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], Gemma X. Crowley[1], Alexandra Rutkowski[1], Amelia C. Lam[1], Elizabeth S. Nazginov[1], Paola Kotemelo[1], Agustina Bacelo[1,2], 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)

ABSTRACT

Accurate and timely diagnosis of oral squamous cell carcinoma (OSCC) is crucial in preventing its progression to advanced stages with a poor 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-acetylneuraminic 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.

33 **INTRODUCTION**

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

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

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

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

78 forefront of early disease prevention and diagnosis[20]. Electrochemical intraoral biodevices capable
79 of detecting salivary biomarkers presently exist, yet they are limited in the scope of molecules they
80 detect and lack adaptability to sense new ones[21]. Biological detection systems based on engineered
81 genetic circuits offer a plug-and-play potential, in which the substitution of one TF by another,
82 plus its cognate operator, is often enough to repurpose a pre-existing sensor. Bioengineering
83 solutions enabling point-of-care diagnostics will greatly benefit patients by improving at-home
84 health monitorization[22]. This approach is especially relevant due to the widespread acceptance and
85 use of consumer technology within society, which has normalized wearable biosensors connected
86 to a phone or a smartwatch[23]. Diagnostic screenings that do not require invasive procedures (i.e. a
87 blood sample or a biopsy) are preferred by patients, which makes saliva a prime biological fluid
88 candidate to be sampled for health and disease markers[24,25]. In this work we focused our attention
89 on N-acetylneuraminic acid (neuraminic acid, Neu5Ac), the most common of the sialic acids
90 originally isolated from salivary glands[26–28]. Neu5Ac fulfills key roles in human physiology[29] as
91 well as in commensal and pathogenic bacteria[26]. Moreover, changes in salivary sialylation might
92 become a suitable early indicator of changes associated with malignant pathologies, given their
93 correlation to periodontal disease (PD) and oral cancer[30]. The oral microbiome of patients with PD
94 include bacteria such as *Tannerella forsythia*, which engage in glycan harvesting behavior: they
95 cleave Neu5Ac molecules from glycoproteins to stealthily evade their host's immune system[31].
96 Additionally, abnormally high concentrations of Neu5Ac have been associated with the onset of
97 oral cancer[32–34]: a group of neoplasms which affect any region in the oral cavity, pharynx, and
98 salivary glands[35]. Oral cancer is often used interchangeably with oral squamous cell carcinoma
99 (OSCC), which accounts for more than 90% of oral malignant epithelial neoplasms[35]. As one of
100 the most prevalent cancers in the world[36], untreated OSCC has very poor prognosis and, as such,

101 early detection is essential for successful treatment[37]. Previous evaluations of sialic acid levels in
102 the saliva of healthy individuals compared to patients with oral cancer provided a guide for the
103 range of salivary Neu5Ac that can be observed. The concentration range identified in the saliva of
104 healthy adults was approximately 21.65 ± 5.71 mg/dl (0.52 - 0.88 mM), while confirmed oral pre-
105 cancer individuals displayed 59.75 ± 7.29 mg/dl (1.7 - 2.2 mM). Histopathologically confirmed
106 OSCC patients showed a wide range of concentrations, 204.85 ± 60.38 mg/dl (4.6 - 8.6 mM),
107 always superior to the other two groups[38]. Additional studies have employed multivariate
108 regression analysis comparing both free and protein-bound salivary sialic acid concentrations
109 between OSCC patients and healthy individuals. While free sialic acid levels were found to be
110 significantly elevated in patients diagnosed with oral cancer, differences in protein-bound sialic
111 acid levels were not found to be statistically significant[39]. These observations are further evidence
112 that total salivary sialic acid concentration can be utilized as a biomarker for OSCC. It is important
113 to note that there is at least one study in which Neu5Ac concentrations in saliva were decreased in
114 OSCC patients, having been hypothesized that a downregulation of the pathways modulating
115 Neu5Ac production might have occurred in these subjects[40]. Beyond its connection to OSCC, it
116 has also been proposed that salivary Neu5Ac could interfere with SARS-Cov-2 (COVID-19)
117 infection, acting as a competitive inhibitor blocking the viral antigens recognized by the host
118 receptors[41]. Each of these factors made quantification of salivary Neu5Ac a desirable target for
119 detection by biosensors.

120 The absence of a native TF capable of detecting Neu5Ac motivated us to construct a
121 synthetic monogenic repressor for its integration in a genetic circuit that would enable the strain
122 bearing it to behave as a WCB. We favor these regulators due to their simplicity, compact nature,
123 and ease of integration into synthetic circuits[8,9,42]. The process of constructing our custom

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

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

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

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

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

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

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

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

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

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

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

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

300 In order to validate our preliminary observations, we systematically compared GFP
301 production over time of MG1655 (pCKT-Sphnx, pHC_DYOLacI-R) cells expressing Sphnx either
302 in the presence or absence of Neu5Ac. *Figure 2a* displays the results of time course experiments
303 over a period of 25 hours. GFP production induced by the presence of Neu5Ac was observed from
304 the first hour of growth, with the Neu5Ac-induced cultures showing a tendency to display higher
305 fluorescence than the repressed ones. At this point our attention was captured by Kunst as a model
306 for a non-functional chimera in opposition to Sphnx. *Supp. Figure 3* shows the behavior of Kunst,

307 which exhibited a tendency to repress P_{lac} at the very beginning of a 40 h time course, yet after the
308 3 h mark did not show any difference between induced and uninduced conditions. A preliminary
309 evaluation of the structural differences between Sphnx and Kunst is included in its own section of
310 *Results and Discussion*.

311 In an effort to validate the *in vivo* observations confirming Sphnx functionality, we
312 analyzed the induction of P_{lac} driving the expression of GFP in MG1655 (pCKT-Sphnx,
313 pHC_DYOLacI-R) by qRT-PCR experiments (*Supp. Figure 4)* performed as indicated in
314 *Materials and Methods*. The expression levels of *sfgfp* calculated as the ratio of the induced
315 condition over the repressed condition were 1.58 ± 0.24 (n = 8, error expressed as SEM),
316 statistically significant and consistent with an induction of the expression in the presence of
317 Neu5Ac.

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

330 in order to calculate the kinetics of the Sphnx-*lacO* and Sphnx-Neu5Ac-*lacO* interactions, yet they
331 represent a promising step forward towards characterizing the interaction between our synthetic
332 TF and its cognate operator.

333 At this juncture of the project, both *in vivo* and *in vitro* experiments strongly suggested that
334 Sphnx was a transcriptional repressor capable of interacting with *lacO* and being de-repressed by
335 neuraminic acid. To better understand the dynamics of the Sphnx-Neu5Ac interaction we analyzed
336 the production of GFP by MG1655 (pCKT-Sphnx, pHC_DYOLacI-R) cultures at 14 hours of
337 growth (*Materials and Methods*) when exposed to different concentrations of neuraminic acid
338 around the average salivary concentration for an adult male (≈1 mM)[66]. The mean fold difference
339 of GFP expression between the repressed and induced conditions was 2.25 ± 0.74 for cultures
340 induced with 0.5 mM Neu5Ac, 2.54 ± 0.52 for cultures induced with 1.0 mM Neu5Ac, 3.81 ± 1.15
341 for cultures induced with 2.0 mM Neu5Ac, and 5.21 ± 2.00 for cultures induced with 5.0 mM
342 Neu5Ac (n = 8, error expressed as SEM). These results were commensurable with those observed
343 for modular chimeric TFs in the presence by their ligands: $SLCP_{GL}$ (LacI-GGBP-OD) induced by
344 glucose[46], CbnR-ABE44898-OD and LmrR-BzdB1_nSP induced by benzoate[15]. *Figure 2c* shows
345 the levels of GFP fluorescence observed in basal and repressed cultures as well as in cultures
346 induced by 0.5, 1.0, 2.0 and 5.0 mM neuraminic acid. These assays corroborated that relative
347 fluorescence between basal (aTc^-/$Neu5Ac^-$) and repressed (aTc^+/$Neu5Ac^-$) conditions were
348 significant. Cultures in which the genetic circuit used to test Sphnx were de-repressed
349 (aTc^+/$Neu5Ac^+$) displayed statistically significant values compared to the repressed cultures when
350 in the presence of 1.0 mM and 2.0 mM Neu5Ac. Induction by 0.5 and 5.0 mM was not significant.
351 It is important to note that relative fluorescence values in the 5.0 mM-induced biological replicas
352 were widely dispersed, suggesting that an overabundance of Neu5Ac might interfere with the

353 stability of the host strain bearing both plasmids, producing some deleterious effect associated to
354 the expression of the heterologous Sphnx and GFP proteins in the assayed conditions, which in
355 practical means may entail the testing of a few sample dilutions when using these biosensors for
356 out-of-the-lab applications. These results showing Sphnx induction are promising, yet they
357 underline the need to refine chimeric TFs resulting from the assembly of pre-existing modules to
358 optimize their induction range as well as gene dose and reporter genes used to study their behaviour
359 *in vivo* and *in vitro*, as it has been previously proposed[14]. Newly created synthetic TFs tend to have
360 limited transfer function and specificity, and thus to require further engineering for their field
361 deployment in biosensors[67]. In this regard, we are optimistic about the future improvement of the
362 chimeras we have assembled: the estimated K_d of SiaP and SatA for neuraminic acid are in the nM
363 range[52,53,55] (calculated *in vitro*), whereas in this work we present data in the mM range (calculated
364 *in vivo*). This suggests that we have not reached the full potential of our LBDs and there is an
365 opportunity to optimize Sphnx's structure to better benefit from the high specificity and affinity
366 of our LBD. It is also important to note that Sphnx is currently at the same point in evolutionary
367 history when DBD-LacI was fused to the ancestral version of its current LBD, long before
368 communication between DBD and LBD had been refined and perfected by natural selection[45]. A
369 systematic exploration of promoter and RBS variants can provide a better solution to compensate
370 for some of these deficiencies[14]. Regarding the use of alternative reporters that may shed more
371 light on the induction profile of Sphnx, we acknowledge that the use of GFP has some inherent
372 limitations in prokaryotes (e.g. widely varying maturation times dependent on strain and metabolic
373 state[68]), which has motivated us to initiate the development of better reporting systems.
374 **Modeling of the 3D structure of chimeric transcription factors**. Another element we are
375 starting to explore is the importance of linkers for the functionality of modular proteins, an often-

376 neglected design element[69]. Since the beginning of our journey characterizing the chimeric TFs,
377 we found extremely intriguing the fact that Sphnx (LacI-LNK2-SiaP, N_t-DBD-EKEKEK-LBD-
378 C_t) and Kunst (LacI-LNK3-SiaP, N_t-DBD-GSGSGS-LBD-C_t), which are nearly identical proteins
379 (*Supp. Materials Appendix*), possessed opposed behaviours when interacting with neuraminic
380 acid. The recent release of the AlphaFold 3 software[70] has enabled us to predict the structures of
381 multimeric chimeric proteins in the presence of their dsDNA operator, allowing us to model the
382 proteins under conditions closer to those found in actual WCBs. Sphnx and Kunst were both
383 modeled as dimers using AlphaFold 3, along with a DNA double helix containing the *lacO* (5'-
384 ATTGTGAGCGGATAACAATT-3' and its complementary sequence) recognized by the DBD-
385 LacI they both carry. Our goal was to predict structures that could help explain why Sphnx binds
386 Neu5Ac while Kunst does not, despite only differing by six amino acids. *Figure 3a* shows the
387 predicted Kunst DBD (green) bound to the DNA double helix (gray) and the formation of a dimer
388 between the PBP (cyan). The structure shows a mostly disordered linking region (orange),
389 including the GSGSGS (LNK3) linker (violet). The predicted Sphnx model (*Figure 3b*), on the
390 other hand, shows that the connecting region adopts a helical structure, including the EKEKEK
391 (LNK2) linker, and appears to form contacts between both the DBD and PBP. Rotating the Kunst
392 structure by 90° (*Figure 3c*) we can observe that the GSGSGS (linker does not interact with the
393 protein or DNA. In contrast, rotating the Sphnx structure by 90° (*Figure 3d*) further shows that the
394 linker forms contacts between the DBD and PBP domains, with interactions occurring across
395 different monomer chains. Further analysis shows that the helical linker in Sphnx forms
396 electrostatic salt bridges between itself and both the DBD and PBP domains. *Figure 3e* shows the
397 acidic residues on the linker (Glu) forming contacts with basic residues (Lys) on the DBD and
398 PBP domains. Specifically, the predicted structure suggests that the residues expected to form salt

399 bridges include Glu67-Lys2 on the DBD and Glu69-Lys152 on the PBP. These salt bridges likely
400 stabilize the helical linker, affecting the flexibility and stability of the polypeptide, an effect that
401 has been previously observed in other proteins[71]. From these predicted structures, we hypothesize
402 that these salt bridges stabilize helicity in the flexible linker in a way that positively affects the
403 PBP function by interacting with the DBD and the PBP, explaining the ligand binding functionality
404 of Sphnx compared to non-binding Kunst protein. Zooming out to see the larger structure (*Figure*
405 *3f*), aromatic residues near the linker salt bridges are visible. These residues are key for testing the
406 new predicted structures and designing new chimeric proteins in the future. We intend to purify
407 Sphnx and Kunst to perform fluorescence analysis of the protein-ligand interactions, including
408 fluorescence quenching and fluorescence anisotropy studies. The construction of 3D models, using
409 the latest version of AlphaFold, will continue to enable us to explore which mutations might
410 enhance ligand binding in a way that was previously unavailable, also allowing us to model
411 conditions closer to the actual multimeric *in vivo* conditions.
412
413
414 **CONCLUSIONS**
415 This work presents MG1655 (pCKT-Sphnx, pHC_DYOLacI-R), the first iteration of a
416 transcription factor-based whole-cell biosensor strain capable of detecting the salivary biomarker
417 N-acetylneuraminic acid and relaying its presence through the production of a fluorescent reporter,
418 the traditional litmus test for a biosensor circuit[14]. The keystone of the genetic circuit enabling the
419 detection of the soluble molecule is a new transcriptional repressor: Sphnx. The apparent absence
420 of a native Neu5Ac-binding transcription factor led the authors to create a custom-made regulator
421 via a straightforward synthetic biology approach that enables the generation of chimeras capable

422 of repressing a known promoter and being de-repressed by a molecule of interest. Neuraminic acid
423 is an attractive analyte to monitor since its enhanced levels in saliva are correlated with conditions
424 that range from periodontal disease[66] (oral microbes harvest sialic acid to evade our immune
425 system[31]) to the onset of oral squamous cell carcinoma[32–34]. To the best of our knowledge, Sphnx
426 is the first published transcriptional repressor capable of being induced by neuraminic acid.

427 This study supports the idea that the ability to construct custom-made TFs, affordably and
428 on-demand, is instrumental in expanding the panoply of small soluble molecules that can be
429 detected by biosensors in a field limited by the availability of sensing modules[14,45]. It is important
430 to note that synthetic TFs such as Sphnx do not represent the closing of a chapter, but the origin
431 of a new source of regulators whose performance can be improved via directed engineering of
432 their ligand binding pockets[72,73], an approach that can now benefit from high-throughput
433 approaches[74]. It is widely acknowledged that biosensor specificity and biosensor-analyte reaction
434 curves need to be optimized for the practical application of the WCB[67]. Once the biosensor strain
435 is optimized we will test it first on synthetic saliva preparations[75] with a gradient of Neu5Ac
436 concentrations and, eventually, with clinical salivary samples. The MG1655 (pCKT-Sphnx,
437 pHC_DYOLacI-R) strain represents the initial step in the obtention of a functional WCB for the
438 detection of neuraminic acid, a necessary launchpad to achieve the goal of eventual clinical
439 deployment of the sensor.

440 The modular assembly of fusion genes such as Sphnx sheds light on the importance of
441 inter-domain communication for multi-domain protein functionality[76] and exemplifies how a
442 reduced number of independent modules can be the playground of natural selection to synergically
443 originate new functionalities in the cell[77]. The results presented in this work have challenged our
444 expectations regarding chimeric transcription factors: while we expected most of them to be able

445 to repress expression from their regulated promoter, we found several instances in which the
446 chimeric product was unable to bind to its cognate operator boxes. Our current hypothesis is that
447 different linkage alternatives condition the 3D structure of the multimodular protein to such an
448 extent that the dimerization step necessary to bind the operator boxes cannot be achieved. We are
449 looking forward to expanding these observations in the future. Our study presents the interesting
450 contrast between the inducibility of Sphnx as opposed to Kunst, and particularly the unexpected
451 role that a potential LBD-mediated stabilization of the linker may play in the de-repressibility of
452 the TF dimer. *In vitro* assays using purified proteins, as well as further modeling simulation
453 analysis and, eventually, the crystallization of these chimeras will contribute to illuminate the
454 complex dynamics necessary for modular proteins to fulfill their function.

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

458

459

460

461

462

463

464 **MATERIALS AND METHODS**

465 **Bacterial strains and growth conditions**. Bacterial strains and plasmids used are listed in *Supp.*
466 *Table 2*. *E. coli* strains were grown at 37°C in LB (Lysogenic Broth[78]) medium unless otherwise
467 indicated. When required, *E. coli* cells were grown in M63 minimal medium[79] using the necessary
468 nutritional supplements and 30 mM glycerol (Sigma, St. Louis, MO) as a carbon source.
469 Antibiotics were added at the following concentrations: 100 $\mu g \cdot ml^{-1}$ ampicillin/carbenicillin and
470 25 $\mu g \cdot ml^{-1}$ chloramphenicol (Sigma, St. Louis, MO). For protein expression from pCKTRBS-
471 derived vectors, cultures were induced with 0.5 $\mu g \cdot ml^{-1}$ anhydrotetracycline (aTc) (Clontech,
472 Mountain View, CA). When experimental conditions required it, cultures were supplemented with
473 N-acetylneuraminic acid (neuraminic acid, Neu5Ac; Sigma, St. Louis, MO) at a 1 mM
474 concentration unless otherwise indicated.

475 Culture growth was routinely monitored in a BioPhotometer D30 (BioSpectrometer Basic Kinetic
476 and Fluorescence; Eppendorf, Hamburg, Germany). When it was required, bacterial growth and
477 GFP production were recorded in a Varioskan LUX Multimode Microplate Reader (Thermo
478 Fisher, Waltham, MA). Cultures were pipetted on a 96-well plate (Flat bottom Polystyrene Black
479 with clear bottom; Corning, Corning, NY) covered with a Breathe-Easy BEM-1 gas-permeable
480 membrane (Diversified Biotech, Dedham, MA) and incubated in a plate reader, where OD_{600} and
481 fluorescence in the range of GFP (excitation 485 nm, emission 528 nm) were either tracked along
482 time courses (up to 40 h) or subjected to point measurements. NEB5-alpha (pHC_DYOLacI-R)
483 and MG1655 were included in every plate as GFP-positive and GFP-negative reference strains,
484 respectively. For standard time course experiments, overnight precultures of the strains were used
485 to inoculate (starting OD_{600} 0.025) fresh growth media (supplemented with aTc and Neu5Ac when
486 indicated) in 96-well plates (180 µl total volume per well). For single point measurements,

487 overnight precultures of the strains (plus aTc and Neu5Ac when required) were used to inoculate
488 5 ml of fresh media supplemented in the same manner as the precultures. After 14 h of growth,
489 180 µl samples of the cultures were transferred to 96-well plates to enable OD_{600} and fluorescence
490 readings.
491 Statistical analysis as well as data visualization were performed with R (https://www.r-
492 project.org/) in the RStudio (RStudio Inc., Boston, MA) and Visual Studio Code environments
493 (Microsoft Corp., Redmond, WA).
494 **Molecular biology techniques.** Molecular biology techniques were performed following
495 commonly used standard protocols and as per manufacturers' instructions[80]. PCR reactions took
496 place in a 6321 Mastercycler PRO Vapo Protect Thermal Cycler (Eppendorf, Hamburg, Germany).
497 Routine separation of DNA and RNA in agarose gels was performed in an RunOne Electrophoresis
498 system (EmbiTec, San Diego, CA). Agarose gels were imaged in an UVP UVSolo Touch (Analytic
499 Jena, Jena, Germany). Plasmid DNA was purified with a Qiaprep Spin Miniprep Kit (Qiagen,
500 Hilden, Germany). DNA fragments were purified with DNA Clean-up and Concentration Kit
501 (Zymo research, Irvine, CA). The oligonucleotides employed for PCR amplification of the cloned
502 fragments and other molecular biology techniques are summarized in *Supp. Table 3* and were
503 supplied by IDT (Coralville, IA). All cloned inserts and DNA fragments were confirmed either via
504 Sanger sequencing[81] performed by Azenta Inc. (Burlington, MA), or Oxford-Nanopore whole-
505 plasmid sequencing performed by Plasmidsaurus (Eugene, OR). Commercially available NEB5-
506 alpha chemically competent cells (NEB, Ipswich, MA) were used for routine transformations.
507 Alternatively, electrocompetent *E. coli* cells were generated and transformed immediately (Gene
508 Pulser; Bio-Rad, Hercules, CA)[80]. Nucleotide sequence analyses were done at the National Center
509 for Biotechnology Information server (www.ncbi.nlm.nih.gov) and Benchling Biology Software

510 (www.benchling.com). Cloning was routinely performed via Gibson assembly[82] unless otherwise
511 indicated.
512 **Domain selection and cloning.** DBD-LacI is located on *lacI* 5' region, from the start codon to
513 base 198, encoding a peptide of 66 amino acids that comprises the N-terminus of our chimeric
514 proteins. This domain was obtained via PCR amplification from plasmid pCKTRBS-LacIwt[15]
515 using the oligonucleotide pair SV00001/SV00002. SatA from *H. ducreyi* (Uniprot Q7VL18) and
516 SiaP from *H. influenzae* (Uniprot P44542) were ordered as synthetic dsDNA fragments (IDT,
517 Coralville, IA). Several variants of the LBDs SatA and SiaP were amplified via PCR so that they
518 would include 5' overhangs including linkers (LNK2-5) or no linker whatsoever (LNK1). SatA
519 was amplified with primers SV00030-SV00034 / SV00004 and SiaP with primers SV00015-
520 SV00019 / JFJ0018.
521 Genes encoding the chimeric TFs were originally assembled in pUC19[83]. The vector was amplified
522 via divergent PCR with primers JFJ00019/JFJ00020, which introduce a consensus RBS for *E. coli*,
523 originating plasmid pUC19RBS. Tripartite Gibson assemblies including linear pUC19RBS, DBD-
524 LacI, and LNK-LBD were performed and the resulting constructions transformed into NEB5-alpha
525 competent cells. After a carbenicillin selection in solid media, a subset of the resulting clones was
526 isolated and their plasmids (pUC19-*Chimera* family, *Supp. Table 2*) purified and sequenced to
527 validate the presence of chimeric TFs. The resulting TF genes cloned into pUC19RBS were
528 amplified using SV00024/SV00025 for SiaP-derived chimeras and SV00024/SV00026 for their
529 SatA counterparts. These primers amplify our TF genes from the 5' DBD-LacI, just downstream
530 of the consensus RBS site, to the stop codon of the LBD. Low copy plasmid pCKTRBS[15] was
531 amplified via divergent PCR using primer pairs SV00022/SV00021 for SiaP-derived chimeras and
532 SV00022/SV00023 for their SatA counterparts. These primers introduce 5' and 3' overhang

533 sequences providing homology to our dsDNA chimeric gene fragments, and including in their 5'
534 side our consensus RBS site. Bipartite Gibson assemblies including linear pCKTRBS and the
535 fusion genes encoding the chimeras were performed, and the resulting assembly products
536 transformed into NEB5-alpha competent cells. After a carbenicillin selection in solid media,
537 resulting clones were isolated and their plasmids sequenced to confirm the presence of chimera-
538 encoding genes. Resulting pCKT-*Chimera* plasmids were purified from NEB5-alpha (pCKT-
539 *Chimera*) strains and electroporated into MG1655 cells. After confirming the presence of the
540 plasmid in the new host, MG1655 (pCKT-*Chimera*) were electroporated with the reporter plasmid
541 pHC_DYOLacI-R. Validated MG1655 (pCKT-*Chimera*, pHC_DYOLacI-R) strains were used for
542 *in vivo* and *in vitro* analysis of the functionality of the novel chimeric TFs when required. The
543 sequences of all the oligonucleotides used in this work can be found in *Supp. Table 3*.
544 **Electrophoretic mobility shift assay (EMSA).** In order to obtain cell extracts enriched in the
545 Sphnx/Siren proteins, NEB5-alpha (pCKT-Sphnx) and NEB5-alpha (pCKT-Siren) cells were
546 grown on LB supplemented with aTc for 5 hours at 37˚C in a shaker incubator. Cells were
547 recovered by centrifugation at 4°C (15' at 4,000 g) and then resuspended in 20 mM Tris-HCl (pH
548 7.0) to be sonicated at 70 amp for 30 seconds (30 cycles of 1 second; Branson Digital Sonifier,
549 Danbury, CN). A subsequent centrifugation at 4°C (20' at 20,000 g) was performed to eliminate
550 cell debris from the sample. The recovered supernatant was then incubated with the DNA probe,
551 a 251 bp dsDNA fragment amplified from plasmid pHC_DYOLacI-R with primers
552 SV00028/SV00029, containing P_{lac} (including *lacO*). Cell extract (5 µl) and DNA probe (5 µl)
553 were incubated together with reaction buffer (20% glycerol, 100 mM KCl, 500 $\mu g \cdot ml^{-1}$ bovine
554 serum albumin, 20mM Tris-HCl pH 7.5). The reaction mixture is incubated for 30' at 30 ˚C. After
555 incubation of the retardation mixtures, 2 µl of Gel loading dye without SDS (NEB, Ipswich, MA)

556 were added to each reaction. Mixtures were subsequently fractionated by electrophoresis in precast
557 6% polyacrylamide gels (Novex TBE Gels; Invitrogen, Waltham, MA) buffered with 0.5X TBE
558 (45 mM Tris borate, 1mM EDTA). Gels were run at constant voltage (100 V) for 90' in a Mini
559 Gel Tank (Invitrogen, Waltham, MA) and afterwards stained for 1 h in 50 ml of 0.5X TBE buffer
560 supplemented with 5 µl of SYBR Gold dye (Invitrogen, Waltham, MA) while gently shaken. Post
561 staining, gels were imaged in a UVP UVSolo Touch (Analytic Jena, Jena, Germany).
562 **qRT-PCR analysis.** Quantitative Real Time Polymerase Chain Reaction (qRT-PCR) was used to
563 assay the transcription of our reporter gene (*sfgfp*). MG1655 (pCKT-*Chimera*, pHCDYOLacI-R)
564 strains were precultured overnight in of 3 ml LB supplemented with the corresponding selection
565 antibiotics. On the next day, 100 µl of overnight preculture were used as inoculum for several
566 cultures, 10 ml each, of fresh LB (plus selection antibiotics). Each one of the cultures was
567 supplemented so that it would represent one of these three different conditions: *Basal* (aTc^-,
568 $Neu5Ac^-$), *Repressed* (aTc^+,$Neu5Ac^-$), and *De-repressed* (aTc^+, $Neu5Ac^+$). Where experimental
569 conditions required it, cultures were supplemented with 0.5 $\mu g \cdot ml^{-1}$ anhydrotetracycline (aTc)
570 (Clontech, Mountain View, CA) and 1 mM neuraminic acid (Neu5Ac; Sigma, St. Louis, MO). For
571 every independent qRT-PCR experiment, three biological replicates were created across each of
572 these three conditions. These cultures were placed in a shaker incubator at 37˚C for 5 hours and
573 pelleted at 4˚C and 4,000 g. Pellets were either frozen at -80˚C for future use or immediately
574 processed for total RNA extraction. Cells from each individual pellet were set for lysis by
575 resuspension in 1 ml of IBI reagent, a phenol and guanidine isothiocyanate mixture (IBI Scientific,
576 Dubuque, IA). 200 µl of chloroform (Fisher Scientific, Pittsburgh, PA) was added to the lysate
577 and the mixture was vigorously shaken to homogeneity, followed by centrifugation (benchtop
578 minifuge) at 4˚C and 14,500 rpm for 15 minutes that led to a phase separation. 500 µl of the upper

579 aqueous phase was mixed with an equal volume of 100% ethanol. At this point, total RNA was
580 purified from DNA/RNA phase with Direct-zol RNA Miniprep Kit (Zymo Research, Irvine, CA)
581 where the TRIzol reagent steps are replaced with the phenol and chloroform phase separation.
582 Total RNA eluted with ddH_2O was then retrotranscribed using qScript Flex cDNA Synthesis Kit
583 (QuantaBio, Beverly, MA). A volume totaling 1 µg RNA was incubated with random hexamers
584 and then supplemented with retrotranscriptase enzyme mix to synthesize cDNA. cDNA resulting
585 from the last step was used as a template for the amplification of a 256 bp fragment with
586 oligonucleotides SV00027A/SV00027 in a CFX96 Real-Time PCR System (Bio-Rad, Hercules,
587 CA). Primers had been designed to amplify a region of pHC_DYOLacI comprising the beginning
588 of *sfgfp*, so that when the tested chimera was being expressed it repressed the expression of GFP
589 unless Neu5Ac was present. Cq values were analyzed as a measure of gene expression using the
590 ΔΔCq ($2^{-\Delta\Delta Ct}$) method[84].
591 **Protein Structure Prediction using AlphaFold 3.** The 3D structures of the target proteins, Sphnx
592 (LacI-LNK2-SiaP, N_t-DBD-EKEKEK-LBD-C_t) and Kunst (LacI-LNK3-SiaP, N_t-DBD-
593 GSGSGS-LBD-C_t), were predicted using AlphaFold 3[70], developed by Google DeepMind
594 (London, UK). The predictions were performed using the publicly available AlphaFold 3 web
595 server (https://golgi.sandbox.google.com/). Both Sphnx and Kunst proteins were modeled as
596 dimers, along with the DNA double helix containing *lacO* (5'-ATTGTGAGCGGATAACAATT-
597 3' and its complementary sequence). Post-prediction, the structures were visualized and further
598 analyzed using VMD[85] to determine key protein-protein interactions.
599

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). 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., G.X.C., A.R., A.C.L., E.S.N., P.K., 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.

Competing interests

The authors declare no competing interests.

621 **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	-

622 **FIGURES**

Figure 1

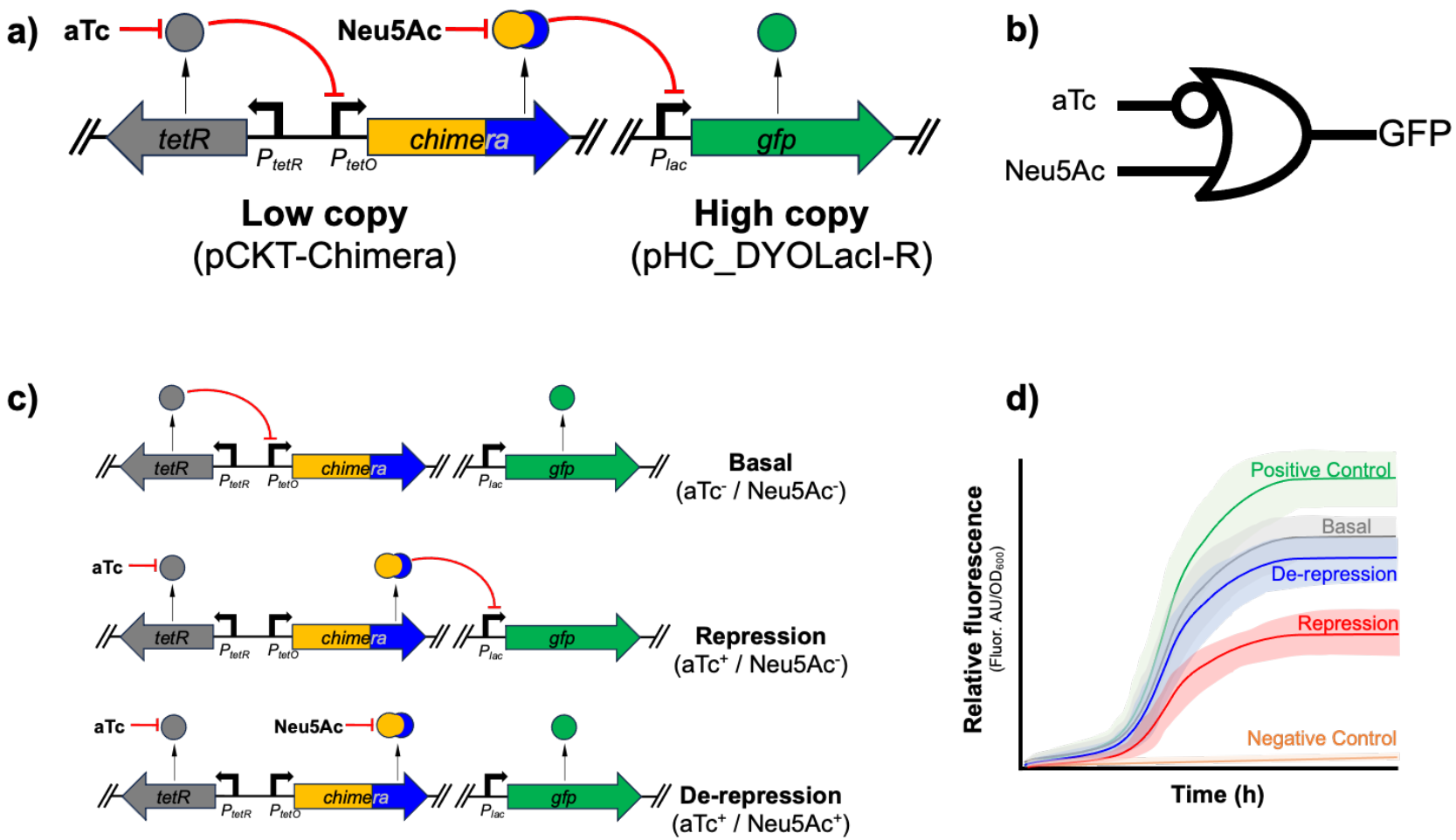
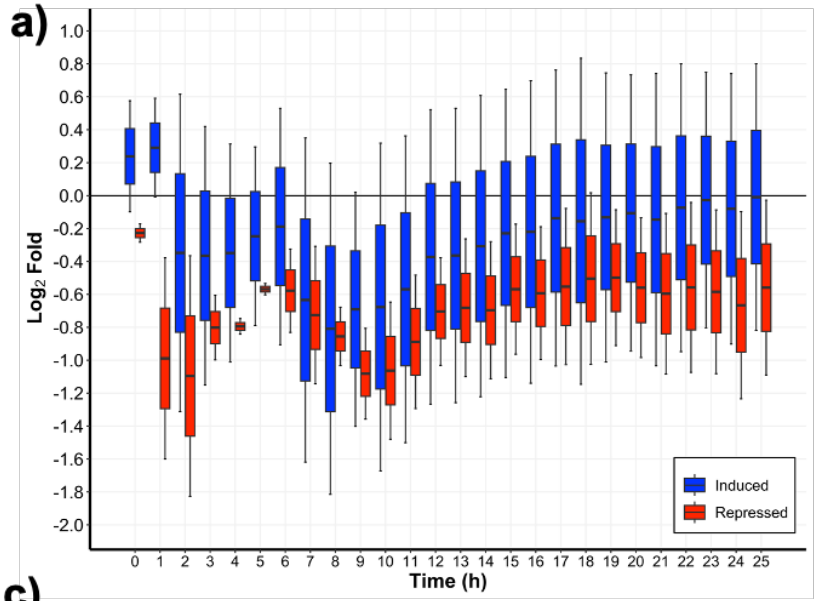
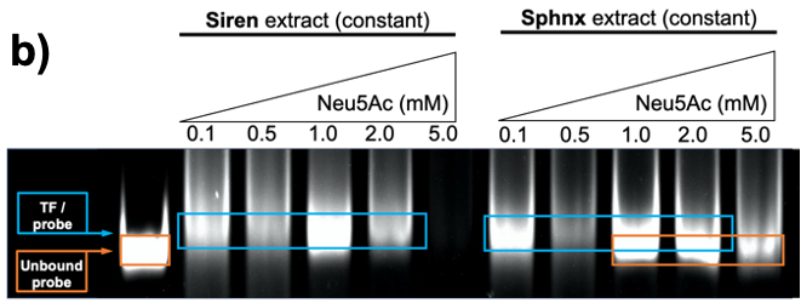


Figure 1. Screening strategy to find functional chimeric transcription factors. a) Biosensor genetic circuit for the *in vivo* testing of the functionality of cloned chimeric TFs. MG1655 (pCKT-*Chimera*, pHC_DYOLacI-R) cells express a chimeric TF under the control of TetR. The chimeric product can repress 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.

623

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

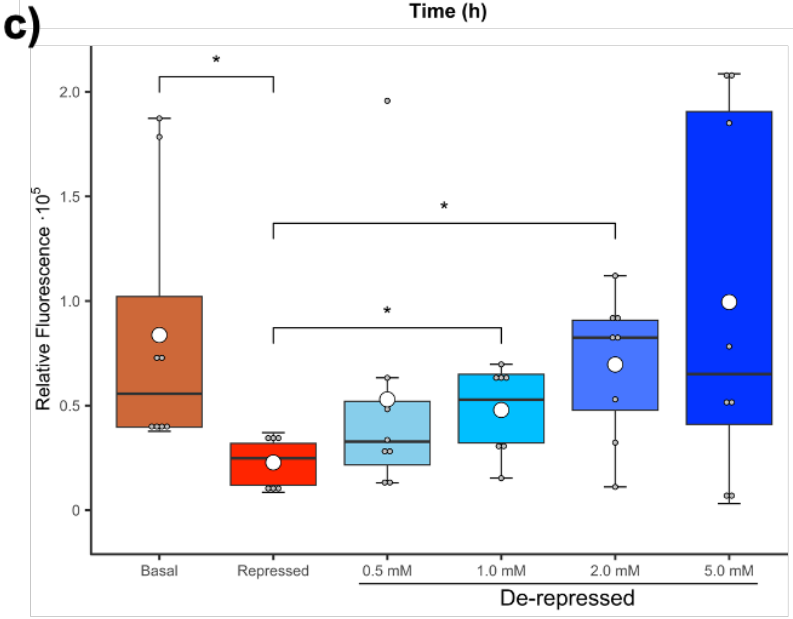

c)
Relative Fluorescence · 10^5
Basal
Repressed
0.5 mM
1.0 mM
2.0 mM
5.0 mM
De-repressed

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) 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 MG1655 (pCKT-Sphnx, pHC_DYOLacI-R) cultures after induction with different concentrations of neuraminic acid and growth in a multiwell plate reader for 14 hours. Promoter activity is measured as 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 *Methods*. Statistically significant differences between selected conditions are marked with stars (p value < 0.05).

625

Figure 3

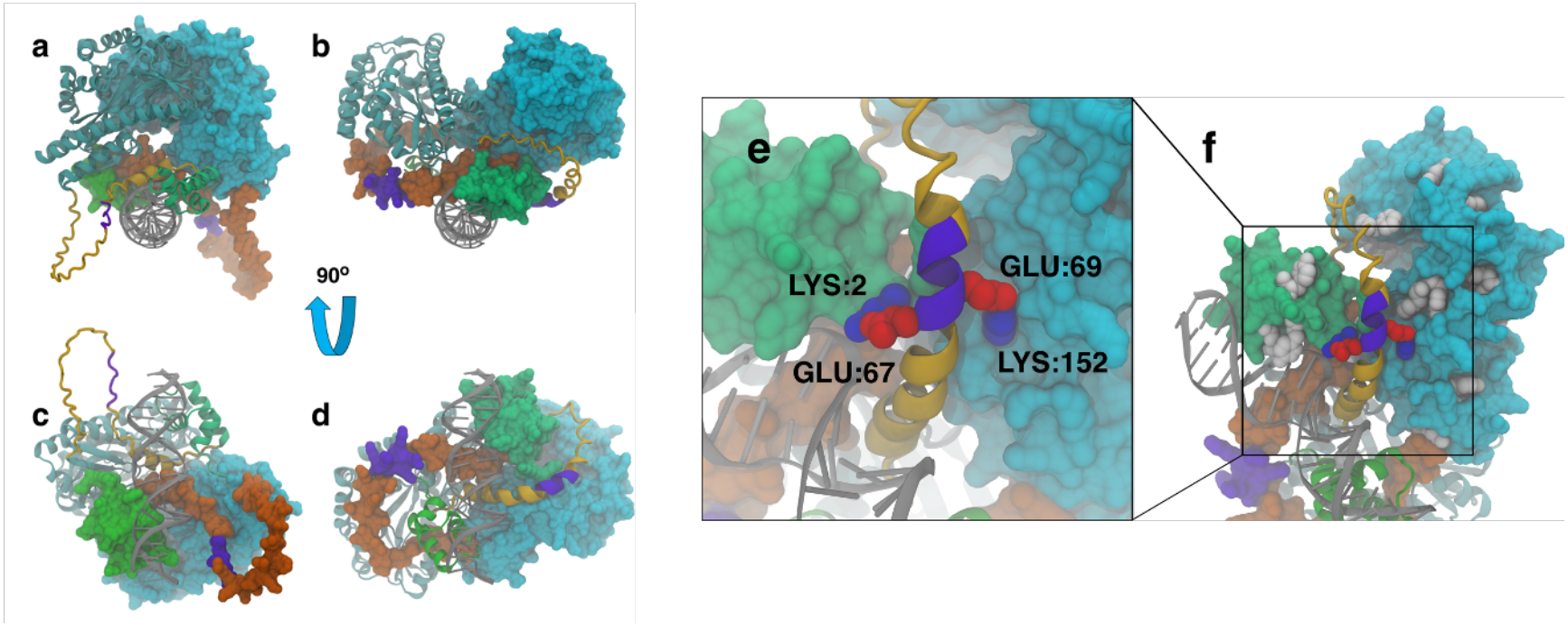


Figure 3. Modeled structures of chimeric protein dimers with double helix DNA obtained with Alphafold 3. For each structure, one monomer is shown using the New Cartoon representation and the other using the space-filling Surf representation. The DBD is shown in *green*, the PBP in *cyan*, and the connecting 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.

627 **BIBLIOGRAPHY**

628 1. Pancrazio, J. J., Whelan, J. P., Borkholder, D. A., Ma, W. & Stenger, D. A. Development and
629 application of cell-based biosensors. *Ann. Biomed. Eng.* **27**, 697–711 (1999).
630 2. Su, L., Jia, W., Hou, C. & Lei, Y. Microbial biosensors: a review. *Biosens. Bioelectron.* **26**,
631 1788–1799 (2011).
632 3. Moraskie, M. *et al.* Microbial whole-cell biosensors: Current applications, challenges, and
633 future perspectives. *Biosens. Bioelectron.* **191**, 113359 (2021).
634 4. O'Connor, G. *et al.* Whole-Cell Biosensors as Tools for the Detection of Quorum-Sensing
635 Molecules: Uses in Diagnostics and the Investigation of the Quorum-Sensing Mechanism.
636 *Adv. Biochem. Eng. Biotechnol.* (2015) doi:10.1007/10_2015_337.
637 5. Chang, H.-J., Voyvodic, P. L., Zúñiga, A. & Bonnet, J. Microbially derived biosensors for
638 diagnosis, monitoring and epidemiology. *Microb. Biotechnol.* **10**, 1031–1035 (2017).
639 6. Wu, Y., Wang, C.-W., Wang, D. & Wei, N. A Whole-Cell Biosensor for Point-of-Care
640 Detection of Waterborne Bacterial Pathogens. *ACS Synth. Biol.* **10**, 333–344 (2021).
641 7. Goode, J. A., Rushworth, J. V. H. & Millner, P. A. Biosensor Regeneration: A Review of
642 Common Techniques and Outcomes. *Langmuir* **31**, 6267–6276 (2015).
643 8. Zhang, F. & Keasling, J. Biosensors and their applications in microbial metabolic
644 engineering. *Trends Microbiol.* **19**, 323–329 (2011).
645 9. He, W., Yuan, S., Zhong, W.-H., Siddikee, M. A. & Dai, C.-C. Application of genetically
646 engineered microbial whole-cell biosensors for combined chemosensing. *Appl. Microbiol.*
647 *Biotechnol.* **100**, 1109–1119 (2016).
648 10. Saltepe, B., Kehribar, E. Ş., Su Yirmibeşoğlu, S. S. & Şafak Şeker, U. Ö. Cellular Biosensors
649 with Engineered Genetic Circuits. *ACS Sens* **3**, 13–26 (2018).

650 11. Berepiki, A., Kent, R., Machado, L. F. M. & Dixon, N. Development of High-Performance
651 Whole Cell Biosensors Aided by Statistical Modeling. *ACS Synth. Biol.* **9**, 576–589 (2020).
652 12. Schleif, R. DNA binding by proteins. *Science* **241**, 1182–1187 (1988).
653 13. Rogers, J. K. *et al.* Synthetic biosensors for precise gene control and real-time monitoring of
654 metabolites. *Nucleic Acids Res.* **43**, 7648–7660 (2015).
655 14. Tellechea-Luzardo, J., Stiebritz, M. T. & Carbonell, P. Transcription factor-based biosensors
656 for screening and dynamic regulation. *Front Bioeng Biotechnol* **11**, 1118702 (2023).
657 15. Juárez, J. F., Lecube-Azpeitia, B., Brown, S. L., Johnston, C. D. & Church, G. M. Biosensor
658 libraries harness large classes of binding domains for construction of allosteric transcriptional
659 regulators. *Nat. Commun.* **9**, 3101 (2018).
660 16. Koch, M., Pandi, A., Borkowski, O., Batista, A. C. & Faulon, J.-L. Custom-made
661 transcriptional biosensors for metabolic engineering. *Curr. Opin. Biotechnol.* **59**, 78–84
662 (2019).
663 17. Kaufman, E. & Lamster, I. B. The diagnostic applications of saliva--a review. *Crit. Rev. Oral*
664 *Biol. Med.* **13**, 197–212 (2002).
665 18. Lee, Y.-H. & Wong, D. T. Saliva: an emerging biofluid for early detection of diseases. *Am.*
666 *J. Dent.* **22**, 241–248 (2009).
667 19. Wren, M. E., Shirtcliff, E. A. & Drury, S. S. Not all biofluids are created equal: chewing over
668 salivary diagnostics and the epigenome. *Clin. Ther.* **37**, 529–539 (2015).
669 20. Khanna, P. & Walt, D. R. Salivary diagnostics using a portable point-of-service platform: a
670 review. *Clin. Ther.* **37**, 498–504 (2015).
671 21. Bandodkar, A. J. & Wang, J. Non-invasive wearable electrochemical sensors: a review.
672 *Trends Biotechnol.* **32**, 363–371 (2014).

673 22. Campuzano, S., Pedrero, M., Torrente-Rodríguez, R. M. & Pingarrón, J. M. Affinity-based
674 wearable electrochemical biosensors: Natural versus biomimetic receptors. *Anal. Sens.*
675 (2022) doi:10.1002/anse.202200087.
676 23. Kim, J., Campbell, A. S., de Ávila, B. E.-F. & Wang, J. Wearable biosensors for healthcare
677 monitoring. *Nat. Biotechnol.* **37**, 389–406 (2019).
678 24. Omar, E. Current concepts and future of noninvasive procedures for diagnosing oral
679 squamous cell carcinoma--a systematic review. *Head Face Med.* **11**, 6 (2015).
680 25. Wang, S., Yang, M., Li, R. & Bai, J. Current advances in noninvasive methods for the
681 diagnosis of oral squamous cell carcinoma: a review. *Eur. J. Med. Res.* **28**, 53 (2023).
682 26. Haines-Menges, B. L., Whitaker, W. B., Lubin, J. B. & Boyd, E. F. Host Sialic Acids: A
683 Delicacy for the Pathogen with Discerning Taste. *Microbiol Spectr* **3**, (2015).
684 27. Blix, G. Concerning the carbohydrate groups of submaxillary mucin. *Hoppe Seylers Z.*
685 *Physiol. Chem.* **240**, 43–54.
686 28. Lundblad, A. Gunnar Blix and his discovery of sialic acids. Fascinating molecules in
687 glycobiology. *Ups. J. Med. Sci.* **120**, 104–112 (2015).
688 29. Röhrig, C. H., Choi, S. S. H. & Baldwin, N. The nutritional role of free sialic acid, a human
689 milk monosaccharide, and its application as a functional food ingredient. *Crit. Rev. Food Sci.*
690 *Nutr.* **57**, 1017–1038 (2017).
691 30. Vajaria, B. N. *et al.* Salivary glyco-sialylation changes monitors oral carcinogenesis.
692 *Glycoconj. J.* **31**, 649–659 (2014).
693 31. Phansopa, C. *et al.* Characterization of a sialate-O-acetylesterase (NanS) from the oral
694 pathogen Tannerella forsythia that enhances sialic acid release by NanH, its cognate sialidase.
695 *Biochem. J* **472**, 157–167 (2015).

696 32. Arthisri, A. S., Sathiyamoorthy, A., Meenakshi, B. & Chandran, C. R. Ratio of Salivary Sialic
697 Acid to Fucose as Tumor Markers in Potentially Malignant Disorders and Oral Cancer.
698 *Contemp. Clin. Dent.* **11**, 131–135 (2020).
699 33. Chaudhari, V., Pradeep, G. L., Prakash, N. & Mahajan, A. M. Estimation of salivary sialic
700 acid in oral premalignancy and oral squamous cell carcinoma. *Contemp. Clin. Dent.* **7**, 451–
701 456 (2016).
702 34. Chittemsetti, S., Manchikatla, P. K. & Guttikonda, V. Estimation of serum sialic acid in oral
703 submucous fibrosis and oral squamous cell carcinoma. *J. Oral Maxillofac. Pathol.* **23**, 156
704 (2019).
705 35. Markopoulos, A. K. Current aspects on oral squamous cell carcinoma. *Open Dent. J.* **6**, 126–
706 130 (2012).
707 36. Coletta, R. D., Yeudall, W. A. & Salo, T. Grand Challenges in Oral Cancers. *Front Oral*
708 *Health* **1**, 3 (2020).
709 37. Warnakulasuriya, S. & Greenspan, J. S. *Textbook of Oral Cancer: Prevention, Diagnosis and*
710 *Management*. (Springer Nature, 2020).
711 38. Jacob, T. V. *et al.* A non-invasive study to estimate and compare salivary sialic acid level as
712 tumor marker in patients with pre-cancer and oral cancer. *J. Cancer Res. Ther.* **12**, 634–639
713 (2016).
714 39. Sanjay, P. R., Hallikeri, K. & Shivashankara, A. R. Evaluation of salivary sialic acid, total
715 protein, and total sugar in oral cancer: a preliminary report. *Indian J. Dent. Res.* **19**, 288–291
716 (2008).
717 40. Sridharan, G., Ramani, P., Patankar, S. & Vijayaraghavan, R. Evaluation of salivary
718 metabolomics in oral leukoplakia and oral squamous cell carcinoma. *J. Oral Pathol. Med.* **48**,

719 299–306 (2019).

720 41. Robson, B. Bioinformatics studies on a function of the SARS-CoV-2 spike glycoprotein as
721 the binding of host sialic acid glycans. *Comput. Biol. Med.* **122**, 103849 (2020).

722 42. Stanton, B. C. *et al.* Systematic transfer of prokaryotic sensors and circuits to mammalian
723 cells. *ACS Synth. Biol.* **3**, 880–891 (2014).

724 43. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol.*
725 *Biol.* **3**, 318–356 (1961).

726 44. Meinhardt, S. *et al.* Novel insights from hybrid LacI/GalR proteins: family-wide functional
727 attributes and biologically significant variation in transcription repression. *Nucleic Acids Res.*
728 **40**, 11139–11154 (2012).

729 45. Fukami-Kobayashi, K., Tateno, Y. & Nishikawa, K. Parallel evolution of ligand specificity
730 between LacI/GalR family repressors and periplasmic sugar-binding proteins. *Mol. Biol.*
731 *Evol.* **20**, 267–277 (2003).

732 46. Sakaguchi-Mikami, A., Taniguchi, A., Sode, K. & Yamazaki, T. Construction of a novel
733 glucose-sensing molecule based on a substrate-binding protein for intracellular sensing.
734 *Biotechnol. Bioeng.* **108**, 725–733 (2011).

735 47. Giuliani, S. E. *et al.* Environment sensing and response mediated by ABC transporters. *BMC*
736 *Genomics* **12 Suppl 1**, S8 (2011).

737 48. Fukamizo, T., Kitaoku, Y. & Suginta, W. Periplasmic solute-binding proteins: Structure
738 classification and chitooligosaccharide recognition. *Int. J. Biol. Macromol.* **128**, 985–993
739 (2019).

740 49. Locher, K. P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat. Struct.*
741 *Mol. Biol.* **23**, 487–493 (2016).

742 50. Post, D. M. B., Mungur, R., Gibson, B. W. & Munson, R. S., Jr. Identification of a novel
743 sialic acid transporter in *Haemophilus ducreyi*. *Infect. Immun.* **73**, 6727–6735 (2005).
744 51. Gruteser, N., Marin, K., Krämer, R. & Thomas, G. H. Sialic acid utilization by the soil
745 bacterium Corynebacterium glutamicum. *FEMS Microbiol. Lett.* **336**, 131–138 (2012).
746 52. Müller, A. *et al.* Conservation of structure and mechanism in primary and secondary
747 transporters exemplified by SiaP, a sialic acid binding virulence factor from *Haemophilus*
748 *influenzae*. *J. Biol. Chem.* **281**, 22212–22222 (2006).
749 53. Severi, E. *et al.* Sialic acid transport in *Haemophilus influenzae* is essential for
750 lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite
751 ATP-independent periplasmic transporter. *Mol. Microbiol.* **58**, 1173–1185 (2005).
752 54. Gangi Setty, T., Cho, C., Govindappa, S., Apicella, M. A. & Ramaswamy, S. Bacterial
753 periplasmic sialic acid-binding proteins exhibit a conserved binding site. *Acta Crystallogr. D*
754 *Biol. Crystallogr.* **70**, 1801–1811 (2014).
755 55. Gangi Setty, T. *et al.* Molecular characterization of the interaction of sialic acid with the
756 periplasmic binding protein from. *J. Biol. Chem.* **293**, 20073–20084 (2018).
757 56. Oertel-Buchheit, P., Reinbolt, J., John, M., Granger-Schnarr, M. & Schnarr, M. A LexA
758 mutant repressor with a relaxed inter-domain linker. *Protein Sci.* **7**, 512–515 (1998).
759 57. Tsirigotaki, A., De Geyter, J., Šoštaric, N., Economou, A. & Karamanou, S. Protein export
760 through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **15**, 21–36 (2017).
761 58. Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13**, 247–261
762 (1995).
763 59. Liu, Q., Schumacher, J., Wan, X., Lou, C. & Wang, B. Orthogonality and Burdens of
764 Heterologous AND Gate Gene Circuits in *E. coli*. *ACS Synth. Biol.* **7**, 553–564 (2018).

765 60. Logic Gates. *The Department of Mathematics and Computer Science at Oxford College of*
766 *Emory University* https://mathcenter.oxford.emory.edu/site/cs170/gates/.
767 61. Lewis, M. The lac repressor. *C. R. Biol.* **328**, 521–548 (2005).
768 62. Deans, T. L., Cantor, C. R. & Collins, J. J. A tunable genetic switch based on RNAi and
769 repressor proteins for regulating gene expression in mammalian cells. *Cell* **130**, 363–372
770 (2007).
771 63. Langdon, R. C., Burr, T., Pagan-Westphal, S. & Hochschild, A. A chimeric activator of
772 transcription that uses two DNA-binding domains to make simultaneous contact with pairs of
773 recognition sites. *Mol. Microbiol.* **41**, 885–896 (2001).
774 64. Durante-Rodríguez, G. *et al.* Identification of a missing link in the evolution of an enzyme
775 into a transcriptional regulator. *PLoS One* **8**, e57518 (2013).
776 65. Durante-Rodríguez, G., Mancheño, J. M., Díaz, E. & Carmona, M. Refactoring the λ phage
777 lytic/lysogenic decision with a synthetic regulator. *Microbiologyopen* **5**, 575–581 (2016).
778 66. Oktay, S. *et al.* Serum and saliva sialic acid in periodontitis patients with and without
779 cardiovascular disease. *Pathophysiol. Haemost. Thromb.* **37**, 67–71 (2010).
780 67. Snoek, T. *et al.* Evolution-guided engineering of small-molecule biosensors. *Nucleic Acids*
781 *Res.* **48**, e3 (2020).
782 68. Hebisch, E., Knebel, J., Landsberg, J., Frey, E. & Leisner, M. High variation of fluorescence
783 protein maturation times in closely related *Escherichia coli* strains. *PLoS One* **8**, e75991
784 (2013).
785 69. Chen, X., Zaro, J. L. & Shen, W.-C. Fusion protein linkers: property, design and functionality.
786 *Adv. Drug Deliv. Rev.* **65**, 1357–1369 (2013).
787 70. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold

788 3. *Nature* (2024) doi:10.1038/s41586-024-07487-w.

789 71. Meuzelaar, H., Vreede, J. & Woutersen, S. Influence of Glu/Arg, Asp/Arg, and Glu/Lys Salt
790 Bridges on α-Helical Stability and Folding Kinetics. *Biophys. J.* **110**, 2328–2341 (2016).

791 72. Raman, S., Taylor, N., Genuth, N., Fields, S. & Church, G. M. Engineering allostery. *Trends*
792 *Genet.* **30**, 521–528 (2014).

793 73. Malisi, C. *et al.* Binding pocket optimization by computational protein design. *PLoS One* **7**,
794 e52505 (2012).

795 74. Lerma Romero, J. A. *et al.* Binding pocket stabilization by high-throughput screening of yeast
796 display libraries. *Front Mol Biosci* **9**, 1023131 (2022).

797 75. Levine, M. J. Development of artificial salivas. *Crit. Rev. Oral Biol. Med.* **4**, 279–286 (1993).

798 76. Persico, M. *et al.* From Protein Communication to Drug Discovery. *Curr. Top. Med. Chem.*
799 **15**, 2019–2031 (2015).

800 77. Long, M., Betrán, E., Thornton, K. & Wang, W. The origin of new genes: glimpses from the
801 young and old. *Nat. Rev. Genet.* **4**, 865–875 (2003).

802 78. Bertani, G. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J.*
803 *Bacteriol.* **186**, 595–600 (2004).

804 79. Miller, J. H. *Experiments in Molecular Genetics*. (1972).

805 80. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual*. (CSHL Press,
806 2001).

807 81. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors.
808 *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463–5467 (1977).

809 82. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases.
810 *Nat. Methods* **6**, 343–345 (2009).

811 83. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host
812 strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
813 84. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time
814 quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).
815 85. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.*
816 **14**, 33–8, 27–8 (1996).

817