Expanding salivary biomarker detection by creating a synthetic neuraminic acid sensor via chimeragenesis

Samuel J. Verzino[1], Sharona A. Priyev[1#], Valeria A. Sánchez Estrada[1#], Ali GholamianMogaddan[1#], Gemma X. Crowley[1], Alexandra Rutkowski[1], Amelia C. Lam[1], Elizabeth S. Nazginov[1], Paola Kotemelo[1], Agustina Bacelo[1,2], Jack D. Flannery[1], Ksenya Gavrilov[1], Desiree T. Sukhram[1], Frank X. Vázquez[2], Javier F. Juárez[1*]

[1]Dept. of Biological Sciences and [2]Dept. of Chemistry, Saint John's University, 8000 Utopia Parkway, Queens, NY (USA)

* Corresponding author.

[#] These authors contributed equally.

ABSTRACT

Accurate and timely diagnosis of oral squamous cell carcinoma (OSCC) is crucial in preventing its progression to advanced stages with poorer prognosis. As such, the construction of sensors capable of detecting previously established disease biomarkers for the early and non-invasive diagnosis of this and many other conditions has enormous therapeutic potential. In this work, we apply synthetic biology techniques for the development of a whole-cell biosensor (WCB) that leverages the physiology of engineered bacteria *in vivo* to promote the expression of an observable effector upon detection of a soluble molecule. To this end, we have constructed a bacterial strain expressing a novel chimeric transcription factor (Sphnx) for the detection of N-acetyl-D-neuraminic acid (Neu5Ac), a salivary biomolecule correlated with the onset of OSCC. This WCB serves as the proof-of-concept of a platform that can eventually be applied to clinical screening panels for a multitude of oral and systemic medical conditions whose biomarkers are present in saliva.

35 **INTRODUCTION**

36 Biosensors are analytical devices in which biological elements convert a physicochemical
37 stimulus into a detectable signal[1,2]. Among the many existing variants, whole-cell biosensors
38 (WCBs) stand out due to their versatility and engineering potential[1]. These are detection systems
39 in which a physiologically active cell sustains an array of biomolecular elements (receptors,
40 transporters, enzymes, regulators, etc.) capable of signal recognition, integration, and production
41 of an observable output[1]. Among the organisms that can serve as WCBs, there are bacterial species
42 which either produce or limit the production of reporter proteins in the presence of a target
43 analyte[3,4]. Microbial WCBs have demonstrated potential in fields such as environmental
44 monitorization, food safety, and medicine[3,5]. Microorganisms can grow rapidly in relatively
45 inexpensive media and produce all the necessary biological elements for analyte detection,
46 providing a low-cost and portable alternative to other detection systems[6]. Unlike conventional
47 sensors that may require large and complex instruments as well as laborious sample processing,
48 WCBs can detect a bioavailable substance directly from the environment with minimal
49 preparation, enabling their incorporation into *in situ* detection hardware that may revolutionize
50 point-of-care diagnosis[6]. Multiplexing strategies, which enable detection of several signals
51 simultaneously, combined with improved safety controls (i.e. biocontainment strategies) increase
52 the opportunities for these biosensors to be deployed in both clinical and diagnostic settings[3,4].
53 Additionally, the robust and versatile nature of bacterial biosensors makes them suitable for use in
54 environments where other technologies would not be able to operate, such as strenuous pH and
55 osmotic conditions[3]. However, the use of WCBs is not exempt from certain drawbacks, such as
56 slow reaction times, potentially impractical dynamic and detection ranges, and sensitivity and

57 selectivity issues. However, each year an increasingly large corpus of publications proposes
58 solutions to these problems, making biosensors more competitive and reliable[7].

59 WCBs benefit from the synergies resulting from the combination of native systems,
60 evolved during millions of years to detect the miscellany of molecules surrounding microbes, with
61 the ever growing toolkit that synthetic biology provides (such as the creation of engineered genetic
62 circuits[8–11]). In WCBs based on transcription factors (TFs) the presence of a soluble analyte
63 becomes associated with a conformational change of a protein, so that the binding state of the
64 polypeptide to DNA gets altered, modifying the transcription profile of a regulated reporter[12].
65 These biosensors benefit the most from the use of monogenic transcriptional repressors[8,9], control
66 systems encoded by a single gene whose product prevents transcription in the absence of a small
67 soluble inducer molecule. The catalog of characterized TFs capable of interacting with small
68 soluble molecules is expanding, yet still limited, which represents one of the main obstacles in the
69 field[13,14]. To expand the panoply of molecules that can be detected, we can employ synthetic
70 biology, whose technologies have enabled us to better understand protein modularity such that we
71 can better engage in regulator customization and *de novo* creation[15,16]. In this paper, we centered
72 our efforts on the detection of a salivary biomarker that, to the best of our knowledge, was not the
73 inducer / inhibitor of any characterized TF and thus required the creation of a synthetic regulator.

74 Saliva is a biological fluid resulting from the secretions of several glands (among them the
75 parotid, submandibular and sublingual), crevicular fluid from the gingiva, cell residues, microbes
76 and associated structures such as biofilms, plus secretions from the respiratory tract[17,18]. Human
77 saliva contains a vast array of metabolites informative of oral and systemic health[19], making its
78 frequent or continuous sampling by portable and wearable equipment extremely coveted.
79 Translation of such sampling technologies to dental practices may increase their role at the

80 forefront of early disease prevention and diagnosis[20]. Electrochemical intraoral biodevices capable
81 of detecting salivary biomarkers presently exist, yet they are limited in the scope of molecules they
82 detect and lack adaptability to sense new ones[21]. Biological detection systems based on engineered
83 genetic circuits offer a plug-and-play potential, in which the substitution of one TF by another,
84 plus its cognate operator, is often enough to repurpose a pre-existing sensor. Bioengineering
85 solutions enabling point-of-care diagnostics will greatly benefit patients by improving at-home
86 health monitorization[22]. This approach is especially relevant due to the widespread acceptance and
87 use of consumer technology within society, which has normalized wearable biosensors connected
88 to a phone or a smartwatch as well as at-home sample collection (e.g. for ancestry information)[23].
89 Diagnostic screenings that do not require invasive procedures (i.e. a blood sample or a biopsy) are
90 preferred by patients, which makes saliva a prime biological fluid candidate to be sampled for
91 health and disease markers[24,25]. In this work we focused our attention on N-acetyl-D-neuraminic
92 acid (neuraminic acid, Neu5Ac), the most common of the sialic acids originally isolated from
93 salivary glands[26–28]. Neu5Ac fulfills key roles in human physiology[29] as well as in commensal and
94 pathogenic bacteria[26]. Moreover, variation in salivary sialylation might become a suitable early
95 indicator of changes associated with malignant pathologies, given their correlation to periodontal
96 disease (PD) and oral cancer[30]. The oral microbiome of patients with PD include bacteria such as
97 *Tannerella forsythia*, which engage in glycan harvesting behavior: they cleave Neu5Ac molecules
98 from glycoproteins to stealthily evade their host's immune system[31]. Additionally, abnormally
99 high concentrations of Neu5Ac have been associated with the onset of oral cancer[32–34]: a group of
100 neoplasms which affect any region in the oral cavity, pharynx, and salivary glands[35]. The term oral
101 cancer is often used interchangeably with oral squamous cell carcinoma (OSCC), which accounts
102 for more than 90% of oral malignant epithelial neoplasms[35]. As one of the most prevalent cancers

103 in the world[36], untreated OSCC has very poor prognosis and, as such, early detection is essential
104 for successful treatment[37]. Previous evaluations of sialic acid levels in the saliva of healthy
105 individuals compared to patients with oral cancer provided a guide for the range of salivary
106 Neu5Ac that can be observed. The concentration range identified in the saliva of healthy adults
107 was approximately 21.65 ± 5.71 mg/dl (0.52 - 0.88 mM), while confirmed oral pre-cancer
108 individuals displayed 59.75 ± 7.29 mg/dl (1.7 - 2.2 mM). Histopathologically confirmed OSCC
109 patients showed a wide range of concentrations, 204.85 ± 60.38 mg/dl (4.6 - 8.6 mM), always
110 superior to the other two groups[38]. Additional studies have employed multivariate regression
111 analysis comparing both free and protein-bound salivary sialic acid concentrations between OSCC
112 patients and healthy individuals. While free sialic acid levels were found to be significantly
113 elevated in patients diagnosed with oral cancer, differences in protein-bound sialic acid levels were
114 not found to be statistically significant[39]. These observations are further evidence that total salivary
115 sialic acid concentration can be utilized as a biomarker for OSCC. It is important to note that there
116 is at least one study in which Neu5Ac concentrations in saliva were decreased in OSCC patients,
117 having been hypothesized that a downregulation of the pathways modulating Neu5Ac production
118 might have occurred in these subjects[40]. Beyond its connection to OSCC, it has also been proposed
119 that salivary Neu5Ac could interfere with SARS-Cov-2 (COVID-19) infection, acting as a
120 competitive inhibitor blocking the viral antigens recognized by the host receptors[41]. Each of these
121 factors made quantification of salivary Neu5Ac a desirable target for detection by biosensors.
122 The absence of a native TF capable of detecting Neu5Ac motivated us to construct a
123 synthetic monogenic repressor for its integration in a genetic circuit that would enable the strain
124 bearing it to behave as a WCB. We favor these regulators due to their simplicity, compact nature,
125 and ease of integration into synthetic circuits[8,9,42]. The process of constructing our custom

126 Neu5Ac-sensing regulators is based on streamlined gene fusion / chimeragenesis methods
127 developed by our laboratory to detect soluble molecules[15] . Here, we describe a Neu5Ac TF-based
128 WCB that employs engineered bacterial machinery to fulfill the functions of the three basic
129 modules present in any detection system: a sensing module, a processing unit, and an actuator. The
130 living WCB can intake a Neu5Ac signal by expressing a novel synthetic transcription factor,
131 Sphnx, that bridges detection and processing (a common feature of bioreceptors[23]). A separate
132 reporter plasmid completes the system providing a gene encoding a fluorescent protein (sfGFP,
133 henceforth stylized as GFP) whose expression is regulated by the synthetic TF Sphnx so that
134 detectable changes in fluorescence occur upon the detection of soluble Neu5Ac. Sphnx is a
135 member of a new combinatorial library of TFs assembled via a simplified chimeragenesis process.
136 Important clues on inter-domain communication of modular proteins have started to be gathered
137 by comparing Sphnx to non-functional members of the TF collection. Protein tertiary structure
138 modeling of functional and non-functional regulators has contributed to establish comparisons
139 between them, helping to elucidate intricate and frequently difficult to predict dynamics that may
140 condition the success or failure of any chimeric TF.
141
142
143 **RESULTS AND DISCUSSION**
144 **Design and Construction of Synthetic Transcription Factors.** The functionality of the WCBs
145 that we built in this work hinged on the obtention of a TF capable of binding Neu5Ac with efficacy
146 and precision. The Neu5Ac-induced monogenic transcriptional repressor we intended to assemble
147 included three modules: a N-terminus (N_t) DNA Binding Domain (DBD), a linker sequence
148 (LNK), and a C-terminus (C_t) Ligand Binding Domain (LBD). The DBD specifically recognizes

149 an operator sequence in the promoter it regulates, driving the expression of the reporter gene
150 (*sfgfp*). The LBD is tasked with the recognition of Neu5Ac, providing specificity to the TF. Finally,
151 the LNK sequence acts as a bridge, providing the right amount of flexibility/rigidity necessary for
152 the transfer of information between DBD and LBD: steric changes in the LBD associated with
153 Neu5Ac binding get translated into steric changes in the DBD effecting a change of its oligomeric
154 state that results in the unbinding of the TF from its operator and thus in a de-repression of
155 transcription. By using a gene fusion / chimeragenesis approach we generated a library of
156 repressors with a common N_t-DBD-LNK-LBD-C_t architecture.

157 Our selected DBD was sourced from the paradigm of transcriptional regulators: the *lac*
158 repressor or LacI[43]. Though naturally induced by lactose/IPTG, we exclusively employ the DBD
159 of LacI recognizing the operator *lacO* present in the P_{lac} promoter (as defined in[15]). We decided to
160 use DBD-LacI because the *lac* repressor originated as a modular protein with two domains, and
161 has been previously used in the construction of several functional chimeras[44–46]. It is important to
162 remark that the exclusive use of DBD-LacI, as opposed to the full protein, ensures that chimeric
163 proteins carrying this domain will recognize their target operator, yet makes them incapable of
164 responding to native inducers.

165 Candidate LBD selection was accomplished through bibliographical research. Since the
166 LBD was the domain that recognized the molecule of interest for which we built the biosensing
167 strain, it could be sourced from any whole protein or protein domain as opposed to being co-opted
168 from a previously documented regulator. One of our main sources of the LBDs we employ in our
169 chimeragenesis pipeline are periplasmic binding proteins (PBPs)[47]. These are standalone subunits
170 of transporters that act as molecular shuttles that assist in the traffic of small solutes from the outer
171 to the inner membrane in Gram negative bacteria[48]. PBPs share a common ancestor with the LBD

172 domains of the LacI / GalR family of regulators[45]. To build the synthetic TF we required for this
173 project, we needed to identify LBD candidates that strongly bound our soluble molecule of interest
174 (Neu5Ac). PBPs SatA and SiaP were selected for their binding affinity and specificity for the
175 transport of Neu5Ac. SatA is a product of the *satABCD* (*s*ialic *a*cid *t*ransport) operon that encodes
176 an ATP Binding Cassette (ABC) transporter[49] from *Haemophilus ducreyi*[50]. An alternative
177 *satABCD* system present in *Corynebacterium glutamicum* has been demonstrated to be essential
178 for its growth using Neu5Ac as sole carbon source[51]. SiaP is a product of the *siaPQM* operon,
179 which encodes a Tripartite ATP-independent periplasmic (TRAP) transporter in *Haemophilus*
180 *influenzae*[52,53]. SiaP was the first TRAP-associated PBP whose structure was elucidated (in *Vibrio*
181 *cholerae*, *Pasteurella multocida* and *Fusobacterium nucleatum*[52,54]). Structural comparisons and
182 thermodynamic studies of SatA and SiaP suggest that similar affinities for Neu5Ac are achieved
183 in the two PBPs through distinct mechanisms: one enthalpically driven and the other entropically
184 driven. SiaP, as well as other TRAP transporters, are hypothesized to command an enthalpically
185 driven ligand binding strategy due to the abundance of charged residues (Arg, Lys, Asp, and Glu)
186 at their binding pocket. SatA, alternatively, is hypothesized to bind its substrate in what is
187 considered an entropic approach because of the numerous hydrophobic and polar residues (Tyr,
188 Phe, Gly, Leu, Asn, and Ser) at its binding pocket[55]. The structure of *H. influenzae* SiaP in the
189 presence of Neu5Ac (PDB 3B50) reveals the ligand bound in a deep cavity with its carboxyl group
190 forming a salt bridge with a highly conserved Arg residue[52]. Initial studies of SiaP isolated from
191 *H. influenzae* displayed a binding K_d of ≈120 nM when affinity for Neu5Ac was tested[53].
192 Subsequent studies established a value closer to K_d ≈58 nM[52]. Similar binding affinity of SatA
193 complexes has been identified with a K_d ≈133 nM[55]. The collection of empirical data supporting

194 the ability of SatA and SiaP to bind Neu5Ac made them ideal candidates to become the LBD of a
195 chimeric TF responsive to Neu5Ac.

196 The final piece of the chimeric gene puzzle consisted on establishing a connection between
197 DBD and LBD domains. This task was performed by elements called linker (LNK) regions that
198 tend to be less structured than the domains they connect[56]. In order to maximize the chances of
199 transmitting the information of Neu5Ac binding to the DBD, so that it can detach from *lacO* and
200 enable transcription, we decided to connect DBD and LBD via two different strategies: i) a direct
201 linkage without a polypeptide linker (condition identify as LNK1 for uniformity purposes), or ii),
202 by one out of four linkers characterized by having different biophysical properties (identified as
203 LNK2 to LNK5; *Supp. Table 1*). It is important to note that while LNK1 does not contain a linker
204 *per se*, the LBDs we chose, by virtue of being PBPs, are proteins that would be exported to the
205 periplasm of the cell under natural circumstances, and thus contain a native signal peptide at their
206 N-terminus. This kind of peptide appears to be unstructured for the most part, barring a short alpha-
207 helix region[57]. We decided to include the signal peptides of both SiaP and SatA due to existent
208 examples of functional chimeras containing similar structures, such as the glucose-inducible
209 $SLCP_{GL}$[46] and the benzoate-inducible CbnR-ABE44898-OD[15] synthetic TFs. At this stage, we had
210 identified all the elements we needed to start assembling our new chimeric TFs.

211 dsDNA fragments encoding the DBD and all LNK-LBD variants were obtained as detailed
212 in the *Materials and Methods* section. To initiate the construction of a library containing all the
213 desired chimeric TFs, we performed several scarless assembly reactions with all the domain-
214 encoding fragments and linearized expression vector pUC19RBS, followed by transformation in
215 NEB5-alpha competent cells. Resulting clones were isolated and individually sequenced. This
216 approach was performed twice. Chimeric combinations that were not present in the libraries were

217 re-attempted to be individually cloned five times with no success, at which point we proceeded to
218 put pause to their construction and move the project forward, subcloning the TF collection we had
219 generated up to that point in time into the low-copy vector pCKTRBS, what generated the set of
220 plasmids we collectively identified as pCKT-*Chimera*. We hypothesized that chimeras
221 unsuccessfully cloned at this stage may be potentially deleterious for the cells. It is important to
222 note that the expression of any heterologous protein is *per se* taxing on bacterial metabolism, and
223 that TFs might be especially burdensome[58,59]. Chimeric constructions including DBD-LacI and
224 SiaP/SatA seem to provoke a certain toxicity to the cell, albeit to different degrees depending on
225 the specific domain combination. Indirect evidence supporting this hypothesis was observed when
226 constructions expressing the chimeric TFs were transformed into *E. coli* MC4100 to assess their
227 behavior in M63 minimal media with glycerol as the sole carbon source. Under these resource-
228 limited conditions, stressed bacteria have more difficulties in thriving, and MC4100 cells bearing
229 pCKT-*Chimera* vectors displayed a very modest and erratic growth consistent with our
230 assumptions (data not shown). *Table 1* contains the chimeric TFs built, assayed, and presented in
231 this study, as well as a simplified interpretation of their functionality when grown in rich media.
232 **Screening of transcription factors**. A functional *in vivo* screening was implemented using GFP
233 as reporter. Plasmid pHC_DYOLacI-R (*Supp. Table 2*) was transformed into candidate strains
234 expressing the chimeric TFs from pCKTRBS vectors under the control of TetR-aTc. These two
235 plasmids expressed in the same strain constituted a genetic circuit (*Figure 1a*) that enabled a
236 controlled expression of the chimera as well as its controlled induction. This circuit behaves as a
237 B-or-not-A universal two input gate[60] (*Figure 1b*), where there are two inputs (A, aTc
238 supplementation; B, Neu5Ac supplementation) and one output (detection of GFP fluorescence).
239 Based on the addition of the inputs we can define three states in the genetic circuit (*Figure 1c*):

240 *Basal* ($aTc^-/Neu5Ac^-$), no input added, GFP is expressed from its promoter in pHC_DYOLacI-R;
241 *Repression* ($aTc^+/Neu5Ac^-$), the presence of aTc induces TetR which is incapable of remaining
242 attached to P_{tetO} and thus the chimera is expressed from its corresponding pCKT-*Chimera* vector
243 repressing the expression of GFP; *De-repression* ($aTc^+/Neu5Ac^+$), the chimera is expressed but if
244 it is functional it is induced by Neu5Ac, de-repressing the promoter driving the expression of GFP.
245 By comparing the behavior of the strain expressing the aforementioned circuit under these
246 conditions, we can rule out non-functional synthetic TFs. GFP-associated fluorescence data
247 displayed in this publication is always expressed as relative fluorescence (Rel.Fluor =
248 Fluorescence arbitrary units / OD_{600}) which includes a correction of fluorescence by the optical
249 density of the culture in order to account for growth disparities among different strains and
250 conditions. In the case that the chimeric protein was not a functional repressor, relative GFP-
251 associated fluorescence (Rel.fluor) could be detected in the *Repression* state ($\Delta Rel.fluor_{basal\text{-}repressed}$
252 tending to 0). On the other hand, if the chimera was able to bind to *lacO* on the reporter promoter
253 yet incapable to recognize Neu5Ac with its LBD, or to transmit a LBD to DBD conformational
254 change resulting in its detachment from the operator, we would have obtained a functional super-
255 repressor[61] unable to respond to its inducer and, thus, not a valid tool to detect Neu5Ac
256 ($\Delta Rel.fluor_{(de\text{-}repressed)\text{-}repressed}$ tending to 0). *Figure 1d* illustrates the expected GFP production along
257 a growth curve for a strain expressing the screening genetic circuit that included an ideal chimera
258 ($\Delta Rel.fluor_{(de\text{-}repressed)\text{-}repressed}$ >0). It is important to note that even though Neu5Ac is a
259 monosaccharide, its supplementation to rich growth media did not significantly increase the OD_{600}
260 of the assayed strains, at least up to the tested 10 mM concentration (*Supp. Figure 1*). It is also
261 important to note that repression by LacI, or LacI-derived chimeras, is rarely absolute so that a
262 certain level of leakage, and thus of background reporter gene expression, is always expected[46,62].

263 All the chimeric TFs included in *Table 1* were screened for their ability to bind DNA and
264 thus to repress the expression of the GFP reporter, as well as to be induced by Neu5Ac enabling
265 production of the fluorescent protein. All members of the collection of MG1655 (pCKT-*Chimera*,
266 pHC_DYOLacI-R) strains bearing the genetic circuit described in *Figure 1* were tested *in vivo* to
267 assess their ability to behave as Neu5Ac biosensors (*Supp. Table 4*). Among the chimeric proteins
268 expressed by this strain collection, we included the TF Lumos, which did not contain any of the
269 expected LNK1-5 options but an unexpected 6 amino acid linker (LNK6) resulting from an
270 abnormal assembly reaction in which the last 18 bases of *DBD-lacI* were duplicated (*Supp. Table*
271 *1*). Given the stable nature of the amino acid string, as well as our agnostic approach to which
272 LNK sequence could be the best performing one, we decided to maintain this construction in our
273 pool. TFs Kunst (LacI-LNK3-SiaP), Scorpio (LacI-LNK4-SiaP), Lumos (LacI-LNK6-SiaP), and
274 Haus (LacI-LNK2-SatA) showed no repression abilities. These TFs seem to not be able to stably
275 bind to their operator boxes in the reporter plasmid, potentially due to an inability to acquire their
276 dimeric form. The strain bearing Haus also presented a diminished growth, which strongly
277 suggests the potential toxicity of this transcription factor for the host cell.
278 Bastet (LacI-LNK5-SiaP) was intriguing due to the possibility that it was able to repress
279 and de-repress for the first 10 hours of growth, yet its behavior requires a deeper study given the
280 observed dispersion in the analyzed data. This observation suggests that the strain might be
281 unstable due to the potential toxicity of the protein. Also of note, Harpy (LacI-LNK1-SatA)
282 seemed to be able to repress at a certain level and potentially de-repress after the 18 hour mark.
283 Another TF that presented this similar and unexpected induction profile but in more consistent
284 basis was Siren (LacI-LNK1-SiaP), which displayed an intriguing function as a proper repressor
285 inducible by Neu5Ac but only in the stationary phase of growth (*Supp. Figure 2*). In the future we

286 intend to study this protein in depth, since its repression/de-repression profile is very attractive for
287 a whole-cell biosensor setup in which a bacterial culture sustaining itself at stationary phase is
288 continuously monitoring an input for an extended period of time. Prolonging the time WCBs are
289 functional is a known challenge associated with these technologies, and engineering of genetic
290 circuits incorporating regulators such as Siren, is one of the strategies that are being explored to
291 account for this problem[3]. A candidate biosensor strain expressing Sphnx (LacI-LNK2-SiaP)
292 exhibited the most suitable repression and induction profile compatible with that of a TF capable
293 of repressing P_{lac} while de-repressing it in a detectable manner upon binding to Neu5Ac. To
294 summarize this initial evaluation of the functionality of our candidate TF-based WCBs, and
295 judging by the difficulty to obtain them, their lack of functionality, and the reduced sample size,
296 SatA-based fusions seem to be less stable and more prone to not be responsive to Neu5Ac. Further
297 studies and the creation of expanded chimeric libraries, however, are required to confirm this
298 hypothesis. At this injunction, we focused our efforts on better characterization of Sphnx, since it
299 was the most promising TF we had assembled (estimated as the relative fluorescence difference
300 under de-repressed and repressed conditions along the growth curve or (ΔRel.fluor$_{\text{(de-repressed)-}}$
301 $_{\text{repressed}}$) (*Supp.Table 4*).

302 In order to validate our preliminary observations, we systematically compared GFP
303 production over time of MG1655 (pCKT-Sphnx, pHC_DYOLacI-R), MG-Sphnx for short, cells
304 expressing Sphnx either in the presence or absence of Neu5Ac. *Figure 2a* displays the results of
305 time course experiments over a period of 25 hours. GFP production induced by the presence of
306 Neu5Ac was observed from the first hour of growth, with the Neu5Ac-induced cultures showing
307 a tendency to display higher fluorescence than the repressed ones. At this point our attention was
308 captured by Kunst as a model for a non-functional chimera in opposition to Sphnx. *Supp. Figure*

309 *3* shows the behavior of Kunst, which exhibited a tendency to repress P_{lac} at the very beginning of
310 a 40 h time course, yet after the 3 h mark did not show any difference between induced and
311 uninduced conditions. A preliminary evaluation of the structural differences between Sphnx and
312 Kunst is included in its own section of *Results and Discussion*.

313 In an effort to validate the *in vivo* observations confirming Sphnx functionality, we
314 analyzed the induction of P_{lac} driving the expression of GFP in MG-Sphnx by qRT-PCR
315 experiments (*Supp. Figure 4)* performed as indicated in *Materials and Methods*. The expression
316 levels of *sfgfp* calculated as the ratio of the induced condition over the repressed condition were
317 1.58 ± 0.24 (n = 8, error expressed as SEM), statistically significant and consistent with an
318 induction of the expression in the presence of Neu5Ac.

319 Further *in vitro* experiments were also performed to confirm the interaction between P_{lac}
320 and the DBD-LacI of Sphnx. Electrophoretic mobility shift assays (EMSA; *Materials and*
321 *Methods*) showed that cell extracts from NEB5-alpha (pCKT-Sphnx) grown for 5 h were capable
322 of retarding the migration of a P_{lac}-*sfgfp* probe compared to control extracts that were not
323 expressing Sphnx (data not shown). Neu5Ac behaved as the inducer of Sphnx, since binding of
324 the protein to the probe was diminished in a concentration-dependent manner in the presence of
325 the sialic acid (*Figure 2b*). On the other hand, cell extracts from NEB5-alpha (pCKT-Siren) grown
326 for 5 h were able to bind the probe but not to unbind in the presence of Neu5Ac, confirming our
327 *in vivo* observations. Previous publications report a similar behavior: DBDs from bacterial
328 transcriptional repressors tend to retain their ability to bind to their operator sequences even when
329 integrated in a new chimeric TF[63–65], another testimony to the high modularity of these proteins.
330 Our initial *in vitro* DNA-protein interaction studies require future validation with purified chimeras
331 in order to calculate the kinetics of the Sphnx-*lacO* and Sphnx-Neu5Ac-*lacO* interactions, yet they

332 represent a promising step forward towards characterizing the interaction between our synthetic
333 TF and its cognate operator.

334 At this juncture of the project, both *in vivo* and *in vitro* experiments strongly suggested that
335 Sphnx was a transcriptional repressor capable of interacting with *lacO* and being de-repressed by
336 neuraminic acid. To better understand the dynamics of the Sphnx-Neu5Ac interaction we analyzed
337 the production of GFP by MG-Sphnx cultures (*Materials and Methods*) when exposed to
338 different concentrations of neuraminic acid around the average
339 salivary concentration for an adult male (≈1 mM)[66]. The mean fold difference
340 of GFP expression between the repressed and induced conditions was 2.25 ± 0.74 for cultures
341 induced with 0.5 mM Neu5Ac, 2.54 ± 0.52 for cultures induced with 1.0 mM Neu5Ac, 3.81 ± 1.15
342 for cultures induced with 2.0 mM Neu5Ac, and 5.21 ± 2.00 for cultures induced with 5.0 mM
343 Neu5Ac (n = 8, error expressed as SEM). These results were commensurable with those observed
344 for modular chimeric TFs in the presence by their ligands: $SLCP_{GL}$ (LacI-GGBP-OD) induced by
345 glucose[46], CbnR-ABE44898-OD and LmrR-BzdB1_nSP induced by benzoate[15]. *Figure 2c* shows
346 the levels of GFP fluorescence observed in basal and repressed cultures as well as in cultures
347 induced by 0.5, 1.0, 2.0 and 5.0 mM neuraminic acid. These assays corroborated that relative
348 fluorescence between basal (aTc^-/$Neu5Ac^-$) and repressed (aTc^+/$Neu5Ac^-$) conditions were
349 significant. Cultures in which the genetic circuit used to test Sphnx were de-repressed
350 (aTc^+/$Neu5Ac^+$) displayed statistically significant values compared to the repressed cultures when
351 in the presence of 1.0 mM and 2.0 mM Neu5Ac. Induction by 0.5 and 5.0 mM was not significant.
352 It is important to note that relative fluorescence values in the 5.0 mM-induced biological replicas
353 were widely dispersed, suggesting that an overabundance of Neu5Ac might interfere with the
354 stability of the host strain bearing both plasmids, producing some deleterious effect associated to

355 the expression of the heterologous Sphnx and GFP proteins in the assayed conditions, which in
356 practical terms means may entail the testing of a few sample dilutions when using these biosensors
357 for out-of-the-lab applications.

358 In order to determine the specificity of Sphnx for Neu5Ac, we analyzed the production of
359 GFP by MG-Sphnx cultures exposed to 1 mM Neu5Ac compared to cultures supplemented with
360 1mM of either of the following analogue compounds: N-acetyl-D-mannosamine (ManNAc), N-
361 acetyl-D-glucosamine (GlcNAc), or N-glycolyl-neuraminic acid (Neu5Gc) (*Materials and*
362 *Methods*). ManNAc and GlcNAc are aminated monosaccharides that are part of the biosynthetic
363 pathway of Neu5Ac, constituting the backbone of the molecule[67–69]. Neu5Gc is a sialic acid that
364 only differs from Neu5Ac in a single hydroxyl group and, interestingly, cannot be synthesized by
365 humans, originating from diet and microbiota production[70–72] Neu5Gc has been postulated as a
366 cancer biomarker due to its proinflammatory effects[73,74]. Molecular docking simulations (*Figure*
367 *3a*) strongly suggested that both Neu5Ac and Neu5Gc could theoretically establish more
368 stabilizing bonds with the amino acids present in the binding pocket of Sphnx-LBD, whereas
369 ManNAc and GlcNAc would be able to enter said pocket yet their interaction with it would be
370 extremely weak. These observations were corroborated *in vivo*. *Figure 3b* displays GFP-
371 fluorescence levels by MG-Sphnx cultures exposed to the different Neu5Ac analogues. Induction
372 by the different compounds is expressed as fold change Rel.fluor (n = 9-12, error expressed as
373 SEM) of each induced culture (aTc^+/$ligand^+$) over a reference repressed condition (aTc^+/$ligand^-$).
374 No statistically significant difference was found between repressed and $ManAc^+$ or $GlcNAc^+$
375 conditions, strongly suggesting that these compounds are not structurally similar enough for Sphnx
376 to bind them. However, $Neu5Gc^+$ cultures displayed a significant induction compared to the
377 reference, presenting on average 62 ± 15% of the $Neu5Ac^+$ induction. Upon further analysis there

378 was no statistically significant difference between Neu5Ac$^+$ and Neu5Gc$^+$ datasets. These findings
379 confirm that Neu5Gc is an inducer of Sphnx, which upon further review of the existing literature
380 was coherent with previous observations that described how SiaP (Sphnx's LBD) was able to bind
381 Neu5Gc *in vitro* with μM affinity[55]

382 Our next objective was to test whether cultures of the WCB could detect Neu5Ac in
383 contrived samples simulating human saliva (*Materials and Methods*). MG-Sphnx cultures mixed
384 with artificial saliva supplemented with Neu5Ac for a final 1 mM concentration returned a
385 Rel.Fluor signal 3.2 ± 0.9-fold (mean ± SEM, n=5) higher than samples of unsupplemented saliva.
386 These data, even though preliminary, are very promising of the capabilities of WCBs, for
387 diagnostic applications using clinical samples.

388 These collective results showing Sphnx induction are promising, yet they underline the
389 need to refine chimeric TFs resulting from the assembly of pre-existing modules to optimize their
390 induction range as well as gene dose and reporter genes used to study their behaviour *in vivo* and
391 *in vitro*, as it has been previously proposed[14]. Newly created synthetic TFs tend to have limited
392 transfer function and specificity, and thus to require further engineering for their field deployment
393 in biosensors[75]. In this regard, we are optimistic about the future improvement of the chimeras we
394 have assembled: the estimated K_d of SiaP and SatA for neuraminic acid are in the nM range[52,53,55]
395 (calculated *in vitro*), whereas in this work we present data in the mM range (calculated *in vivo*).
396 This suggests that we have not reached the full potential of our LBDs and there is an opportunity
397 to optimize Sphnx's structure to better benefit from the high specificity and affinity of our LBD.
398 It is also important to note that Sphnx is currently at the same point in evolutionary history when
399 DBD-LacI was fused to the ancestral version of its current LBD, long before communication
400 between DBD and LBD had been refined and perfected by natural selection[45]. A systematic

401 exploration of promoter and RBS variants can provide a better solution to compensate for some of
402 these deficiencies[14]. Regarding the use of alternative reporters that may shed more light on the
403 induction profile of Sphnx, we acknowledge that the use of GFP has some inherent limitations in
404 prokaryotes (e.g. widely varying maturation times dependent on strain and metabolic state[76]),
405 which has motivated us to initiate the development of better reporting systems.

406 **Modeling of the 3D structure of chimeric transcription factors**. Another element we are
407 starting to explore is the importance of linkers for the functionality of modular proteins, an often-
408 neglected design element[77]. Since the beginning of our journey characterizing the chimeric TFs,
409 we found extremely intriguing the fact that Sphnx (LacI-LNK2-SiaP, N_t-DBD-EKEKEK-LBD-
410 C_t) and Kunst (LacI-LNK3-SiaP, N_t-DBD-GSGSGS-LBD-C_t), which are nearly identical proteins
411 (*Supp. Materials Appendix*), possessed opposed behaviours when interacting with neuraminic
412 acid. The recent release of the AlphaFold 3 software[78] has enabled us to predict the structures of
413 multimeric chimeric proteins in the presence of their dsDNA operator, allowing us to model the
414 proteins under conditions closer to those found in actual WCBs. Sphnx and Kunst were both
415 modeled as dimers using AlphaFold 3, along with a DNA double helix containing the *lacO* (5'-
416 ATTGTGAGCGGATAACAATT-3' and its complementary sequence) recognized by the DBD-
417 LacI they both carry. Our goal was to predict structures that could help explain why Sphnx binds
418 Neu5Ac while Kunst does not, despite only differing by six amino acids. *Figure 4a* shows the
419 predicted Kunst DBD (green) bound to the DNA double helix (gray) and the formation of a dimer
420 between the PBP (cyan). The structure shows a mostly disordered linking region (orange),
421 including the GSGSGS (LNK3) linker (violet). The predicted Sphnx model (*Figure 4b*), on the
422 other hand, shows that the connecting region adopts a helical structure, including the EKEKEK
423 (LNK2) linker, and appears to form contacts between both the DBD and PBP. Rotating the Kunst

structure by 90° (*Figure 4c*) we can observe that the GSGSGS (linker does not interact with the protein or DNA. In contrast, rotating the Sphnx structure by 90° (*Figure 4d*) further shows that the linker forms contacts between the DBD and PBP domains, with interactions occurring across different monomer chains. Further analysis shows that the helical linker in Sphnx forms electrostatic salt bridges between itself and both the DBD and PBP domains. *Figure 4e* shows the acidic residues on the linker (Glu) forming contacts with basic residues (Lys) on the DBD and PBP domains. Specifically, the predicted structure suggests that the residues expected to form salt bridges include Glu67-Lys2 on the DBD and Glu69-Lys152 on the PBP. These salt bridges likely stabilize the helical linker, affecting the flexibility and stability of the polypeptide, an effect that has been previously observed in other proteins[79]. From these predicted structures, we hypothesize that these salt bridges stabilize helicity in the flexible linker in a way that positively affects the PBP function by interacting with the DBD and the PBP, explaining the ligand binding functionality of Sphnx compared to non-binding Kunst protein. Zooming out to see the larger structure (*Figure 4f*), aromatic residues near the linker salt bridges are visible. These residues are key for testing the new predicted structures and designing new chimeric proteins in the future. We intend to purify Sphnx and Kunst to perform fluorescence analysis of the protein-ligand interactions, including fluorescence quenching and fluorescence anisotropy studies. The construction of 3D models, using the latest version of AlphaFold, will enable us to explore which mutations might enhance ligand binding in a way that was previously unavailable, also allowing us to model conditions closer to the actual multimeric *in vivo* structures.

CONCLUSIONS

446 This work presents MG1655 (pCKT-Sphnx, pHC_DYOLacI-R), MG-Sphnx, the first
447 iteration of a transcription factor-based whole-cell biosensor strain capable of detecting the
448 salivary biomarker N-acetyl-D-neuraminic acid and relaying its presence through the production
449 of a fluorescent reporter, the traditional litmus test for a biosensor circuit[14]. The keystone of the
450 genetic circuit enabling the detection of the soluble molecule is a new transcriptional repressor:
451 Sphnx. The apparent absence of a native Neu5Ac-binding transcription factor led the authors to
452 create a custom-made regulator via a straightforward synthetic biology approach that enables the
453 generation of chimeras capable of repressing a known promoter and being de-repressed by a
454 molecule of interest. Neuraminic acid is an attractive analyte to monitor since its enhanced levels
455 in saliva are correlated with conditions that range from periodontal disease[66] (oral microbes harvest
456 sialic acid to evade our immune system[31]) to the onset of oral squamous cell carcinoma[32–34]. To
457 the best of our knowledge, Sphnx is the first published transcriptional repressor capable of being
458 induced by N-acetyl-D-neuraminic acid. Its ability to detect N-glycolyl-D-neuraminic as well
459 makes MG-Sphnx a strong biosensor for the detection of sialic acid imbalances associated with
460 multiple dysbiotic, inflammatory, and neoplasic conditions[74,80–82].

461 This study supports the idea that the ability to construct custom-made TFs, affordably and
462 on-demand, is instrumental in expanding the panoply of small soluble molecules that can be
463 detected by biosensors in a field limited by the availability of sensing modules[14,45]. It is important
464 to note that synthetic TFs such as Sphnx do not represent the closing of a chapter, but the origin
465 of a new source of regulators whose performance can be improved via directed engineering of
466 their ligand binding pockets[83,84], an approach that can now benefit from high-throughput
467 approaches[85]. It is widely acknowledged that biosensor specificity and biosensor-analyte reaction
468 curves need to be optimized for the practical application of the WCB[75]. We have obtained

469 promising results on the functionality of MG-Sphnx to detect N-acetyl-D-neuraminic in
470 preparations of artificial saliva[86], however, further experiments are needed to tune the parameters
471 of the assay to detect a wide range of biomarker concentrations and, eventually, move on to test it
472 with clinical salivary samples. The MG-Sphnx strain represents the initial step in the obtention of
473 a functional WCB for the detection of neuraminic acid, a necessary launchpad to achieve the goal
474 of eventual clinical deployment of the sensor.

475 The modular assembly of fusion genes such as Sphnx sheds light on the importance of
476 inter-domain communication for multi-domain protein functionality[87] and exemplifies how a
477 reduced number of independent modules can be the playground of natural selection to synergically
478 originate new functionalities in the cell[88]. The results presented in this work have challenged our
479 expectations regarding chimeric transcription factors: while we expected most of them to be able
480 to repress expression from their regulated promoter, we found several instances in which the
481 chimeric product was unable to bind to its cognate operator boxes. Our current hypothesis is that
482 different linkage alternatives condition the 3D structure of the multimodular protein to such an
483 extent that the dimerization step necessary to bind the operator boxes cannot be achieved. We are
484 looking forward to expanding these observations in the future. Our study presents the interesting
485 contrast between the inducibility of Sphnx as opposed to Kunst, and particularly the unexpected
486 role that a potential LBD-mediated stabilization of the linker may play in the de-repressibility of
487 the TF dimer. *In vitro* assays using purified proteins, as well as further modeling simulation
488 analysis will contribute to illuminate the complex dynamics necessary for modular proteins to
489 fulfill their function. The eventual crystallization of these chimeras will provide the final clues on
490 the matter.

491 In summary, this work presents a new biosensor for N-acetyl-D-neuraminic acid, a salivary
492 biomarker relevant to oral health, which is based on a synthetic TF whose construction adds to our
493 understanding on the role that different domains play in the functionality of modular proteins.

494 **MATERIALS AND METHODS**

495 **Bacterial strains and growth conditions**. Bacterial strains and plasmids used are listed in *Supp.*
496 *Table 2*. *E. coli* strains were grown at 37°C in LB (Lysogenic Broth[89]) medium unless otherwise
497 indicated. When required, *E. coli* cells were grown in M63 minimal medium[90] using the necessary
498 nutritional supplements and 30 mM glycerol (Sigma, St. Louis, MO) as a carbon source.
499 Antibiotics were added at the following concentrations: 100 $\mu g \cdot ml^{-1}$ ampicillin/carbenicillin and
500 25 $\mu g \cdot ml^{-1}$ chloramphenicol (Sigma, St. Louis, MO). For protein expression from pCKTRBS-
501 derived vectors, cultures were induced with 0.5 $\mu g \cdot ml^{-1}$ anhydrotetracycline (aTc) (Clontech,
502 Mountain View, CA). When experimental conditions required it, cultures were supplemented with
503 N-acetyl-D-neuraminic acid (neuraminic acid, Neu5Ac; Sigma, St. Louis, MO), N-acetyl-D-
504 mannosamine (ManNAc; TCI Chemicals, Portland, OR), N-acetyl-D-glucosamine (GlcNAc; TCI
505 Chemicals, Portland, OR), or N-glycolyl-neuraminic acid (Neu5Gc; AFB Bioscience-VWR,
506 Radnor, PA) for a 1 mM concentration unless otherwise indicated.

507 Culture growth was routinely monitored in a BioPhotometer D30 (BioSpectrometer Basic Kinetic
508 and Fluorescence; Eppendorf, Hamburg, Germany). When it was required, bacterial growth and
509 GFP production were recorded in a Varioskan LUX Multimode Microplate Reader (Thermo
510 Fisher, Waltham, MA). Cultures were pipetted on a 96-well plate (Flat bottom Polystyrene Black
511 with clear bottom; Corning, Corning, NY) covered with a Breathe-Easy BEM-1 gas-permeable
512 membrane (Diversified Biotech, Dedham, MA) and incubated in a plate reader, where OD_{600} and
513 fluorescence in the range of GFP (excitation 485 nm, emission 528 nm) were either tracked along
514 time courses (up to 40 h) or subjected to point measurements. NEB5-alpha (pHC_DYOLacI-R)
515 and MG1655 were included in every plate as GFP-positive and GFP-negative reference strains,
516 respectively. For standard time course experiments, overnight precultures of the strains were used

517 to inoculate (starting OD_{600} 0.025) fresh growth media (supplemented with aTc and Neu5Ac when
518 indicated) in 96-well plates (180 µl total volume per well). For the single point measurements
519 described in *Figure 2c* in which the MG-Sphnx is exposed to different concentrations of Neu5Ac,
520 5 ml overnight precultures of the strains (plus aTc and Neu5Ac when required) were used to
521 inoculate 5 ml of fresh media supplemented in the same manner as the precultures. After 14 h of
522 growth, 180 µl samples of the cultures were transferred to 96-well plates to enable OD_{600} and
523 fluorescence readings. For the single point measurements described in *Figure 3b* in which the MG-
524 Sphnx is exposed to 1mM Neu5Ac or 1mM of analogue compounds (ManNAc, GlcNAc,
525 Neu5Gc), smaller 2 ml overnight precultures of the strain (plus aTc and Neu5Ac / ManNAc /
526 GlcNAc / Neu5Gc when required) were used to innoculate 2 ml of fresh media supplemented as
527 the precultures. After 20 h of growth, 180 µl samples of the cultures were transferred to 96-well
528 plates to enable OD_{600} and fluorescence readings. For the experiments using artificial saliva a 5 ml
529 overnight preculture of MG-Sphnx supplemented with aTc was used to inoculate (with 20 µl) a
530 mixture of 750 µl of fresh media (plus aTc) and 250 µl of artificial saliva (artificial saliva with
531 mucin, stabilized, Ref. 1700-0316; Pickering Laboratories, Mountain View, CA). The mix was
532 incubated for 22 h at 37°C in an orbital shaker to ensure proper mixing and favor strain growth.
533 This prolonged incubation seems to be key for the homogenization of the sample, since initial
534 attempts to grow the mixture of culture and artificial saliva under the more gentle shaking
535 conditions provided by the plate reader failed, presumably due to the viscosity of the saliva. Once
536 the incubation was finished 180 µl samples of the cultures were transferred to 96-well plates to
537 enable OD_{600} and fluorescence readings as described above for endpoint measurements. Incubation
538 of the samples in the plate reader, after the endpoint measurement, with up to 48 h continuous
539 OD_{600} and fluorescence monitorization did not improve the quality of the data.

540 Statistical analysis as well as data visualization were performed with R (https://www.r-
541 project.org/) in the RStudio (RStudio Inc., Boston, MA) and Visual Studio Code environments
542 (Microsoft Corp., Redmond, WA).

543 **Molecular biology techniques.** Molecular biology techniques were performed following
544 commonly used standard protocols and as per manufacturers' instructions[91]. PCR reactions took
545 place in a 6321 Mastercycler PRO Vapo Protect Thermal Cycler (Eppendorf, Hamburg, Germany).
546 Routine separation of DNA and RNA in agarose gels was performed in an RunOne Electrophoresis
547 system (EmbiTec, San Diego, CA). Agarose gels were imaged in an UVP UVSolo Touch (Analytic
548 Jena, Jena, Germany). Plasmid DNA was purified with a Qiaprep Spin Miniprep Kit (Qiagen,
549 Hilden, Germany). DNA fragments were purified with DNA Clean-up and Concentration Kit
550 (Zymo research, Irvine, CA). The oligonucleotides employed for PCR amplification of the cloned
551 fragments and other molecular biology techniques are summarized in *Supp. Table 3* and were
552 supplied by IDT (Coralville, IA). All cloned inserts and DNA fragments were confirmed either via
553 Sanger sequencing[92] performed by Azenta Inc. (Burlington, MA), or Oxford-Nanopore whole-
554 plasmid sequencing performed by Plasmidsaurus (Eugene, OR). Commercially available NEB5-
555 alpha chemically competent cells (NEB, Ipswich, MA) were used for routine transformations.
556 Alternatively, electrocompetent *E. coli* cells were generated and transformed immediately (Gene
557 Pulser; Bio-Rad, Hercules, CA)[91]. Nucleotide sequence analyses were done at the National Center
558 for Biotechnology Information server (www.ncbi.nlm.nih.gov) and Benchling Biology Software
559 (www.benchling.com). Cloning was routinely performed via Gibson assembly[93] unless otherwise
560 indicated.

561 **Domain selection and cloning.** DBD-LacI is located on *lacI* 5' region, from the start codon to
562 base 198, encoding a peptide of 66 amino acids that comprises the N-terminus of our chimeric

563 proteins. This domain was obtained via PCR amplification from plasmid pCKTRBS-LacIwt[15]
564 using the oligonucleotide pair SV00001/SV00002. SatA from *H. ducreyi* (Uniprot Q7VL18) and
565 SiaP from *H. influenzae* (Uniprot P44542) were ordered as synthetic dsDNA fragments (IDT,
566 Coralville, IA). Several variants of the LBDs SatA and SiaP were amplified via PCR so that they
567 would include 5' overhangs including linkers (LNK2-5) or no linker whatsoever (LNK1). SatA
568 was amplified with primers SV00030-SV00034 / SV00004 and SiaP with primers SV00015-
569 SV00019 / JFJ0018.
570 Genes encoding the chimeric TFs were originally assembled in pUC19[94]. The vector was amplified
571 via divergent PCR with primers JFJ00019/JFJ00020, which introduce a consensus RBS for *E. coli*,
572 originating plasmid pUC19RBS. Tripartite Gibson assemblies including linear pUC19RBS, DBD-
573 LacI, and LNK-LBD were performed and the resulting constructions transformed into NEB5-alpha
574 competent cells. After a carbenicillin selection in solid media, a subset of the resulting clones was
575 isolated and their plasmids (pUC19-*Chimera* family, *Supp. Table 2*) purified and sequenced to
576 validate the presence of chimeric TFs. The resulting TF genes cloned into pUC19RBS were
577 amplified using SV00024/SV00025 for SiaP-derived chimeras and SV00024/SV00026 for their
578 SatA counterparts. These primers amplify our TF genes from the 5' DBD-LacI, just downstream
579 of the consensus RBS site, to the stop codon of the LBD. Low copy plasmid pCKTRBS[15] was
580 amplified via divergent PCR using primer pairs SV00022/SV00021 for SiaP-derived chimeras and
581 SV00022/SV00023 for their SatA counterparts. These primers introduce 5' and 3' overhang
582 sequences providing homology to our dsDNA chimeric gene fragments, and including in their 5'
583 side our consensus RBS site. Bipartite Gibson assemblies including linear pCKTRBS and the
584 fusion genes encoding the chimeras were performed, and the resulting assembly products
585 transformed into NEB5-alpha competent cells. After a carbenicillin selection in solid media,

586 resulting clones were isolated and their plasmids sequenced to confirm the presence of chimera-
587 encoding genes. Resulting pCKT-*Chimera* plasmids were purified from NEB5-alpha (pCKT-
588 *Chimera*) strains and electroporated into MG1655 cells. After confirming the presence of the
589 plasmid in the new host, MG1655 (pCKT-*Chimera*) were electroporated with the reporter plasmid
590 pHC_DYOLacI-R. Validated MG1655 (pCKT-*Chimera*, pHC_DYOLacI-R) strains were used for
591 *in vivo* and *in vitro* analysis of the functionality of the novel chimeric TFs when required. The
592 sequences of all the oligonucleotides used in this work can be found in *Supp. Table 3*.
593 **Electrophoretic mobility shift assay (EMSA).** In order to obtain cell extracts enriched in the
594 Sphnx/Siren proteins, NEB5-alpha (pCKT-Sphnx) and NEB5-alpha (pCKT-Siren) cells were
595 grown on LB supplemented with aTc for 5 hours at 37˚C in a shaker incubator. Cells were
596 recovered by centrifugation at 4°C (15' at 4,000 g) and then resuspended in 20 mM Tris-HCl (pH
597 7.0) to be sonicated at 70 amp for 30 seconds (30 cycles of 1 second; Branson Digital Sonifier,
598 Danbury, CN). A subsequent centrifugation at 4°C (20' at 20,000 g) was performed to eliminate
599 cell debris from the sample. The recovered supernatant was then incubated with the DNA probe,
600 a 251 bp dsDNA fragment amplified from plasmid pHC_DYOLacI-R with primers
601 SV00028/SV00029, containing P_{lac} (including *lacO*). Cell extract (5 µl) and DNA probe (5 µl)
602 were incubated together with reaction buffer (20% glycerol, 100 mM KCl, 500 $\mu g \cdot ml^{-1}$ bovine
603 serum albumin, 20mM Tris-HCl pH 7.5). The reaction mixture is incubated for 30' at 30 ˚C. After
604 incubation of the retardation mixtures, 2 µl of Gel loading dye without SDS (NEB, Ipswich, MA)
605 were added to each reaction. Mixtures were subsequently fractionated by electrophoresis in precast
606 6% polyacrylamide gels (Novex TBE Gels; Invitrogen, Waltham, MA) buffered with 0.5X TBE
607 (45 mM Tris borate, 1mM EDTA). Gels were run at constant voltage (100 V) for 90' in a Mini
608 Gel Tank (Invitrogen, Waltham, MA) and afterwards stained for 1 h in 50 ml of 0.5X TBE buffer

609 supplemented with 5 µl of SYBR Gold dye (Invitrogen, Waltham, MA) while gently shaken. Post
610 staining, gels were imaged in a UVP UVSolo Touch (Analytic Jena, Jena, Germany).
611 **qRT-PCR analysis.** Quantitative Real Time Polymerase Chain Reaction (qRT-PCR) was used to
612 assay the transcription of our reporter gene (*sfgfp*). MG1655 (pCKT-*Chimera*, pHCDYOLacI-R)
613 strains were precultured overnight in of 3 ml LB supplemented with the corresponding selection
614 antibiotics. On the next day, 100 µl of overnight preculture were used as inoculum for several
615 cultures, 10 ml each, of fresh LB (plus selection antibiotics). Each one of the cultures was
616 supplemented so that it would represent one of these three different conditions: *Basal* (aTc^-,
617 $Neu5Ac^-$), *Repressed* (aTc^+,$Neu5Ac^-$), and *De-repressed* (aTc^+, $Neu5Ac^+$). Where experimental
618 conditions required it, cultures were supplemented with 0.5 $\mu g \cdot ml^{-1}$ anhydrotetracycline (aTc)
619 (Clontech, Mountain View, CA) and 1 mM neuraminic acid (Neu5Ac; Sigma, St. Louis, MO). For
620 every independent qRT-PCR experiment, three biological replicates were created across each of
621 these three conditions. These cultures were placed in a shaker incubator at 37˚C for 5 hours and
622 pelleted at 4˚C and 4,000 g. Pellets were either frozen at -80˚C for future use or immediately
623 processed for total RNA extraction. Cells from each individual pellet were set for lysis by
624 resuspension in 1 ml of IBI reagent, a phenol and guanidine isothiocyanate mixture (IBI Scientific,
625 Dubuque, IA). 200 µl of chloroform (Fisher Scientific, Pittsburgh, PA) was added to the lysate
626 and the mixture was vigorously shaken to homogeneity, followed by centrifugation (benchtop
627 minifuge) at 4˚C and 14,500 rpm for 15 minutes that led to a phase separation. 500 µl of the upper
628 aqueous phase was mixed with an equal volume of 100% ethanol. At this point, total RNA was
629 purified from DNA/RNA phase with Direct-zol RNA Miniprep Kit (Zymo Research, Irvine, CA)
630 where the TRIzol reagent steps are replaced with the phenol and chloroform phase separation.
631 Total RNA eluted with ddH_2O was then retrotranscribed using qScript Flex cDNA Synthesis Kit

632 (QuantaBio, Beverly, MA). A volume totaling 1 µg RNA was incubated with random hexamers
633 and then supplemented with retrotranscriptase enzyme mix to synthesize cDNA. cDNA resulting
634 from the last step was used as a template for the amplification of a 256 bp fragment with
635 oligonucleotides SV00027A/SV00027 in a CFX96 Real-Time PCR System (Bio-Rad, Hercules,
636 CA). Primers had been designed to amplify a region of pHC_DYOLacI comprising the beginning
637 of *sfgfp*, so that when the tested chimera was being expressed it repressed the expression of GFP
638 unless Neu5Ac was present. Cq values were analyzed as a measure of gene expression using the
639 ΔΔCq ($2^{-\Delta\Delta Ct}$) method[95].

640 **Protein Structure Prediction using AlphaFold 3.** The 3D structures of the target proteins, Sphnx
641 (LacI-LNK2-SiaP, N_t-DBD-EKEKEK-LBD-C_t) and Kunst (LacI-LNK3-SiaP, N_t-DBD-
642 GSGSGS-LBD-C_t), were predicted using AlphaFold 3[78], developed by Google DeepMind
643 (London, UK). The predictions were performed using the publicly available AlphaFold 3 web
644 server (https://golgi.sandbox.google.com/). Both Sphnx and Kunst proteins were modeled as
645 dimers, along with the DNA double helix containing *lacO* (5'-ATTGTGAGCGGATAACAATT-
646 3' and its complementary sequence). Post-prediction, the structures were visualized and further
647 analyzed using VMD[96] to determine key protein-protein interactions.

648 **Molecular docking simulations.** Sphnx (LacI-LNK2-SiaP) and Kunst 3D structures were used to
649 forecast the primary binding patterns between the TF and a collection of putative ligands: Neu5Ac
650 (PubChem CID 439197), ManNAc (PubChem CID 439281), GlcNAc (PubChem CID 439174),
651 and Neu5Gc (PubChem CID 440001). First, water molecules and heteroatoms were removed from
652 the PDB files for Sphnx and Kunst 3D structure. Using AutoDock4[94], Kollman charges were
653 supplemented to the 3D structure of both our ligands (see PubChem reference above) and proteins,
654 with polar hydrogen charges being also added to the transcription factor structures. Once ligands

655 and receptors were prepared, a grid box delineating the region of the whole protein where the
656 docking would take place was established. Employing the molecular docking simulation protocols
657 provided by the AutoDock 4 suite, binding affinities were assessed. Post-docking, visualization
658 software (Pymol, Schrödinger Inc., New York, NY; Discovery Studio, Biovia, San Diego, CA)
659 were used to identify interactions between Sphnx / Kunst and the aforementioned ligands.
660
661

Acknowledgments and Funding

This material is based upon work supported by St. John's University (SJU) via J.F.J.'s starting package as well as St. John's SEED grants (2020-2021, 2021-2022, 2024-2025). Analysis and visualization of predicted structures was performed by F.X.V and A.B. on hardware resources supported by the National Institutes of Health under award number NIH SC2GM131992. This work is solely the responsibility of the authors and does not necessarily represent the official views of St. John's University. We want to acknowledge L. Bryan, D. Howarth, T. Mustacchio, M. Ruggiu, and W. Munyao (SJU) for their support and time sharing equipment, reagents, and protocols.

Author contributions

S.J.V. and J.F.J. developed the concepts, conceived of the study plan, and wrote the manuscript. S.J.V. designed and performed the experiments with assistance from S.A.P., V.A.S.E., A.GM., G.X.C., A.R., A.C.L., E.S.N., P.K., J.D.F., K.G., and D.T.S. J.F.J provided project administration, supervision, resources, and data analysis, the latter with V.A.S.E. and S.J.V.'s assistance. A.B. performed 3D structure modeling and analysis of protein interactions under the supervision of F.X.V., who wrote the corresponding section of the manuscript. A.GM. perform molecular docking simulations.

Competing interests

The authors declare no competing interests.

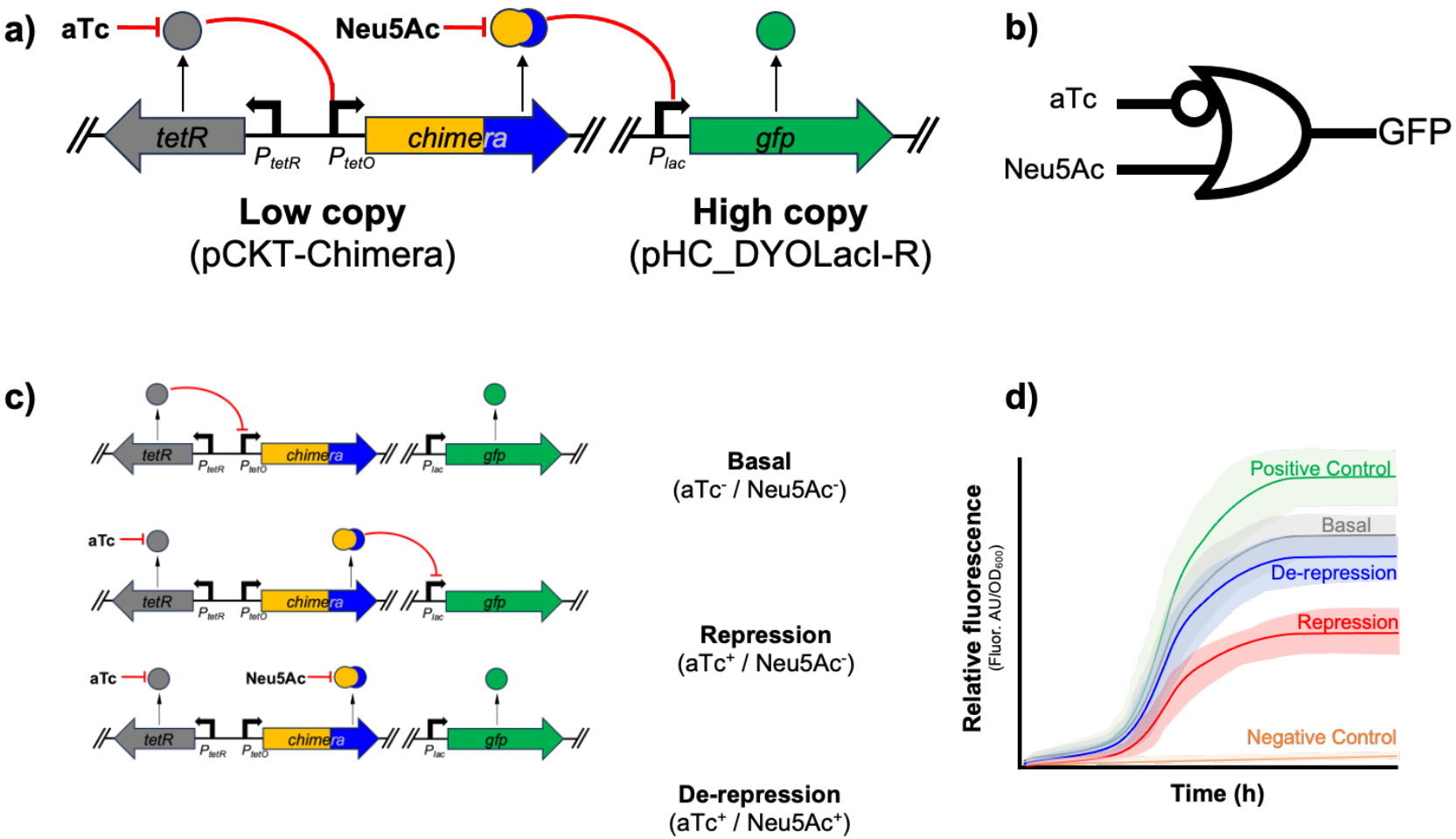
685 **TABLES**

Table 1. Chimeric transcription factors discussed in this work.

Name		TF behaviour		
Common	**Systematic**	**Repressor**	**De-repressible**	**Comments**
Siren	LacI-LNK1-SiaP	Yes	Yes	Functional only in late stationary phase
Sphnx	LacI-LNK2-SiaP	Yes	Yes	Biosensor candidate
Kunst	LacI-LNK3-SiaP	Yes	No	-
Scorpio	LacI-LNK4-SiaP	No	No	-
Bastet	LacI-LNK5-SiaP	No	No	Repression in early hours of growth.
Lumos	LacI-LNK6-SiaP	No	No	Duplicated sequence acts as Linker.
Harpy	LacI-LNK1-SatA	Yes	Unreliable	Unstable strain
Haus	LacI-LNK2-SatA	No	No	-

686 **FIGURES**

Figure 1



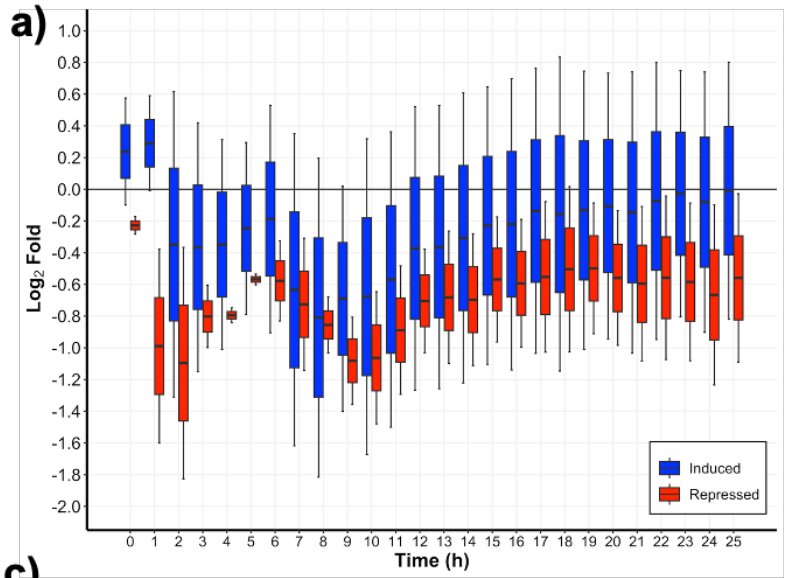
687

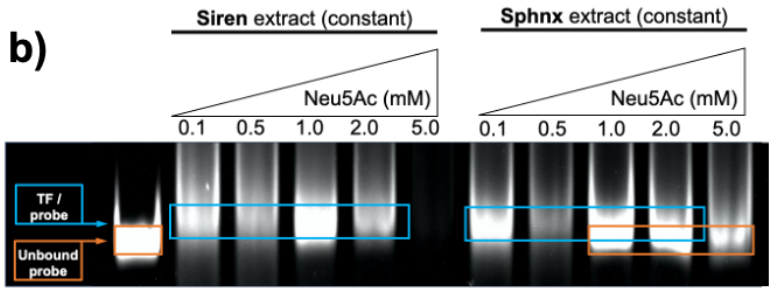
688

Figure 1. Screening strategy to find functional chimeric transcription factors. a) Biosensor genetic circuit for the *in vivo* testing of the functionality of cloned chimeric TFs. MG1655 (pCKT-*Chimera*, pHC_DYOLacI-R) cells express a chimeric TF under the control of TetR. The chimeric product can repress in *trans* the expression of a GFP reporter unless Neu5Ac is present. *tetR* / TetR (*grey arrow / circle*); chimeric TF gene / protein (*gold-blue arrow / circle*); *gfp* /GFP (*green arrow / circle*); anhydrotetracycline, aTc; neuraminic acid, Neu5Ac. **b)** Standard representation of the B-or-not-A universal two input gate that governs the expression of GFP in the circuit. **c)** States of the biosensor genetic circuit (top to bottom). *Basal*, the absence of aTc (aTc^-) enables TeR to repress the expression of the chimeric repressor: regardless of the presence or absence of Neu5Ac GFP is expressed, and a fluorescent signal is identified. *Repressed*, the presence of aTc (aTc^+) induces TetR which is unable to repress the promoter driving the transcription of the chimeric TF, which in turn represses the promoter driving the expression of *gfp* so that no fluorescent signal is detected. *De-repressed*, the presence of aTc (aTc^+) ensures the production of the chimeric TF which, if functional, gets in turn induced by Neu5Ac, so that it cannot repress *gfp* and the GFP-associated fluorescence is detected again. **d)** Expected profile of fluorescent production over time by a bacterial culture expressing the biosensor genetic circuit and a functional chimeric TF. Relative fluorescence (GFP-associated fluorescence corrected by cell density measured as OD_{600}). *Positive control* (*green line*), corresponds to the maximum GFP-associated fluorescence that can be observed, corresponding to a NEB5-alpha (pHC_DYOLacI-R) strain in which no repressor of the promoter driving GFP expression is present. *Negative control* (*orange line*), displays background fluorescence by a strain that does not carry *gfp*. *Basal* (*grey line*), *Repressed* (*red line*), *De-repressed* (*blue line*): GFP production associated to a MG1655 (pCKT-*Chimera*, pHC_DYOLacI-R) culture expressing a chimera in the basal, repressed, and de-repressed states, respectively.

689

Figure 2


a)
Log2 Fold
Time (h)
Induced
Repressed


b)
Siren extract (constant)
Sphnx extract (constant)
Neu5Ac (mM)
0.1 0.5 1.0 2.0 5.0
TF / probe
Unbound probe

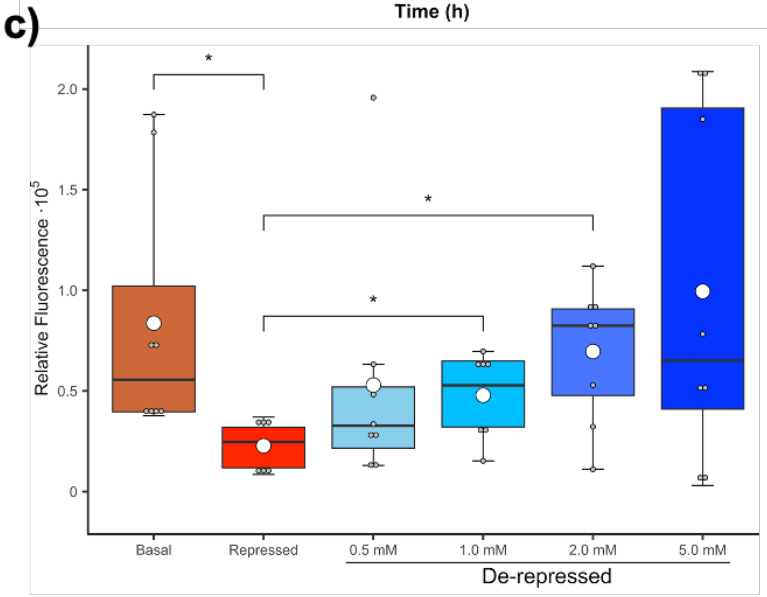

c)
Relative Fluorescence · 10^5
Basal
Repressed
0.5 mM
1.0 mM
2.0 mM
5.0 mM
De-repressed

Figure 2. *In vivo* validation of the functionality of Sphnx (LacI-LNK2-SiaP). a) Time course showing relative fluorescence of MG1655 (pCKT-Sphnx, pHC_DYOLacI-R), MG-Sphnx, cells grown in a multiwell plate reader. Promoter activity is measured as log_2 fold of the relative fluorescence (GFP-associated fluorescence in arbitrary units / OD_{600}) of the strain growing in the de-repressed (aTc^+, $Neu5Ac^+$; *blue boxplots*) or repressed (aTc^+, $Neu5Ac^-$; *red boxplots*) states compared to the basal expression of the reporter (aTc^-, $Neu5Ac^-$). Boxplots with whiskers represent data dispersion of the average values of biological replicas (n = 8-10). It can be appreciated how in order to approach the basal activity (log_2 fold = 0), the addition of neuraminic acid is necessary when the chimera gets expressed by the supplementation of aTc. Growth conditions and fluorescence assays performed as described in *Materials and Methods*.

b) Sphnx and Siren protein interact with their cognate promoter P_{lac}. EMSA assays were performed as indicated under *Materials and Methods*. Interaction between NEB5-alpha (pCKT-*Chimera*) cell extracts expressing either Sphnx or Siren chimeric TFs and a dsDNA probe (251 bp) including the P_{lac} promoter. A fixed concentration of cell extract and DNA probe was exposed to growing concentrations of neuraminic acid (Neu5Ac). Lane numbers refer to the Neu5Ac concentration (mM) used for each reaction. The positions in the gel associated to unbound and bound probes are identified with orange and blue dashed rectangles, respectively.

c) Relative fluorescence of MG-Sphnx cultures after induction with different concentrations of neuraminic acid and growth in a multiwell plate reader for 14 hours. Promoter activity is measured as Rel.fluor, relative fluorescence (GFP-associated fluorescence in arbitrary units / OD_{600}) of the strain growing in the *basal* (aTc^-, $Neu5Ac^-$; *brown*), repressed (aTc^+, $Neu5Ac^-$; *red*) or de-repressed (aTc^+, $Neu5Ac^+$; *blue shades*) conditions. The concentrations of neuraminic acid (mM) added to the de-repressed conditions are specified under the corresponding plots. Boxplots with whiskers represent data dispersion of the average values of biological replicas (n = 8). The average value for every condition corresponds to a white circle. Growth conditions, induction protocol, and fluorescence assays performed as described in *Materials and Methods*. Statistically significant differences between selected conditions are marked with stars (p value < 0.05).

690

691

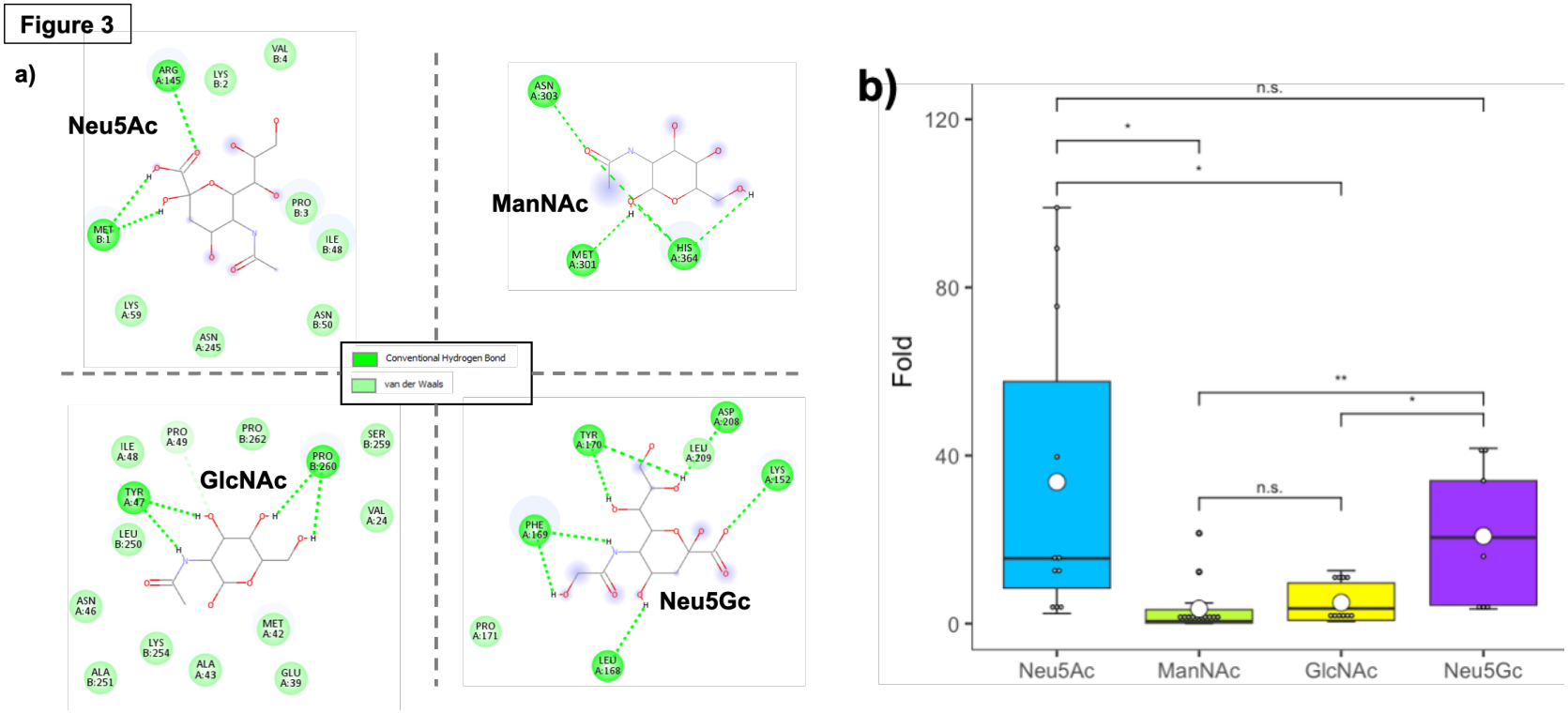


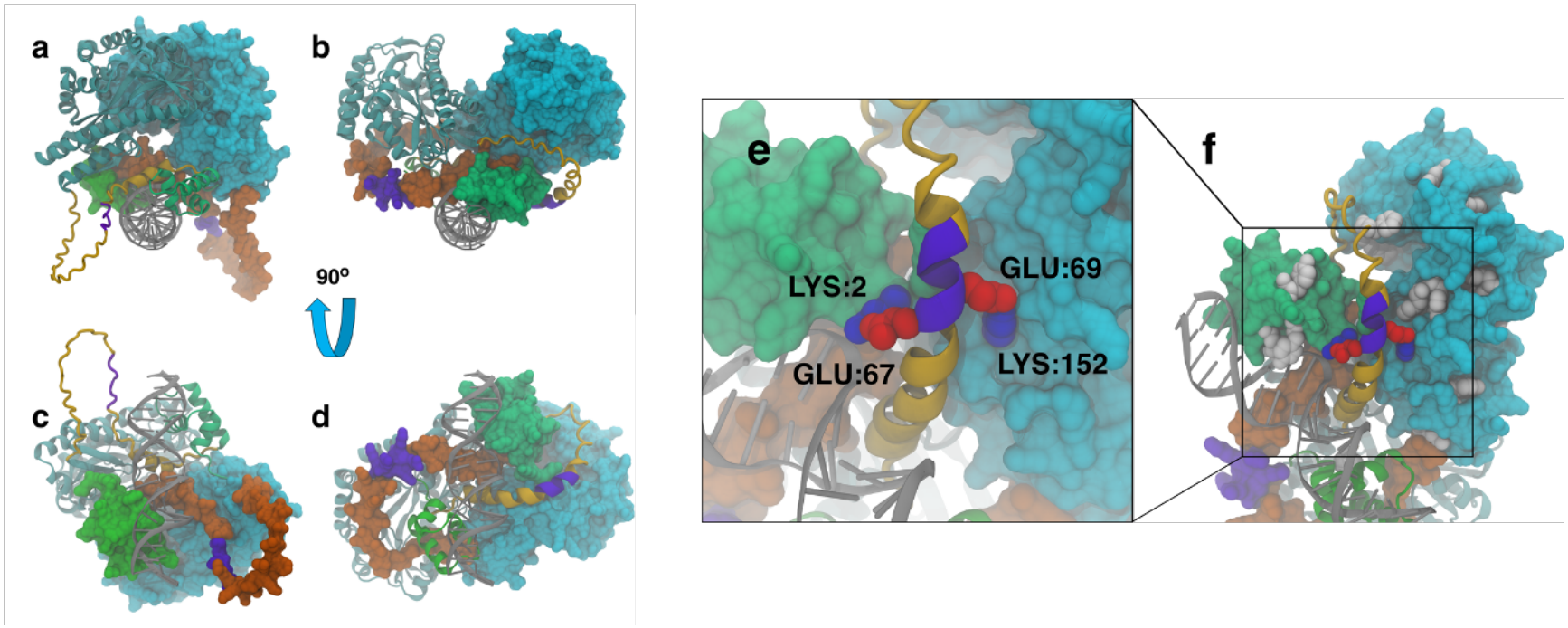
Figure 3 Interaction of Sphnx (LacI-LNK2-SiaP) with different ligands. Study of the interaction of Sphnx with N-acetyl-D-neuraminic acid (Neu5Ac), N-acetyl-D-mannosamine (ManNAc), N-acetyl-D-glucosamine (GlcNAc), and N-glycolyl-neuraminic acid (Neu5Gc).

a) Molecular docking simulations for 3D models of Sphnx and ligand molecules displaying stabilizing van der Waals forces and hydrogen bonds between the binding pocket of the regulator and the ligands. Images generated in Discovery Studio (*Materials and Methods*).

b) Relative fluorescence of MG1655 (pCKT-Sphnx, pHC_DYOLacI-R), MG-Sphnx, cultures after induction 1 mM of one of the following compounds: Neu5Ac (*blue*), ManNAc (*green*), GlcNAc (*yellow*), or Neu5Gc (*purple*) and growth in a shaker incubator with a subsequent transfer to a plate-reader. Promoter activity is measured as fold change of Rel.Fluor over a culture in the repressed condition (aTc^+, $ligand^-$). Boxplots with whiskers represent data dispersion of the average values of biological replicas (n = 9-12). The average value for every condition corresponds to a white circle. Growth conditions, induction protocol, and fluorescence assays performed as described in *Materials and Methods*. Statistically significant differences between selected conditions are marked with stars (* p value < 0.05; ** p value < 0.01; n.s. non-significant).

693

Figure 4



694

Figure 4. Modeled structures of chimeric protein dimers with double helix DNA obtained with Alphafold 3. For each structure, one monomer is shown using the New Cartoon representation and the other using the space-filling Surf representation. The DBD is shown in *green*, the PBP in *cyan*, and the connecting chain in *orange*, with the engineered linker in *violet*. The Kunst (**a**) and Sphnx (**b**) dimers are shown looking down the axis of the double helix, along with the rotated Kunst (**c**) and Sphnx (**d**) structures. The next panels contain the zoomed-in helical protein (**e**) linker (*violet*) in contact with the DBD (green) and PBP through the formation of salt bridges, with acidic residues shown in red and basic residues in blue. The zoomed-out protein (**f**) highlights the location of aromatic residues (*white*) that can be used to explore the validity of predicted structures using fluorescence anisotropy.

695 **BIBLIOGRAPHY**

696 1. Pancrazio, J. J., Whelan, J. P., Borkholder, D. A., Ma, W. & Stenger, D. A. Development
697 and application of cell-based biosensors. *Ann. Biomed. Eng.* **27**, 697–711 (1999).
698 2. Su, L., Jia, W., Hou, C. & Lei, Y. Microbial biosensors: a review. *Biosens. Bioelectron.* **26**,
699 1788–1799 (2011).
700 3. Moraskie, M. *et al.* Microbial whole-cell biosensors: Current applications, challenges, and
701 future perspectives. *Biosens. Bioelectron.* **191**, 113359 (2021).
702 4. O'Connor, G. *et al.* Whole-Cell Biosensors as Tools for the Detection of Quorum-Sensing
703 Molecules: Uses in Diagnostics and the Investigation of the Quorum-Sensing Mechanism.
704 *Adv. Biochem. Eng. Biotechnol.* (2015) doi:10.1007/10_2015_337.
705 5. Chang, H.-J., Voyvodic, P. L., Zúñiga, A. & Bonnet, J. Microbially derived biosensors for
706 diagnosis, monitoring and epidemiology. *Microb. Biotechnol.* **10**, 1031–1035 (2017).
707 6. Wu, Y., Wang, C.-W., Wang, D. & Wei, N. A Whole-Cell Biosensor for Point-of-Care
708 Detection of Waterborne Bacterial Pathogens. *ACS Synth. Biol.* **10**, 333–344 (2021).
709 7. Goode, J. A., Rushworth, J. V. H. & Millner, P. A. Biosensor Regeneration: A Review of
710 Common Techniques and Outcomes. *Langmuir* **31**, 6267–6276 (2015).
711 8. Zhang, F. & Keasling, J. Biosensors and their applications in microbial metabolic
712 engineering. *Trends Microbiol.* **19**, 323–329 (2011).
713 9. He, W., Yuan, S., Zhong, W.-H., Siddikee, M. A. & Dai, C.-C. Application of genetically
714 engineered microbial whole-cell biosensors for combined chemosensing. *Appl. Microbiol.*
715 *Biotechnol.* **100**, 1109–1119 (2016).
716 10. Saltepe, B., Kehribar, E. Ş., Su Yirmibeşoğlu, S. S. & Şafak Şeker, U. Ö. Cellular
717 Biosensors with Engineered Genetic Circuits. *ACS Sens* **3**, 13–26 (2018).

718 11. Berepiki, A., Kent, R., Machado, L. F. M. & Dixon, N. Development of High-Performance
719 Whole Cell Biosensors Aided by Statistical Modeling. *ACS Synth. Biol.* **9**, 576–589 (2020).
720 12. Schleif, R. DNA binding by proteins. *Science* **241**, 1182–1187 (1988).
721 13. Rogers, J. K. *et al.* Synthetic biosensors for precise gene control and real-time monitoring of
722 metabolites. *Nucleic Acids Res.* **43**, 7648–7660 (2015).
723 14. Tellechea-Luzardo, J., Stiebritz, M. T. & Carbonell, P. Transcription factor-based
724 biosensors for screening and dynamic regulation. *Front Bioeng Biotechnol* **11**, 1118702
725 (2023).
726 15. Juárez, J. F., Lecube-Azpeitia, B., Brown, S. L., Johnston, C. D. & Church, G. M. Biosensor
727 libraries harness large classes of binding domains for construction of allosteric
728 transcriptional regulators. *Nat. Commun.* **9**, 3101 (2018).
729 16. Koch, M., Pandi, A., Borkowski, O., Batista, A. C. & Faulon, J.-L. Custom-made
730 transcriptional biosensors for metabolic engineering. *Curr. Opin. Biotechnol.* **59**, 78–84
731 (2019).
732 17. Kaufman, E. & Lamster, I. B. The diagnostic applications of saliva--a review. *Crit. Rev.*
733 *Oral Biol. Med.* **13**, 197–212 (2002).
734 18. Lee, Y.-H. & Wong, D. T. Saliva: an emerging biofluid for early detection of diseases. *Am.*
735 *J. Dent.* **22**, 241–248 (2009).
736 19. Wren, M. E., Shirtcliff, E. A. & Drury, S. S. Not all biofluids are created equal: chewing
737 over salivary diagnostics and the epigenome. *Clin. Ther.* **37**, 529–539 (2015).
738 20. Khanna, P. & Walt, D. R. Salivary diagnostics using a portable point-of-service platform: a
739 review. *Clin. Ther.* **37**, 498–504 (2015).
740 21. Bandodkar, A. J. & Wang, J. Non-invasive wearable electrochemical sensors: a review.

741 *Trends Biotechnol.* **32**, 363–371 (2014).
742 22. Campuzano, S., Pedrero, M., Torrente-Rodríguez, R. M. & Pingarrón, J. M. Affinity-based
743 wearable electrochemical biosensors: Natural versus biomimetic receptors. *Anal. Sens.*
744 (2022) doi:10.1002/anse.202200087.
745 23. Kim, J., Campbell, A. S., de Ávila, B. E.-F. & Wang, J. Wearable biosensors for healthcare
746 monitoring. *Nat. Biotechnol.* **37**, 389–406 (2019).
747 24. Omar, E. Current concepts and future of noninvasive procedures for diagnosing oral
748 squamous cell carcinoma--a systematic review. *Head Face Med.* **11**, 6 (2015).
749 25. Wang, S., Yang, M., Li, R. & Bai, J. Current advances in noninvasive methods for the
750 diagnosis of oral squamous cell carcinoma: a review. *Eur. J. Med. Res.* **28**, 53 (2023).
751 26. Haines-Menges, B. L., Whitaker, W. B., Lubin, J. B. & Boyd, E. F. Host Sialic Acids: A
752 Delicacy for the Pathogen with Discerning Taste. *Microbiol Spectr* **3**, (2015).
753 27. Blix, G. Concerning the carbohydrate groups of submaxillary mucin. *Hoppe Seylers Z.*
754 *Physiol. Chem.* **240**, 43–54.
755 28. Lundblad, A. Gunnar Blix and his discovery of sialic acids. Fascinating molecules in
756 glycobiology. *Ups. J. Med. Sci.* **120**, 104–112 (2015).
757 29. Röhrig, C. H., Choi, S. S. H. & Baldwin, N. The nutritional role of free sialic acid, a human
758 milk monosaccharide, and its application as a functional food ingredient. *Crit. Rev. Food*
759 *Sci. Nutr.* **57**, 1017–1038 (2017).
760 30. Vajaria, B. N. *et al.* Salivary glyco-sialylation changes monitors oral carcinogenesis.
761 *Glycoconj. J.* **31**, 649–659 (2014).
762 31. Phansopa, C. *et al.* Characterization of a sialate-O-acetylesterase (NanS) from the oral
763 pathogen Tannerella forsythia that enhances sialic acid release by NanH, its cognate

764 sialidase. *Biochem. J* **472**, 157–167 (2015).

765 32. Arthisri, A. S., Sathiyamoorthy, A., Meenakshi, B. & Chandran, C. R. Ratio of Salivary
766 Sialic Acid to Fucose as Tumor Markers in Potentially Malignant Disorders and Oral
767 Cancer. *Contemp. Clin. Dent.* **11**, 131–135 (2020).

768 33. Chaudhari, V., Pradeep, G. L., Prakash, N. & Mahajan, A. M. Estimation of salivary sialic
769 acid in oral premalignancy and oral squamous cell carcinoma. *Contemp. Clin. Dent.* **7**, 451–
770 456 (2016).

771 34. Chittemsetti, S., Manchikatla, P. K. & Guttikonda, V. Estimation of serum sialic acid in oral
772 submucous fibrosis and oral squamous cell carcinoma. *J. Oral Maxillofac. Pathol.* **23**, 156
773 (2019).

774 35. Markopoulos, A. K. Current aspects on oral squamous cell carcinoma. *Open Dent. J.* **6**,
775 126–130 (2012).

776 36. Coletta, R. D., Yeudall, W. A. & Salo, T. Grand Challenges in Oral Cancers. *Front Oral*
777 *Health* **1**, 3 (2020).

778 37. Warnakulasuriya, S. & Greenspan, J. S. *Textbook of Oral Cancer: Prevention, Diagnosis*
779 *and Management*. (Springer Nature, 2020).

780 38. Jacob, T. V. *et al.* A non-invasive study to estimate and compare salivary sialic acid level as
781 tumor marker in patients with pre-cancer and oral cancer. *J. Cancer Res. Ther.* **12**, 634–639
782 (2016).

783 39. Sanjay, P. R., Hallikeri, K. & Shivashankara, A. R. Evaluation of salivary sialic acid, total
784 protein, and total sugar in oral cancer: a preliminary report. *Indian J. Dent. Res.* **19**, 288–
785 291 (2008).

786 40. Sridharan, G., Ramani, P., Patankar, S. & Vijayaraghavan, R. Evaluation of salivary

787 metabolomics in oral leukoplakia and oral squamous cell carcinoma. *J. Oral Pathol. Med.*
788 **48**, 299–306 (2019).
789 41. Robson, B. Bioinformatics studies on a function of the SARS-CoV-2 spike glycoprotein as
790 the binding of host sialic acid glycans. *Comput. Biol. Med.* **122**, 103849 (2020).
791 42. Stanton, B. C. *et al.* Systematic transfer of prokaryotic sensors and circuits to mammalian
792 cells. *ACS Synth. Biol.* **3**, 880–891 (2014).
793 43. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol.*
794 *Biol.* **3**, 318–356 (1961).
795 44. Meinhardt, S. *et al.* Novel insights from hybrid LacI/GalR proteins: family-wide functional
796 attributes and biologically significant variation in transcription repression. *Nucleic Acids*
797 *Res.* **40**, 11139–11154 (2012).
798 45. Fukami-Kobayashi, K., Tateno, Y. & Nishikawa, K. Parallel evolution of ligand specificity
799 between LacI/GalR family repressors and periplasmic sugar-binding proteins. *Mol. Biol.*
800 *Evol.* **20**, 267–277 (2003).
801 46. Sakaguchi-Mikami, A., Taniguchi, A., Sode, K. & Yamazaki, T. Construction of a novel
802 glucose-sensing molecule based on a substrate-binding protein for intracellular sensing.
803 *Biotechnol. Bioeng.* **108**, 725–733 (2011).
804 47. Giuliani, S. E. *et al.* Environment sensing and response mediated by ABC transporters.
805 *BMC Genomics* **12 Suppl 1**, S8 (2011).
806 48. Fukamizo, T., Kitaoku, Y. & Suginta, W. Periplasmic solute-binding proteins: Structure
807 classification and chitooligosaccharide recognition. *Int. J. Biol. Macromol.* **128**, 985–993
808 (2019).
809 49. Locher, K. P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat.*

810 *Struct. Mol. Biol.* **23**, 487–493 (2016).
811 50. Post, D. M. B., Mungur, R., Gibson, B. W. & Munson, R. S., Jr. Identification of a novel
812 sialic acid transporter in Haemophilus ducreyi. *Infect. Immun.* **73**, 6727–6735 (2005).
813 51. Gruteser, N., Marin, K., Krämer, R. & Thomas, G. H. Sialic acid utilization by the soil
814 bacterium Corynebacterium glutamicum. *FEMS Microbiol. Lett.* **336**, 131–138 (2012).
815 52. Müller, A. *et al.* Conservation of structure and mechanism in primary and secondary
816 transporters exemplified by SiaP, a sialic acid binding virulence factor from Haemophilus
817 influenzae. *J. Biol. Chem.* **281**, 22212–22222 (2006).
818 53. Severi, E. *et al.* Sialic acid transport in Haemophilus influenzae is essential for
819 lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite
820 ATP-independent periplasmic transporter. *Mol Microbiol* **58**, 1173–1185 (2005).
821 54. Gangi Setty, T., Cho, C., Govindappa, S., Apicella, M. A. & Ramaswamy, S. Bacterial
822 periplasmic sialic acid-binding proteins exhibit a conserved binding site. *Acta Crystallogr.*
823 *D Biol. Crystallogr.* **70**, 1801–1811 (2014).
824 55. Gangi Setty, T. *et al.* Molecular characterization of the interaction of sialic acid with the
825 periplasmic binding protein from. *J Biol Chem* **293**, 20073–20084 (2018).
826 56. Oertel-Buchheit, P., Reinbolt, J., John, M., Granger-Schnarr, M. & Schnarr, M. A LexA
827 mutant repressor with a relaxed inter-domain linker. *Protein Sci.* **7**, 512–515 (1998).
828 57. Tsirigotaki, A., De Geyter, J., Šoštaric, N., Economou, A. & Karamanou, S. Protein export
829 through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **15**, 21–36 (2017).
830 58. Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13**, 247–
831 261 (1995).
832 59. Liu, Q., Schumacher, J., Wan, X., Lou, C. & Wang, B. Orthogonality and Burdens of

833 Heterologous AND Gate Gene Circuits in E. coli. *ACS Synth. Biol.* **7**, 553–564 (2018).
834 60. Logic Gates. *The Department of Mathematics and Computer Science at Oxford College of*
835 *Emory University* https://mathcenter.oxford.emory.edu/site/cs170/gates/.
836 61. Lewis, M. The lac repressor. *C. R. Biol.* **328**, 521–548 (2005).
837 62. Deans, T. L., Cantor, C. R. & Collins, J. J. A tunable genetic switch based on RNAi and
838 repressor proteins for regulating gene expression in mammalian cells. *Cell* **130**, 363–372
839 (2007).
840 63. Langdon, R. C., Burr, T., Pagan-Westphal, S. & Hochschild, A. A chimeric activator of
841 transcription that uses two DNA-binding domains to make simultaneous contact with pairs
842 of recognition sites. *Mol. Microbiol.* **41**, 885–896 (2001).
843 64. Durante-Rodríguez, G. *et al.* Identification of a missing link in the evolution of an enzyme
844 into a transcriptional regulator. *PLoS One* **8**, e57518 (2013).
845 65. Durante-Rodríguez, G., Mancheño, J. M., Díaz, E. & Carmona, M. Refactoring the λ phage
846 lytic/lysogenic decision with a synthetic regulator. *Microbiologyopen* **5**, 575–581 (2016).
847 66. Oktay, S. *et al.* Serum and saliva sialic acid in periodontitis patients with and without
848 cardiovascular disease. *Pathophysiol. Haemost. Thromb.* **37**, 67–71 (2010).
849 67. Jacobs, C. L. *et al.* Substrate specificity of the sialic acid biosynthetic pathway.
850 *Biochemistry* **40**, 12864–12874 (2001).
851 68. Chen, X. & Varki, A. Advances in the biology and chemistry of sialic acids. *ACS Chem*
852 *Biol* **5**, 163–176 (2010).
853 69. Mo, Y. *et al.* Rational Design of -Acetylglucosamine-2-epimerase and -Acetylneuraminic
854 Lyase for Efficient -Acetylneuraminic Acid Biosynthesis. *J Agric Food Chem* **73**, 5320–
855 5327 (2025).

856 70. Varki, A. Colloquium paper: uniquely human evolution of sialic acid genetics and biology.
857 *Proc Natl Acad Sci U S A* **107 Suppl 2**, 8939–8946 (2010).
858 71. Kooner, A. S., Yu, H. & Chen, X. Synthesis of -Glycolylneuraminic Acid (Neu5Gc) and Its
859 Glycosides. *Front Immunol* **10**, 2004 (2019).
860 72. Dunker, K. *et al.* Human-specific evolutionary genetic loss of addition of a single oxygen
861 atom from sialic acids increases hydrophobicity of cells and proteins. *Carbohydr Res* **552**,
862 109469 (2025).
863 73. Samraj, A. N., Läubli, H., Varki, N. & Varki, A. Involvement of a non-human sialic Acid in
864 human cancer. *Front Oncol* **4**, 33 (2014).
865 74. Pearce, O. M. T. & Läubli, H. Sialic acids in cancer biology and immunity. *Glycobiology*
866 **26**, 111–128 (2016).
867 75. Snoek, T. *et al.* Evolution-guided engineering of small-molecule biosensors. *Nucleic Acids*
868 *Res.* **48**, e3 (2020).
869 76. Hebisch, E., Knebel, J., Landsberg, J., Frey, E. & Leisner, M. High variation of
870 fluorescence protein maturation times in closely related Escherichia coli strains. *PLoS One*
871 **8**, e75991 (2013).
872 77. Chen, X., Zaro, J. L. & Shen, W.-C. Fusion protein linkers: property, design and
873 functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–1369 (2013).
874 78. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with
875 AlphaFold 3. *Nature* (2024) doi:10.1038/s41586-024-07487-w.
876 79. Meuzelaar, H., Vreede, J. & Woutersen, S. Influence of Glu/Arg, Asp/Arg, and Glu/Lys
877 Salt Bridges on α-Helical Stability and Folding Kinetics. *Biophys. J.* **110**, 2328–2341
878 (2016).

879 80. Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. Diversity of microbial
880 sialic acid metabolism. *Microbiol Mol Biol Rev* **68**, 132–153 (2004).
881 81. Virion, Z. *et al.* Sialic acid mediated mechanical activation of β adrenergic receptors by
882 bacterial pili. *Nat Commun* **10**, 4752 (2019).
883 82. Clark, N. D. *et al.* Functional and structural analyses reveal that a dual domain sialidase
884 protects bacteria from complement killing through desialylation of complement factors.
885 *PLoS Pathog* **19**, e1011674 (2023).
886 83. Raman, S., Taylor, N., Genuth, N., Fields, S. & Church, G. M. Engineering allostery.
887 *Trends Genet.* **30**, 521–528 (2014).
888 84. Malisi, C. *et al.* Binding pocket optimization by computational protein design. *PLoS One* **7**,
889 e52505 (2012).
890 85. Lerma Romero, J. A. *et al.* Binding pocket stabilization by high-throughput screening of
891 yeast display libraries. *Front Mol Biosci* **9**, 1023131 (2022).
892 86. Levine, M. J. Development of artificial salivas. *Crit. Rev. Oral Biol. Med.* **4**, 279–286
893 (1993).
894 87. Persico, M. *et al.* From Protein Communication to Drug Discovery. *Curr. Top. Med. Chem.*
895 **15**, 2019–2031 (2015).
896 88. Long, M., Betrán, E., Thornton, K. & Wang, W. The origin of new genes: glimpses from
897 the young and old. *Nat. Rev. Genet.* **4**, 865–875 (2003).
898 89. Bertani, G. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J.*
899 *Bacteriol.* **186**, 595–600 (2004).
900 90. Miller, J. H. *Experiments in Molecular Genetics*. (1972).
901 91. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual*. (CSHL Press,

902 2001).

903 92. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating

904 inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).

905 93. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred

906 kilobases. *Nat. Methods* **6**, 343–345 (2009).

907 94. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host

908 strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119

909 (1985).

910 95. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time

911 quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).

912 96. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.*

913 **14**, 33–8, 27–8 (1996).

914