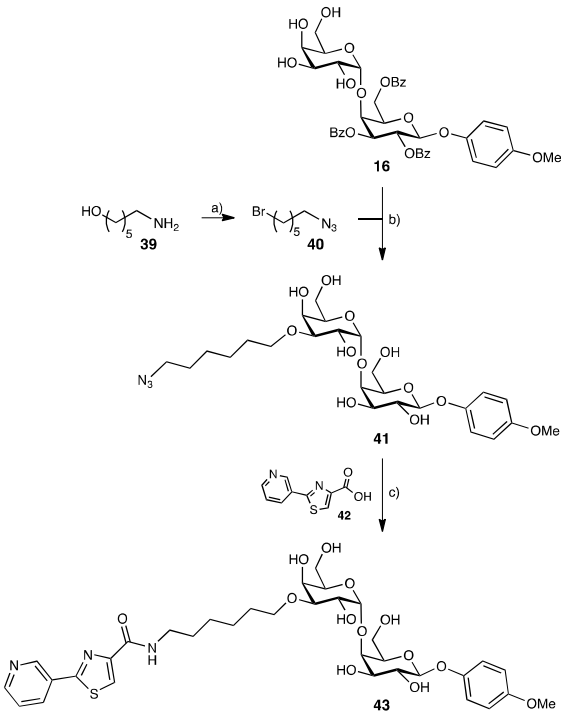
Entry	Cpd	R	IC50 (µM)	rIC50
1	**4**	H	382	1.00
2	**24a**		352	0.92
3	**24b**		348	0.91
4	**24c**		454	1.19
5	**24d**		389	1.02
6	**24e**		336	0.88
7	**24f**		267	0.70
8	**24g**		225	0.59
9	**24h**		323	0.84
10	**24i**		204	0.53

Reductive amination with amines **22a-c** (scheme 1) and **36** (obtained from the ester **33** by direct amidation with 2,2'-(ethylenedioxy)bis(ethylamine) and subsequent deprotection provided test compounds **34a-c** and **38**.



Scheme 3. a) OsO_4, $NaIO_4$, 2,6-lutidine, dioxane/water 3:1, rt, overnight, 70%; b) i. amine **22a** and **22c**, $NaBH_3CN$, MeOH, rt; ii. MeONa, MeOH, rt, 23-26% or i. amine **22b,** $NaBH(AcO)_3$, DCE, rt, 63%; ii. MeONa, MeOH, rt, 4h, 93%; c) 2,2'-(ethylenedioxy)bis(ethylamine), DMF, 70 °C, 45 min, 53%; d) **36**, $NaBH(AcO)_3$, DCE, rt, 2h, 67%; ii. MeONa, MeOH, rt, 2h, 97%.

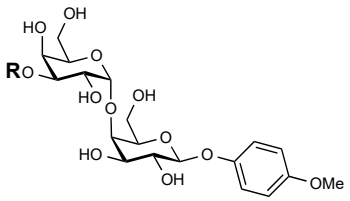
In the amine series, the elongation of the carbon linkers by two carbons led to a 10-fold improvement in affinity. Obviously, with antagonist **34b** (Table 2, entry 3) exhibiting the elongated linker a better positioning of fragment **6** becomes possible, whereas the linker in **34a** (Table 2, entry 2) is just too short. On the other side, a further elongation of the linker (→ **34c** & **38**) resulted in a loss of affinity. One plausible reason is that the secondary amine of the linker, which is charged at physiological pH, generates repulsion with Lys172, which is part of the rim of the binding site and faces the 3'-position of the galabiose (Figure S1).

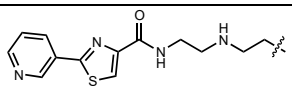
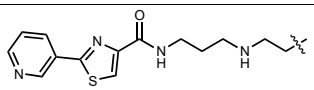
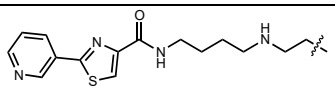
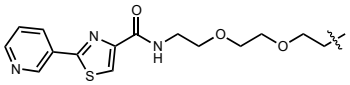
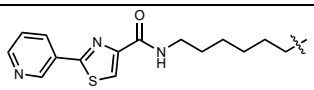


Scheme 4. a) i. TfN_3, $CuSO_4$, $NaHCO_3$, MeOH, water, 15 min, 88%; ii. CBr_4, PPh_3, DCM, 73%; b) i. Bu_2SnO, toluene, reflux, 2h; ii. 4, TBABr, toluene, 75 °C, 2 d; iii. MeONa, MeOH, rt, 5h, 40% over 3 steps; c) i. Pd/C 10%, H_2, MeOH, rt, 97%; ii. HOBt, HBTU, DIPEA, DMF, rt, 1.5h, 59%.

To further improve affinity of antagonist **34b**, the bioisostere **43** was synthesized (Scheme 4), where the charged amino group is replaced by a neutral methylene group. This should reduce the costs of desolvation but could beneficially or non-beneficially influence the linker length. First, the bromoazide **40** was synthesized starting from 6-amino-1-hexanol (**39**) by diazo-transfer [29] followed by Appel bromination [36]. Regioselective alkylation of the 3'-position of **16** with bromide **40** afforded **41**. Finally, reduction of the azide to the amine followed by HOBt/HBTU-promoted amide bond formation with **42** [37] yielded **43**. Affinity measurements, however, revealed only a moderate increase in affinity (Table 2, entry 6) could be reached.

Table 2. Affinity values from the competitive binding assay for the secondary amines library



Entry	Cpd	R	IC_{50} (µM)	rIC_{50}
1	**4**	H	382	1
2	**34a**		2’100	5.5
3	**34b**		197	0.51
4	**34c**		332	0.87
5	**38**		733	1.92
6	**43**		144	0.38

To verify our interpretations, compound **43** was co-crystallized with PapG-II and studied by X-ray. Disappointingly, the pyridylthiazolyl moiety is outcompeted by crystal contacts and cannot be visualized in the electron density map (not shown). At the same time, the binding epitopes of compounds **43** and **4** were determined by STD-NMR (Figure 4). Evidently, the interaction is mainly driven by the galabiose core, with the aromatic moieties at the reducing and non-reducing end interacting only loosely. Interestingly, while the 4-methoxyphenyl in the anomeric position is able to increase the affinity of **4** as compared to galabiose, the same does not hold true for the pyridylthiazolyl moiety in the 3’-position of **43**, despite a comparable contribution to the binding epitope is suggested by the STD-NMR data. Most likely this is due to the high flexibility of the linear linker, related to high entropy costs. Attempts to confirm this interpretation by isothermal titration calorimetry (ITC) experiments have been initiated.

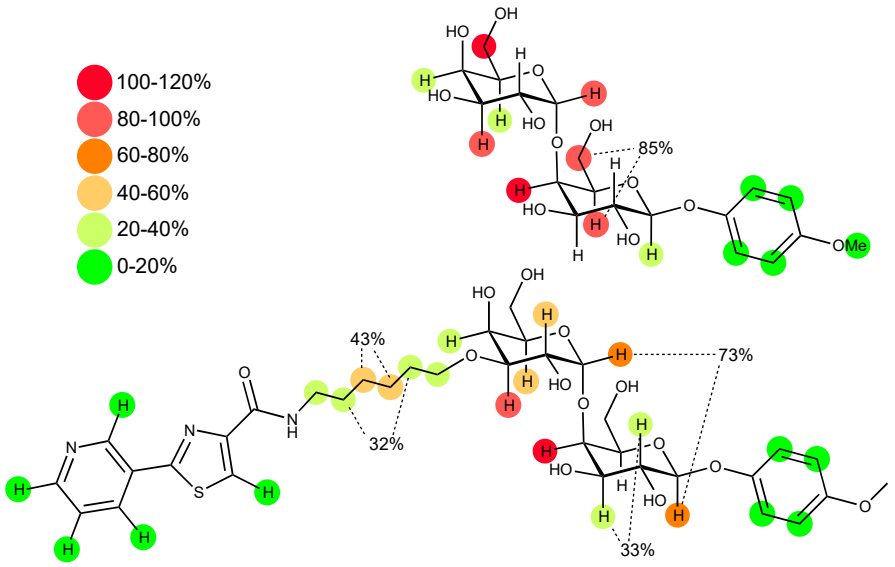


Figure 4. Binding epitopes of compounds **4** (above) and **43** (below). Due to the large overlap of the signals in the NMR spectra, not all protons can be assigned. Overlapping signals are indicated by dashed lines and the overall saturation is indicated.

Conclusion

In conclusion, we successfully screened a library of fragments by NMR and identified compound **6** as second-site ligand suitable for linking to the first-site ligand **4**. In total, fourteen antagonists linked to the 3'-position to fragment **6** were synthesized and their binding affinities for PapG-II determined. By a bioisosteric substitution of the linker of the most potent compound **34b**, antagonist **43** with a approximately 3-fold increased affinity as compared to the lead structure **4** was discovered. Likely, the rather low improvement in affinity is due to the high flexibility of the linker, causing a high entropy penalty, a hypothesis we are currently studying by ITC. In addition, structural information would help to identify appropriately rigidified linkers. So far, X-ray studies were compromized by crystal contacts and therefore new crystallization conditions are currently under investigation. We are convinced that by following up on our results, highly potent ligands can be discovered, as we already showed when targeting MAG and E-selectin [20,21].

References

[1] N. Sharon, *Trends Biochem. Sci* **1993**, *18*, 221-226.

[2] H. Hoschützky, F. Lottspeich, K. Jann. *Infect. Immun.* **1989**, *57*, 76-81.

[3] J. R. Johnson, A. L. Stell, N. Kaster, C. Fasching, T. T. O'Bryan. *Infect Immun* **2001**, *69*, 2318–2327.

[4] N. Strömberg, B.-I. Marklund, B. Lund, D. Ilver, A. Hamers, W. Gaastra, A. Karlsson, S. Normark. *EMBO J.* **1990**, *9*, 2001–2010.

[5] N. Strömberg, P.-G. Nyholm, I. Pascher, S. Normark. *Proc. Natl. Acad. Sci. USA* **1991***, 88,* 9340–9344.

[6] S. D. Manning, L. Zhang, B. Foxman, A. Spindler, P. Tallman, C. F. Marrs. *Clin. Diagn. Lab. Immunol.* **2001**, *8*, 637–640.

[7] I. M. Johanson, K. Plos, B. I. Marklund, C. Svanborg. *Microb. Pathogenesis* **1993**, *15*, 121-129.

[8] J. A. Roberts, B. I. Marklund, D. Ilver, D. Haslam, M. B. Kaack, G. Baskin, M. Louis, R. Möllby, J. Winberg, S. Normark. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11889-11893.

[9] C.C. Tseng, J.J. Huang, M.C. Wang, A.B. Wu, W. C. Ko, W.C. Chen, J.J. Wu. *Kidney int.* **2007**, *71*, 764-770.

[10] J. R. Johnson. *papG* Alleles among *Escherichia coli Infect. Immun.* **1998**, *66*, 4568-4571.

[11] J. B. Hill, J. S. Sheffield, D. D. McIntire, G. D. Wendel. *Obstet. Gynecol.* **2005**, *105*, 18-23.

[12] A. Larsson, J. Ohlsson, K. W. Dodson, S. J. Hultgren, U. Nilsson, J. Kihlberg. *Bioorg. Med. Chem.* **2003**, *11*, 2255-2261.

[13] [Urinary Tract Infection, Chapter 11] Litwin M. S., Saigal C. S., editors. *Urologic Diseases in America.* US Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases. Washington, DC: US Government Printing Office, **2012**; NIH Publication No. 12-7865.

[14] K. Gupta, D. Scholes, W. E. Stamm. *J. Amer. Med. Assoc.* **1999**, *281*, 736-738.

[15] L. G. Greer, S. W. Roberts, J. S. Sheffield, V. L. Rogers, J. B. Hill, D. D. Mcintire, G. D. Wendel. *Infect. Dis. Obstet. Gynecol.* **2008** ID 891426.

[16] N. Sharon. *Biochim. Biophys. Acta* **2006**, *1760*, 527-537.

[17] K. W. Dodson, J. S. Pinkner, T. Rose, G. Magnusson, S. J. Hultgren, G. Waksman. *Cell* **2001**, *105*, 733 – 743.

[18] R. Striker, U. Nilsson, A. Stonecipher, G. Magnusson, S. J. Hultgren. *Molec. Biol.* **1995**, *16*, 1021-1029.

[19] J. Ohlsson, J. Jass, B. E. Uhlin, J. Kihlberg, U. J. Nilsson. *ChemBioChem* **2002**, *3*, 772-779.

[20] Shelke, S.; Cutting, B.; Jiang, X.; Koliwer-Brandl, H.; Strasser, D.; Schwardt, O.; Kelm, S.; Ernst, B. *Angew. Chem., Int. Ed.* **2010**, *49*, 5721−5725.

[21] Jonas Egger, Céline Weckerle, Brian Cutting, Oliver Schwardt, Said Rabbani, Katrin Lemme, and Beat Ernst *J. Am. Chem. Soc.* **2013**, *135*, 9820–9828.

[22] M. Congreve, R. Carr, C. Murray, H. Jhoti. *Drug Discov Today* **2003**, *8*, 876-7.

[23] W. Jahnke, L. B. Perez, C. G. Paris, A. Strauss, G. Fendrich, C. M. Nalin. Second-Site NMR Screening with a Spin-Labeled First Ligand *J. Am. Chem. Soc.* **2000**, *122*, 7394-7395.

[24] J. Ohlsson, J. Jass, B. E. Uhlin, J. Kihlberg, U. J. Nilsson. *ChemBioChem* **2002**, *3*, 772-779.

[25] Bertini, I.; Luchinat, C.; Parigi, G.; Pierattelli, R. ChemBioChem **2005**, *6*, 1536-1549.

[26] T. G. Davies, M. Hyvönen (Eds.). *Fragment-Based drug Discovery and X-ray Crystallography*. Springer **2012**, XII.

[27] F. Himo, T. Lovell, R. Hilgraf, V. V. Rostovtsev, L. Noodleman, K. B. Sharpless, V. V. Fokin. *J. Am. Chem. Soc.*, **2005**, *127*, 210-216.

[28] P. Gillespie et al., US. Patent 2007/0167622 A1, 2007.

[29] A.Titz, Z. Radic, O. Schwardt, B. Ernst. *Tetr. Lett.* **2006**, *47*, 2383-2385.

[30] D. B. Dess, J. C. Martin. Readily accessible 12-I-5 oxidant for the conversion of primary and secondary alcohols to aldehydes and ketones. *J. Org. Chem.* **1983**, *48*, 4155-4156.

[31] E. J. Corey, P. L. Fuchs. *Tetrahedron Lett.* **1972**, *13*, 3769-3772.

[32] J. Ohlsson, G. Magnusson. *Carbohyd. Res.* **2000**, *329*, 49-55.

[33] W. Zerong. Zemplén Deacetylation. *Comprehensive Organic Name Reactions and Reagents*. John Wiley & Sons **2010**, 691, 3123-3128.

[34] S. Rabbani, X. Jiang, O. Schwardt, B. Ernst. *Anal. Biochem.* **2010**, *407*, 188-195.

[35] J. Ohlsson, J. Jass, B. E. Uhlin, J. Kihlberg, U. J. Nilsson. *ChemBioChem* **2002**, *3*, 772-779.

[36] R. Appel. *Angew. Chem.* **1975**, *14*, 801-811.

[37] S. Schlimme, A.-T. Hauser, V. Carafa, R. Heinke, S. Kannan, D. A. Stolfa, S. Cellamare, A. Carotti, L. Altucci, M. Jung, W. Sippl. *ChemMedChem* **2011**, *6*, 1193-1198.

Supporting Information

3.2.3 New PapG-II Antagonists by a Fragment-Based Approach

Giulio Navarra, Pascal Zihlmann, Brigitte Fiege, Said Rabbani, Rachel Hevey, Beat Ernst

[a] Institute of Molecular Pharmacy, Pharmacenter, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland

[b] Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Table of Contents

1. General methods:

All NMR experiments were performed on a Bruker AVANCE III 500 MHz NMR spectrometer equipped with a BBO room temperature probe head with Z-gradients at

a temperature of 298 K. Spectra were acquired and processed with Topspin 2.1 (Bruker BioSpin, Switzerland) and analyzed with MestReNova 6.1.1 (Mestrelab Research, Spain). Assignment of ^{1}H and ^{13}C NMR spectra was achieved using 2D methods (COSY, HSQC, TOCSY and HMBC). Chemical shifts are expressed in ppm using residual $CHCl_3$, CHD_2OD, H_2O, and TMS as references. Optical rotations were measured on a Perkin-Elmer Polarimeter 341. IR spectra were recorded on a Perkin Elmer Spectrum One FT-IR spectrometer as KBr pellets, or thin films. Electron spray ionization mass spectra (ESI-MS) were obtained on a Waters micromass ZQ. Reactions were monitored by TLC using glass plates coated with silica gel 60 F254 (Merck) and visualized by using UV light and/or by heating to 150 °C for 5 min with aq. $KMnO_4$ solution or a molybdate solution (a 0.02 M solution of ammonium cerium sulfate dihydrate and ammonium molybdate tetrahydrate in aq. 10% H_2SO_4). Column chromatography was performed on a CombiFlash Companion (Teledyne-ISCO, Inc.) using RediSep® normal phase disposable flash columns (silica gel) or handmade reversed phase columns packed with LiChroprep® RP-18 (Merck, 40-63 µm). LC-MS separations were carried out using Sunfire C18 columns (analytical: 2.1 × 50 mm, 3.5 µm; preparative: 19 × 150 mm, 5.0 µm) on a Waters 2525 LC, equipped with a Waters 2996 photodiode array and a Waters micromass ZQ for detection. Hydrogenation reactions were performed in a shaking apparatus (Parr Instruments Company, Moline, Illinois, USA) in 250 mL or 500 mL bottles with H_2 pressure as stated, or in conventional flasks equipped with a balloon. Solvents and phosphate buffer solutions were purchased from Fluka or Acros. Solvents were dried prior to use where indicated. Methanol (MeOH) was dried by refluxing with sodium methoxide and distilling. Dry MeOH was stored on activated (400 °C under high vacuum) MS 3Å. Dichloromethane (DCM) was dried by filtration through activated Al_2O_3 (Fluka, type 5016 A basic, heated at 400 °C under vacuum for 1 h and then cooled under argon) and stored over activated MS 3Å. Tetrahydrofurane (THF) was dried by distillation from sodium/benzophenone. Dry *N,N*-dimethylformamide (DMF) was purchased from Acros.

2. NMR Experiments

2.1 Protein preparation

A deuterated phosphate buffer pH 7.4 (corrected for deuterium effect) containing 150 mM NaCl was prepared (= NMR buffer). Lyophilized PapG-II protein was dissolved in NMR buffer. After centrifugation at 13000 RPM in an Eppendorf 5427 R centrifuge, the supernatant was collected and the concentration of the protein was determined by UV absorption (NanoDrop ND-100 spectrophotometer, NanoDrop Technologies, USA) as µM or mg/ml. Alternatively, a 4 mg/mL stock in phosphate buffer 20 mM, containing 150 mM NaCl was diluted with NMR buffer. A 1D NMR spectrum of 15 µM PapG-II in NMR buffer was recorded to control the correct folding of the protein.

2.2 NMR fragment screening

Stocks of the cocktail mixtures of fragments (2 to 6 compounds per mixture) were prepared in deuterated DMSO (DMSO-*d6*) at 10 mM final concentration per fragment. NMR samples for the screening were prepared in 3 mm tubes (Hilgenberg, Germany) by adding the cocktail mixture at a final concentration of 250 µM per fragment to either NMR buffer, or to PapG-II at a final concentration of 15 µM in NMR buffer. TSP-*d4* (3-(trimethylsilyl)-2,2',3,3'-tetradeuteropropionic acid, Armar Chemicals, Switzerland) was added as an internal reference. Identical samples, except for the absence of PapG-II protein, were prepared in 5 mm tubes (Bruker BioSpin, Switzerland) and were used for recording the reference spectra of the fragments.

$T_{1\rho}$ spinlock-filtered experiments [S1] were measured with a standard Bruker pulse sequence with excitation sculpting (es) for suppression of residual water [S2] and modified by insertion of a continuous wave spinlock pulse between the 90° observe pulse and the es sequence. For the fragment screening, $T_{1\rho}$ experiments of the NMR samples in absence and presence of PapG-II protein were measured with spinlock pulses of 20 ms and 200 ms at RF field strength of 2.6 kHz. Each experiment was recorded with 512 scans, a spectral width of 12 ppm, 32k data points, an acquisition time of 2.72 s and an interscan delay of 10 s to ensure efficient longitudinal relaxation between the experiments. The difference in signal intensity at 20 ms and 200 ms spinlock pulse length was recorded in absence ($I_{red}^{Ref}=I_{200\,ms}^{Ref}/I_{20\,ms}^{Ref}$) and in presence of PapG-II ($I_{red}^{PapG}=I_{200\,ms}^{PapG}/I_{20\,ms}^{PapG}$). The percent of signal reduction ***R*** due to the protein was calculated according to equation 1:

$$\boldsymbol{R}=\frac{I^{PapG}_{200\,ms}/I^{PapG}_{20\,ms}}{I^{Ref}_{200\,ms}/I^{Ref}_{20\,ms}}*100=\frac{I^{PapG}_{red}}{I^{Ref}_{red}}*100 \qquad \text{eq. 1}$$

When $\boldsymbol{R} \geq 10\%$ the fragment was considered as a binder.

For the validation of initial hits from cocktail mixtures and for the additional spin-spin-relaxation experiments with spin-labeled first-site ligand **15**, samples with single fragments were prepared. PapG-II was present at 20 µM concentration in NMR buffer. A 20 mM stock solution of the fragment to test in DMSO-*d6* was prepared and added to a final sample concentration of 500 µM, resulting in a final DMSO-*d6* concentration of 2.5%. The same experimental scheme as described for the screening was used and the signal was recorded at 20 ms and 200 ms spinlock time, in absence and presence of spin-labeled compound **15** at 3 mM concentration. For each experiment, 256 scans were recorded. The ***R***-value was calculated as shown in equation 2:

$$\boldsymbol{R}=\frac{I^{SL}_{200\,ms}/I^{SL}_{20\,ms}}{I^{PapG}_{200\,ms}/I^{PapG}_{20\,ms}}*100=\frac{I^{SL}_{red}}{I^{PapG}_{red}}*100 \qquad \text{eq. 2}$$

where $I_{red}^{SL}=I_{200\,ms}^{SL}/I_{20\,ms}SL$ is the ratio between signal intensity of the fragment at 20 ms and 200 ms spinlock pulse length in presence of spin-labeled compound and of PapG. If the ***R***-value was significantly increased, the fragment was considered as a second-site ligand.

2.3 $T_{1\rho}$ experiments with reference ligand 4

In case of the reference ligand **4**, whole $T_{1\rho}$ relaxation curves were measured with spinlock pulses from 20 ms to 250 ms. The relative intensity reduction of each ligand peak was monitored as a function of the spinlock time and fitted to equation 3 with Prism 5 (GraphPad Software Inc., San Diego, U.S.A.):

$$I(t)=I_0*\exp^{(-T_{1\rho}*t)} \qquad \text{eq. 3}$$

where $I(t)$ is the peak intensity after spinlock time t and $T_{1\rho}$ is the fitted $T_{1\rho}$ relaxation rate constant.

For determination of the K_D value of compound **4**, whole $T_{1\rho}$ relaxation curves were measured for a sample of 7.6 µM PapG-II in 20 mM HEPES-*d18* with increasing

concentrations of compound **4**. This buffer was chosen because our previous affinity assays were always run in HEPES buffer. For the screening, phosphate buffer was used for economic reasons and ease of preparation. The binding in both buffers was compared by polymer assay [S3] and showed similar values (not shown). The protein concentration was not held constant during the titration and was reduced to 7.5 µM (250 µM compound **4**), 7.4 µM (500 µM compound **4**), 7.3 µM (750 µM compound **4**) and 7.2 µM (1 mM compound **4**). A sample of 1 mM compound **4** in absence of protein in D_2O was measured to obtain the $T_{1\rho}$ relaxation rate of the free ligand. $T_{1\rho}$ relaxation rate constants as a function of the ligand excess were obtained as described above and fitted to a one-site binding model with Prism 5 (GraphPad Software Inc., San Diego, U.S.A.):

$$T_{1\rho}(c)=\frac{T_{1\rho,free}*c}{K_D'+c} \qquad \text{eq. 4}$$

where c is the ligand excess and $T_{1\rho,free}$ is the $T_{1\rho}$ relaxation rate constant of the free ligand measured in absence of protein. The K_D was calculated from the unit-less K_D' by multiplication with the protein concentration.

2.4 STD NMR experiments

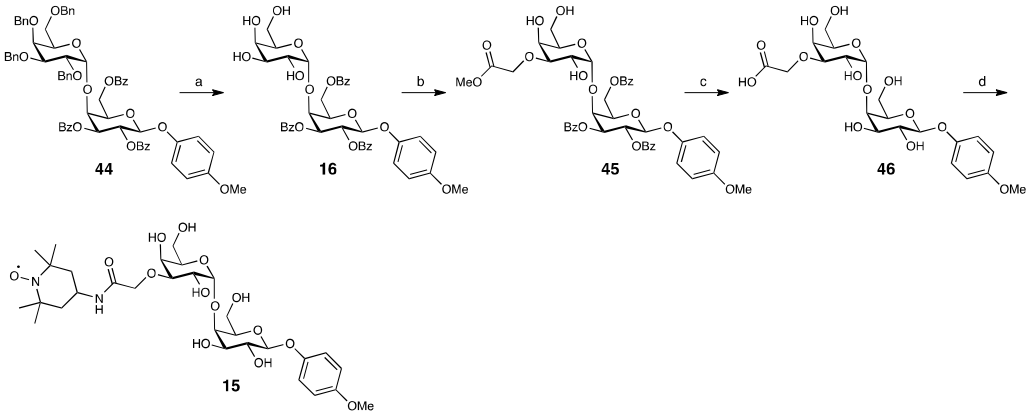
STD NMR experiments [S4,S5] were measured for determination of binding epitopes of fragment hits, reference ligand **4**, and compound **42**. Samples contained 40 µM PapG-II with 2 mM compound **4** or 0.7 mM compound **42** in NMR buffer. 0.1 mM TSP-*d4* was added as internal reference. A standard Bruker pulse sequence (stddiffesgp.3) with interleaved acquisition of on- and off-resonance spectra and with an excitation sculpting sequence for water suppression was used. A 30 ms spinlock filter at 2.6 kHz RF field strength for suppression of protein signals was applied. A train of 50 ms Gaussian shaped pulses at a power level of 45 dB corresponding to an RF field strength of 114.2 Hz was used for selective irradiation of the protein. The on-resonance was set to 0.5 ppm and the off-resonance to 300 ppm. A saturation time of 20 s and an acquisition time of 1.4 s were applied resulting in a total relaxation time of 21.4 s for the off-resonance experiment. For the determination of the binding epitopes, the T_1 relaxation rate constants of the free ligands were determined by measuring T_1 inversion recovery experiments with 10 delay times (0.1, 0.2, 0.3, 0.5, 0.8, 1, 1.5, 2, 3 and 5 s) and a relaxation delay of 20 s for ligand samples in absence of protein (2 mM compound **4** in D_2O, or 1 mM compound **42** in NMR buffer). T_1

relaxation rate constants were obtained by fitting the normalized signal intensity (intensity on the first experiment with 0.1 s delay set to −1) as a function of the inversion recovery delay to equation 5:

$$I(t)=A(1-B*\exp^{(-t/T1)}) \qquad \text{eq. 5}$$

where $I(t)$ is the signal intensity after delay time t, and A and B are correction factors. Absolute STD effects, defined as % of signal intensity in the difference (off – on) spectrum relative to the off-resonance spectrum, were determined by manual scaling for each non-overlapping peak. Absolute STD effects were then divided by the T_1 relaxation rate constants and further normalized (setting the largest quotient to 100%) to yield the STD binding epitope free of any longitudinal relaxation bias [S6].

3. Synthetic procedures



Scheme S1. a) Pd/C, H_2, AcOH, rt, 3 h, 77%; b. i) Bu_2SnO, C_6H_6, reflux, 4 h, ii) $BrCH_2COOMe$, TBABr, C_6H_6, 75 °C overnight, 34%; c. i) MeONa, MeOH, rt, 2d, ii) NaOH 0.5 M, MeOH, rt, overnight, 79%; d. 4-NH_2-TEMPO, HBTU, HOBt, DIPEA, rt, 1h, 48%.

4-Methoxyphenyl α-D-galactopyranosyl-(1→4)-2,3,6-*O*-benzoyl-β-D-galactopyranoside (16):

Compound **44**[S8] (424 mg, 0.38 mmol) was dissolved in AcOH (8 mL) under argon. $Pd(OH)_2/C$ (10-20% w/w, 100 mg) was added. Hydrogen gas was provided by a balloon. The mixture was vigorously stirred for 2 h, then the hydrogen atmosphere was replaced with argon. The suspension was diluted with MeOH and filtered through celite. The filtrate was concentrated and the residue applied to a silica gel column, eluting with DCM/MeOH gradient, to yield 220 mg (77%) of **16**.

ESI-MS: *m/z*: Calcd for $C_{40}H_{40}NaO_{15}^+$ $[M+Na]^+$: 783.2, found 783.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 8.15 – 8.02 (m, 2H, Ar), 8.01 – 7.91 (m, 4H, Ar), 7.71 – 7.61 (m, 1H, Ar), 7.59 – 7.47 (m, 4H), 7.44 – 7.31 (m, 4H), 6.98 – 6.88 (m, 2H, Ar), 6.69 – 6.53 (m, 2H, Ar), 5.99 – 5.84 (m, 1H, H-2), 5.57 (d, *J* = 10.5 Hz, 1H, H-3), 5.41 (d, *J* = 7.8 Hz, 1H, H-1), 5.02 – 4.94 (m, 2H, H-1', H-6a), 4.76 (dd, *J* = 11.6, 3.8 Hz, 1H, H-6b), 4.60 (s, 1H, H-4), 4.46 (dd, *J* = 7.6, 4.4 Hz, 1H, H-5), 4.19 (s, 1H, H-5'), 4.04 (s, 1H, H-4'), 4.00 (d, *J* = 10.5 Hz, 1H, H-3'), 3.87 (d, *J* = 10.1 Hz, 1H, H-2'), 3.67 (s, 3H, OCH_3), 3.27 (dd, *J* = 10.7, 4.9 Hz, 2H, H-6'a, H-6'b).

^{13}C NMR (126 MHz, CD_3OD): δ 167.53 (C=O), 167.41 (C=O), 167.13 (C=O), 156.94 (Ar), 152.39 (Ar), 134.67 (Ar), 134.56 (Ar), 134.43 (Ar), 131.21 (Ar), 130.91 (Ar), 130.73 (Ar), 130.56 (Ar), 130.47 (Ar), 129.65 (Ar), 129.63 (Ar), 129.58 (Ar), 119.35 (Ar), 115.46 (Ar), 102.92 (C-1'), 101.59 (C-1), 76.45 (C-4), 75.08 (C-3), 74.86 (C-5), 71.93 (C-5'), 71.31 (C-4', C-2), 70.98 (C-3'), 70.48 (C-2'), 64.56 (C-6), 62.18 (C-6'), 55.96 (OCH_3).

4-Methoxyphenyl [3-*O*-(methoxycarbonylmethyl)-α-D-galactopyranosyl]-(1→4)-2,3,6-*O*-benzoyl-β-D-galactopyranoside (45):

Compound **16** (190 mg, 0.25 mmol) was dissolved in benzene (25 mL). Bu_2SnO (68.4 mg, 0.275 mmol) was added and the mixture was refluxed in a Dean-Stark apparatus, removing 20 mL of solvent. Methyl bromoacetate (0.118 mL, 1.249 mmol) and TBABr (40.3 mg, 0.125 mmol) were added and the flask was sealed. The mixture was stirred at 75 °C overnight, then the solvent was removed. The residue was chromatographed on silica (PE/EtOAc gradient), to yield 71 mg of **45** (34%).

ESI-MS: *m/z*: Calcd for $C_{43}H_{44}NaO_{17}^+$ $[M+Na]^+$: 855.3, found 855.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.10 – 8.05 (m, 2H, Ar), 7.96 (m, 4H, Ar), 7.61 (t, *J* = 7.4 Hz, 1H), 7.55 – 7.45 (m, 4H, Ar), 7.38 (m, 4H, Ar), 6.97 – 6.92 (m, 2H, Ar), 6.70 – 6.64 (m, 2H, Ar), 5.93 (dd, *J* = 10.6, 7.9 Hz, 1H, H-2), 5.35 (dd, *J* = 10.6, 2.8 Hz, 1H, H-3), 5.17 (d, *J* = 7.9 Hz, 1H, H-1), 5.12 (d, *J* = 3.8 Hz, 1H, H-1'), 4.91 (dd, *J* = 11.5, 7.4 Hz, 1H, H-6a), 4.78 (dd, *J* = 11.5, 6.1 Hz, 1H, H-6b), 4.54 (d, *J* = 2.8 Hz, 1H, H-4), 4.47 (d, *J* = 17.4 Hz, 1H, -OC*H*$_2$COOMe), 4.29 – 4.20 (m, 2H, -OC*H*$_2$COOMe, H-5), 4.17 – 4.08 (m, 3H, H-2', H-5', H-4'), 3.82 (s, 3H, -OCH_3), 3.78 (dd, *J* = 3.0, 10.0 Hz, 1H, H-3'), 3.71 (s, 3H, -OCH_3), 3.44 (dt, *J* = 12.0, 3.9 Hz,

1H, H-6a), 3.35 (ddd, *J* = 12.0, 9.0, 5.1 Hz, 1H, H-6b), 3.09 (d, *J* = 5.4 Hz, 1H, 4'-OH), 2.53 (dd, *J* = 8.9, 3.9 Hz, 1H, 6'-OH).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 133.36 (Ar), 129.90 (Ar), 129.81 (Ar), 129.73 (Ar), 128.65 (Ar), 128.56 (Ar), 128.48 (Ar), 118.80 (Ar), 114.46 (Ar), 101.11 (C-1), 100.97 (C-1'), 81.44, 75.06, 73.72, 73.04, 70.20, 69.95, 69.59, 68.45, 68.02, 67.44, 62.96, 62.28, 57.19, 55.59, 52.59, 42.97, 40.88, 38.00, 25.52.

4-Methoxyphenyl [3-O-(hydroxycarbonylmethyl)-α-D-galactopyranosyl]-(1→4)-β-D-galactopyranoside (46):

Compound **45** (71 mg, 0.085 mmol) was dissolved in dry methanol (1 mL), then 1 M MeONa was added (16 µL) with stirring. After 20 h stirring at rt, additional 10 µL of 1 M MeONa were added. After 30 h, a further 20 µL aliquot of 1 M MeONa was added. After 48 h the mixture was evaporated to reduce the solvent volume, then 0.5 M NaOH was added at rt, and the mixture was stirred overnight. The mixture was acidified to pH 3 with amberlite IR-120, filtered and evaporated. Two chromatographies (RP, C18, water/ACN + 0.1 % TFA) yielded 34 mg of **46** (79%).

ESI-MS: *m/z*: Calcd for $C_{21}H_{30}NaO_{14}^+$ $[M+Na]^+$: 529.2, found 529.2.

^{1}H NMR (500 MHz, CD_3OD): δ =7.10 – 7.04 (m, 2H, Ar), 6.89 – 6.84 (m, 2H, Ar), 5.04 (d, *J* = 3.9 Hz, 1H, H-1'), 4.83 (d, *J* = 7.6 Hz, 1H, H-1), 4.40 (d, *J* = 16.9 Hz, 1H, -OC*H*HCOOMe), 4.35 - 4.28 (m, 2H, OC*H*HCOOMe, H-), 4.16 (d, *J* = 2.1 Hz, 1H, H-4'), 4.08 (d, *J* = 3.0 Hz, 1H, H-4), 4.01 (dd, *J* = 10.2, 3.9 Hz, 1H, H-2'), 3.95 – 3.85 (m, 1H), 3.83 – 3.78 (m, 1H), 3.77 (s, 3H, $-OCH_3$), 3.76 – 3.69 (m, 3H, H-3', H-2), 3.65 (dd, *J* = 10.1, 3.1 Hz, 1H, H-3).

^{13}C NMR (126 MHz, CD_3OD): δ = 175.37 (C=O), 156.78 (Ar), 153.01 (Ar), 119.26 (Ar), 115.51 (Ar), 103.99, 102.64, 81.08, 79.68, 76.25, 74.67, 72.66, 70.11, 68.45, 62.61, 61.12, 56.07.

4-Methoxyphenyl 3-[(1-oxyl-2,2,6,6-tetramethylpiperidine-4-amino)carbonyl]methyl-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (15):

HBTU (7.2 mg, 0.019 mmol), HOBt (5 mg, 0.038 mmol) and compound **46** (6.4 mg, 0.013 mmol) were dissolved in dry DMF (0.3 mL). After shaking for 10 min, 4-amino-TEMPO was added (3.2 mg, 0.019 mmol), followed by DIPEA (7 µL, 0.038 mmol). After 1h, water was added and the solvent was removed under reduced

pressure. The residue was purified by silica gel chromatography (DCM/MeOH gradient), yielding **15** as reddish oil.

ESI-MS: *m/z*: Calcd for $C_{30}H_{49}N_2O_{14}^+$ $[M+Na]^+$: 661.3, found 661.4 (reduced form).

NMR data were measured after treatment with sodium ascorbate, in order to convert the *N*-oxyl group to an *N*-hydroxyl group.

^{1}H NMR (500 MHz, D_2O): δ = 7.18 – 7.13 (m, 2H, Ar), 7.04 – 6.93 (m, 2H, Ar), 5.06 (d, *J* = 4.0 Hz, 1H, H-1'), 5.05 (d, *J* = 7.1 Hz, 1H, H-1), 4.36 (t, *J* = 6.4 Hz, 1H), 4.30 – 4.19 (m, 3H, H-4'), 4.16 (d, *J* = 16.0 Hz, 1H), 4.14 (d, *J* = 2.7 Hz, 1H), 4.02 (dd, *J* = 10.4, 3.9 Hz, 1H, H-2'), 3.96 – 3.79 (m, 9H, H-2, H-3', OCH_2CO, OCH_3), 3.76 (d, *J* = 6.3 Hz, 2H), 2.02 – 1.89 (m, 2H, 2 TEMPO CH*H*), 1.59 (dd, *J* = 22.8, 12.3 Hz, 2H, 2 TEMPO CH*H*), 1.31 – 1.13 (m, 12H, 4 CH_3).

^{13}C NMR (126 MHz, D_2O): δ = 193.33 (C=O), 154.76 (Ar), 118.41 (Ar), 115.04 (Ar), 101.68 (C-1'), 100.28 (C-1), 78.78, 77.47, 75.26, 72.28, 70.72, 67.60, 60.51, 55.84, 43.60 (TEMPO CH_2), 40.53, 30.15 (2 CH_3), 19.76 (2 CH_3).

General procedure for the synthesis of compounds 17a and 17b:

A solution of compound **16** in toluene (2 mL) was degassed with ultrasound under gentle vacuum for 10 min. To this solution, dibutyltin oxide (1.2 eq) was added under argon. The mixture was then refluxed for 2 to 4 h, with azeotropic removal of water. The residual toluene (approx 0.5 mL) was evaporated and the residue was dried under high vacuum for 2 h. Afterwards, it was dissolved in 0.5 mL of dry toluene. To this solution, the bromoalkyne was added, followed by TBABr (0.65 eq). The mixture was stirred in a sealed flask at 75 °C (15 to 72 h). The mixture was evaporated, re-dissolved in DCM and applied to a silica gel column. The product was eluted with a DCM/MeOH gradient.

4-Methoxyphenyl 3'-*O*-(propargyl)-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (17a):

Yield: 51.0 mg (60%)

$[\alpha]_D^{20}$ +32.2 (*c* 0.46, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{43}H_{42}NaO_{15}^+$ $[M+Na]^+$: 821.2, found 821.3.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.07 (d, *J* = 7.3 Hz, 2H, Ar), 7.98 (dd, *J* = 14.1, 7.4 Hz, 4H, Ar), 7.61 (t, *J* = 7.4 Hz, 1H, Ar), 7.51 (m, 4H, Ar), 7.38 (m, 4H, Ar), 6.96 (m,

2H, Ar), 6.68 (m, 2H, Ar), 5.96 (dd, *J* = 10.5, 7.9 Hz, 1H, H-2), 5.38 (dd, *J* = 10.6, 2.7 Hz, 1H, H-3), 5.19 (d, *J* = 7.8 Hz, 1H, H-1), 5.10 (d, *J* = 3.5 Hz, 1H, H-1'), 4.88 (dd, *J* = 11.4, 7.4 Hz, 1H, H-6a), 4.76 (dd, *J* = 11.4, 6.1 Hz, 1H, H-6b), 4.55 (d, *J* = 2.6 Hz, 1H, H-4), 4.43 (m, 2H, HCCC*H*$_2$-), 4.29 (s, 1H, H-4'), 4.25 (t, *J* = 6.7 Hz, 1H, H-5), 4.15 (t, *J* = 4.2 Hz, 1H, H-5'), 4.07 - 3.96 (m, 2H, H-2', H-3'), 3.72 (s, 3H), 3.49 – 3.33 (m, 2H, H-6'a, H-6'b), 2.72 (s, 1H), 2.66 (t, *J* = 2.2 Hz, 1H, *H*CCH$_2$-), 2.46 (d, *J* = 6.6 Hz, 1H, 2'-OH), 2.29 – 2.23 (m, 1H, 6'-OH).

^{13}C NMR (126 MHz, CDCl$_3$): δ = 166.12 (C=O), 166.02 (C=O), 165.71 (C=O), 155.72 (Ar), 151.12 (Ar), 133.78 (Ar), 133.44 (Ar), 133.41 (Ar), 129.91 (Ar), 129.81 (Ar), 129.73 (Ar), 129.55 (Ar), 129.20 (Ar), 128.91 (Ar), 128.68 (Ar), 128.57 (Ar), 128.48 (Ar), 118.84 (Ar), 114.49 (Ar), 101.12 (C-1'), 100.79 (C-1), 79.67 (alkyne quart.), 77.71 (C-3'), 75.98 (H*C*CCH$_2$), 74.64 (C-4), 73.70 (C-3), 73.01 (C-5), 70.32 (C-5'), 69.54 (C-2), 68.66 (C-4'), 68.34 (C-2'), 62.86 (C-6'), 62.38 (C-6), 57.74 (HCC*C*H$_2$), 55.60 (OCH$_3$).

4-Methoxyphenyl 3'-*O*-(4-pentynyl)-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (17b):

Yield: 30.0 mg (51%)

$[\alpha]_D^{20}$ 34.5 (*c* 0.6, CHCl$_3$).

ESI-MS: *m/z*: Calcd for $C_{45}H_{46}NaO_{15}^+$ $[M+Na]^+$: 849.3, found 849.3.

^{1}H NMR (500 MHz, CDCl$_3$): δ = 8.09 – 8.04 (m, 2H, Ar), 7.97 (dd, *J* = 11.3, 4.2 Hz, 2H, Ar), 7.61 (t, *J* = 7.4 Hz, 1H, Ar), 7.56 - 7.46 (m, 4H, Ar), 7.39 (dd, *J* = 16.9, 8.0 Hz, 4H, Ar), 6.99 – 6.92 (m, 2H, Ar), 6.72 – 6.61 (m, 2H, Ar), 5.95 (dd, *J* = 10.6, 7.9 Hz, 1H, H-2), 5.36 (dd, *J* = 10.6, 2.9 Hz, 1H, H-3), 5.18 (d, *J* = 7.8 Hz, 1H, H-1), 5.08 (d, *J* = 3.8 Hz, 1H, H-1'), 4.89 (dd, *J* = 11.4, 7.4 Hz, 1H, H-6a), 4.78 (dd, *J* = 11.4, 6.3 Hz, 1H, H-6b), 4.54 (d, *J* = 2.8 Hz, 1H, H-4), 4.26 – 4.22 (m, 2H, H-5, H-4'), 4.15 – 4.07 (m, 1H, H-5'), 4.02 - 3.96 (m, 1H, H-2'), 3.92 – 3.86 (m, 1H, OCH*H*CH$_2$), 3.80 – 3.73 (m, 2H, OC*H*HCH$_2$), 3.72 (s, 3H, OCH$_3$), 3.36 (dt, *J* = 11.6, 3.7 Hz, 1H, H6'a), 3.33 – 3.25 (m, 1H, H6'b), 2.44 (d, *J* = 6.35, 1H, 2'-OH), 2.41 – 2.36 (m, 2H, HCCC*H*$_2$CH$_2$), 2.24 (dd, *J* = 8.2, 4.3 Hz, 1H, 6'-OH), 2.07 (t, *J* = 2.6 Hz, 1H, *H*CCCH$_2$CH$_2$), 1.94 – 1.80 (m, 2H, HCCH$_2$C*H*$_2$).

^{13}C NMR (126 MHz, CDCl$_3$): δ = 166.12 (C=O), 166.05 (C=O), 165.76 (C=O), 155.72 (Ar), 151.14 (Ar), 133.78 (Ar), 133.46 (Ar), 133.40 (Ar), 129.88 (Ar), 129.81

(Ar), 129.73 (Ar), 129.57 (Ar), 129.19 (Ar), 128.97 (Ar), 128.68 (Ar), 128.57 (Ar), 128.50 (Ar), 118.84 (Ar), 114.49 (Ar), 101.16 (C-1), 100.87 (C-1'), 83.63 (alkyne quart.), 78.15 (O*C*H_2CH_2), 74.72 (C-4), 73.73 (C-3), 73.01 (C4'), 70.09 , 69.55 (C-2), 69.41, 68.39 (C-2'), 68.12 , 67.98 (C-5), 62.82 (C-6'), 62.19 (C-6), 55.60 (OCH_3), 28.11 (HCCCH_2*C*H_2), 15.24 (HCC*C*H_2CH_2).

General procedure for the synthesis of amines 22a-c:

Compound **21** [20] was suspended in the amine of choice and heated at 70 °C with stirring under argon. After 45 min the mixture was evaporated (under high vacuum, when needed), and then the residue was chromatographed on silica, eluting with DCM/MeOH/water/NH_4OHc (8:2:1:0.2).

***N*-(2-Aminoethyl)-2-(pyridin-3-yl)thiazole-4-carboxamide (22a):**

Compound **21** (88 mg, 0.376 mmol) and ethylendiamine (1mL). Yield 75 mg (80%).

ESI-MS: *m/z*: Calcd for $C_{11}H_{13}N_4OS^+$ $[M]^+$: 249.1, found 248.8.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.24 (d, *J* = 2.0 Hz, 1H, Ar), 8.65 (dd, *J* = 4.9, 1.4 Hz, 1H, Ar), 8.47 – 8.40 (m, 1H, Ar), 8.29 (s, 1H, thiazole), 7.57 (dd, *J* = 8.0, 4.9 Hz, 1H, Ar), 3.53 (t, *J* = 6.3 Hz, 2H, NHC*H*$_2$CH$_2$NH$_2$), 2.91 (t, *J* = 6.3 Hz, 2H, NHCH$_2$C*H*$_2$NH$_2$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.09 (C=O), 163.65 (Ar), 152.17 (Ar), 151.83 (Ar), 148.13 (Ar), 136.00 (Ar), 130.82 (Ar), 125.75 (Ar), 125.71 (Ar), 42.67, 42.05.

***N*-(3-Aminopropyl)-2-(pyridin-3-yl)thiazole-4-carboxamide (22b):**

Compound **21** (100 mg, 0.427 mmol) and 1,3-diaminopropane (0.8 mL). Yield 112 mg (quant).

ESI-MS: *m/z*: Calcd for $C_{12}H_{15}N_4OS^+$ $[M]^+$: 263.1, found 262.8.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.26 (d, *J* = 1.7 Hz, 1H, Ar), 8.67 (dd, *J* = 4.9, 1.5 Hz, 1H, Ar), 8.52 – 8.39 (m, 1H, Ar), 8.30 (s, 1H, thiazole), 7.65 – 7.55 (m, 1H, Ar), 3.53 (t, *J* = 6.8 Hz, 2H, NHC*H*$_2$CH$_2$CH$_2$NH$_2$), 2.75 (t, *J* = 6.9 Hz, 2H, NHCH$_2$CH$_2$C*H*$_2$NH$_2$), 1.83 (p, *J* = 6.8 Hz, 2H, NHCH$_2$C*H*$_2$CH$_2$NH$_2$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.10 (C=O), 163.38 (Ar), 152.29 (Ar), 151.83 (Ar), 148.11 (Ar), 135.99 (Ar), 130.82 (Ar), 125.71 (Ar), 125.59 (Ar), 39.73, 37.79, 33.49.

***N*-(4-Aminobutyl)-2-(pyridin-3-yl)thiazole-4-carboxamide (22c):**

Compound **21** (88 mg, 0.376 mmol) and 1,4-diaminobutane (1 mL). Yield 92.5 mg (89%).

ESI-MS: *m/z*: Calcd for $C_{13}H_{17}N_4OS$ 277.11, found 276.88.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.26 (d, *J* = 1.8 Hz, 1H, Ar), 8.67 (dd, *J* = 4.9, 1.5 Hz, 1H, Ar), 8.51 – 8.43 (m, 1H, Ar), 8.29 (s, 1H, thiazole), 7.60 (dd, *J* = 8.0, 4.9 Hz, 1H, Ar), 3.47 (t, *J* = 7.1 Hz, 2H, $NHCH_2CH_2CH_2CH_2NH_2$), 2.72 (t, *J* = 7.2 Hz, 2H, $NHCH_2CH_2CH_2CH_2NH_2$), 1.78 – 1.66 (m, 2H, $NHCH_2CH_2CH_2CH_2NH_2$), 1.65 – 1.52 (m, 2H, $NHCH_2CH_2CH_2CH_2NH_2$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.08 (C=O), 163.20 (Ar), 152.35 (Ar), 151.82 (Ar), 148.12 (Ar), 136.00 (Ar), 130.83 (Ar), 125.71 (Ar), 125.53 (Ar), 42.19, 40.28, 31.00, 28.05.

General procedure for the synthesis of azides 23a-b:

The amine (1 eq), $NaHCO_3$ (4 eq), and $CuSO_4$ pentahydrate (0.04 eq) were dissolved/suspended in water ([amine] = 1.29 mM). To this solution/suspension, a freshly prepared 2 M solution of triflyl azide in toluene was added dropwise, with vigorous stirring ar rt (2-4 eq). Methanol was added to yield a homogeneous system (2 to 4 mL). After completion, solvents were evaporated and the residue was chromatographed on silica gel, eluting with DCM/MeOH gradient.

***N*-(2-Azidoethyl)-2-(pyridin-3-yl)thiazole-4-carboxamide (23a):**

Compound **22a** (75 mg, 0.3 mmol). Yield 80 mg (97%).

ESI-MS: *m/z*: Calcd for $C_{11}H_{11}N_6OS^+$ $[M]^+$: 275.1, found 274.7.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.25 (d, *J* = 2.0 Hz, 1H, Ar), 8.67 (dd, *J* = 4.9, 1.4 Hz, 1H, Ar), 8.50 – 8.43 (m, 1H, Ar), 8.32 (s, 1H, thiazole), 7.59 (dd, *J* = 8.0, 4.9 Hz, 1H, Ar), 3.65 (t, *J* = 5.9 Hz, 2H, $NHCH_2CH_2N_3$), 3.56 (t, *J* = 6.0 Hz, 2H, $NHCH_2CH_2N_3$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.13 (C=O), 163.49 (Ar), 151.95 (Ar), 151.80 (Ar), 148.09 (Ar), 135.98 (Ar), 130.77 (Ar), 125.90 (Ar), 125.68 (Ar), 51.46 ($NHCH_2CH_2N_3$), 39.96 ($NHCH_2CH_2N_3$).

***N*-(3-Azidopropyl)-2-(pyridin-3-yl)thiazole-4-carboxamide (23b):**

Compound **22b** (42 mg, 0.16 mmol). Yield 41 mg (89%).

ESI-MS: *m/z*: Calcd for $C_{12}H_{13}N_6OS^+$ $[M]^+$: 289.1, found 289.0.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.23 (s, 1H, Ar), 8.65 (d, *J* = 4.3 Hz, 1H, Ar), 8.46 – 8.39 (m, 1H, Ar), 8.28 (s, 1H, thiazole), 7.57 (dd, *J* = 8.0, 4.9 Hz, 1H, Ar), 3.53 (t, *J* = 6.9 Hz, 2H, NHC*H*$_2$CH$_2$CH$_2$N$_3$), 3.45 (t, *J* = 6.7 Hz, 2H, NHCH$_2$CH$_2$C*H*$_2$N$_3$), 1.93 (p, *J* = 6.8 Hz, 2H, NHCH$_2$C*H*$_2$CH$_2$N$_3$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.07 (C=O), 163.31 (Ar), 152.19 (Ar), 151.81 (Ar), 148.10 (Ar), 135.95 (Ar), 130.78 (Ar), 125.69 (Ar), 125.66 (Ar), 50.33 (NHCH$_2$CH$_2$*C*H$_2$N$_3$), 38.07 (NH*C*H$_2$CH$_2$CH$_2$N$_3$), 29.97 (NHCH$_2$*C*H$_2$CH$_2$N$_3$).

_N_-(4-Azidobutyl)-2-(pyridin-3-yl)thiazole-4-carboxamide (23c):

Compound **22c** (51 mg, 0.19 mmol). Yield 56 mg (quant).

ESI-MS: *m/z*: Calcd for $C_{13}H_{15}N_6OS^+$ $[M]^+$: 303.1, found 303.0.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.23 (s, 1H, Ar), 8.66 (s, 1H, Ar), 8.46 – 8.35 (m, 1H, Ar), 8.28 (s, 1H, thiazole), 7.57 (dd, *J* = 7.9, 4.9 Hz, 1H, Ar), 3.47 (t, *J* = 6.8 Hz, 2H, NHC*H*$_2$CH$_2$CH$_2$*C*H$_2$N$_3$), 3.37 (t, *J* = 6.6 Hz, 2H, NHCH$_2$CH$_2$CH$_2$C*H*$_2$N$_3$), 1.79 – 1.72 (m, 2H, NHCH$_2$C*H*$_2$CH$_2$CH$_2$N$_3$), 1.72 – 1.65 (m, 2H, NHCH$_2$CH$_2$CH$_2$C*H*$_2$N$_3$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.06 (C=O), 163.22 (Ar), 152.26 (Ar), 151.79 (Ar), 148.10 (Ar), 135.97 (Ar), 125.71 (Ar), 125.59 (Ar), 52.19 (NHCH$_2$CH$_2$CH$_2$*C*H$_2$N$_3$), 39.96 (NH*C*H$_2$CH$_2$CH$_2$CH$_2$N$_3$), 27.95 and 27.40 (NHCH$_2$*C*H$_2$*C*H$_2$CH$_2$N$_3$).

General procedure for the synthesis of compounds 18a-b:

Compound **17a** or **17b** (1 eq) was dissolved in dry MeOH (1 mL). To this solution, 10 µL of a freshly prepared 1M NaOMe solution were added under argon, with stirring at rt. The mixture was stirred until TLC showed completion (≈ 6 h) and then neutralized with amberlyst 15. The mixture was filtered, concentrated and chromatographed on silica (DCM/MeOH gradient) to yield the corresponding deprotected disaccharide.

4-Methoxyphenyl 3'-*O*-(propargyl)-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (18a):

Compound **17a** (51 mg, 0.064 mmol). Yield: 27.5 mg (88.5%).

$[\alpha]_D^{20}$ +12.0 (*c* 0.55, MeOH).

ESI-MS: *m/z*: Calcd for $C_{22}H_{30}NaO_{12}^+$ $[M+Na]^+$: 509.2, found 509.2.

^{1}H NMR (500 MHz, CD_3OD): δ = 7.12 – 7.02 (m, 2H), 6.92 – 6.82 (m, 2H), 5.04 (d, *J* = 3.9 Hz, 1H, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.40 (d, *J* = 2.3 Hz, 2H, HCCC*H*$_2$), 4.35 (t, *J* = 6.1 Hz, 1H, H-5'), 4.20 (d, *J* = 2.6 Hz, 1H, H-4'), 4.10 (d, *J* = 2.9 Hz, 1H, H-4), 3.95 (dd, *J* = 10.2, 3.8 Hz, 1H, H-2'), 3.93 – 3.88 (m, 1H, H-5), 3.86 (dd, *J* = 10.2, 3.0 Hz, 1H, H-3'), 3.83 – 3.75 (m, 7H, H-2, OCH_3, H-6'b, H-6a, H-6b), 3.73 (dd, *J* = 11.1, 5.5 Hz, 1H, H-6'a), 3.67 (dd, *J* = 10.1, 3.0 Hz, 1H, H-3), 2.88 (t, *J* = 2.3 Hz, 1H, *H*CCCH$_2$).

^{13}C NMR (126 MHz, CD_3OD): δ = 156.81 (Ar), 152.99 (Ar), 119.31 (Ar), 115.54 (Ar), 104.03 (C-1), 102.59 (C-1'), 81.13 (alkyne quart.), 79.06 (C-4), 78.58 (C-3'), 76.25, 75.92 (H*C*CCH$_2$), 74.64 (C-3), 72.66, 72.61 (C-5'), 69.96, 68.21 (C-4'), 62.62, 60.93, 57.91 (HCC*C*H$_2$), 56.08 (OCH_3).

4-Methoxyphenyl 3'-*O*-(penten-4-yl)-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (18b):

Compound **17b** (30 mg, 0.036 mmol). Yield: 16.3 mg (87.3%).

$[\alpha]_D^{20}$ +16.2 *(c* 0.37, MeOH).

ESI-MS: *m/z*: Calcd for $C_{24}H_{34}NaO_{12}^+$ $[M+Na]^+$: 537.2, found 537.2.

^{1}H NMR (500 MHz, CD_3OD): δ = 7.11 – 7.04 (m, 2H), 6.91 – 6.83 (m, 2H), 5.02 (d, *J* = 3.9 Hz, 1H, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.33 (t, *J* = 6.1 Hz, 1H, H-5'), 4.15 (d, *J* = 2.6 Hz, 1H, H-4'), 4.09 (d, *J* = 3.0 Hz, 1H, H-4), 3.94 – 3.86 (m, 2H, H-2', H-6a), 3.85 – 3.71 (m, 9H, H-2, H-5, H-6'a-b, H-6b, OCH_3, HCCCH$_2$CH$_2$CH*H*), 3.70 – 3.64 (m, 2H, H-3, HCCCH$_2$CH$_2$C*H*H), 3.61 (dd, *J* = 10.2, 3.0 Hz, 1H, H-3'), 2.36 (tt, *J* = 7.0, 2.4 Hz, 1H, HCCC*H*$_2$CH$_2$CH$_2$), 2.23 (t, *J* = 2.6 Hz, 1H, *H*CCCH$_2$CH$_2$CH$_2$), 1.91 – 1.78 (m, 1H, HCCCH$_2$C*H*$_2$CH$_2$).

^{13}C NMR (126 MHz, CD_3OD): δ = 156.81 (Ar), 152.99 (Ar), 119.28 (Ar), 115.54 (Ar), 104.02 (C-1), 102.64 (C-1'), 84.89 (alkyne quart), 79.64 (C-3'), 79.09 (C-4), 76.31, 74.61, 72.65, 69.92, 69.67, 69.30, 67.96 (C-4'), 62.68, 60.95, 56.09, 30.10 (HCCCH$_2$*C*H$_2$CH$_2$), 15.89 (HCC*C*H$_2$CH$_2$CH$_2$).

General procedure for the synthesis of compounds 24a-f:

The appropriate terminal alkyne (1 eq) and the azide (1.1 eq) were dissolved in a 2:1 mixture of *t*-BuOH and water. Aqueous solutions of sodium ascorbate and $CuSO_4$ were sequentially added (0.2 and 0.1 eq, respectively). The bright yellow suspension

was stirred vigorously ar rt until TLC (DCM/MeOH, 9:1) showed completion (3-5 h). Whenever needed, additional aliquots of sodium ascorbate were added. Purification: Chromatography on silica (DCM/MeOH/water/NH_4OH, 8:2:0.5:0.1).

3-*O*-[(1-(2-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)ethyl)-1,2,3-triazol-4-yl) methyl]-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (24a):

Compound **18a** (14 mg, 0.029 mmol) and azide **23a** (8.67 mg, 0.032 mmol). Yield: 13.7 mg (63%).

$[\alpha]_D^{20}$ +12.1 (*c* 0.69, MeOH).

ESI-MS: *m/z*: Calcd for $C_{33}H_{40}N_6NaO_{13}^+$ $[M+Na]^+$: 783.2, found 783.3.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.20 (d, *J* = 1.8 Hz, 1H, Ar), 8.68 – 8.62 (m, 1H, Ar), 8.43 – 8.37 (m, 1H, Ar), 8.28 (s, 1H, thiazole), 8.07 (s, 1H, triazole), 7.57 (dd, *J* = 8.0, 4.9 Hz, 1H, Ar), 7.08 – 7.00 (m, 2H, Ar), 6.87 – 6.79 (m, 2H, Ar), 5.04 (d, *J* = 3.9 Hz, 1H, H-1'), 4.85 (d, *J* = 12.8 Hz, 1H, OC*H*H-triazole), 4.82 (d, *J* = 7.6 Hz, 1H, H-1), 4.77 (d, *J* = 12.5 Hz, 1H, OCH*H*-triazole), 4.69 (t, *J* = 5.9 Hz, 2H, $NHCH_2CH_2$), 4.31 (t, *J* = 6.1 Hz, 1H, H-5'), 4.18 (d, *J* = 2.4 Hz, 1H, H-4'), 4.08 (d, *J* = 2.9 Hz, 1H, H-4), 3.97 (dd, *J* = 10.2, 3.8 Hz, 1H, H-2'), 3.92 (t, *J* = 5.9 Hz, 2H, $NHCH_2CH_2$), 3.86 (dt, *J* = 10.5, 5.4 Hz, 1H, H-5), 3.83 – 3.70 (m, 10H, H-2, H-3, H6a-b, H-3'. H-6'a-b, OCH_3), 3.66 (dd, *J* = 10.1, 3.0 Hz, 1H, H-3).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.09 (C=O), 163.50 (Ar), 156.71 (Ar), 153.00 (Ar), 151.83 (Ar), 151.73 (Ar), 148.09 (Ar), 146.51 (Ar), 136.01 (Ar), 130.73 (Ar), 126.14 (Ar), 125.73 (Ar), 125.52 (Ar), 119.22 (Ar), 115.50 (Ar), 103.97 (C-1), 102.63 (C-1'), 79.39 (C-4), 79.23, 76.18, 74.69 (C-3), 72.82, 72.64, 69.84 (C-2'), 67.80 (C-4'), 63.46 (O*C*H_2-triazole), 62.74, 61.12, 56.06 (OCH_3), 50.65 ($NHCH_2CH_2$), 40.73 ($NHCH_2CH_2$).

3-*O*-[(1-(3-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)propyl)-1,2,3-triazol-4-yl) methyl]-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (24b):

Compound **18a** (8.1 mg, 0.017 mmol) and azide **23b** (6.72 mg, 0.023 mmol). Yield: 12.4 mg (96%).

$[\alpha]_D^{20}$ +11.3 (*c* 0.62, MeOH).

ESI-MS: *m/z*: Calcd for $C_{34}H_{43}N_6NaO_{13}^+$ $[M+Na]^+$: 775.3, found 775.5.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.25 (d, *J* = 1.8 Hz, 1H, Ar), 8.68 (dd, *J* = 4.9, 1.3 Hz, 1H, Ar), 8.48 – 8.43 (m, 1H, Ar), 8.30 (s, 1H, thiazole), 8.10 (s, 1H, triazole), 7.60 (dd, *J* = 8.0, 4.9 Hz, 1H, Ar), 7.07 – 7.01 (m, 2H, Ar), 6.85 – 6.79 (m, 2H, Ar), 5.04 (d, *J* = 3.9 Hz, 1H, H-1'), 4.83 (d, *J* = 12.8 Hz, 1H, OC*H*H-triazole), 4.82 (d, *J* = 7.4 Hz, 1H, H-1), 4.76 (d, *J* = 12.5 Hz, 1H, OCH*H*-triazole), 4.54 (t, *J* = 6.8 Hz, 2H, NCH_2CH$_2$CH$_2$NH), 4.31 (t, *J* = 6.1 Hz, 1H, H-5'), 4.17 (d, *J* = 2.3 Hz, 1H, H-4'), 4.08 (d, *J* = 2.9 Hz, 1H, H-4), 3.98 (dd, *J* = 10.2, 3.9 Hz, 1H, H-2'), 3.90 – 3.83 (m, 1H, H-5), 3.82 – 3.71 (m, 10H, 10H, H-2, H-3, H6a-b, H-3'. H-6'a-b, OCH_3), 3.66 (dd, *J* = 10.1, 3.0 Hz, 1H, H-3), 3.50 (t, *J* = 6.7 Hz, 2H, NCH$_2$CH$_2$CH_2NH), 2.29 (p, *J* = 6.8 Hz, 2H, NCH$_2$CH_2CH$_2$NH).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.07 (C=O), 163.42 (Ar), 156.71 (Ar), 152.99 (Ar), 152.04 (Ar), 151.85 (Ar), 148.13 (Ar), 146.40 (Ar), 136.04 (Ar), 130.80 (Ar), 125.83 (Ar), 125.74 (Ar), 125.37 (Ar), 119.22 (Ar), 115.49 (Ar), 103.97 (C-1), 102.63 (C-1'), 79.34 (, 79.17, 76.19, 74.68 (C-3), 72.78 (C-5'), 72.63, 69.85 (C-2), 67.76 (C-4'), 63.43 (O*C*H$_2$-triazole), 62.70, 61.07 (C-5), 56.05 (OCH_3), 49.37 (N*C*H$_2$CH$_2$CH$_2$NH), 37.74 (NCH$_2$CH$_2$*C*H$_2$NH), 31.17 (NCH$_2$*C*H$_2$CH$_2$NH).

3-*O*-[3(1-(2-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)ethyl)-1,2,3-triazol-4-yl) propyl]-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (24c):

Compound **18b** (8.5 mg, 0.017 mmol) and azide **23a** (5.07 mg, 0.019 mmol). Yield: 9.4 mg (72%).

$[\alpha]_D^{20}$ +11.9 (*c* 0.45, MeOH).

ESI-MS: *m/z*: Calcd for $C_{35}H_{45}N_6O_{13}S^+$ $[M+Na]^+$: 789.3, found 789.3.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.27 (s, 1H, Ar), 8.73 (s, 1H, Ar), 8.43 (d, *J* = 7.9 Hz, 1H, Ar), 8.30 (s, 1H, thiazole), 7.85 (s, 1H, Ar), 7.62 (s, 1H, Ar), 7.10 – 7.01 (m, 2H, Ar), 6.88 – 6.80 (m, 2H, Ar), 5.02 (d, *J* = 3.9 Hz, 1H, H-1'), 4.85 (d, *J* = 7.7 Hz, 1H, H-1), 4.71 – 4.62 (m, 2H, NCH_2CH$_2$NH), 4.25 (t, *J* = 6.1 Hz, 2H, H-5'), 4.08 (d, *J* = 2.9 Hz, 2H, H-4), 4.05 (d, *J* = 2.3 Hz, 2H, H-4'), 3.96 – 3.88 (m, 4H, NCH$_2$CH_2NH, H-2', H-5), 3.86 – 3.73 (m, 7H, OCH_3, H-6'a, H-2, H6a-b), 3.72 – 3.58 (m, 3H, H-3, H-6'b, CH$_2$CH$_2$C*H*HO), 3.49 (dd, *J* = 10.2, 2.9 Hz, 1H, H-3'), 3.45 – 3.39 (m, 1H, CH$_2$CH$_2$CH*H*O), 2.90 – 2.80 (m, 2H, CH_2CH$_2$CH$_2$O), 2.02 – 1.81 (m, 2H, CH$_2$CH_2CH$_2$O).

^{13}C NMR (126 MHz, CD_3OD): δ = 163.36 (C=O), 156.70 (Ar), 152.97 (Ar), 151.82 (Ar), 151.70 (Ar), 148.04 (Ar), 135.95 (Ar), 126.18 (Ar), 119.17 (Ar), 115.50 (Ar), 103.94 (C-1'), 102.75 (C-1), 79.62 (C-3'), 79.50 (C-4), 76.20, 74.67, 72.80, 72.67, 69.72, 68.27 ($CH_2CH_2CH_2O$), 67.45 (C-4'), 62.78, 61.17, 56.04, 50.38, 40.57, 30.56 ($CH_2CH_2CH_2O$), 22.36 ($CH_2CH_2CH_2O$).

3-*O*-[3(1-(3-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)propyl)-1,2,3-triazol-4-yl)propyl]-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (23d):

Compound **18b** (9.3 mg, 0.018 mmol) and azide **23b** (7.30 mg, 0.025 mmol). Yield: 11.5 mg (79%).

$[\alpha]_D^{20}$ +9.0 (*c* 0.3, MeOH).

ESI-MS: *m/z*: Calcd for $C_{36}H_{47}N_6O_{13}S^+$ $[M+Na]^+$: 803.3, found 803.5.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.28 (s, 1H, Ar), 8.70 (s, 1H, Ar), 8.47 (d, *J* = 8.0 Hz, 1H, Ar), 8.30 (s, 1H, thiazole), 7.88 (s, 1H, Ar), 7.67 – 7.53 (m, 1H, Ar), 7.11 – 6.98 (m, 2H, Ar), 6.87 – 6.79 (m, 2H, Ar), 5.04 (d, *J* = 3.9 Hz, 1H, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.50 (t, *J* = 6.8 Hz, 2H, $NHCH_2CH_2CH_2N$), 4.30 (t, *J* = 6.2 Hz, 1H, H-5'), 4.13 (d, *J* = 2.4 Hz, 1H, H-4'), 4.09 (d, *J* = 2.9 Hz, 1H, H-4), 3.95 (dd, *J* = 10.2, 3.9 Hz, 1H, H-2'), 3.94 – 3.88 (m, 1H, H-5), 3.86 – 3.64 (m, 10H, OCH_3, H-2, H-3, $OCHHCH_2CH_2$, H-6'a-b, H-6a-b), 3.56 (dd, *J* = 10.2, 3.0 Hz, 1H, H-3'), 3.53 – 3.43 (m, 3H, $OCHHCH_2CH_2$, $NHCH_2CH_2CH_2$), 2.84 (t, *J* = 7.2 Hz, 2H, $OCH_2CH_2CH_2$), 2.26 (p, *J* = 6.7 Hz, 2H, $NHCH_2CH_2CH_2$), 2.02 – 1.85 (m, 2H, $OCH_2CH_2CH_2$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.12 (C=O), 163.37 (Ar), 156.74 (Ar), 152.98 (Ar), 152.05 (Ar), 151.85 (Ar), 148.68 (Ar), 148.14 (Ar), 136.03 (Ar), 125.83 (Ar), 123.96 (Ar), 119.22 (Ar), 115.52 (Ar), 103.98 (C-1), 102.76 (C-1'), 79.57 (C-3'), 79.50 (C-4), 76.25, 74.68, 72.79, 72.67, 69.82 (C-2'), 68.68, 67.59 (C-4'), 62.76, 61.14, 56.06 (OCH_3), 49.41 ($NHCH_2CH_2CH_2N$), 37.89 ($NHCH_2CH_2CH_2$), 31.16 ($NHCH_2CH_2CH_2$), 30.56 ($OCH_2CH_2CH_2$), 22.54 ($OCH_2CH_2CH_2$).

3-*O*-[(1-(4-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)butyl)-1,2,3-triazol-4-yl) methyl]-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (24e)

Compound **18a** (9.8 mg, 0.020 mmol) and azide **23c** (8.83 mg, 0.029 mmol). Yield: 12.5 mg (79%).

$[\alpha]_D^{20}$ +11.1 (*c* 0.63, MeOH).

ESI-MS: *m/z*: Calcd for $C_{35}H_{45}N_6O_{13}S^+$ $[M+Na]^+$: 789.3, found 789.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.25 (s, 1H, Ar), 8.67 (d, *J* = 4.5 Hz, 1H, Ar), 8.46 (d, *J* = 8.0 Hz, 1H, Ar), 8.29 (s, 1H, thiazole), 8.05 (s, 1H, triazole), 7.60 (dd, *J* = 8.0, 4.9 Hz, 1H, Ar), 7.10 – 7.02 (m, 1H, Ar), 6.87 – 6.82 (m, 1H, Ar), 5.04 (d, *J* = 3.9 Hz, 1H, H-1'), 4.85 (d, overlaps with water peak, H-1) 4.83 (d, *J* = 12.4 Hz, 1H, OC*H*H), 4.78 (d, *J* = 12.4 Hz, 1H, OC*H*H), 4.49 (t, *J* = 7.0 Hz, 2H, $NHCH_2CH_2CH_2CH_2N$), 4.31 (t, *J* = 6.1 Hz, 1H, H-5'), 4.18 (d, *J* = 2.5 Hz, 1H, H-4'), 4.08 (d, *J* = 2.9 Hz, 1H, H-4), 3.98 (dd, *J* = 10.2, 3.8 Hz, 1H, H-2), 3.90 – 3.83 (m, 1H, H-5), 3.82 – 3.71 (m, 9H, H-2, H-6a-b, H-3', H-6'a-b, OCH_3), 3.66 (dd, *J* = 10.1, 3.0 Hz, 1H, H-3), 3.48 (t, *J* = 6.9 Hz, 2H, $NHCH_2CH_2CH_2CH_2N$), 2.06 – 1.97 (m, 2H, $NHCH_2CH_2CH_2CH_2N$), 1.71 – 1.63 (m, 2H, $NHCH_2CH_2CH_2CH_2N$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.10 (C=O), 163.31 (Ar), 156.74 (Ar), 153.00 (Ar), 152.21 (Ar), 151.82 (Ar), 148.12 (Ar), 146.43 (Ar), 136.04 (Ar), 130.82 (Ar), 125.70 (Ar), 125.07 (Ar), 119.22 (Ar), 115.52 (Ar), 103.95 (C-1), 102.67 (C-1'), 79.43 (C-4), 79.29, 76.18, 74.69 (C-3), 72.82, 72.65, 69.85 (C-2'), 67.79 (C-4'), 63.50 (OCH_2), 62.70, 61.07, 56.07, 50.99 ($NHCH_2CH_2CH_2CH_2N$), 39.66 ($NHCH_2CH_2CH_2CH_2N$), 28.67, 27.56.

3-*O*-[3-(1-(4-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)butyl)-1,2,3-triazol-4-yl)propyl]-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (24f):

Compound **18b** (6.9 mg, 0.013 mmol) and azide **23c** (5.25 mg, 0.017 mmol). Yield: 7.2 (66%).

$[\alpha]_D^{20}$ +14.9 (*c* 0.36, MeOH).

ESI-MS: *m/z*: Calcd for $C_{37}H_{48}N_6NaO_{13}S^+$ $[M+Na]^+$: 839.3, found 839.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.26 (d, *J* = 1.6 Hz, 1H, Ar), 8.74 (t, *J* = 6.0 Hz, 1H, CONH), 8.67 (d, *J* = 3.8 Hz, 1H, Ar), 8.49 – 8.43 (m, 1H, Ar), 8.30 (s, 1H, thiazole), 7.83 (s, 1H, triazole), 7.60 (dd, *J* = 7.9, 4.9 Hz, 1H, Ar), 7.13 – 7.01 (m, 2H, Ar), 6.91 – 6.76 (m, 2H, Ar), 5.03 (d, *J* = 3.9 Hz, 1H, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.44 (t, *J* = 7.0 Hz, 2H, $NHCH_2CH_2CH_2CH_2N$), 4.30 (t, *J* = 6.2 Hz, 1H, H-5'), 4.14 (d, *J* = 2.5 Hz, 1H, H-4'), 4.09 (d, *J* = 2.9 Hz, 1H, H-4), 3.95 (dd, *J* = 10.3, 3.9 Hz, 1H, H-2'), 3.92 – 3.89 (m, 1H, H-6a), 3.85 – 3.69 (m, 10H, OC*H*HCH_2CH_2, OCH_3, H-2, H-5, H-6a, H-6'a-b, OH), 3.66 (dd, *J* = 10.1, 3.0 Hz, 1H, H-3), 3.57 (dd,

J = 10.2, 3.0 Hz, 1H, H-3'), 3.53 - 3.44 (m, 3H, NHC*H*$_2$CH$_2$CH$_2$CH$_2$N, OCH*H*CH$_2$CH$_2$), 2.87 (t, *J* = 7.2 Hz, 2H, OCH$_2$CH$_2$C*H*$_2$), 2.11 – 1.86 (m, 4H, NHCH$_2$CH$_2$C*H*$_2$CH$_2$N, OCH$_2$C*H*$_2$CH$_2$), 1.75 – 1.52 (m, 2H, NHCH$_2$C*H*$_2$CH$_2$CH$_2$N).
^{13}C NMR (126 MHz, CD$_3$OD): δ = 166.09 (C=O), 163.37 (Ar), 156.75 (Ar), 153.00 (Ar), 152.25 (Ar), 151.81 (Ar), 148.65 (Ar), 148.12 (Ar), 136.04 (Ar), 130.84 (Ar), 125.71 (Ar), 123.72 (Ar), 119.21 (Ar), 115.52 (Ar), 103.99 (C-1), 102.76 (C-1'), 101.41, 79.59 (C-3'), 79.50 (C-4), 76.25, 74.67 (C-3), 72.78 (C-5'), 72.67, 69.81 (C-2'), 68.69 (O*C*H$_2$CH$_2$CH$_2$), 67.59 (C-4'), 62.75, 61.13, 56.06 (OCH$_3$), 50.87 (NHCH$_2$CH$_2$CH$_2$*C*H$_2$N), 39.82, 39.69, 30.62, 28.73, 27.64 (NHCH$_2$*C*H$_2$CH$_2$CH$_2$N), 22.53 (NHCH$_2$CH$_2$CH$_2$*C*H$_2$N).

Phenyl 3-*O*-3-[(*tert*-butyldimethylsilyl)oxy]propyl-2,4,6-tri-*O*-(4-methoxybenzyl)-1-thio-β-D-galactopyranoside (26):

Compound **25** (389 mg, 0.615 mmol) was dissolved in DMF and cooled to 0 °C. NaH (60% dispersion in mineral oil, 52 mg, 1.29 mmol) was added in 2 portions with stirring under argon. The reaction was allowed to warm to rt, then a solution of (3-bromopropoxy)-(*tert*-butyl)dimethylsilane (342.5 mg, 1.532 mmol) in 1 mL of dry DMF was added with stirring under argon. After 1h, additional 10 mg of NaH were added, and this addition was repeated twice (every 30 min). After 2 h, the mixture was poured on ice and extracted 3 times with DCM. The organic phases were collected, dried, filtered and evaporated. DMF was coevaporated with xylenes. The residue was chromatographed on silica (EtOAc/PE gradient), to yield 350 mg (71%) of **26**.
$[\alpha]_D^{20}$ -2.8 (*c* 2.53, CHCl$_3$).
ESI-MS: *m/z*: Calcd for $C_{45}H_{60}NaO_9SSi^+$ $[M+Na]^+$: 827.4, found 827.4.
^{1}H NMR (500 MHz, CDCl$_3$): δ = 7.56 – 7.50 (m, 2H, Ar), 7.36 – 7.30 (m, 2H, Ar), 7.25 – 7.22 (m, 2H, Ar), 7.21 – 7.15 (m, 5H, Ar), 6.89 – 6.81 (m, 6H, Ar), 4.84 (d, *J* = 11.1 Hz, 1H, benzyl-CH), 4.68 (d, *J* = 9.9 Hz, 1H, benzyl-CH), 4.65 (d, *J* = 9.9 Hz, 1H, benzyl-CH), 4.59 (d, *J* = 9.7 Hz, 1H, H-1), 4.49 (d, *J* = 11.2 Hz, 1H, benzyl-CH), 4.40 (d, *J* = 11.3 Hz, 1H, benzyl-CH), 4.33 (d, *J* = 11.3 Hz, 1H, benzyl-CH), 3.92 (d, *J* = 2.6 Hz, 1H, H-4), 3.82 (s, 3H, OCH$_3$), 3.80 – 3.64 (m, 11H, H-2, OC*H*$_2$CH$_2$C*H*$_2$OTBMDS, 2 OCH$_3$), 3.61 – 3.52 (m, 3H, H-5, H-6a-b), 3.39 (dd, *J* = 9.2, 2.7 Hz, 1H, H-3), 1.84 (p, *J* = 6.4 Hz, 2H, OCH$_2$C*H*$_2$CH$_2$OTBMDS), 1.55 (s, 3H), 0.88 (s, 9H, *t*-butyl), 0.03 (d, *J* = 2.7 Hz, 6H, Si(C*H*$_3$)$_2$).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 159.30 (Ar), 159.08 (Ar), 134.47 (Ar), 131.35 (Ar), 131.07 (Ar), 130.67 (Ar), 130.04 (Ar), 129.92 (Ar), 129.59 (Ar), 129.42 (Ar), 128.74 (Ar), 126.88 (Ar), 113.82 (Ar), 113.76 (Ar), 113.57 (Ar), 87.80 (C-1), 84.89 (C-3), 75.20 (benzyl-CH_2), 73.97 (benzyl-CH_2), 73.22 (benzyl-CH_2), 72.89 (C-4), 68.53, 67.77, 60.17, 55.29, 33.69 ($OCH_2CH_2CH_2OTBMDS$), 25.96 ($C(CH_3)_3$), 18.31 ($C(CH_3)_3$), -5.27 ($Si(CH_3)_2$).

Phenyl 3-*O*-(3-carbonylpropyl)-2,4,6-tri-*O*-(4-methoxybenzyl)-1-thio-β-D-galactopyranoside (27):

Compound **26** (314 mg, 0.39 mmol) was dissolved in 1 M TBAF in THF (2 mL) at rt and stirred for 1h under argon (TLC: 50% EtOAc/PE), to remove the TBDMS protective group. The mixture was then diluted with EtOAc, washed with water and brine, dried on Na_2SO_4, filtered and concentrated. The residue was chromatographed on silica (PE/EtOAc gradient) to yield 238 mg (88%) of the corresponding free alcohol.

$[\alpha]_D^{20}$ -2.2 (*c* 0.6, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{39}H_{46}NaO_9S^+$ $[M+Na]^+$: 713.3, found 713.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.55 – 7.50 (m, 2H, Ar), 7.37 – 7.32 (m, 2H, Ar), 7.27 – 7.14 (m, 7H, Ar), 6.91 – 6.81 (m, 6H, Ar), 4.79 (d, *J* = 11.0 Hz, 1H, benzyl-CH), 4.73 (d, *J* = 9.8 Hz, 1H, benzyl-CH), 4.60 (t, *J* = 9.6 Hz, 2H, H-1 and benzyl-CH), 4.50 (d, *J* = 11.0 Hz, 1H, benzyl-CH), 4.42 (d, *J* = 11.3 Hz, 1H, benzyl-CH), 4.35 (d, *J* = 11.3 Hz, 1H, benzyl-CH), 3.95 (d, *J* = 2.5 Hz, 1H, H-4), 3.88 – 3.69 (m, 14H, $OCH_2CH_2CH_2OH$, H-2, 3xOCH_3), 3.62 – 3.59 (m, 1H), 3.63 – 3.53 (m, 3H, H-5, H-6a-b), 3.39 (dd, *J* = 9.2, 2.6 Hz, 1H, H-3), 1.90 – 1.82 (p, *J* = 5 Hz, 2H, $OCH_2CH_2CH_2OH$).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 159.50 (Ar), 134.45 (Ar), 131.45 (Ar), 130.95 (Ar), 130.61 (Ar), 130.06 (Ar), 129.81 (Ar), 129.59 (Ar), 128.94 (Ar), 127.13 (Ar), 114.05 (Ar), 113.99 (Ar), 113.94 (Ar), 113.77 (Ar), 88.01 (C-1), 84.85 (C-3), 77.41, 77.16, 77.09, 76.97, 76.91, 75.32 (benzyl-CH_2), 74.31 (benzyl-CH_2), 73.40 (benzyl-CH_2), 72.94 (C-4), 69.53, 68.41, 61.51, 55.43, 32.78 ($OCH_2CH_2CH_2OH$).

The alcohol (233 mg, 0.337 mmol) was dissolved in dry DCM (5 mL). To the resulting solution, Dess-Martin periodinane (174.5 mg, 0.411 mmol) was added in one portion at rt with stirring under argon. After 40 min (TLC: 50% EtOAc/PE), the mixture was diluted with DCM (10 mL), and 4 mL of $NaHCO_3$ satd and 250 mg of

$Na_2S_2O_3$ pentahydrate were added. The mixture was stirred for 5 min, then it was transferred to a separating funnel and washed with $NaHCO_3$ satd and brine. The organic phase was dried on Na_2SO_4, filtered and concentrated. The residue was chromatographed on silica (EtOAc/PE gradient) to yield 201 mg (87%) of **27**.

$[\alpha]_D^{20}$ +1.00 (*c* 0.35, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{39}H_{44}NaO_9S^+$ $[M+Na]^+$: 711.3, found 711.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 9.76 (t, *J* = 1.7 Hz, 1H, CHO), 7.54 (dd, *J* = 6.4, 3.1 Hz, 2H, Ar), 7.32 (d, *J* = 8.6 Hz, 2H, Ar), 7.26 – 7.14 (m, 7H, Ar), 6.89 – 6.80 (m, 6H, Ar), 4.74 (d, *J* = 11.2 Hz, 1H, benzyl-CH), 4.71 (d, *J* = 10.0 Hz, 1H, benzyl-CH), 4.58 (d, *J* = 9.7 Hz, 1H, H-1), 4.55 (d, *J* = 10.0 Hz, 1H, benzyl-CH), 4.48 (d, *J* = 11.2 Hz, 1H, benzyl-CH), 4.41 (d, *J* = 11.3 Hz, 1H, benzyl-CH), 4.35 (d, *J* = 11.3 Hz, 1H, benzyl-CH), 3.98 – 3.87 (m, 3H, $HCOCH_2CH_2O$, H-4), 3.81 (s, 3H, OCH_3), 3.79 – 3.85 (m, 10H, H-2, 2 OCH_3), 3.61 – 3.52 (m, 3H, H-5, H-6a-b), 3.39 (dd, *J* = 9.2, 2.8 Hz, 1H, H-3), 2.69 – 2.56 (m, 2H, $HCOCH_2CH_2O$).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 201.00 (CHO), 159.56 (Ar), 159.34 (Ar), 134.53 (Ar), 131.56 (Ar), 131.11 (Ar), 130.73 (Ar), 130.15 (Ar), 130.08 (Ar), 129.85 (Ar), 129.67 (Ar), 129.01 (Ar), 127.21 (Ar), 114.04 (Ar), 113.80 (Ar), 88.07 (C-1), 85.17 (C-3), 77.51, 77.26, 77.16, 77.01, 75.38 (benzyl-CH_2), 74.29 (benzyl-CH_2), 73.46, 73.24, 68.52, 64.89 ($HCOCH_2CH_2O$), 55.51, 44.43 ($HCOCH_2CH_2O$).

Phenyl 3-*O*-(3-propynyl)-2,4,6-tri-*O*-(4-methoxybenzyl)-1-thio-β-D-galactopyranoside (28):

PPh_3 (289.4 mg, 1.10 mmol) and CBr_4 were dissolved in dry DCM (4 mL) and cooled to 0 °C. To this mixture, a solution of compound **27** in dry DCM (5 mL) was added dropwise at 0 °C with stirring under argon. The mixture was stirred for 30 min at 0 °C, then diluted with DCM and transferred to a separating funnel. The organic phase was washed with water. The water phase was extracted twice with DCM. The combined organic fractions were dried on Na_2SO_4, filtered and concentrated. The residue was chromatographed on silica (EtOAc/PE gradient) to yield 186 mg (80%) of the corresponding ω,ω-dibromoalkene, according to ESI-MS analysis (ESI-MS: *m/z*: Calcd for $C_{40}H_{44}Br_2NaO_8S$: 867.1, found 867.2). This compound was immediately used for the next step. To a solution of the dibromoalkene (186 mg, 0.22 mmol) in dry THF (5 mL) kept under argon and cooled to -70 °C, *n*-BuLi (1.5 M in hexanes, 0.36

mL) was added dropwise. During the addition, the temperature raised to -61 °C, then it was kept at -70 °C for 20 min, after which it was raised to -40 - -30 °C during 25 min. Afterwards, it was cooled again to -70 °C for ice addition (excess). The temperature was allowed to reach rt and the mixture was stirred for 1 h, then diluted with water and extracted 3 times with Et_2O. The combined organic phases were dried on Na_2SO_4, filtered and evaporated. The residue was chromatographed on silica (EtOAc/PE gradient) to yield 134 mg (89%) of **28**.

$[\alpha]_D^{20}$ +1.0 (*c* 0.2, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{40}H_{44}NaO_8S^+$ $[M+Na]^+$: 707.3, found 707.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 7.56 – 7.51 (m, 2H), 7.36 – 7.31 (m, 2H), 7.26 (t, *J* = 4.3 Hz, 3H), 7.22 – 7.15 (m, 5H), 6.89 – 6.79 (m, 6H), 4.86 (d, *J* = 11.2 Hz, 1H, benzyl-CH), 4.70 (d, *J* = 10.0 Hz, 1H, benzyl-CH), 4.67 (d, *J* = 9.9 Hz, 1H, benzyl-CH), 4.58 (d, *J* = 9.7 Hz, 1H, H-1), 4.53 (d, *J* = 11.2 Hz, 1H, benzyl-CH), 4.40 (d, *J* = 11.3 Hz, 1H, benzyl-CH), 4.33 (d, *J* = 11.3 Hz, 1H, benzyl-CH), 3.93 (d, *J* = 2.7 Hz, 1H, H-4), 3.84 – 3.81 (m, 4H, OCH_3, H-2), 3.80 – 3.77 (m, 6H, 2 OCH_3), 3.74 (t, *J* = 6.6 Hz, 2H, $HCCCH_2CH_2O$), 3.60 – 3.51 (m, 3H, H-5, H-6a-b), 3.41 (dd, *J* = 9.2, 2.8 Hz, 1H, H-3), 2.47 (td, *J* = 6.6, 2.6 Hz, 2H, $HCCCH_2CH_2O$), 1.97 (t, *J* = 2.6 Hz, 1H, $HCCCH_2CH_2O$).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 159.33 (Ar), 134.38 (Ar), 131.38 (Ar), 131.01 (Ar), 130.62 (Ar), 129.98 (Ar), 129.60 (Ar), 128.77 (Ar), 126.96 (Ar), 113.81 (Ar), 113.58 (Ar), 87.84 (C-1), 84.98 (C-3), 81.55 , 77.29, 77.22, 77.03, 76.96, 76.78, 75.23 (benzyl-CH_2), 73.99 (benzyl-CH_2), 73.23 (benzyl-CH_2), 73.01, 69.56 ($HCCCH_2CH_2O$), 69.07 ($HCCCH_2CH_2O$), 68.46, 55.31, 20.47 ($HCCCH_2CH_2O$).

4-Methoxyphenyl 3-*O*-(3-butynyl)-2,4,6-(4-methoxyphenyl)-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (30):

Compound **28** (134 mg, 0.196 mmol), 4-methoxyphenyl 2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (141 mg, 0.235 mmol) and NIS (48.5 mg, 0.215 mmol) were dissolved in dry DCM (2 mL). To this solution, dry Et_2O (4 mL) was added, and the mixture was cooled to -55 °C under argon. TMSOTf (5 µL) was added with stirring. After 1h, triethylamine (0.1 mL) was added. The mixture was stirred for 1 h at -55 °C, then the temperature was increased to rt. The mixture was diluted with DCM, washed with $Na_2S_2O_3$ satd and $NaHCO_3$ satd, dried on Na_2SO_4, filtered and concentrated. The

residue was chromatographed on silica (EtOAc/PE gradient) to yield 130 mg (57%) of **30**.

ESI-MS: *m/z*: Calcd for $C_{68}H_{68}NaO_{18}^{+}$ $[M+Na]^{+}$: 1195.4, found 1195.8.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.04 – 7.99 (m, 2H, Ar), 7.95 (d, *J* = 7.4 Hz, 4H, Ar), 7.60 (t, *J* = 7.4 Hz, 1H, Ar), 7.52 – 7.40 (m, 4H, Ar), 7.35 (t, *J* = 7.8 Hz, 2H, Ar), 7.26 (dd, *J* = 15.4, 7.6 Hz, 5H, Ar), 7.17 (dd, *J* = 8.6, 3.0 Hz, 4H, Ar), 6.98 – 6.92 (m, 2H, Ar), 6.87 – 6.82 (m, 2H, Ar), 6.81 – 6.75 (m, 2H, Ar), 6.72 – 6.68 (m, 2H, Ar), 6.68 – 6.63 (m, 2H, Ar), 5.96 (dd, *J* = 10.5, 7.8 Hz, 1H, H-2), 5.25 (dd, *J* = 10.6, 2.9 Hz, 1H, H-3), 5.12 (d, *J* = 7.8 Hz, 1H, H-1), 4.87 – 4.71 (m, 5H, H-1', H-6a-b, 2 benzyl-CH), 4.62 (d, *J* = 11.5 Hz, 1H, benzyl-CH), 4.41 (d, *J* = 2.8 Hz, 1H, H-4), 4.39 (d, *J* = 10.7 Hz, 1H, benzyl-CH), 4.31 (dd, *J* = 9.4, 5.0 Hz, 1H, H-5'), 4.14 (t, *J* = 6.6 Hz, 1H, H-5), 4.08 – 4.03 (m, 2H, benzyl-CH, H-4'), 4.02 – 3.91 (m, 3H, benzyl-CH, H-2', H-3'), 3.87 – 3.81 (m, 2H, $HCCCH_2C\mathit{H}_2O$), 3.78 (s, 3H, OCH_3), 3.76 (s, 3H, OCH_3), 3.71 (s, 3H, OCH_3), 3.63 (s, 3H, OCH_3), 3.36 (t, *J* = 8.9 Hz, 1H, H-6'b), 2.90 (dd, *J* = 8.3, 4.9 Hz, 1H, H-6'a), 2.54 – 2.48 (m, 2H, $HCCC\mathit{H}_2CH_2O$), 1.98 (t, *J* = 2.6 Hz, 1H, $\mathit{H}CCCH_2CH_2O$).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 166.47 (C=O), 166.08 (C=O), 165.33 (C=O), 159.07 (Ar), 155.51 (Ar), 151.29 (Ar), 133.32 (Ar), 133.19 (Ar), 131.25 (Ar), 130.56 (Ar), 130.48 (Ar), 130.01 (Ar), 129.94 (Ar), 129.84 (Ar), 129.73 (Ar), 129.66 (Ar), 129.55 (Ar), 129.32 (Ar), 129.25 (Ar), 129.08 (Ar), 128.48 (Ar), 128.40 (Ar), 118.76 (Ar), 114.41 (Ar), 113.70 (Ar), 113.63 (Ar), 113.57 (Ar), 113.50 (Ar), 101.43 (C-1'), 101.01 (C-1), 81.84, 79.55, 75.73, 75.18, 74.52, 74.16, 73.60, 73.16, 72.63 (benzyl-CH_2), 69.76 (C-5'), 69.57, 69.48, 68.54 ($HCCCH_2\mathit{C}H_2O$), 67.13 (C-6'), 62.82 (C-6), 55.58 (OCH_3), 55.25 (OCH_3), 55.06 (OCH_3), 20.58 ($HCC\mathit{C}H_2CH_2O$).

4-Methoxyphenyl 3-*O*-(3-butynyl)-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (31):

Compound **30** (116 mg, 0.099mmol) was dissolved in 9 mL of a 10 % TFA solution in dry DCM. After 5 min, the reaction was complete (TLC: DCM/MeOH 95:5). The mixture was carefully poured on excess $NaHCO_3$ satd. The water phase was extracted 4 times with DCM, the organic fractions were collected, dried on Na_2SO_4, filtered and concentrated. The residue was chromatographed on silica gel (DCM/MeOH gradient)

to yield 80 mg (quant) of the intermediate 4-methoxyphenyl 3-*O*-(3-butynyl)-α-D-galactopyranosyl-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside.

$[\alpha]_D^{20}$ +36.5 (*c* 0.38, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{44}H_{44}NaO_{15}^+$ $[M+Na]^+$: 835.3, found 835.3.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 8.09 – 8.05 (m, 2H, Ar), 8.01 – 7.95 (m, 4H, Ar), 7.61 (t, *J* = 7.4 Hz, 1H, Ar), 7.57 – 7.46 (m, 4H, Ar), 7.39 (dd, *J* = 16.6, 8.2 Hz, 4H, Ar), 6.98 – 6.93 (m, 2H, Ar), 6.71 – 6.66 (m, 2H, Ar), 5.95 (dd, *J* = 10.6, 7.9 Hz, 1H, H-2), 5.35 (dd, *J* = 10.6, 2.9 Hz, 1H, H-3), 5.18 (d, *J* = 7.9 Hz, 1H, H-1), 5.09 (d, *J* = 3.8 Hz, 1H, H-1'), 4.90 (dd, *J* = 11.4, 7.4 Hz, 1H, H-6a), 4.78 (dd, *J* = 11.4, 6.3 Hz, 1H, H-6b), 4.54 (d, *J* = 2.8 Hz, 1H, H-4), 4.28 – 4.21 (m, 2H, H-4', H-5), 4.12 (t, *J* = 4.2 Hz, 1H, H-5'), 4.02 (dd, *J* = 9.9, 3.7 Hz, 1H, H-2'), 3.88 (dd, *J* = 9.9, 3.1 Hz, 1H, H-3'), 3.84 (t, *J* = 6.1 Hz, 2H, $HCCCH_2CH_2O$), 3.73 (s, 3H, OCH_3), 3.35 (dd, *J* = 11.9, 3.9 Hz, 1H, H-6'a), 3.29 (dd, *J* = 11.9, 5.2 Hz, 1H, H-6'b), 2.56 (td, *J* = 6.1, 2.0 Hz, 2H, $HCCCH_2CH_2O$), 2.18 (t, *J* = 2.6 Hz, 1H, $HCCCH_2CH_2O$).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 166.12 (C=O), 166.04 (C=O), 165.76 (C=O), 155.73 (Ar), 151.12 (Ar), 133.79 (Ar), 133.47 (Ar), 133.40 (Ar), 129.88 (Ar), 129.81 (Ar), 129.73 (Ar), 129.56 (Ar), 129.18 (Ar), 128.96 (Ar), 128.68 (Ar), 128.57 (Ar), 128.50 (Ar), 118.83 (Ar), 114.49 (Ar), 101.14 (C-1'), 100.87 (C-1), 81.90 (alkyne quart), 78.22 (C-3'), 74.69 (C-4), 73.73 (C-3), 72.99, 70.34 ($HCCCH_2CH_2O$), 70.05 (C-5'), 69.54 (C-2), 68.26, 67.44 ($HCCCH_2CH_2O$), 62.79 (C-6'), 62.17 (C-6), 55.60 (OCH_3), 20.57 ($HCCCH_2CH_2O$).

The intermediate (80 mg, 0.098 mmol) was dissolved in dry MeOH (3 mL). To the resulting solution, a freshly prepared 1 M NaOMe solution was added (30 µL). The reaction mixture was stirred for 2 h and then neutralized with Amberlyst 15, filtered and evaporated. The residue was chromatographed on silica (DCM/MeOH gradient) to yield 42 mg (85%) of **31**.

$[\alpha]_D^{20}$ +13.3 (*c* 1.39, MeOH).

ESI-MS: *m/z*: Calcd for $C_{23}H_{32}NaO_{12}^+$ $[M+Na]^+$: 523.2, found 523.2.

^{1}H NMR (500 MHz, CD_3OD): δ = 7.15 – 7.02 (m, 2H, Ar), 6.92 – 6.83 (m, 2H, Ar), 5.03 (d, *J* = 3.9 Hz, 1H, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.33 (t, *J* = 6.2 Hz, 1H, H-5'), 4.16 (d, *J* = 2.5 Hz, 1H, H-4'), 4.09 (d, *J* = 3.0 Hz, 1H, H-4), 3.94 – 3.87 (m, 2H, H-2', H-6a), 3.86 – 3.70 (m, 10H, $HCCCH_2CH_2O$, OCH_3, H-2, H-6'a-b, H-5, H-6b), 3.70 – 3.64 (m, 2H, H-3, H-3'), 2.62 – 2.47 (m, 2H, $HCCCH_2CH_2O$), 2.31 (t, *J* = 2.7 Hz, 1H, $HCCCH_2CH_2O$).

^{13}C NMR (126 MHz, CD_3OD): δ = 156.80 (Ar), 153.00 (Ar), 119.28 (Ar), 115.55 (Ar), 104.02 (C-1), 102.60 (C-1'), 82.18 (alkyne quart), 79.76, 79.22 (C-4), 76.30, 74.61, 72.64, 70.69 (H*C*CCH$_2$CH$_2$O), 69.90 (C-2'), 69.25 (C-6'), 68.15 (C-4'), 62.66, 61.01 (C-6), 56.09 (OCH$_3$), 20.73 (HCC*C*H$_2$CH$_2$O).

3-O-[2-(1-(2-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)ethyl)-1,2,3-triazol-4-yl) ethyl]-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (24g):

Compounds **31** (8.3 mg, 0.017 mmol) and **23a** (5.9 mg, 0.022 mmol) were dissolved in a mixture of *t*-BuOH (0.12 mL) and water (0.05 mL). A solution of sodium ascorbate in water and a solution of $CuSO_4$ in water were successively added (0.3 and 0.1 eq respectively). The bright yellow suspension was stirred vigorously at rt for 1 h, and then the mixture was chromatographed on a RP C18 column (water/ACN gradient) to yield 6.6 mg (51%) of **24g**.

$[\alpha]_D^{20}$ +11.3 (*c* 0.66, MeOH).

ESI-MS: *m/z*: Calcd for $C_{34}H_{43}N_6O_{13}S^+$ $[M+Na]^+$: 775.3, found 775.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.22 (s, 1H, Ar), 8.66 (d, *J* = 4.4 Hz, 1H, Ar), 8.46 – 8.37 (m, 1H, Ar), 8.29 (d, *J* = 1.5 Hz, 1H, thiazole), 7.96 (s, 1H, Ar, triazole), 7.62 – 7.52 (m, 1H, Ar), 7.13 – 7.01 (m, 2H, Ar), 6.84 (d, *J* = 9.0 Hz, 2H, Ar), 5.02 (d, *J* = 3.9 Hz, 1H, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.66 (t, *J* = 5.9 Hz, 2H, NC*H*$_2$CH$_2$NH), 4.29 (t, *J* = 6.2 Hz, 1H, H-5'), 4.14 (d, *J* = 2.3 Hz, 1H, H-4'), 4.08 (d, *J* = 2.9 Hz, 1H, H-4), 4.00 – 3.86 (m, 5H, OC*H*HCH$_2$, H-2', NCH$_2$C*H*$_2$NH, H-6a), 3.85 – 3.70 (m, 9H, H-2, H-5, H-6b, H-6'a-b, OCH*H*CH$_2$, OCH$_3$), 3.66 (dd, *J* = 10.1, 2.9 Hz, 1H, H-3), 3.61 (dd, *J* = 10.2, 3.0 Hz, 1H, H-3'), 3.01 (t, *J* = 6.2 Hz, 2H, OCH$_2$C*H*$_2$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.12 (C=O), 163.45 (Ar), 156.75 (Ar), 153.00 (Ar), 151.85 (Ar), 151.76 (Ar), 148.10 (Ar), 146.79 (Ar), 136.02 (Ar), 130.74 (Ar), 126.11 (Ar), 125.76 (Ar), 124.82 (Ar), 119.21 (Ar), 115.52 (Ar), 103.95 (C-1), 102.69 (C-1'), 101.42, 79.77 (C-3'), 79.47, 76.24, 74.65 (C-3), 72.68 (C-5'), 69.81, 69.25, 67.72, 62.71 (C-6'), 61.12 (C-6), 56.07 (OCH$_3$), 50.48, 40.64, 27.29 (OCH$_2$C*H*$_2$).

3-O-[2-(1-(3-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)propyl)-1,2,3-triazol-4-yl) ethyl]-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (24h):

Compounds **31** (12.3 mg, 0.025 mmol) and **23b** (8.5 mg, 0.029 mmol) were dissolved in a mixture of *t*-BuOH (0.12 mL) and water (0.05 mL). A solution of sodium ascorbate in water and a solution of $CuSO_4$ in water were successively added (0.3 and 0.1 eq respectively). The bright yellow suspension was stirred vigorously at rt for 1 h. As the azide was consumed before the alkyne, 1 mg of compound **23b** was added. After additional 30 min the mixture was chromatographed on a RP C18 column (water/ACN gradient), to yield 14 mg (72%) of **24h**.

$[\alpha]_D^{20}$ +12.1 (*c* 0.70, MeOH).

ESI-MS: *m/z*: Calcd for $C_{35}H_{45}N_6O_{13}S$ 789.3, found 789.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.26 (d, *J* = 1.8 Hz, 1H, Ar), 8.67 (dd, *J* = 4.9, 1.5 Hz, 1H, Ar), 8.49 – 8.43 (m, 1H, Ar), 8.30 (s, 1H, thiazole), 8.01 (s, 1H, triazole), 7.60 (dd, *J* = 8.0, 4.9 Hz, 1H, Ar), 7.08 – 7.00 (m, 2H, Ar), 6.86 – 6.78 (m, 2H, Ar), 5.04 (d, *J* = 3.9 Hz, 1H, H-1'), 4.83 (d, *J* = 7.6 Hz, 1H, H-1), 4.50 (t, *J* = 6.8 Hz, 2H, NC*H*$_2$CH$_2$CH$_2$NH), 4.31 (t, *J* = 6.2 Hz, 1H, H-5'), 4.15 (d, *J* = 2.3 Hz, 1H, H-4'), 4.09 (d, *J* = 3.0 Hz, 1H, H-4), 4.02 - 3.94 (m, 2H, H-2', OC*H*HCH$_2$), 3.94 – 3.86 (m, 1H, H-6a), 3.83 – 3.71 (m, 9H, H-2, H-5, H-6b, H-6'ab, OCH*H*CH$_2$, OCH$_3$), 3.66 (dd, *J* = 10.1, 3.0 Hz, 1H, H-3), 3.63 (dd, *J* = 10.2, 3.0 Hz, 1H, H-3'), 3.50 (t, *J* = 6.6 Hz, 2H, NCH$_2$CH$_2$C*H*$_2$NH), 2.99 (t, *J* = 6.3 Hz, 2H, OCH$_2$C*H*$_2$), 2.27 (p, *J* = 6.7 Hz, 2H, NCH$_2$C*H*$_2$CH$_2$NH).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.06 (Ar), 163.37 (Ar), 156.71 (Ar), 152.96 (Ar), 152.01 (Ar), 151.83 (Ar), 148.12 (Ar), 146.73 (Ar), 136.02 (Ar), 130.76 (Ar), 125.83 (Ar), 125.73 (Ar), 124.62 (Ar), 119.19 (Ar), 115.49 (Ar), 103.94 (C-1), 102.67 (C-1'), 79.73 (C-3'), 79.37 (C-4), 76.22, 74.63, 72.69, 69.82, 69.13, 67.70 (C-4'), 62.68, 61.07, 56.04 (OCH$_3$), 37.83 (NH*C*H$_2$CH$_2$CH$_2$), 31.10 (NHCH$_2$*C*H$_2$CH$_2$), 27.29 (*C*H$_2$CH$_2$O).

3-*O*-[2-(1-(4-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)butyl)-1,2,3-triazol-4-yl) ethyl]-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (24i):

Compounds **31** (10.1 mg, 0.020 mmol) and **23c** (9.1 mg, 0.030 mmol) were dissolved in a mixture of *t*-BuOH (0.12 mL) and water (0.05 mL). A solution of sodium ascorbate in water and a solution of $CuSO_4$ in water were successively added (0.3 and 0.1 eq respectively). The bright yellow suspension was stirred vigorously at rt for 40 min, and then the mixture was chromatographed on a RP C18 column (water/ACN gradient) to yield 7.7 mg (47.5%) of **24i**.

$[\alpha]_D^{20}$ +12.2 (*c* 0.77, MeOH).

ESI-MS: *m/z*: Calcd for $C_{36}H_{46}N_6NaO_{13}S^+$ $[M+Na]^+$: 825.3, found 825.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.25 (s, 1H, Ar), 8.67 (d, *J* = 4.3 Hz, 1H, Ar), 8.46 (dt, *J* = 8.0, 1.7 Hz, 1H, Ar), 8.29 (s, 1H, thiazole), 7.96 (s, 1H, triazole), 7.59 (dd, *J* = 8.0, 4.9 Hz, 1H, Ar), 7.10 – 7.00 (m, 2H, Ar), 6.88 – 6.79 (m, 2H, Ar), 5.03 (d, *J* = 3.9 Hz, 1H, H-1'), 4.83 (d, *J* = 7.6 Hz, 1H, H-1), 4.45 (t, *J* = 7.0 Hz, 2H, $NCH_2CH_2CH_2CH_2NH$), 4.31 (t, *J* = 6.1 Hz, 1H, H-5'), 4.17 (d, *J* = 2.4 Hz, 1H, H-4'), 4.08 (d, *J* = 2.9 Hz, 1H, H-4), 4.00 (dt, *J* = 9.4, 6.1 Hz, 1H, $OCHHCH_2$), 3.95 (dd, *J* = 10.2, 3.9 Hz, 1H, H-2'), 3.90 (dd, *J* = 10.7, 6.8 Hz, 1H, H-6a), 3.84 – 3.70 (m, 9H, H-2, H-5, H-6b, H-6'a-b, OCH_3, $OCHHCH_2$), 3.68 – 3.62 (m, 2H, H-3, H-3'), 3.48 (t, *J* = 6.9 Hz, 2H, $NCH_2CH_2CH_2CH_2NH$), 3.02 (t, *J* = 6.2 Hz, 2H, OCH_2CH_2), 2.06 – 1.92 (m, 2H, $NCH_2CH_2CH_2CH_2NH$), 1.73 – 1.56 (m, 2H, $NCH_2CH_2CH_2CH_2NH$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.10 (C=O), 163.30 (C=O), 156.75 (Ar), 152.99 (Ar), 152.22 (Ar), 151.81 (Ar), 148.13 (Ar), 146.76 (Ar), 136.03 (Ar), 125.74 (Ar), 125.68 (Ar), 124.38 (Ar), 119.23 (Ar), 115.52 (Ar), 103.98 (C-1), 102.69 (C-1'), 101.41, 79.75, 79.40 (C-4), 76.25, 74.65, 72.67 (C-5'), 69.83 (C-2'), 69.15 (OCH_2CH_2), 67.71 (C-4'), 62.71 (C-6'), 61.10 (C-6), 56.06 (OCH_3), 50.89 ($NCH_2CH_2CH_2CH_2NH$), 39.71 ($NCH_2CH_2CH_2CH_2NH$), 28.68 ($NCH_2CH_2CH_2CH_2NH$), 27.61 ($NCH_2CH_2CH_2CH_2NH$), 27.31 (OCH_2CH_2).

4-Methoxyphenyl 2,4,6-tri-*O*-acetyl-3-*O*-allyl-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (32): Ref. [S7].

4-Methoxyphenyl 2,4,6-tri-*O*-acetyl-3-*O*-carbonylmethyl-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (33):

To a solution of compound **32** (349 mg, 0.377 mmol) in dioxane (2.6 mL) were added water (0.86 mL), 2,6-lutidine (0.088 mL, 0.753 mmol), OsO4 (2.5% in *t*-BuOH, 0.235 mL, 0.0190 mmol), and $NaIO_4$ (330.2 mg, 1.544 mmol). The dense, white suspension was stirred at rt under argon and monitored by TLC. After 8 h, $NaIO_4$ (110 mg 0.514 mmol) and water (0.2 µL) were added. After overnight stirring, $NaIO_4$ (53 mg 0.25 mmol) was added. After further 5 h, water (20 mL) and DCM (50 mL) were added. The organic layer was separated, and the water layer was extracted three times with DCM. The combined organic layers were dried over Na_2SO_4, filtered, and evaporated. The residue was chromatographed on silica (5:2:3 PE/DCM/EtOAc → EtOAc) to yield **33** (246 mg, 70%).

$[\alpha]_D^{20}$ +40.4 (*c* 0.45, $CHCl_3$).

ESI-MS: *m/z*: Calcd for $C_{48}H_{48}NaO_{19}^+$ $[M+Na]^+$: 951.3, found 951.4.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 9.63 (s, 1H, CHO), 8.09 – 8.04 (m, 2H, Ar), 7.98 – 7.93 (m, 4H, Ar), 7.63 (t, *J* = 7.4 Hz, 1H, Ar), 7.55 – 7.45 (m, 4H, Ar), 7.38 (dd, *J* = 14.8, 7.5 Hz, 4H, Ar), 6.99 – 6.95 (m, 2H, Ar), 6.71 – 6.66 (m, 2H, Ar), 5.95 (dd, *J* = 10.6, 7.8 Hz, 1H, H-2), 5.52 (d, *J* = 1.9 Hz, 1H, H-4'), 5.35 (dd, *J* = 10.6, 2.9 Hz, 1H, H-3), 5.25 – 5.15 (m, 3H, H-1, H-1', H-2'), 4.82 (dd, *J* = 11.3, 7.7 Hz, 1H, H-6a), 4.59 – 4.51 (m, 2H, H-6b, H-5'), 4.49 (d, *J* = 2.6 Hz, 1H, H-4), 4.32 (d, *J* = 17.5 Hz, 1H, HCOC*H*HO), 4.27 – 4.20 (m, 1H, H-5), 4.16 (d, *J* = 16.4 Hz, 1H, HCOCH*H*O), 4.09 (dd, *J* = 10.5, 3.2 Hz, 1H, H-3'), 3.88 (dd, *J* = 10.9, 8.1 Hz, 1H, H-6'a), 3.73 (s, 3H, $ArOCH_3$), 3.57 (dd, *J* = 10.9, 5.9 Hz, 1H, H-6'b), 2.20 (s, 3H, Ac), 2.09 (s, 3H, Ac), 1.91 (s, 3H, Ac).

^{13}C NMR (126 MHz, CD_3OD): δ = 199.56 (CHO), 170.77 (C=O), 170.45 (C=O), 170.28 (C=O), 166.38 (C=O), 166.16 (C=O), 165.62 (C=O), 155.88 (Ar), 151.24 (Ar), 133.89 (Ar), 133.68 (Ar), 133.58 (Ar), 130.02 (Ar), 129.92 (Ar), 129.83 (Ar), 128.83 (Ar), 128.76 (Ar), 128.65 (Ar), 118.91 (Ar), 114.64 (Ar), 101.24 (C-1), 99.05 (C-1'), 75.96 (C-4), 75.31 (HCO*C*H_2), 74.76 (C-3'), 73.71 (C-3), 72.95 (C-5), 69.99 (C-2'), 69.45 (C-2), 67.64 (C-5'), 66.72 (C-4'), 62.64 (C-6), 61.02 (C-6'), 55.73 ($ArOCH_3$), 21.10 (Ac), 20.85 (Ac), 20.72 (Ac).

3-*O*-{2-[(4-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)butyl)amino]-ethyl}-α-D-galactopyranosyl-(1→4)-2,3,6-tri-*O*-benzoyl-β-D-galactopyranoside (34c):

Compounds **32** (20 mg, 0.022 mmol) and **22c** (18.5 mg, 0.067 mmol) were dissolved in MeOH and $NaBH_3CN$ (1 M in THF, 0.03 µL, 0.03 mmol) was added at rt with stirring, followed by 0.033 µL of AcOH. A mixture of products formed, among which also partial deprotection products. The mixture was evaporated after 1 h, and then re-dissolved in dry MeOH (1 mL). To this solution, 1 M MeONa in MeOH was added until pH > 10. The mixture was stirred overnight, then neutralized with AcOH and evaporated. The residue was purified by column chromatography on silica (DCM/MeOH/water/NH_4OH, 80:30:5:1). The silica was pre-washed extensively with this mixture before chromatography. As this procedure did not eliminate all impurities, the collected fractions were re-purified by RP on C18 column (water/ACN gradient). The product **34c** was isolated in 26% overall yield (4.2 mg).

$[\alpha]_D^{20}$ +12.3 (*c* 0.42, MeOH).

ESI-MS: *m/z*: Calcd for $C_{34}H_{47}N_4O_{13}S^+$ $[M]^+$: 751.3, found 751.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.36 (s, 1H, Ar), 8.89 – 8.69 (m, 1H, Ar), 8.58 (t, *J* = 8.5 Hz, 1H, Ar), 8.35 (s, 1H, thiazole), 7.73 (s, 1H, Ar), 7.10 – 7.03 (m, 2H, Ar), 6.89 – 6.80 (m, 2H, Ar), 5.05 (d, *J* = 3.9 Hz, 1H, H-1'), 4.85 (d, *J* = 7.5 Hz, 1H, H-1), 4.38 (t, *J* = 6.1 Hz, 1H, H-5'), 4.20 (d, *J* = 2.6 Hz, 1H, H-4'), 4.10 (d, *J* = 2.9 Hz, 1H, H-4), 4.04 – 3.94 (m, 2H, OC*H*HCH_2, H-2'), 3.92 – 3.64 (m, 12H, H-2, H-3, H-5, H-6a-b, H-3', H-6'a-b, OCH*H*CH_2, OCH_3), 3.53 (t, *J* = 6.1 Hz, 2H, CONHC$H_2CH_2CH_2CH_2$NH), 3.31 – 3.24 (m, 2H, OCH_2CH_2NH), 3.22 – 3.13 (m, 2H, $NHCH_2CH_2CH_2CH_2$NH), 1.89 – 1.85 (m, 4H, $NHCH_2CH_2CH_2CH_2$NH).

^{13}C NMR (126 MHz, CD_3OD): δ = 165.63 (C=O), 163.45 (C=O), 156.78 (Ar), 152.99 (Ar), 152.16 (Ar), 150.62 (Ar), 146.98 (Ar), 137.28 (Ar), 126.25 (Ar), 119.10 (Ar), 115.54 (Ar), 103.95 (C-1), 102.14 (C-1'), 101.40, 79.86, 78.86 (C-4), 76.37, 74.47, 72.62, 72.14 (C-5'), 69.98 (C-2'), 68.00 (C-4'), 64.96 (O*C*H_2CH_2), 62.43 (C-6'), 61.01 (C-6), 56.07 (OCH_3), 49.00 (OCH_2CH_2NH), 48.40 ($NHCH_2CH_2CH_2CH_2$NH), 39.42 (CONH*C*$H_2CH_2CH_2CH_2$NH), 27.76, 24.56.

***N*-(2-(2-(2-aminoethoxy)ethoxy)ethyl)-2-(pyridin-3-yl)thiazole-4-carboxamide (36):**

Compound **33** (66.0 mg, 0.282 mmol) was dissolved in dry DMF (0.1 mL) and 2,2'-(ethylenedioxy)bis(ethylamine) (0.1 mL) and the mixture was heated at 70 °C under argon with stirring for 45 min. The solvents were evaporated and the residue was

purified on silica (DCM/MeOH/H_2O/NH_4OH, 80:20:2.5:1) yielding **36** (50.0 mg, 53%).

ESI-MS: *m/z*: Calcd for $C_{15}H_{21}N_4O_3S^+$ $[M]^+$: 337.1, found 337.0.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.24 (dd, *J* = 2.2, 0.6 Hz, 1H, py), 8.68 (dd, *J* = 4.9, 1.5 Hz, 1H, py), 8.47 – 8.42 (m, 1H, py), 8.31 (s, 1H, thiazole), 7.60 (ddd, *J* = 8.0, 4.9, 0.7 Hz, 1H, py), 3.77 – 3.70 (m, 4H), 3.70 – 3.62 (m, 4H), 3.57 (t, *J* = 5.3 Hz, 2H), 2.82 (t, *J* = 5.3 Hz, 2H).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.10 (C=O), 163.17 (Ar), 152.10 (Ar), 151.85 (Ar), 148.05 (Ar), 135.92 (Ar), 130.71 (Ar), 125.72 (Ar), 125.69 (Ar), 72.69, 71.37, 71.35, 70.59, 41.89 and 40.29 (CH_2NH_2 and CONH*C*H_2).

3-*O*-(1-Oxo-1-(2-(pyridin-3-yl)thiazol-4-yl)-5,8-dioxa-2,11-diazatridecan-13-yl)-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (37):

Compounds **33** (24.8 mg, 0.027 mmol) and **36** (13.9 mg, 0.041 mmol) were dissolved in DCE. $NaBH(OAc)_3$ (7.9 mg, 0.037 mmol) was added and the mixture was stirred at rt under argon for 2 h (TLC: 9:1 DCM/MeOH). Afterwards, the reaction mixture was applied to a silica column and chromatographed with DCM/MeOH gradient, to yield 22.2 mg of protected product (67%, slightly contaminated with some byproducts), which was used immediately in the next step. 22 mg (0.018 mmol) of the intermediate were dissolved in dry MeOH (0.5 mL) under argon. To this mixture, 10 µL of 1 M MeONa were added with stirring. After 2 h, 10 additional µL of 1 M MeONa were added. After a total of 4 h reaction time, the mixture was neutralized with AcOH and evaporated. The residue was applied to a RP C-18 column and chromatographed (water/ACN + 0.1% TFA gradient) to yield 15.8 mg (97%) of **38**.

$[\alpha]_D^{20}$ +9.2 (*c* 0.79, MeOH).

ESI-MS: *m/z*: Calcd for $C_{36}H_{51}N_4O_{15}S^+$ $[M]^+$: 811.3, found 811.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.34 (s, 1H, Ar), 8.75 (s, 1H, Ar), 8.58 (d, *J* = 8.1 Hz, 1H, Ar), 8.37 (s, 1H, thaizole), 7.73 (dd, *J* = 8.0, 5.1 Hz, 1H, Ar), 7.11 – 6.99 (m, 2H, Ar), 6.90 – 6.81 (m, 2H, Ar), 5.05 (d, *J* = 3.9 Hz, 1H, H-1'), 4.86 (d, *J* = 7.5 Hz, 1H, H-1), 4.36 (t, *J* = 6.1 Hz, 1H, H-5'), 4.18 (d, *J* = 2.6 Hz, 1H, H-4'), 4.10 (d, *J* = 3.0 Hz, 1H, H-4), 4.03 – 3.98 (m, 1H, OC*H*HCH_2), 3.97 (dd, *J* = 10.2, 3.9 Hz, 1H, H-2'), 3.92 – 3.63 (m, 22H, H-2, H-3, H-3', H-5, H-6a-b, H6'a-b, 4xOCH_2,

pyranoseOCH*H*CH_2, OCH_3), 3.40 – 3.34 (m, 1H, pyranoseOCH_2C*H*HNH), 3.32 – 3.25 (m, 5H, pyranoseOCH_2CH*H*NH, CH_2NHCH_2).

^{13}C NMR (126 MHz, CD_3OD): δ = 165.55, 163.30, 156.77, 152.99, 152.18, 150.46, 146.81, 137.56, 126.34, 119.09, 115.55, 103.94 (C-1), 102.23 (C-1'), 79.91, 79.04 (C-4), 76.36, 74.50, 72.64, 72.21 (C-5'), 71.50, 71.38, 70.67, 69.95 (C-2'), 68.02 (C-4'), 66.89, 64.97, 62.47 (C-6'), 61.05 (C-6), 56.07 (OCH_3), 40.29.

6-Azido-1-bromohexane (40):

Preparation of the triflyl azide stock solution: To a solution of sodium azide (545 mg, 8.38 mmol) in water (1.37 mL) was added toluene (1.37 mL). The mixture was cooled to 0 °C with vigorous stirring. After dropwise addition of triflyc anhydride (896 µL, 4.19 mmol) and further vigorous stirring for 30 min at 0 °C, the temperature was raised to 10 °C and the biphasic mixture was stirred for another 2 h. A saturated aqueous solution of sodium hydrogencarbonate was added dropwise until gas evolution ceased. The two phases were separated and the aqueous layer was extracted with toluene (2 x 1.37 mL). The combined organic layers were used in the subsequent diazo transfer reactions.

Diazo Transfer: The amine, 6-amino-1-hexanol (50 mg, 0.43 mmol), sodium hydrogencarbonate (143.4 mg, 1.71 mmol) and copper(II) sulfate pentahydrate (4.3 mg, 0.02 mmol) were dissolved in water (0.6 mL). Triflic azide stock solution (2 M, 1 mL) was added, followed by the addition of methanol (3.9 mL) to yield a homogeneous system. Subsequently, the blue mixture was stirred vigorously at rt for 15 min. The solvents were removed in vacuo with a rotary evaporator keeping the temperature strictly below 30 °C. The residue was purified by chromatography on silica gel (gradient PE/EtOAc) to yield 54 mg of the azido alcohol (88%).

The product was dissolved in DCM (1 mL) and added dropwise to a cooled solution (0 °C) of CBr_4 (150 mg, 0.45 mmol) and PPh_3 (98.9 mg, 0.38 mmol) in DCM (4 mL). The solution was stirred at 0 °C for 1 h, then evaporated (bath at 30 °C). The residue was purified on silica (PE/Et_2O, gradient 5% to 15%) to yield 57 mg (73 %) of **40**.

^{1}H NMR (500 MHz, $CDCl_3$)*:* δ = 3.41 (t, *J* = 6.7 Hz, 2H), 3.28 (t, *J* = 6.9 Hz, 2H), 1.96 – 1.79 (m, 2H), 1.62 (m, 2H), 1.53 – 1.35 (m, 4H).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 51.33, 33.64, 32.56, 28.70, 27.69, 25.91.

3-*O*-(6-Azidohexanyl)-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (41):

Compound **16** (24 mg, 0.032 mmol) was dissolved in toluene (2 mL) and degassed with ultrasound under light vacuum. Bu_2SnO (9 mg, 0.038 mmol) was added and the mixture was stirred under argon at reflux for 2 h, with azeotropic removal of water. The residual toluene was evaporated and the white solid was dried under high vacuum for 1 h. It was then dissolved in a solution of **40** (57 mg) in dry toluene (0.5 mL) under argon. The resulting mixture was stirred for 2 d at 75 °C under argon in a sealed flask. Afterwards, the solution was cooled to rt and applied to a silica column. It was eluted with a DCM/MeOH gradient to yield 13 mg (47%) of the 2,3,6-tri-*O*-benzoyl protected intermediate. The latter was dissolved in dry MeOH (0.5 mL). To the resulting solution, 5 µL of a freshly prepared solution of 1 M MeONa in MeOH were added. The mixture was stirred at rt under argon for 5 h, before the solvent was evaporated. The residue was purified on silica (DCM/MeOH gradient) to yield 13.7 mg of **41** (85%).

$[\alpha]_D^{20}$ +12.7 (*c* 1.05, MeOH).

ESI-MS: *m/z*: Calcd for $C_{25}H_{39}N_3NaO_{12}^+$ $[M+Na]^+$: 596.2, found 596.3.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 6.98 (d, *J* = 9.0 Hz, 2H), 6.78 (d, *J* = 9.1 Hz, 2H), 5.11 (d, *J* = 2.5 Hz, 1H, H-1'), 4.74 (d, *J* = 7.5 Hz, 1H, H-1), 4.60 – 4.40 (m, 1H, H-5', 1 OH), 4.21 (m, 2H, 2 OH), 4.12 (s, 1H, H-4), 4.02 (m, 2H, H-4', H-2'), 3.96 – 3.86 (m, 2H, H-6a, H-5), 3.84 – 3.71 (m, 7H, OCH_3, H-6'a, H-6b, H-2), 3.71 – 3.54 (m, 3H, H-6'b, H-3, H-3'), 3.48 (m, 1H, OC*H*H$(CH_2)_5N_3$), 3.32 (s, 1H, 4'-OH), 3.24 (t, *J* = 6.8 Hz, 2H, O$(CH_2)_5$CH_2N_3), 1.64 – 1.52 (m, 4H), 1.45 – 1.31 (m, 4H).

^{13}C NMR (126 MHz, $CDCl_3$): δ = 155.40 (Ar), 151.30 (Ar), 118.42 (Ar), 114.56 (Ar), 102.60 (C-1'), 101.39 (C-1), 79.92 (C-4), 77.72 (C-3'), 73.66, 73.54, 72.00, 71.70, 69.86 (O*C*$H_2(CH_2)_5N_3$), 68.30, 67.06, 62.32, 60.07, 55.61 (OCH_3), 51.33 (CH_2N_3), 29.68, 28.70, 26.48, 25.50.

3-*O*-(6-(((2-(3-Pyridyl)-thiazol-4-yl)carbonyl)amino)-hex-6-yl-α-D-galactopyranosyl-(1→4)-β-D-galactopyranoside (43):

Compound **41** (13.7 mg, 0.024 mmol) was dissolved in 0.5 mL of MeOH under argon. To this solution, Pd/C (10%, 6 mg) was added. The mixture was hydrogenated (1 bar H_2) for 40 min, afterwards it was diluted with MeOH and filtered through a PTFE filter (0.45 µm). The filtrate was concentrated under reduced pressure, to afford

12.7 mg of the amine compound. A mixture of HBTU (13 mg, 0.035 mmol), HOBt (9.4 mg, 0.061 mmol) and **42** (7.2 mg, 0.035 mmol) was dissolved in dry DMF (0.5 mL) under argon. After shaking for 10 min, this solution was added to the amine compound under argon. To this solution was added DIPEA (12 µL, 0.023 mmol). Solvents were removed under vacuum after 1.5 h. The residue was purified on silica gel (DCM/MeOH gradient). The fractions containing the product were further purified by RP C-18 column (water/ACN gradient) to afford **43** (10 mg, 59%).

$[\alpha]_D^{20}$ +12.5 (*c* 0.5, MeOH).

ESI-MS: *m/z*: Calcd for $C_{34}H_{46}N_3O_{13}S^+$ $[M+Na]^+$: 736.3, found 736.4.

^{1}H NMR (500 MHz, CD_3OD): δ = 9.25 (d, *J* = 2.0 Hz, 1H, pyridine-H-1), 8.67 (dd, *J* = 4.9, 1.3 Hz, 1H, pyridine-H-5), 8.45 (dt, *J* = 8.0, 1.8 Hz, 1H, pyridine-H-3), 8.29 (s, 1H, thiazole-CH), 7.59 (dd, *J* = 8.0, 4.9 Hz, 1H, pyridine-H-4), 7.13 – 7.02 (m, 2H, 2 x phenyl-CH), 6.89 – 6.79 (m, 2H, 2 x phenyl-CH), 5.02 (d, *J* = 3.9 Hz, 1H, H-1'), 4.84 (d, *J* = 7.6 Hz, 1H, H-1), 4.31 (t, *J* = 6.2 Hz, 1H, H-5'), 4.14 (d, *J* = 2.6 Hz, 1H, H-4'), 4.08 (d, *J* = 3.0 Hz, 1H, H-4), 3.94 – 3.86 (m, 2H, H-6a, H-2'), 3.84 – 3.69 (m, 9H, OCH_3, H-2, H-5, H-6b, H-6'a, H-6'b, OC*H*H$(CH_2)_5$N), 3.66 (dd, *J* = 10.1, 3.0 Hz, 1H, H-3), 3.62 – 3.55 (m, 2H, H-3', OCH*H*$(CH_2)_5$N), 3.46 (t, *J* = 7.1 Hz, 2H, CONH(C*H*$_2$)), 1.75 – 1.63 (m, 4H, $OCH_2CH_2(CH_2)_2CH_2CH_2N$), 1.56 – 1.40 (m, 4H, $O(CH_2)_2CH_2CH_2(CH_2)_2N$).

^{13}C NMR (126 MHz, CD_3OD): δ = 166.07 (C=O), 163.17 (Ar), 156.78 (Ar), 152.99 (Ar), 152.37 (Ar), 151.80 (Ar), 148.12 (Ar), 136.03 (Ar), 130.85 (Ar), 125.73 (Ar), 125.53 (Ar), 119.27 (Ar), 115.52 (Ar), 104.01 (C-1), 102.70 (C-1'), 79.46, 79.19 (C-4), 76.28, 74.64 (C-3), 72.74, 72.64, 70.71 (O*C*H$_2$$(CH_2)_5$N), 69.91 (C-2'), 67.94 (C-4'), 62.71 (C-2'), 60.98 (C-6), 56.07 (OCH_3), 40.47, 30.89 (OCH_2*C*H$_2$$(CH_2)_2CH_2CH_2N$), 30.61 ($OCH_2CH_2(CH_2)_2$*C*H$_2CH_2N$), 27.89 ($O(CH_2)_2$*C*H$_2CH_2(CH_2)_2N$), 26.84 ($O(CH_2)_2CH_2$*C*H$_2(CH_2)_2N$).

3. References

[S1] P. J. Hajduk, E. T. Olejniczak, S. W. Fesik. *J. Am. Chem. Soc.* **1997**, *119*, 12257-12261.

[S2] T. L. Hwang, A. J. Shaka. *J. Magn. Reson. A* **1995**, *112*, 275-279.

[S3] S. Rabbani, X. Jiang, O. Schwardt, B. Ernst. *Anal. Biochem*. **2010**, *407*, 188-195.

[S4] M. Mayer, B. Meyer. *Angew Chem Int Ed Engl* **1999**, *38*, 1784-1788.

[S5] M. Mayer, B. Meyer. *J. Am. Chem. Soc.* **2001**, *123*, 6108-6117.

[S6] S. Kemper, M. K. Patel, J. C. Errey, B. G. Davis, J. A. Jones, T. D. Claridge. *J. Magn. Reson.* **2010**, *203*, 1-10.

[S7] J. Ohlsson, J. Jass, B. E. Uhlin, J. Kihlberg, U. J. Nilsson. *ChemBioChem* **2002**, 3, 772-779.

[S8] J. Ohlsson, G. Magnusson. *Carbohydr. Res.* **2000**, *329*, 49 - 55.

3.2.4 Paper 5

Carbohydrate–Lectin Interactions: An unexpected contribution to affinity

This paper describes the study of the interaction of PapG-II with the carbohydrate epitopes of sialyl galactosyl globoside (SGG), globotetraosylceramide (GbO4), and with the glycomimetic 4-methoxyphenyl β-galabiose (**4**). The SGG epitope bears a Neu5Acα(2-3)Galβ(1-3) extension at the non-reducing end and exhibits a 5-fold increase in affinity in respect to the GbO4 epitope. As the additional disaccharide moiety lies outside the known PapG-II binding pocket, the superior binding was unexpected. By studying crystal structures, thermodynamic fingerprints, and solution conformations, an unanticipated entropy-driven contribution to the binding was identified.

Contribution to the project:

Giulio Navarra synthesized compound **4**, interpreted the results from isothermal titration calorimetry (ITC), the crystal structures, the molecular dynamics simulation, and wrote the manuscript, with the exception of the supporting information on ITC, on protein expression, and on crystallography.

This paper was published in *ChemBioChem*:

Giulio Navarra, Pascal Zihlmann, Roman P. Jakob, K. Stangier, Roland C. Preston, Said Rabbani, Martin Smiesko, Bea Wagner, Timm Maier, and Beat Ernst

Reprinted with permission from Navarra et al. *ChemBioChem* **2017**, *18*, 539-544.

ChemPubSoc Europe
DOI: 10.1002/cbic.201600615
CHEMBIOCHEM Full Papers

Carbohydrate–Lectin Interactions: An Unexpected Contribution to Affinity

Giulio Navarra,[a] Pascal Zihlmann,[a] Roman P. Jakob,[b] Katja Stangier,[a] Roland C. Preston,[a] Said Rabbani,[a] Martin Smiesko,[a] Bea Wagner,[a] Timm Maier,[b] and Beat Ernst*[a]

Uropathogenic E coli exploit PapG-II adhesin for infecting host cells of the kidney; the expression of PapG-II at the tip of bacterial pili correlates with the onset of pyelonephritis in humans, a potentially life-threatening condition. It was envisaged that blocking PapG-II (and thus bacterial adhesion) would provide a viable therapeutic alternative to conventional antibiotic treatment. In our search for potent PapG-II antagonists, we observed an increase in affinity when tetrasaccharide 1, the natural ligand of PapG-II in human kidneys, was elongated to hexasaccharide 2, even though the additional Siaa(2–3)Gal extension is not in direct contact with the lectin. ITC studies suggest that the increased affinity results from partial desolvation of nonbinding regions of the hexasaccharide; this is ultimately responsible for perturbation of the outer hydration layers. Our results are in agreement with previous observations and suggest a general mechanism for modulating carbohydrate–protein interactions based on nonbinding regions of the ligand.

Introduction

In numerous bacterial infections, adhesins mediate interactions with host cells and thereby function as important virulence factors.[1] In this process, glycans (as parts of glycoproteins or glycolipids) on host cell surfaces function as ligands for bacterial adhesins.[2,3] A classical example is the fimbrial adhesin FimH of uropathogenic Escherichia coli (UPEC), which binds to and infects urothelial cells, a process initiating urinary tract infections (UTI).[4] When UPEC strains cause infections in the upper urinary tract, they take advantage of an additional adhesin, PapG, which exists in three molecular variants (classes I–III)[5–9] classified according to their slightly different agglutination patterns.[7,8] The PapG-II variant is of particular medical interest because of its strong association with pyelonephritis in humans,[10–13] a potentially life-threatening disease and a frequent complication during pregnancy.[14] In addition, there is evidence of the involvement of PapG-II in bacteremia.[15] PapG-II preferentially binds to the globoside GbO4,[16] which is present in the human upper urinary tract. Its binding epitope is the tetrasaccharide 1, which consists of a galabiose core (Galα(1–4)Gal) flanked by a b(1–3)-linked N-acetyl glucosamine residue on the nonreducing and a b(1–4)-linked glucose residue on the reducing end.[11] Stapleton et al. reported increased affinity for the natural sialosyl galactosyl globoside (SGG), which bears the P blood group antigen Neu5Acα(2–3)Galb(1–3)GalNAcb(1–3)Galα(1–4)Galb(1–4)Glc (SSEA4, 2).[17] Individuals expressing SGG in the urinary tract are more susceptible to recurrent UTI. Clinical isolates from such individuals exhibit increased numbers of bacteria expressing the pap genes, which encode PapG proteins.[17]

The current treatment of pyelonephritis with antibiotics[18] has become increasingly inefficacious because of antibiotic resistance.[18–20] Therefore, new therapeutic options are of urgent importance. A promising new strategy to prevent colonization of the kidneys is blocking the initial bacterial adhesion mediated by PapG-II by using soluble antagonists (antiadhesive therapy). As such, PapG-II antagonists are not bactericidal and therefore do not exert any selection pressure, and the probability of developing resistance is minimized.[21]

Despite detailed structural knowledge of the PapG-II lectin domain (apo form: PDB ID 1J8S; co-crystallized with epitope 1 of GbO4: PDB ID 1J8R),[22] the design of small-molecule antagonists has been only partly successful. One possible reason is the extended hydrogen-bond network that is indispensable for binding of the central galabiose. Structural modifications disrupted this network, thereby resulting in substantial reduction in affinity.[23–29]

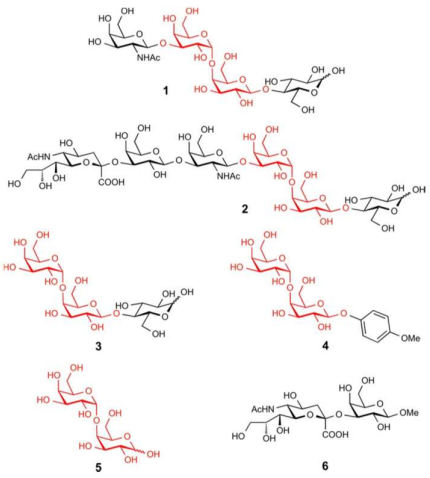
The goal of the present work was to answer questions regarding the individual contributions to binding of the various monosaccharide moieties of tetrasaccharide 1, hexasaccharide 2, and the p-methoxyphenyl aglycone in antagonist 4. Furthermore, the question why almost identical affinities were reported for 1 and trisaccharide 3 was addressed (Scheme 1).

[a] Dr. G. Navarra, Dr. P. Zihlmann, Dr. K. Stangier, Dr. R. C. Preston, Dr. S. Rabbani, Dr. M. Smiesko, B. Wagner, Prof. Dr. B. Ernst
Institute of Molecular Pharmacy, Pharmacenter, University of Basel
Klingelbergstrasse 50, 4056 Basel (Switzerland)
E-mail: beat.ernst@unibas.ch

[b] Dr. R. P. Jakob, Prof. Dr. T. Maier
Biozentrum, Focal Area Structural Biology, University of Basel
Klingelbergstrasse 70, 4056 Basel (Switzerland)

Supporting information for this article can be found under: http://dx.doi.org/10.1002/cbic.201600615.

ChemPubSoc Europe
CHEMBIOCHEM Full Papers



Scheme 1. The tetrasaccharide epitope (**1**) of GbO4, hexasaccharide epitope (**2**) of SSEA4, trisaccharide epitope **3** of GbO3, pseudotrisaccharide **4**, galabiose (Gal(α1-4)Gal, **5**), and disaccharide Neu5Ac(α2-3)GalβMe (**6**). The galabiose core is highlighted in red.

Results and Discussion

For a molecular understanding of the binding affinities of oligosaccharides **1**–**6**, the N-terminal lectin domain (LD) of PapG-II (PapG-II$_{LD}$: residues 1–196) from the clinical UPEC isolate BI47 (from the University Hospital Basel, Switzerland) was produced in genetically modified *E. coli.*, and dissociation constants (K_D) and thermodynamic fingerprints were determined by isothermal titration calorimetry (ITC). Because adaption to the bioactive conformation can impose substantial entropy costs and thereby influence binding affinity, solution conformations of ligands were analyzed by molecular dynamics and compared to the PapG-II-bound conformation obtained by X-ray crystallography. Finally, by co-crystallizing of PapG-II with ligands **1**, **2**, and **4**, a detailed view of the H-bond network was obtained.

Isothermal titration calorimetry

As affinity data are available only for the trimethylsilylethyl (TMSEt) glycosides of **1** and **3**[16] (and relative affinities for hexasaccharide **2**),[17] the thermodynamic profiles of **1**–**6** binding to PapG-II$_{LD}$ were determined by ITC. Because of the low solubility of PapG-II$_{LD}$ and the low binding affinities for all test compounds, obtained *c* values were lower than 1.0, thus requiring the stoichiometry (*N* value) to be set at 1.[30] Whereas the introduction of the disaccharide unit Neu5Acα(2–3)Galβ(1–3) at the nonreducing end of tetrasaccharide **1** (→**2**) led to a threefold improvement of affinity, the isolated disaccharide methyl Neu5Acα(2–3)Galβ(1–3)[31] (**6**) turned out to be inactive (dissociation constant $K_D > 10$ mM). The K_D values of **1** and **4** determined by ITC correlate with published values, although these were obtained in a different assay format and for the corresponding TMSEt derivatives.[16] However, for unknown reasons, we obtained a twofold lower affinity for **3**. Hexasaccharide **2** ($K_D = 21.9$ μM) was the most potent PapG-II ligand. The interaction is strongly enthalpy driven ($\Delta H^\circ_{obs} = -53.9$ kJ mol^{-1}), partly compensated for by an unfavorable entropy ($-T\Delta S^\circ_{obs} = 27.3$ kJ mol^{-1}). Removal of the terminal disaccharide unit (→tetrasaccharide **1**) resulted in a threefold reduction in affinity. Besides a slightly improved enthalpy ($\Delta\Delta H^\circ_{obs\,2\to1} = -1.9$ kJ mol^{-1}), a substantial, nonbeneficial entropy ($T\Delta\Delta S^\circ_{obs\,2\to1} = 4.3$ kJ mol^{-1}) was responsible for the reduction of affinity.

Removal of the Gal*N*Ac moiety at the nonreducing end of **1** (→**3**) led to a substantial improvement in enthalpy ($\Delta\Delta H^\circ_{obs\,1\to3} = -6.4$ kJ mol^{-1}). This, however, was overcompensated for by a severe entropy penalty ($T\Delta\Delta S^\circ_{obs\,1\to3} = 8.4$ kJ mol^{-1}). As a result, **3** showed a twofold reduction in affinity compared to **1**.

Further shortening (removal of the reducing-end glucose residue; **3**→**5**, galabiose) decreased the affinity by nearly one order of magnitude, almost exclusively due to a loss in entropy ($T\Delta\Delta S^\circ_{obs\,3\to5} = 4.5$ kJ mol^{-1}). Surprisingly, the enthalpy term remained almost unchanged ($\Delta\Delta H^\circ_{obs\,3\to5} = 0.9$ kJ mol^{-1}). When a 4-methoxyphenyl aglycone was introduced to galabiose (**5**; →**4**),[16] a tenfold improvement of affinity was observed, mainly as a consequence of the improved entropy ($T\Delta\Delta S^\circ_{obs\,5\to4} = -7.7$ kJ mol^{-1}), accompanied by a minor, nonbeneficial enthalpy change ($\Delta\Delta H^\circ_{obs\,5\to4} = 1.8$ kJ mol^{-1}). Although trisaccharide **3** and the galabiose derivative **4** showed similar affinities, they exhibited markedly different thermodynamic profiles: enthalpy was more beneficial for **3** ($\Delta\Delta H^\circ_{obs\,3\to4} = 2.7$ kJ mol^{-1}), and entropy was more beneficial for **4** ($T\Delta\Delta S^\circ_{obs\,3\to4} = -3.2$ kJ mol^{-1}). This enthalpy–entropy compensation reflects the "classical" hydrophobic effect.[32] The introduction of a lipophilic aglycone in place of the polar glucose (**3**→**4**) decrease the enthalpy and increase the entropy. The ITC results (Table 1 and Figure 1) reveal an interesting trend: with decreasing size of the oligosaccharide, enthalpy gains while entropy penalties increase. This, at first surprising, trend reflects the high desolvation cost of the various hydroxyl groups of carbohydrates.[33] Reduction

Table 1. Affinity and ITC data for saccharides **1**–**6**. Entries are listed in order of ΔG°_{obs} value. Literature K_D values[16] and the IC_{50} value[24] were available for the corresponding TMSEt glycosides.[16] Compounds **1**–**4** were measured in triplicate and globally fitted in SEDPHAT software. *N* was set to 1 for all measurements.

Cpd	K_D [μM]	ΔG°_{obs} [kJ mol^{-1}]	ΔH°_{obs} [kJ mol^{-1}]	$-T\Delta S^\circ_{obs}$ [kJ mol^{-1}]	K_D[16] [μM]	IC_{50}[24] [μM]
2	21.9	−26.6	−53.9	27.3	–	–
1	59.1	−24.1	−55.8	31.6	66	–
4	105.1	−22.7	−59.5	36.8	140	110
3	129.4	−22.2	−62.2	40.0	78	–
5	1160	−16.8	−61.3	44.5	–	–
6	>10000	–	–	–	–	–

ChemPubSoc Europe
CHEMBIOCHEM Full Papers

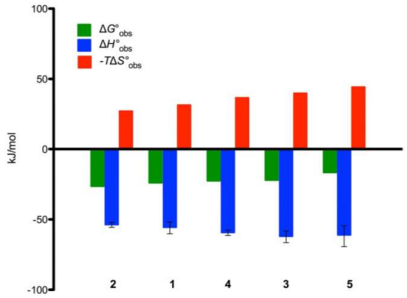


Figure 1. Thermodynamic parameters of saccharides **1**–**5** determined by direct ITC titrations (mean ± SD, $n = 3$).

in oligosaccharide size goes hand-in-hand with a decreasing number of hydroxyl groups to be desolvated. As a result, reduction in enthalpy cost is accompanied by a simultaneous smaller entropy gain (because of the lower number of water molecules released to bulk). The only deviation from this trend was **4** (aromatic aglycone in place of the reducing terminal glucose).

Although similar observations have been made for oligosaccharides binding to the lectin domains of *Clostridium stercorarium* and *Thermotoga maritima*,[34] an explanation for this effect is not available, probably because multiple binding modes could not be excluded. As PapG adhesins strictly require the Galα(1–4)Gal core for binding affinity,[35] multiple binding modes can be excluded in the present study.

Molecular dynamics simulations

Ligands pre-organized in the bioactive conformation benefit from improved target affinity.[36] In the case of PapG-II ligands, the central galabiose core forms an extended H-bond network with the target protein.[22] Extension of galabiose **5** on the reducing and/or nonreducing end would influence its conformation, thus the entropy would be affected. In order to determine whether the affinity increase observed upon chain elongation (from disaccharide **5** to hexasaccharide **2**) was a conformational effect, the conformation of galabiose **5** and the galabiose core of **1**, **2**, and **4** were analyzed by molecular dynamics (MD) simulation. The results (Figure 2) clearly indicate similar solution conformations and therefore comparable entropy penalties related to the adoption of the bound conformation of the galabiose core.

Obviously, ligand binding can affect protein dynamics and thus affinity.[42] However, the interaction of PapG-II with its ligands has been described as a "rigid body type" interaction and is not accompanied by substantial conformational changes.[22]

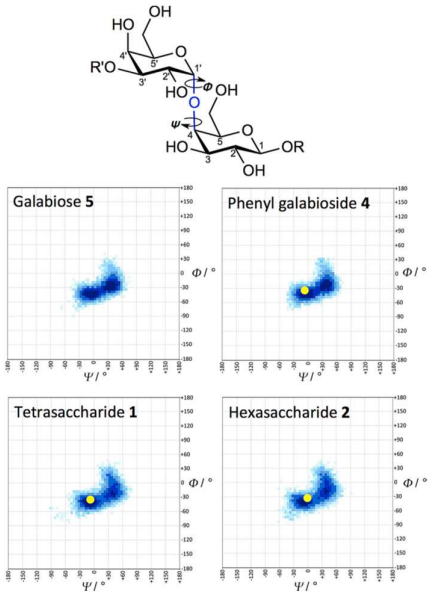


Figure 2. Evaluation of Ψ and Φ dihedral angles (indicated, top) for galabiose (**5**) and the galabiose cores of **1**, **2**, and **4**. Blue: calculated by MD simulation; yellow: values from crystal co-structures with PapG-II$_{LD}$ (Figure 3).

X-ray crystallographic analysis of PapG-II/ligand complexes

In order to rationalize the improved affinity of **2** over **1** and the comparable affinities of trisaccharide **3** and the pseudotrisaccharide **4**, the crystal structures of PapG-II$_{LD}$ in complex with **2** and **4** were determined and compared to PapG-II$_{LD}$ co-crystallized with tetrasaccharide **1** (PDB ID: 1J8R).[22] Whereas our crystal structures were obtained with PapG-II$_{LD}$ from the patient *E. coli* isolate BI47, the previous study used PapG-II$_{LD}$ from the *E. coli* isolate AD110.[22] The two proteins differ at eight residues, all of which are outside the sphingolipid binding site. Overall, the resulting crystal structures of PapG-II from the two isolates are virtually identical (Figure S1 in the Supporting Information), and thus the variations between them are very unlikely to have an influence on binding affinity. This observation is supported by the similarity of data among previous studies and ours.

Crystal structure of apoPapG-II$_{LD}$ from E. coli *BI47* (Figure 3 A). Two crystal forms were obtained, with diffraction to 1.74 and 2.50 Å (Table S1). The structures were solved by molecular replacement by using the apo structure of PapG-II$_{LD}$ from *E. coli* AD110 (PDB ID: 1J8S). The protein shows an elongated β-barrel "jellyroll" fold, typical for bacterial lectins. The overall confor-

ChemPubSoc Europe
CHEMBIOCHEM Full Papers

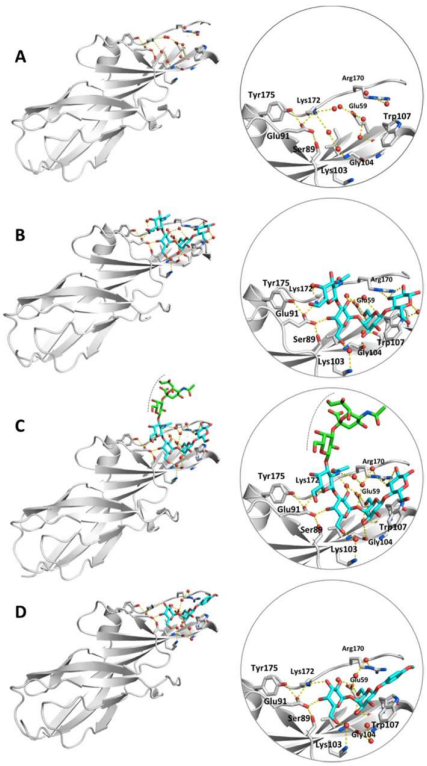


Figure 3. Crystal structures of protein/ligand complexes. A) apo PapG-II_{LD}; B) complex with **1** (epitope of GbO4);[22] C) complex with **2** (epitope of SSEA4); D) complex with *p*-methoxyphenyl β-D-galabiose (**4**). The protein is shown in gray; ligand parts in contact with the protein are cyan, solvent-exposed ligand parts are green; dashed line in (C) indicates crystal contacts to a symmetry-related PapG molecule.

mation of PapG-II_{LD} from isolate BI47 was virtually identical to that of *E. coli* AD110 (rmsd only 0.4 Å; Figure S1).

*PapG-II_{LD} co-crystallized with tetrasaccharide **1*** (Figure 3B). Dodson et al. published the first crystal structure of **1** in complex with PapG-II_{LD} from *E. coli* AD110.[22] The ligand binds in its V-shape solution conformation (Figure 2), with one leg comprising Gal*N*Acβ(1–3)Gal, and the other being formed by the second Galβ(1–4)-linked to a Glc unit. The central galabiose core forms a dense network of H-bonds and water-mediated interactions involving Gly104, Glu59, Lys106, Lys172, Glu91, and Lys103, as well as apolar interactions with Trp107, Ile61, Leu102, and the aliphatic part of the Lys172 side-chain. The Glc moiety forms polar interactions with $N_{\varepsilon 1}$ of Trp107 and with the charged guanidinium moiety of Arg170, as well as apolar interactions with Trp107. Finally, the Gal*N*Ac unit is involved in indirect, water-mediated interactions with Glu59, Lys172, Tyr175, and Arg92 (not shown).

*PapG-II_{LD} co-crystallized with the hexasaccharide **2*** (Figure 3C). In order to identify the reason for the difference in binding affinities between **1** (epitope of GbO4) and **2** (epitope of SSEA4), we resolved co-crystal structure of PapG-II_{LD} with **2** in two space groups, to 1.73 and 1.80 Å resolution (Table S1). The structure of the ligand-bound form of the protein is strikingly similar to that of the apo protein (rmsd = 0.4 Å, Figure S4), and binding-site residues adopt equivalent conformations. In crystals of both space groups, **2** binds in an almost identical horseshoe-like conformation (Figure S2) and is involved in contacts to a neighboring PapG molecule in the crystal. These contacts are due to crystal packing. Formation of these contacts in solution is excluded by the ITC measurements (Figure S6, Table S2). There is no considerable difference in the interaction patterns of **2** and **1** (Figure 3B, C). Thus, for both oligosaccharides, similar polar interactions are observed (two water-mediated interactions and contact with Trp107). Moreover, the solvent-accessible area of PapG-II_{LD} buried upon ligand-binding was calculated by using PDBePISA.[37] It is practically identical for **2** and **1** (487 and 489 $Å^2$, respectively), thus indicating that the additional Neu5Acα(2–3)Galβ disaccharide does not directly contribute to PapG-II_{LD} binding.

*Structure of PapG-II_{LD} co-crystallized with 4-methoxyphenyl β-D-galabiose (**4**)* (Figure 3D). When **4** was co-crystalized with PapG-II_{LD}, two crystal forms were obtained (1.45 and 1.5 Å resolution; Table S1). Again, the overall conformation is very similar to the structures described above (rmsd = 0.6 Å). In the bound conformation, residues of the galabiose core involved in ligand binding are in identical orientations for **1** and **2** (Figure S3). The aromatic ring is close to Trp 107 and Arg170. However, because of the observed binding mode we can exclude a strong cation–π interaction with the guanidinium moiety for the latter (Figure S5).[41] Moreover, electron-deficient analogues of **4** have shown comparable affinities.[29] According to our ITC data, the aromatic ring increases the affinity by improving the entropy, thus suggesting that the polar interactions established by the glucose unit in **2** and **3** are largely compensated for by desolvation cost, and that the main contribution of the glucose unit to binding arise from lipophilic interactions established by the hydrophobic β-face. Thus, substitution of the glucose by 4-methoxyphenyl (→**4**) preserved this interaction, but with reduced desolvation cost.

As the binding pocket of PapG-II has been shown to accommodate up to four saccharide units (Figure 3B), the increased affinity of tetrasaccharide **1** over trisaccharide **3** is attributable to both enthalpic and entropic effects. Thus, the increased number of ligand–protein interactions established by **1** has a beneficial enthalpic effect, but this is overcompensated for by the desolvation penalty ($\Delta\Delta H^\circ_{obs\,\mathbf{3}\to\mathbf{1}} = 6.4\ kJ\,mol^{-1}$). On the other hand, as binding of **1** increases the number of water molecules from the solvation shell released to bulk, a beneficial entropy ($T\Delta\Delta S^\circ_{obs\,\mathbf{3}\to\mathbf{1}} = -8.4\ kJ\,mol^{-1}$) is observed.

ChemPubSoc Europe

CHEMBIOCHEM Full Papers

The higher affinity of **2** over **1** was unexpected (Figure 3 B, C). How can the additional disaccharide moiety, which is outside the binding pocket, induce a threefold increase in affinity? Because the solvation layers of the polyamphiphilic surfaces[38] of protein and ligand can exhibit an extension of up to 10 Å,[39] the nonbinding disaccharide moiety of **2** and the neighboring protein surface are, for steric reasons, forced to release water molecules from the outer solvation shells to bulk water.

Obviously, partial desolvation of ligand and protein is not free. However, as outer water layers feel only a marginal electrostatic influence of the polar surfaces of ligand and PapG-II, the enthalpy penalty is rather small ($\Delta\Delta H^{\circ}_{obs\,1\rightarrow 2} = 1.9\ kJ\,mol^{-1}$). As a result, the extension by a nonbinding disaccharide on the nonreducing end of tetrasaccharide **1** (→**2**) contributes substantially to binding ($\Delta\Delta G^{\circ}_{obs\,1\rightarrow 2} = -2.5\ kJ\,mol^{-1}$).

Conclusion

Uropathogenic *E. coli* exploit PapG-II adhesin for infecting host cells in the kidney. They preferentially bind to tetrasaccharide **1** (part of the globoside GbO4),[16] which is expressed in the human upper urinary tract. In order to understand the cause of the higher affinity exhibited by hexasaccharide **2** in comparison to tetrasaccharide **1** (as well as the comparable affinities of trisaccharide **3** and the simplified mimic **4**), ITC, X-ray crystallography and MD simulation were employed. Elongation of **1** at its nonreducing terminus with the disaccharide Neu5Acα(2–3)Galβ (→**2**) led to substantially improved affinity, although no affinity was found for disaccharide **6** on its own. Furthermore, analysis of the structure of PapG-II_{LD} co-crystallized with **2** provided no evidence for a direct interaction of the additional Neu5Acα(2–3)Galβ moiety with the protein. Supported by ITC data, our explanation is based on the presence of extended solvation shells (up to 10 Å)[38,39] at the surface of both protein and ligand. Upon complex formation, the disaccharide moiety is close to the PapG-II surface, thus forcing both ligand and protein to release water molecules from their outer solvation shells, thereby resulting in an improvement in the entropy of 4.3 $kJ\,mol^{-1}$. On the other hand, the penalty for this desolvation is rather small, because the electrostatic forces exerted by the distant polar surfaces of ligand and protein are also rather small (1.9 $kJ\,mol^{-1}$). As a result, an unexpected threefold improvement of affinity was obtained. A similar observation was made for the equipotent trisaccharide **3** and its mimic **4** in the entropy ($T\Delta\Delta S^{\circ}_{obs\,4\rightarrow 3} = 3.2\ kJ\,mol^{-1}$). Although the phenyl ring at the reducing end of **4** has been shown to play a crucial role in PapG affinity,[23,28] it is largely solvent exposed. Again, the proximity of the protein surface forces partial desolvation. In contrast, the glucose moiety in **3** establishes H-bond interactions with PapG-II, thus leading to an enthalpy improvement of comparable size ($\Delta\Delta H^{\circ}_{obs\,3\rightarrow 4} = -2.7$).

By combining ITC data with structural information from X-ray crystallography and MD simulation, the molecular basis for a series of carbohydrate–lectin interactions was revealed. Of special interest is the proximity effect of nonbinding parts of hexasaccharide **2**. The space between the disaccharide moiety at the nonreducing end of **2** and the protein surface forces partial desolvation and thus leads to a substantially improved entropy. A similar effect has been observed for other protein–carbohydrate interactions. The carbohydrate-binding module (CBM) of *Thermotoga maritima* can host up to five saccharide units. Its affinities for laminari-pentaose and laminari-hexaose are identical, but the longer oligosaccharide binds with a smaller enthalpy gain and a smaller entropy penalty.[34] When *xylo*-oligosaccharides bind to CBM6-1 of *Clostridium stercorarium*, the affinity increases from xylobiose to xylopentaose, as a result of reduced entropy penalties that offset the increased enthalpy penalties. Notably, xylopentaose shows tighter binding than xylotetraose, although the CBM hosts up to only four saccharide units. Finally, a much higher affinity for sialyl Lewis^{x} compared to Lewis^{x} was found for the Lewis-binding Norovirus VA207, despite no apparent contact of the additional sialyl moiety with the binding site.[40]

The data presented here contribute to a deeper understanding of ligand recognition by PapG-II, and will guide future medicinal chemistry work. Furthermore, one additional example is provided—what appears as a Nature's strategy to improve affinity of carbohydrate–lectin interactions—namely the entropy-driven contribution of nonbinding saccharide moieties to binding.

Experimental Section

Oligosaccharides 1–6: Compounds **4** and **6** were synthesized as described in refs. [24] and [31], respectively. All other compounds were purchased from Elicityl (Crolles, France).

Protein expression and purification: PapG-II protein was produced in *E. coli* AD494(DE3) clones containing the constructs PapGII_{LD}-6His or PapG-II_{LD} and purified by dialysis and affinity chromatography. Detailed procedures are in the Supporting Information.

Isothermal titration calorimetry: ITC experiments were performed with a VP-ITC instrument (GE Healthcare). Detailed procedures are in the Supporting Information.

Crystallography: X-ray crystallographic data were measured at SLS beamlines X06DA and X06SA (Swiss Light Source, Paul Scherrer Institute, Switzerland) at 100 K. Crystallization procedures, data processing, and structure determination are in the Supporting Information.

Acknowledgements

This work was supported by the Swiss National Science Foundation Grant R'EQUIP 145023 and partly funded by the SNF: Marie Heim-Vögtlin—Promotion of Women. The clinical uropathogenic E. coli *(UPEC) isolate BI47 was obtained from the University Hospital Basel, Switzerland.*

Keywords: antiadhesive · PapG · pyelonephritis · thermodynamics · X-ray diffraction

[1] *Bacterial Adherence. Receptors and Recognition, Vol. 6*, (Ed.: E. H. Beachey), Chapman and Hall, London, **1980**.
[2] K.-A. Karlsson, *Curr. Opin. Struct. Biol.* **1995**, *5*, 622–635.

ChemPubSoc Europe

CHEMBIOCHEM Full Papers

[3] *Microbial Lectins and Agglutinins: Properties and Biological Activity* (Ed.: D. Mirelman) Wiley, New York, **1986**.
[4] J. J. Martinez, M. A. Mulvey, J. D. Schilling, J. S. Pinkner, S. J. Hultgren, *EMBO J.* **2000**, *19*, 2803 – 2812.
[5] H. Hoschützky, F. Lottspeich, K. Jann, *Infect. Immun.* **1989**, *57*, 76 – 81.
[6] J. R. Johnson, A. L. Stell, N. Kaster, C. Fasching, T. T. O'Bryan, *Infect. Immun.* **2001**, *69*, 2318 – 2327.
[7] N. Strömberg, B.-I. Marklund, B. Lund, D. Ilver, A. Hamers, W. Gaastra, K.-A. Karlsson, S. Normark, *EMBO J.* **1990**, *9*, 2001 – 2010.
[8] N. Strömberg, P.-G. Nyholm, I. Pascher, S. Normark, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 9340 – 9344.
[9] S. D. Manning, L. Zhang, B. Foxman, A. Spindler, P. Tallman, C. F. Marrs, *Clin. Diagn. Lab. Immunol.* **2001**, *8*, 637 – 640.
[10] I.-M. Johanson, K. Plos, B.-I. Marklund, C. Svanborg, *Microb. Pathog.* **1993**, *15*, 121 – 129.
[11] J. A. Roberts, B.-I. Marklund, D. Ilver, D. Haslam, M. B. Kaack, G. Baskin, M. Louis, R. Möllby, J. Winberg, S. Normark, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 11889 – 11893.
[12] C. C. Tseng, J. J. Huang, M. C. Wang, A. B. Wu, W. C. Ko, W. C. Chen, J. J. Wu, *Kidney Int.* **2007**, *71*, 764 – 770.
[13] J. R. Johnson, *Infect. Immun.* **1998**, *66*, 4568 – 4571.
[14] J. B. Hill, J. S. Sheffield, D. D. McIntire, G. D. Wendel, Jr., *Obstet. Gynecol.* **2005**, *105*, 18 – 23.
[15] G. Otto, G. Magnusson, M. Svensson, J. Braconier, C. Svanborg, *Clin. Infect. Dis.* **2001**, *32*, 1523 – 1531.
[16] A. Larsson, J. Ohlsson, K. W. Dodson, S. J. Hultgren, U. Nilsson, J. Kihlberg, *Bioorg. Med. Chem.* **2003**, *11*, 2255 – 2261.
[17] A. E. Stapleton, M. R. Stroud, S. I. Hakomori, W. E. Stamm, *Infect. Immun.* **1998**, *66*, 3856 – 3861.
[18] K. Gupta, T. M. Hooton, K. G. Naber, B. Wullt, R. Colgan, L. G. Miller, G. J. Moran, L. E. Nicolle, R. Raz, A. J. Schaeffer, D. E. Soper, *Clin. Infect. Dis.* **2011**, *52*, 561 – 564.
[19] K. Gupta, D. Scholes, W. E. Stamm, *JAMA J. Am. Med. Assoc.* **1999**, *281*, 736 – 738.
[20] L. G. Greer, S. W. Roberts, J. S. Sheffield, V. L. Rogers, J. B. Hill, D. D. Mcintire, G. D. Wendel, *Infect. Dis. Obstet. Gynecol.* **2008**, 891426.
[21] N. Sharon, *Biochim. Biophys. Acta Gen. Subj.* **2006**, *1760*, 527 – 537.
[22] K. W. Dodson, J. S. Pinkner, T. Rose, G. Magnusson, S. J. Hultgren, G. Waksman, *Cell* **2001**, *105*, 733 – 743.
[23] J. Kihlberg, S. J. Hultgren, S. Normark, G. Magnusson, *J. Am. Chem. Soc.* **1989**, *111*, 6364 – 6368.
[24] J. Ohlsson, J. Jass, B. E. Uhlin, J. Kihlberg, U. J. Nilsson, *ChemBioChem* **2002**, *3*, 772 – 779.
[25] H. C. Hansen, G. Magnusson, *Carbohydr. Res.* **1999**, *322*, 166 – 180.
[26] J. Kihlberg, T. Frejd, K. Jansson, G. Magnusson, *Carbohydr. Res.* **1986**, *152*, 113 – 130.
[27] U. Nilsson, R. Johansson, G. Magnusson, *Chem. Eur. J.* **1996**, *2*, 295 – 302.
[28] H. C. Hansen, G. Magnusson, *Carbohydr. Res.* **1998**, *307*, 233 – 242.
[29] J. Ohlsson, A. Larsson, S. Haataja, J. Alajääski, P. Stenlund, J. S. Pinkner, S. J. Hultgren, J. Finne, J. Kihlberg, U. J. Nilsson, *Org. Biomol. Chem.* **2005**, *3*, 886 – 900.
[30] a) J. Tellinghuisen, *Anal. Biochem.* **2008**, *373*, 395 – 397; b) W. B. Turnbull, A. H. Daranas, *J. Am. Chem. Soc.* **2003**, *125*, 14859 – 14866.
[31] J. Thiem. B. Sauerbrei. *Angew. Chem. Int. Ed. Engl.* **1991**. *30*. 1503 – 1505; *Angew. Chem.* **1991**, *103*, 1521 – 1523.
[32] S. G. Krimmer, M. Betz, A. Heine, G. Klebe, *ChemMedChem* **2014**, *9*, 833 – 846.
[33] S. Cabani, P. Gianni, V. Mollica, L. Lepori, *J. Solution Chem.* **1981**, *10*, 563 – 595.
[34] a) A. Lammerts van Bueren, A. B. Boraston, *J. Mol. Biol.* **2004**, *340*, 869 – 879; b) A. B. Boraston, T. J. Revett, C. M. Boraston, D. Nurizzo, G. J. Davies, *Structure* **2003**, *11*, 665 – 675; c) A. B. Boraston, D. Nurizzo, V. Notenboom, V. Ducros, D. R. Rose, D. G. Kilburn, G. J. Davies, *J. Mol. Biol.* **2002**, *319*, 1143 – 1156.
[35] S. B. Svenson, H. Hultberg, G. Kailenius, T. K. Korhonen, R. Mdliby, J. Winberg, *Infection* **1983**, *11*, 61 – 67.
[36] F. P. C. Binder, K. Lemme, R. C. Preston, B. Ernst, *Angew. Chem. Int. Ed.* **2012**, *51*, 7327 – 7331; *Angew. Chem.* **2012**, *124*, 7440 – 7444, and references cited therein.
[37] E. Krissinel, K. Henrick, *J. Mol. Biol.* **2007**, *372*, 774 – 797.
[38] R. U. Lemieux, *Acc. Chem. Res.* **1996**, *29*, 373 – 380.
[39] S. Ebbinghaus, S. J. Kim, M. Heyden, X. Yu, U. Heugen, M. Gruebele, D. M. Leitner, M. Havenith, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20749 – 20752.
[40] Y. Chen, M. Tan, M. Xia, N. Hao, X. C. Zhang, P. Huang, X. Jiang, X. Li, Z. Rao, *PLOS Pathog.* **2011**, *7*, e1002152.
[41] P. B. Crowley, A. Golovin, *Proteins* **2005**, *59*, 231 – 239.
[42] C. Diehl, O. Engström, T. Delaine, M. Håkansson, S. Genheden, K. Modig, H. Leffler, U. Ryde, U. J. Nilsson, M. Akke, *J. Am. Chem. Soc.* **2010**, *132*, 14577 – 14589.

Manuscript received: November 15, 2016
Accepted Article published: January 11, 2017
Final Article published: February 7, 2017

