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 detection systems resulting from an analytical approach in which biological
35 elements convert a physicochemical stimulus into a detectable signal[1]. Among them, Whole Cell
36 Biosensors (WCBs) - detection systems in which a physiologically active cell sustains an array of
37 biomolecular elements (receptors, channels, enzymes, regulators, etc.) capable of detecting and
38 processing a signal and produce a detectable output - stand out due to their versatility and
39 engineering potential[1]. Microbial WCBs in particular have shown their potential in fields ranging
40 from environmental monitorization and food testing, to medical applications[2]. Microorganisms
41 can grow rapidly in relatively inexpensive media and produce all the necessary biological elements
42 for analyte detection. Unlike conventional sensors that may require large and complex instruments
43 and laborious sample processing, WCBs can detect a bioavailable substance directly from the
44 environment with minimal preparation, enabling their incorporation into *in situ* detection methods
45 that can revolutionize point-of-care early detection diagnosis[3].

46 WCBs benefit from the synergies resulting from the combination of native systems,
47 evolved during millions of years to detect the miscellany of molecules surrounding microbes, with
48 the ever growing toolkit that synthetic biology provides (such as the creation of synthetic genetic
49 circuits[4–7]). Frequently, the presence of a soluble analyte becomes associated with a
50 conformational change of a protein, so that its binding state to DNA gets altered, modifying the
51 transcription profile of a regulated reporter[8]. These transcription factor (TF)-based WCBs benefit
52 the most from the use of monogenic transcriptional repressors[4,5], control systems encoded by a
53 single gene whose product prevents transcription in the absence of a small soluble inducer
54 molecule. The catalog of characterized TFs capable of interacting with small soluble molecules is
55 expanding yet still limited[9]. The advent of synthetic biology has enabled us to better understand

56 protein modularity such that we can better engage in regulator customization and *de novo* creation,
57 to expand the panoply of molecules that can be detected[10,11]. In this paper, we centered our efforts
58 on the detection of a salivary biomarker that, to the best of our knowledge, was not the inducer of
59 any characterized TF.

60 Human saliva contains a vast array of metabolites informative of oral and systemic health[12],
61 making its continuous sampling by benchtop, portable, and wearable equipment extremely
62 coveted. A dental practice able to utilize such sampling technologies may increase its role at the
63 forefront of early disease prevention and diagnosis. Electrochemical intraoral biodevices capable
64 of detecting salivary biomarkers presently exist, yet they are limited in the scope of molecules they
65 detect and lack adaptability to sense new ones[13]. Bioengineering solutions enabling point-of-care
66 diagnostics will greatly benefit patients by improving at-home health monotorization[14]. Diagnostic
67 screenings that do not require invasive procedures (i.e. a blood sample or a biopsy) are preferred
68 by patients, which makes saliva a prime biological fluid candidate to be sampled for health and
69 disease markers[15,16]. In this work we focused our attention on N-acetylneuraminic acid (neuraminic
70 acid, Neu5Ac), the most common of the sialic acids originally isolated from salivary glands[17–19].
71 Neu5Ac fulfills key roles in human physiology[20] as well as in commensal and pathogenic
72 bacteria[17]. Moreover, changes in salivary sialylation might become a suitable early indicator of
73 changes associated with malignant pathologies, given their correlation to periodontal disease (PD)
74 and oral squamous cell carcinoma (OSCC)[21]. The oral microbiome of patients with PD include
75 bacteria such as *Tannerella forsythia*, which engage in glycan harvesting behavior: they cleave
76 Neu5Ac molecules from glycoproteins to stealthily evade their host's immune system[22].
77 Additionally, abnormally high concentrations of Neu5Ac have been associated with the onset of
78 oral cancer[23,24]: a group of neoplasms which affect any region in the oral cavity, pharynx, and

79 salivary glands[25]. Oral cancer is often used interchangeably with oral squamous cell carcinoma
80 (OSCC), which accounts for more than 90% of oral malignant epithelial neoplasms[25]. As one of
81 the most prevalent cancers in the world[26], untreated OSCC has very poor prognosis and, as such,
82 early detection is essential for successful treatment remission[27]. Recently, it has also been proposed
83 that salivary Neu5Ac could interfere with SARS-Cov-2 (COVID-19) infection, acting as a
84 competitive inhibitor blocking the viral antigens recognized by the host receptors[28]. Each of these
85 factors made quantification of salivary Neu5Ac a desirable target for detection by biosensors.

86 The absence of a native TF capable of detecting Neu5Ac motivated us to construct a
87 synthetic monogenic repressor for its integration in a synthetic circuit that would enable the strain
88 bearing it to behave as a WCB. We favor these regulators due to their simplicity, compact nature,
89 and ease of integration into synthetic circuits[4,5,29]. The process of constructing our custom
90 Neu5Ac-sensing regulators is based on streamlined gene fusion / chimeragenesis methods
91 developed by our laboratory to detect soluble molecules[10] . Here, we describe a Neu5Ac TF-based
92 WCB that employs engineered bacterial machinery to fulfill the functions of the three basic
93 modules present in any detection system: a sensing module, a processing unit, and an actuator. The
94 living WCB can intake a Neu5Ac signal by expressing a novel synthetic transcription factor,
95 Sphnx, that bridges detection and processing. A separate reporter plasmid completes the system
96 providing a gene encoding a fluorescent protein (sfGFP, henceforth stylized as GFP) whose
97 expression is regulated by the synthetic TF Sphnx so that detectable changes in fluorescence occur
98 upon the detection of soluble Neu5Ac. Sphnx is a member of a new combinatorial library of TFs
99 assembled via a simplified chimeragenesis process. Important clues on inter-domain
100 communication of modular proteins have started to be gathered by comparing Sphnx to non-
101 functional members of the TF collection. Protein tertiary structure modeling of functional and non-

102 functional regulators has contributed to establish comparisons between them helping to elucidate
103 intricate and frequently difficult to predict dynamics that may condition the success or failure of
104 any chimeric TF.
105
106
107 **RESULTS AND DISCUSSION**
108 **Design and Construction of Synthetic Transcription Factors.** The functionality of the WCBs
109 that we have built in this work hinges on the obtention of a TF capable of binding Neu5Ac with
110 efficacy and precision. The Neu5Ac-induced monogenic transcriptional repressor we intended to
111 assemble included three modules: a N-terminus (N_t) DNA Binding Domain (DBD), a linker
112 sequence (LNK), and a C-terminus (C_t) Ligand Binding Domain (LBD). The DBD specifically
113 recognizes an operator sequence in the promoter it regulates, driving the expression of the reporter
114 gene (*sfgfp*). The LBD is tasked with the recognition of Neu5Ac, providing specificity to the TF.
115 Finally, the LNK sequence act as a bridge, providing the right amount of flexibility / rigidity
116 necessary for information transfer between DBD and LBD: steric changes in the LBD associated
117 with Neu5Ac binding get translated into steric changes in the DBD effecting a change of its
118 oligomeric state that results in the unbinding of the TF from its operator and thus in a de-repression
119 of transcription. By using a gene fusion / chimeragenesis approach we generated a library of
120 repressors with a common N_t-DBD-LNK-LBD-C_t architecture.
121 Our selected DBD was sourced from the paradigm of transcriptional regulators: the *lac*
122 repressor or LacI[30]. Though naturally induced by lactose / IPTG, we exclusively employ the DBD
123 of LacI recognizing the operator *lacO* present in the P_{lac} promoter (as defined in[10]). We decided to
124 use DBD-LacI because the *lac* repressor originated as a modular protein with two domains, and

125 has been previously used in the construction of several functional chimeras[31–33]. It is important to
126 remark that the exclusive use of DBD-LacI, as opposed to the full protein, ensures that chimeric
127 proteins carrying this domain will recognize their target operator, yet avoid responding to native
128 inducers.

129 Candidate LBD selection is accomplished through bibliographical research. Since the LBD
130 is the domain that recognizes the molecule of interest for which we build the biosensing strain, it
131 can be sourced from any whole protein or protein domain as opposed to being co-opted from a
132 previously documented regulator. One of our main sources of the LBDs we employ in our
133 chimeragenesis pipeline are periplasmic binding proteins (PBPs)[34]. These are standalone subunits
134 of transporters that share a common ancestor with the LBD domains of the LacI / GalR family of
135 regulators[32]. To build the synthetic TF we required for this project, we needed to identify LBD
136 candidates that strongly bind our soluble molecule of interest (Neu5Ac). PBPs SatA and SiaP were
137 selected for their binding affinity and specificity for the transport of Neu5Ac. SatA is a product of
138 the *satABCD* (<u>s</u>ialic <u>a</u>cid <u>t</u>ransport) operon that encodes an ATP Binding Cassette (ABC)
139 transporter[35] from *Haemophilus ducreyi*[36]. An alternative *satABCD* system present in
140 *Corynebacterium glutamicum* has been demonstrated to be essential for its growth using Neu5Ac
141 as sole carbon source[37]. SiaP is a product of the *siaPQM* operon, which encodes a Tripartite ATP-
142 independent periplasmic (TRAP) transporter in *Haemophilus influenzae*[38,39]. SiaP was the first
143 TRAP-associated PBP whose structure was elucidated (*Vibrio cholerae*, *Pasteurella multocida*
144 and *Fusobacterium nucleatum*[38,40]). Structural comparisons and thermodynamic studies of SatA
145 and SiaP suggest that similar affinities for Neu5Ac are achieved in the two PBPs through distinct
146 mechanisms: one enthalpically driven and the other entropically driven. SiaP, as well as other
147 TRAP transporters, are hypothesized to command an enthalpically driven ligand binding strategy

148 due to the abundance of charged residues (Arg, Lys, Asp, and Glu) at their binding pocket. SatA,
149 alternatively, is hypothesized to bind its substrate in what is considered an entropic approach
150 because of the numerous hydrophobic and polar residues (Tyr, Phe, Gly, Leu, Asn, and Ser) at its
151 binding pocket[41]. The structure of *H. influenzae* SiaP in the presence of Neu5Ac (PDB 3B50)
152 reveals the ligand bound in a deep cavity with its carboxyl group forming a salt bridge with a
153 highly conserved Arg residue[38]. Initial studies of SiaP isolated from *H. influenzae* displayed a
154 binding K_d of ≈120 nM when affinity for Neu5Ac was tested[39]. Subsequent studies established a
155 value closer to K_d ≈58 nM[38]. Similar binding affinity of SatA complexes has been identified with
156 a K_d ≈133 nM[41]. The collection of empirical data supporting the ability of SatA and SiaP to bind
157 Neu5Ac made them ideal candidates to become the LBD of a chimeric TF responsive to Neu5Ac.
158 The final piece of the chimeric gene puzzle consisted on establishing a connection between
159 DBD and LBD domains. This task was performed by elements called linker (LNK) regions that
160 tend to be less structured than the domains they connect[42]. In order to maximize the chances of
161 transmitting the information of Neu5Ac binding to the DBD, so that it can detach from *lacO* and
162 enable transcription, we decided to connect both domains via two different strategies: directly
163 without a true LNK (condition identify as LNK1 for uniformity purposes) or by one out of four
164 LNKs characterized by having different biophysical properties (identified as LNK2 to LNK5,
165 *Supp. Table 1*). It is important to note that while LNK1 does not contain a true linker, the LBDs
166 we chose, by virtue of being PBPs, are proteins that would be exported to the periplasm of the cell
167 under natural circumstances, and thus contain a native signal peptide at their N-terminus. This kind
168 of peptide appears to be unstructured for the most part, barring a short alpha-helix region[43]. We
169 decided to include the signal peptides of both SiaP and SatA due to existent examples of functional
170 chimeras containing similar structures, such as the glucose-inducible $SLCP_{GL}$[33] and the benzoate-

171 inducible CbnR-ABE44898-OD[10] TFs. At this stage, we had identified all the elements we needed
172 to start assembling our new chimeric TFs.

173 dsDNA fragments encoding the DBD and all LNK-LBD variants were obtained as detailed
174 in the *Materials and Methods* section. To initiate the construction of a library containing all the
175 desired chimeric TFs, we performed several Gibson assembly reactions with all the domain-
176 encoding fragments and linearized expression vector pUC19RBS, followed by transformation in
177 NEB5-alpha competent cells. Resulting clones were isolated and individually sequenced. This
178 approach was performed twice. Chimeric combinations that were not present in the libraries were
179 re-attempted to be individually cloned five times with no success, at which point we proceeded to
180 put pause to their construction and move the project forward, subcloning the TF collection we had
181 generated up to that point in time into the low-copy vector pCKTRBS. It is important to note that
182 the expression of any heterologous protein is *per se* taxing on bacterial metabolism, and that TFs
183 might be especially burdensome[44,45]. Chimeric constructions including DBD-LacI and SiaP / SatA
184 seem to provoke a certain toxicity to the cell, albeit to different degrees depending on the specific
185 domain combination. Indirect evidence supporting this hypothesis was observed when
186 constructions expressing the chimeric TFs were transformed into *E. coli* MC4100 to assess their
187 behavior in M63 minimal media with glycerol as the sole carbon source. Under these resource-
188 limited conditions, stressed bacteria have more difficulties in thriving, and MC4100 cells bearing
189 pCKT-*Chimera* vectors displayed a very modest and erratic growth consistent with our
190 assumptions (data not shown). *Table 1* contains the chimeric TFs built and assayed presented in
191 this study, as well as a simplified interpretation of their functionality when grown in rich media.

192 **Screening of transcription factors**. A functional *in vivo* screening was implemented using GFP
193 as reporter. Plasmid pHC_DYOLacI-R (*Supp. Table 2*) was transformed into candidate strains

194 expressing the chimeric TFs from pCKTRBS vectors under the control of TetR-aTc. This
195 constitutes a genetic circuit (*Figure 1a*) that enables a controlled expression of the chimera as well
196 as its controlled induction. This circuit behaves as a B-or-not-A universal two input gate[46] (*Figure*
197 *1b*), where there are two inputs (A, aTc supplementation; B, Neu5Ac supplementation) and one
198 output (detection of GFP fluorescence). Based on the addition of the inputs we can define three
199 states in the genetic circuit (*Figure 1c*): *Basal* (aTc^-/$Neu5Ac^-$), no input added, GFP is expressed
200 from its promoter in pHC_DYOLacI-R; *Repression* (aTc^+/$Neu5Ac^-$), the presence of aTc induces
201 TetR which is incapable of remaining attached to P_{tetO} and thus the chimera is expressed from its
202 pCKT vector repressing the expression of GFP; *De-repression* (aTc^+/$Neu5Ac^+$), the chimera is
203 expressed but if it is functional it is induced by Neu5Ac, de-repressing the promoter driving the
204 expression of GFP. By comparing the behavior of the strain expressing the aforementioned circuit
205 under these conditions, we can rule out non-functional synthetic TFs. In the case that the chimeric
206 protein was not a functional repressor, GFP-associated fluorescence could be detected in the
207 *Repression* state. On the other hand, if the chimera was able to bind to *lacO* on the reporter
208 promoter yet incapable to recognize Neu5Ac with its LBD, or to transmit a LBD to DBD
209 conformational change resulting in its detachment from the operator, we would have obtained a
210 functional super-repressor[47] unable to respond to its inducer and, thus, not a valid tool to detect
211 Neu5Ac. *Figure 1d* illustrates the expected GFP production along a growth curve for a strain
212 expressing the screening genetic circuit that included an ideal chimera. It is important to note that
213 even though Neu5Ac is a monosaccharide, its supplementation to growth media did not
214 significantly increase the OD_{600} of the assayed strains, at least up to the tested 10 mM concentration
215 (*Supp. Figure 1*). Regardless of this observation, GFP-associated fluorescence data displayed in
216 this publication is always expressed as relative fluorescence (arbitrary units) corrected by the

217 OD_{600} of the culture, in order to account for growth disparities among different strains and
218 conditions. It is also important to note that repression by LacI, or LacI-derived chimeras, is rarely
219 absolute so that a certain level of leakage, and thus of background reporter gene expression, is
220 always expected to some extent[33,48].

221 All the chimeric TFs included in *Table 1* were screened for their ability to bind DNA and
222 thus to repress the expression of the GFP reporter, as well as to be induced by Neu5Ac enabling
223 production of the fluorescent protein. All members of the collection of MG1655 (pCKT-*Chimera*,
224 pHC_DYOLacI-R) strains bearing the genetic circuit described in *Figure 1* were tested *in vivo* to
225 assess their ability to behave as Neu5Ac biosensors. Among the chimeric proteins expressed by
226 this strain collection we include the TF Lumos, which did not contain any of the expected no-
227 linker / linker options but an unexpected 6 amino acid linker (LNK6) resulting from an abnormal
228 assembly reaction in which the last 18 bases of *DBD-lacI* were duplicated (*Supp. Table 1*). Given
229 the stable nature of the amino acid string, as well as our agnostic approach to which LNK sequence
230 could be the best performing one, we decided to maintain this construction in our pool. TFs Kunst
231 (LacI-LNK3-SiaP), Bastet (LacI-LNK5-SiaP), Harpy (LacI-LNK1-SatA), and Haus (LacI-LNK2-
232 SatA) showed no de-repression abilities when Neu5Ac was supplemented. Further assaying their
233 behavior, our attention was captured by Kunst. *Supp. Figure 2* shows the behavior of Kunst, which
234 only exhibits a tendency to repress P_{lac} at the very beginning of a 40 h time course, yet after the 3
235 h mark does not show any difference between induced and uninduced conditions. A preliminary
236 evaluation of the structural differences between Sphnx and Kunst is included in its own section of
237 *Results and Discussion*. Another TF that presented an unexpected induction profile was Siren
238 (LacI-LNK1-SiaP), which displayed an intriguing function as a proper repressor inducible by
239 Neu5Ac but only in the stationary phase of growth (*Supp. Figure 3*). In the future we intend to

240 study this protein in depth, since its repression / de-repression profile is very attractive for a whole-
241 cell biosensor setup in which a bacterial culture sustaining itself at stationary phase is continuously
242 monitoring an input for an extended period of time. The TF Lumos (LacI-LNK6-SiaP) provided
243 some preliminary data supporting its functionality, while a candidate biosensor strain expressing
244 Sphnx (LacI-LNK2-SiaP) exhibited the most suitable repression and induction profile compatible
245 with that of a TF capable of binding to Neu5Ac. Even though our sample size is extremely reduced
246 - judging by the difficulty to obtain them as well as by their lack of functionality- SatA-based
247 fusions seem to be less stable and more prone to Neu5Ac non-responsiveness. Further studies and
248 the creation of bigger chimeric libraries, however, are required to confirm this hypothesis. At this
249 injunction, we focused our efforts on better characterization of the most promising TF we had
250 assembled: Sphnx.

251 In order to validate our preliminary observations, we systematically compared GFP
252 production over time of MG1655 (pCKT-Sphnx, pHC_DYOLacI-R) cells expressing Sphnx either
253 in the presence or absence of Neu5Ac. *Figure 2a* displays the results of time course experiments
254 over a period of 25 hours. GFP production induced by the presence of Neu5Ac was observed from
255 the first hour of growth, with the Neu5Ac-induced cultures showing a tendency to display higher
256 fluorescence than repressed ones.

257 As a away to confirm these *in vivo* observations, we analyzed the induction of P_{lac} driving
258 the expression of GFP in MG1655 (pCKT-Sphnx, pHC_DYOLacI-R) by qRT-PCR experiments
259 performed as indicated in *Materials and Methods*. The expression levels of *sfgfp* calculated as the
260 ratio of the induced condition over the repressed condition were 1.58 ± 0.24 ($n = 8$, error expressed
261 as SEM) statistically significant and consistent with an induction of the expression in the presence

262 of Neu5Ac (*Supp. Figure 4a)*. However, the degree of induction varied widely across biological
263 replicas (*Supp. Figure 4b)*.

264 Further *in vitro* experiments were also performed to confirm the interaction between P_{lac}
265 and the DBD-LacI of Sphnx. Electrophoretic mobility shift assays (EMSA, *Materials and*
266 *Methods*) showed that cell extracts from NEB5-alpha (pCKT-Sphnx) grown for 5 h were capable
267 of retarding the migration of a P_{lac}-*sfgfp* probe compared to control extracts that were not
268 expressing Sphnx (data not shown). Neu5Ac behaved as the inducer of Sphnx, since binding of
269 the protein to the probe was diminished in a concentration-dependent manner in the presence of
270 the sialic acid (*Figure 2b*). On the other hand, cell extracts from NEB5-alpha (pCKT-Siren) grown
271 for 5 h were able to bind the probe but not to unbind in the presence of Neu5Ac, confirming our
272 *in vivo* observations. Previous publications report a similar behaviour: DBDs from bacterial
273 transcriptional repressors tend to retain their ability to bind to their operator sequences even when
274 integrated in a new chimeric TF[49–51], another testimony to the high modularity of these proteins.
275 Our initial *in vitro* DNA-protein interaction studies require future validation with purified chimeras
276 in order to calculate the kinetics of the Sphnx-*lacO* and Sphnx-Neu5Ac-*lacO* interactions, yet they
277 represent a promising step forward towards characterizing the interaction between our synthetic
278 TF and its cognate operator.

279 At this juncture of the project, both *in vivo* and *in vitro* experiments strongly suggested that
280 Sphnx was a transcriptional repressor capable of interacting with *lacO* and being de-repressed by
281 neuraminic acid. To better understand the dynamics of the Sphnx-Neu5Ac interaction we analyzed
282 the production of GFP by MG1655 (pCKT-Sphnx, pHC_DYOLacI-R) cultures at 14 hours of
283 growth (*Materials and Methods*) when exposed to different concentrations of neuraminic acid
284 around the average for an adult male (≈1 mM)[52]. The average fold difference of GFP expression

285 between the repressed and induced conditions was 2.25 ± 0.74 for cultures induced with 0.5 mM
286 Neu5Ac, 2.54 ± 0.52 for cultures induced with 1.0 mM Neu5Ac, 3.81 ± 1.15 for cultures induced
287 with 2.0 mM Neu5Ac, and 5.21 ± 2.00 for cultures induced with 5.0 mM Neu5Ac (n = 8, error
288 expressed as SEM). These results were commensurable with those observed for $SLCP_{GL}$ (LacI-
289 GGBP-OD)[33], as well as for CbnR-ABE44898-OD and LmrR-BzdB1_nSP[10], all of them modular
290 chimeric TFs. *Figure 2c* shows the levels of GFP fluorescence observed in basal and repressed
291 cultures as well as in cultures induced by 0.5, 1.0, 2.0 and 5.0 mM neuraminic acid. These assays
292 corroborated that relative fluorescence between basal (aTc^-/$Neu5Ac^-$) and repressed
293 (aTc^+/$Neu5Ac^-$) conditions were significant. Cultures in which the genetic circuit used to test
294 Sphnx were de-repressed (aTc^+/$Neu5Ac^+$) displayed statistically significant values compared to
295 the repressed cultures when in the presence of 1.0 mM and 2.0 mM Neu5Ac. Induction by 0.5 and
296 5.0 mM was not significant. It is important to note that relative fluorescence values in the 5.0 mM-
297 induced biological replicas were widely dispersed, suggesting that an overabundance of Neu5Ac
298 might interfere with the stability of the host strain producing some deleterious effect associated to
299 the expression of the heterologous Sphnx and GFP proteins in the assayed conditions. These results
300 showing Sphnx induction are promising, yet they underline the need to refine chimeric TFs
301 resulting from the assembly of pre-existing modules to optimize their induction range as well as
302 gene dose and reporter genes used to study their behaviour *in vivo* and *in vitro*. Newly created
303 synthetic TFs tend to have limited transfer function and specificity, and thus to require further
304 engineering for their field deployment in biosensors[53]. In this regard, we are optimistic about the
305 future improvement of the chimeras we create: the estimated K_d of SiaP and SatA for neuraminic
306 acid are in the nM range[38,39,41] (calculated *in vitro*), whereas in this work we present data in the
307 mM range (*in vivo*). This suggests that we have not reached the full potential of our LBDs and

308 there is an opportunity to optimize Sphnx's structure to better benefit from the high specificity and
309 affinity of our LBD. It is also important to note that Sphnx is currently at the same point in
310 evolutionary history when DBD-LacI was fused to the ancestral version of its current LBD, long
311 before communication between DBD and LBD had been refined and perfected by natural
312 selection[32]. Regarding the use of alternative reporters that may shed more light on the induction
313 profile of Sphnx, we acknowledge that the use of GFP has some inherent limitations for their use
314 of prokaryotes (e.g. extremely varying maturation times dependent on strain and metabolic state[54]),
315 which has motivated us to initiate the development of better reporting systems.

316 **Modeling of the 3D structure of chimeric transcription factors**. Another element we are
317 starting to explore is the importance of linkers for the functionality of modular proteins, an often-
318 neglected design element[55]. Since the beginning of our journey characterizing the chimeric TFs,
319 we found extremely intriguing the fact that Sphnx (LacI-LNK2-SiaP, N_t-DBD-EKEKEK-LBD-
320 C_t) and Kunst (LacI-LNK3-SiaP, N_t-DBD-GSGSGS-LBD-C_t), which are nearly identical proteins
321 (*Supp. Materials Appendix*), possessed opposed behaviours when interacting with neuraminic
322 acid. The recent release of AlphaFold 3[56] has enabled us to predict the structures of multimeric
323 chimeric proteins in the presence of their dsDNA operator, allowing us to model the proteins under
324 conditions closer to those found in actual WCBs. Sphnx and Kunst were both modeled as dimers
325 using AlphaFold 3, along with a DNA double helix containing the *lacO* (5'-
326 ATTGTGAGCGGATAACAATT-3' and its complementary sequence) recognized by the DBD-
327 LacI they both carry. Our goal was to predict structures that could help explain why Sphnx binds
328 Neu5Ac while Kunst does not, despite only differing by six amino acids. *Figure 3a* shows the
329 predicted Kunst DBD (green) bound to the DNA double helix (gray) and the formation of a dimer
330 between the PBP (cyan). The structure shows a mostly disordered linking region (orange),

331 including the GSGSGS (LNK3) linker (violet). The predicted Sphnx model (*Figure 3b*), on the
332 other hand, shows that the connecting region adopts a helical structure, including the EKEKEK
333 (LNK2) linker, and appears to form contacts between both the DBD and PBP. Rotating the Kunst
334 structure by 90° (*Figure 3c*) we can observe that the GSGSGS (linker does not interact with the
335 protein or DNA. In contrast, rotating the Sphnx structure by 90° (*Figure 3d*) further shows that the
336 linker forms contacts between the DBD and PBP domains, with interactions occurring across
337 different monomer chains. Further analysis shows that the helical linker in Sphnx forms
338 electrostatic salt bridges between itself and both the DBD and PBP domains. *Figure 3e* shows the
339 acidic residues on the linker (Glu) forming contacts with basic residues (Lys) on the DBD and
340 PBP domains. Specifically, the predicted structure suggests that the residues expected to form salt
341 bridges include Glu67-Lys2 on the DBD and Glu69-Lys152 on the PBP. These salt bridges likely
342 stabilize the helical linker, affecting the flexibility and stability of the polypeptide, an effect that
343 has been previously observed in other proteins[57]. From these predicted structures, we hypothesize
344 that these salt bridges stabilize helicity in the flexible linker in a way that positively affects the
345 PBP function by interacting with the DBD and the PBP, explaining the ligand binding functionality
346 of Sphnx compared to non-binding Kunst protein. Zooming out to see the larger structure (*Figure*
347 *3f*), aromatic residues near the linker salt bridges are visible. These residues are key for testing the
348 new predicted structures and designing new chimeric proteins in the future. We intend to purify
349 Sphnx and Kunst to perform fluorescence analysis of the protein-ligand interactions, including
350 fluorescence quenching and fluorescence anisotropy studies. The construction of 3D models, using
351 the latest version of AlphaFold, will continue to enable us to explore which mutations might
352 enhance ligand binding in a way that was previously unavailable, also allowing us to model
353 conditions closer to the actual multimeric *in vivo* conditions.

354

355

356 **CONCLUSIONS**

357 This work presents MG1655 (pCKT-Sphnx, pHC_DYOLacI-R), the first iteration of a
358 transcription factor-based whole-cell biosensor strain capable of detecting the salivary biomarker
359 neuraminic acid and relay its presence through the production of a fluorescent reporter. The
360 keystone of the genetic circuit enabling the detection of the soluble molecule is a new
361 transcriptional repressor: Sphnx. The apparent absence of a native Neu5Ac-binding transcription
362 factor led the authors to create a custom-made regulator via a straightforward synthetic biology
363 approach that enables the generation of chimeras capable of repressing a known promoter and
364 being de-repressed by a molecule of interest. Neuraminic acid is an attractive analyte to monitor
365 since its enhanced levels in saliva are correlated with conditions that range from periodontal
366 disease[52] (oral microbes harvest sialic acid to evade our immune system[22]) to the onset of oral
367 squamous cell carcinoma[23,24]. To the best of our knowledge, Sphnx is the first published
368 transcriptional repressor capable of being induced by neuraminic acid.

369 This study supports the idea that the ability to construct custom-made TFs, affordably and
370 on-demand, is instrumental in expanding the panoply of small soluble molecules that can be
371 detected by biosensors in a field limited by the availability of sensing modules[32]. It is important to
372 note that synthetic TFs such as Sphnx are not the closing of a chapter, but the origin of a new
373 source of regulators whose performance can be improved via directed engineering of their ligand
374 binding pockets[58]. It is widely acknowledged that biosensor specificity and biosensor-analyte
375 reaction curves need to be optimized for the practical application of the WCB[53]. Once the biosensor
376 strain is optimized we will test it first on synthetic saliva[59] preparations with a gradient of Neu5Ac

377 concentrations and, eventually, with clinical salivary samples. The MG1655 (pCKT-Sphnx,
378 pHC_DYOLacI-R) strain represents the first step in the obtention of a functional WCB for the
379 detection of neuraminic acid, a necessary launchpad to achieve the goal of eventual clinical
380 deployment of the sensor.

381 Conclusively, the modular assembly of fusion genes sheds light on the importance of inter-
382 domain communication for multi-domain protein functionality[60]. Our study presents the
383 interesting contrast between the inducibility of Sphnx as opposed to Kunst, and particularly the
384 unexpected role that a potential LBD-mediated stabilization of the linker may play in the de-
385 repressibility of the TF dimer. *In vitro* assays using purified proteins, as well as further modeling
386 simulation analysis and, eventually, the crystallization of these proteins will contribute in
387 illuminating the complex dynamics necessary for modular proteins to fulfill their function.

388 In summary, this work presents a new biosensor for N-acetylneuraminic acid, a salivary
389 biomarker relevant to oral health, which is based on a synthetic TF whose construction contributes
390 to illuminating the role of different domains in the functionality of modular proteins.

391

392

393

394

395

396

397 **MATERIALS AND METHODS**

398 **Bacterial strains and growth conditions**. Bacterial strains and plasmids used are listed in *Supp.*
399 *Table 2*. *E. coli* strains were grown at 37°C in LB (Lysogenic Broth[61]) medium unless otherwise
400 indicated. When required, *E. coli* cells were grown in M63 minimal medium[62] using the necessary
401 nutritional supplements and 30 mM glycerol (Sigma, St. Louis, MO) as a carbon source.
402 Antibiotics were added at the following concentrations: 100 $\mu g \cdot ml^{-1}$ ampicillin/carbenicillin and
403 25 $\mu g \cdot ml^{-1}$ chloramphenicol (Sigma, St. Louis, MO). For protein expression from pCKTRBS-
404 derived vectors, cultures were induced with 0.5 $\mu g \cdot ml^{-1}$ anhydrotetracycline (aTc) (Clontech,
405 Mountain View, CA). When experimental conditions required it, cultures were supplemented with
406 N-acetylneuraminic acid (neuraminic acid, Neu5Ac; Sigma, St. Louis, MO) at a 1 mM
407 concentration unless otherwise indicated.

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

420 overnight precultures of the strains (plus aTc and Neu5Ac when required) were used to inoculate
421 5 ml of fresh media supplemented in the same manner as the precultures. After 14 h of growth,
422 180 µl samples of the cultures were transferred to 96-well plates to enable OD_{600} and fluorescence
423 readings.
424 Statistical analysis as well as data visualization were performed with R (https://www.r-
425 project.org/) in the RStudio (RStudio Inc., Boston, MA) and Visual Studio Code environments
426 (Microsoft Corp., Redmond, WA).
427 **Molecular biology techniques.** Molecular biology techniques were performed following
428 commonly used standard protocols and as per manufacturers' instructions[63]. PCR reactions took
429 place in a 6321 Mastercycler PRO Vapo Protect Thermal Cycler (Eppendorf, Hamburg, Germany).
430 Routine separation of DNA and RNA in agarose gels was performed in an RunOne Electrophoresis
431 system (EmbiTec, San Diego, CA). Agarose gels were imaged in an UVP UVSolo Touch (Analytic
432 Jena, Jena, Germany). Plasmid DNA was purified with a Qiaprep Spin Miniprep Kit (Qiagen,
433 Hilden, Germany). DNA fragments were purified with DNA Clean-up and Concentration Kit
434 (Zymo research, Irvine, CA). The oligonucleotides employed for PCR amplification of the cloned
435 fragments and other molecular biology techniques are summarized in *Supp. Table 3* and were
436 supplied by IDT (Coralville, IA). All cloned inserts and DNA fragments were confirmed either via
437 Sanger sequencing[64] performed by Azenta Inc. (Burlington, MA), or Oxford-Nanopore whole-
438 plasmid sequencing performed by Plasmidsaurus (Eugene, OR). Commercially available NEB5-
439 alpha chemically competent cells (NEB, Ipswich, MA) were used for routine transformations.
440 Alternatively, electrocompetent *E. coli* cells were generated and transformed immediately (Gene
441 Pulser; Bio-Rad, Hercules, CA)[63]. Nucleotide sequence analyses were done at the National Center
442 for Biotechnology Information server (www.ncbi.nlm.nih.gov) and Benchling Biology Software

443 (www.benchling.com). Cloning was routinely performed via Gibson assembly[65] unless otherwise
444 indicated.
445 **Domain selection and cloning.** DBD-LacI is located on *lacI* 5' region, from the start codon to
446 base 198, encoding a peptide of 66 amino acids that comprises the N-terminus of our chimeric
447 proteins. This domain was obtained via PCR amplification from plasmid pCKTRBS-LacIwt[10]
448 using the oligonucleotide pair SV00001 / SV00002. SatA from *H. ducreyi* (Uniprot Q7VL18) and
449 SiaP from *H. influenzae* (Uniprot P44542) were ordered as synthetic dsDNA fragments (IDT,
450 Coralville, IA). Several variants of the LBDs SatA and SiaP were amplified via PCR so that they
451 would include 5' overhangs including linkers (LNK2-5) or no linker whatsoever (LNK1). SatA
452 was amplified with primers SV00030-SV00034 / SV00004 and SiaP with primers SV00015-
453 SV00019 / JFJ0018.
454 Genes encoding the chimeric TFs were originally assembled in pUC19[66]. The vector was amplified
455 via divergent PCR with primers JFJ00019 / JFJ00020, which introduce a consensus RBS for *E.*
456 *coli*, originating plasmid pUC19RBS. Tripartite Gibson assemblies including linear pUC19RBS,
457 DBD-LacI, and LNK-LBD were performed and the resulting constructions transformed into
458 NEB5-alpha competent cells. After a carbenicillin selection in solid media, a subset of the resulting
459 clones was isolated and their plasmids (pUC19-*Chimera* family, *Supp. Table 2*) purified and
460 sequenced to validate the presence of chimeric TFs. The resulting TF genes cloned into
461 pUC19RBS were amplified using SV00024 / SV00025 for SiaP-derived chimeras and SV00024 /
462 SV00026 for their SatA counterparts. These primers amplify our TF genes from the 5' DBD-LacI,
463 just downstream of the consensus RBS site, to the stop codon of the LBD. Low copy plasmid
464 pCKTRBS[10] was amplified via divergent PCR using primer pairs SV00022 / SV00021 for SiaP-
465 derived chimeras and SV00022 / SV00023 for their SatA counterparts. These primers introduce 5'

466 and 3' overhang sequences providing homology to our dsDNA chimeric gene fragments, and
467 including in their 5' side our consensus RBS site. Bipartite Gibson assemblies including linear
468 pCKTRBS and the fusion genes encoding the chimeras were performed, and the resulting
469 assembly products transformed into NEB5-alpha competent cells. After a carbenicillin selection
470 in solid media, resulting clones were isolated and their plasmids sequenced to confirm the presence
471 of chimera-encoding genes. Resulting pCKT-*Chimera* plasmids were purified from NEB5-alpha
472 (pCKT-*Chimera*) strains and electroporated into MG1655 cells. After confirming the presence of
473 the plasmid in the new host, MG1655 (pCKT-*Chimera*) were electroporated with the reporter
474 plasmid pHC_DYOLacI-R. Validated MG1655 (pCKT-*Chimera*, pHC_DYOLacI-R) strains were
475 used for *in vivo* and *in vitro* analysis of the functionality of the novel chimeric TFs. The sequences
476 of all the oligonucleotides used in this work can be found in *Supp. Table 3*.
477 **Electrophoretic mobility shift assay (EMSA).** In order to obtain cell extracts enriched in the
478 Sphnx / Siren proteins, NEB5-alpha (pCKT-Sphnx) and NEB5-alpha (pCKT-Siren) cells were
479 grown on LB supplemented with aTc for 5 hours at 37˚C in a shaker incubator. Cells were
480 recovered by centrifugation at 4°C (15' at 4,000 g) and then resuspended in 20 mM Tris-HCl (pH
481 7.0) to be sonicated at 70 amp for 30 seconds (30 cycles of 1 second; Branson Digital Sonifier,
482 Danbury, CN). A subsequent centrifugation at 4°C (20' at 20,000 g) was performed to eliminate
483 cell debris from the sample. The recovered supernatant was then incubated with the DNA probe,
484 a 251 bp dsDNA fragment amplified from plasmid pHC_DYOLacI-R with primers SV00028 /
485 SV00029, containing P_{lac} (including *lacO*). Cell extract (5 µl) and DNA probe (5 µl) were
486 incubated together with reaction buffer (20% glycerol, 100 mM KCl, 500 $\mu g \cdot ml^{-1}$ bovine serum
487 albumin, 20mM Tris-HCl pH 7.5). The reaction mixture is incubated for 30' at 30 ˚C. After
488 incubation of the retardation mixtures, 2 µl of Gel loading dye without SDS (NEB, Ipswich, MA)

489 were added to each reaction. Mixtures were subsequently fractionated by electrophoresis in precast
490 6% polyacrylamide gels (Novex TBE Gels; Invitrogen, Waltham, MA) buffered with 0.5X TBE
491 (45 mM Tris borate, 1mM EDTA). Gels were run at constant voltage (100 V) for 90' in a Mini
492 Gel Tank (Invitrogen, Waltham, MA) and afterwards stained for 1 h in 50 ml of 0.5X TBE buffer
493 supplemented with 5 µl of SYBR Gold dye (Invitrogen, Waltham, MA) while gently shaken. Post
494 staining, gels were imaged in a UVP UVSolo Touch (Analytic Jena, Jena, Germany).
495 **qRT-PCR analysis.** Quantitative Real Time Polymerase Chain Reaction (qRT-PCR) was used to
496 assay the transcription of our reporter gene (*sfgfp*). MG1655 (pCKT-*Chimera*, pHCDYOLacI-R)
497 strains were precultured overnight in of 3 ml LB supplemented with the corresponding selection
498 antibiotics. On the next day, 100 µl of overnight preculture were used as inoculum for several
499 cultures, 10 ml each, of fresh LB (plus selection antibiotics). Each one of the cultures was
500 supplemented so that it would represent one of these three different conditions: *Basal* (aTc^-,
501 $Neu5Ac^-$), *Repressed* (aTc^+,$Neu5Ac^-$), and *De-repressed* (aTc^+, $Neu5Ac^+$). Where experimental
502 conditions required it, cultures were supplemented with 0.5 $\mu g \cdot ml^{-1}$ anhydrotetracycline (aTc)
503 (Clontech, Mountain View, CA) and 1 mM neuraminic acid (Neu5Ac; Sigma, St. Louis, MO). For
504 every independent qRT-PCR experiment, three biological replicates were created across each of
505 these three conditions. These cultures were placed in a shaker incubator at 37˚C for 5 hours and
506 pelleted at 4˚C and 4,000 g. Pellets were either frozen at -80˚C for future use or immediately
507 processed for total RNA extraction. Cells from each individual pellet were set for lysis by
508 resuspendion in 1 ml of IBI reagent, a phenol and guanidine isothiocyanate mixture (IBI Scientific,
509 Dubuque, IA). 200 µl of chloroform (Fisher Scientific, Pittsburgh, PA) was added to the lysate
510 and the mixture was vigorously shaken to homogeneity, followed by centrifugation (benchtop
511 minifuge) at 4˚C and 14,500 rpm for 15 minutes that led to a phase separation. 500 µl of the upper

512 aqueous phase was mixed with an equal volume of 100% ethanol. At this point, total RNA was
513 purified from DNA / RNA phase with Direct-zol RNA Miniprep Kit (Zymo Research, Irvine, CA)
514 where the TRIzol reagent steps are replaced with the phenol and chloroform phase separation.
515 Total RNA eluted with ddH_2O was then retrotranscribed using qScript Flex cDNA Synthesis Kit
516 (QuantaBio, Beverly, MA). A volume totaling 1 µg RNA was incubated with random hexamers
517 and then supplemented with retrotranscriptase enzyme mix to synthesize cDNA. cDNA resulting
518 from the last step was used as a template for the amplification of a 256 bp fragment with
519 oligonucleotides SV00027A / SV00027 in a CFX96 Real-Time PCR System (Bio-Rad, Hercules,
520 CA). Primers had been designed to amplify a region of pHC_DYOLacI comprising the beginning
521 of *sfgfp*, so that when the tested chimera was being expressed it repressed the expression of GFP
522 unless Neu5Ac was present. Cq values were analyzed as a measure of gene expression using the
523 ΔΔCq ($2^{-\Delta\Delta Ct}$) method[67].
524 **Protein Structure Prediction using AlphaFold 3.** The 3D structures of the target proteins, Sphnx
525 (LacI-LNK2-SiaP, N_t-DBD-EKEKEK-LBD-C_t) and Kunst (LacI-LNK3-SiaP, N_t-DBD-
526 GSGSGS-LBD-C_t), were predicted using AlphaFold 3[56], developed by Google DeepMind
527 (London, UK). The predictions were performed using the publicly available AlphaFold 3 web
528 server (https://golgi.sandbox.google.com/). Both Sphnx and Kunst proteins were modeled as
529 dimers, along with the DNA double helix containing *lacO* (5'-ATTGTGAGCGGATAACAATT-
530 3' and its complementary sequence). Post-prediction, the structures were visualized and further
531 analyzed using VMD[68] to determine key protein-protein interactions.
532

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.

554 **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	-
Bastet	LacI-LNK5-SiaP	Yes	No	-
Lumos	LacI-LNK6-SiaP	Yes	Yes	Duplicated sequence acts as Linker.
Harpy	LacI-LNK1-SatA	Yes	Unreliable	Unstable strain
Haus	LacI-LNK2-SatA	Yes	No	-

555

Figure 1

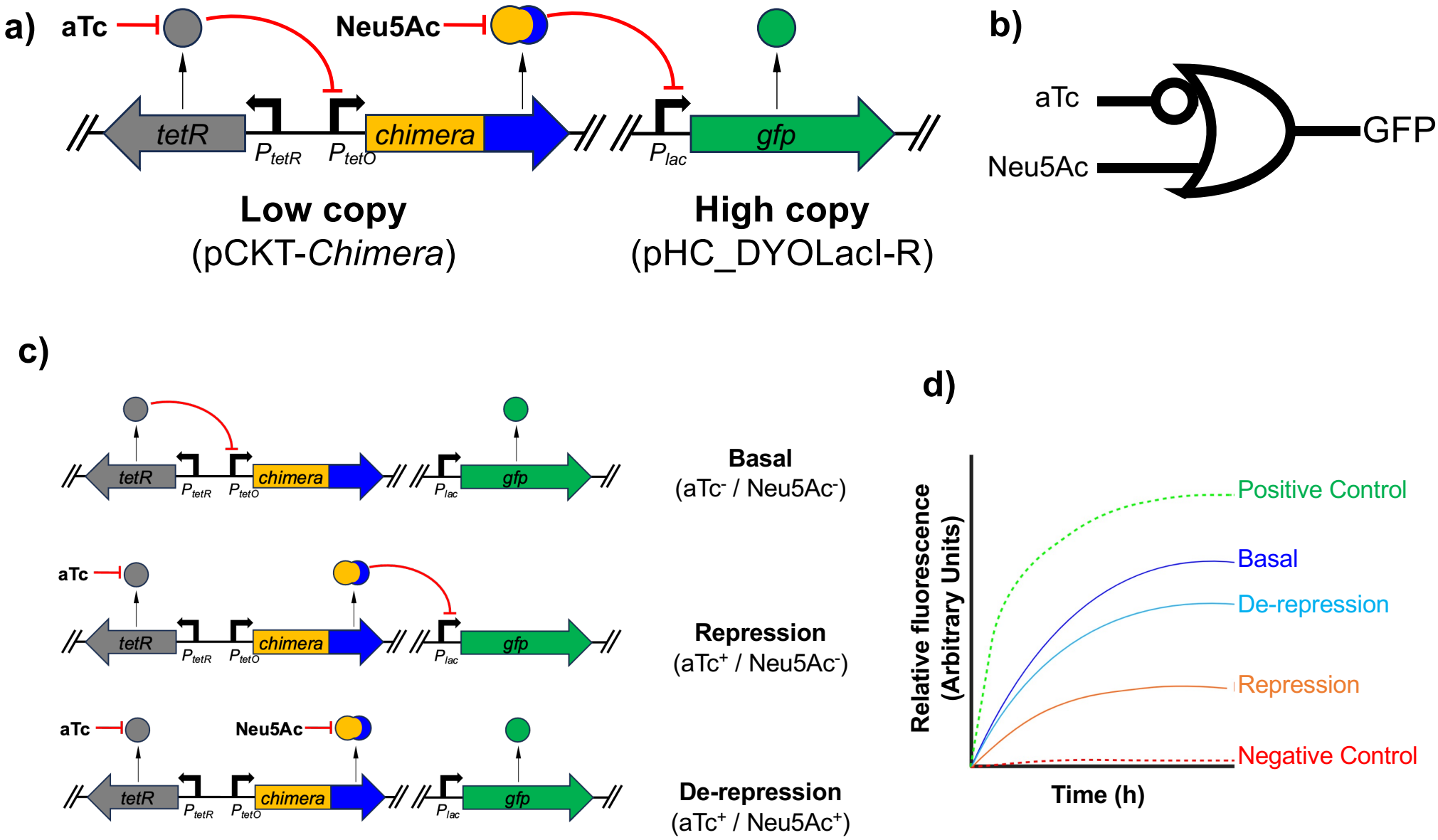
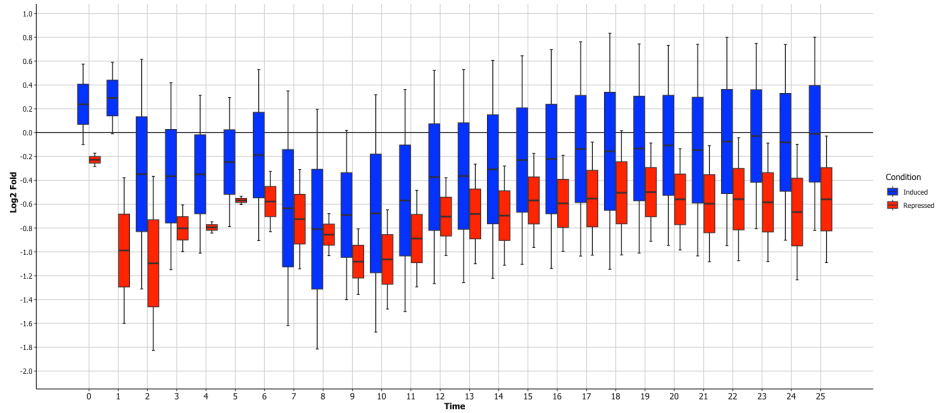

a)
aTc
Neu5Ac
tetR
chimera
gfp
P_{tetR}
P_{tetO}
P_{lac}
Low copy
(pCKT-*Chimera*)
High copy
(pHC_DYOLacI-R)
b)
aTc
Neu5Ac
GFP
c)
Basal
(aTc⁻ / Neu5Ac⁻)
Repression
(aTc⁺ / Neu5Ac⁻)
De-repression
(aTc⁺ / Neu5Ac⁺)
d)
Relative fluorescence
(Arbitrary Units)
Time (h)
Positive Control
Basal
De-repression
Repression
Negative Control

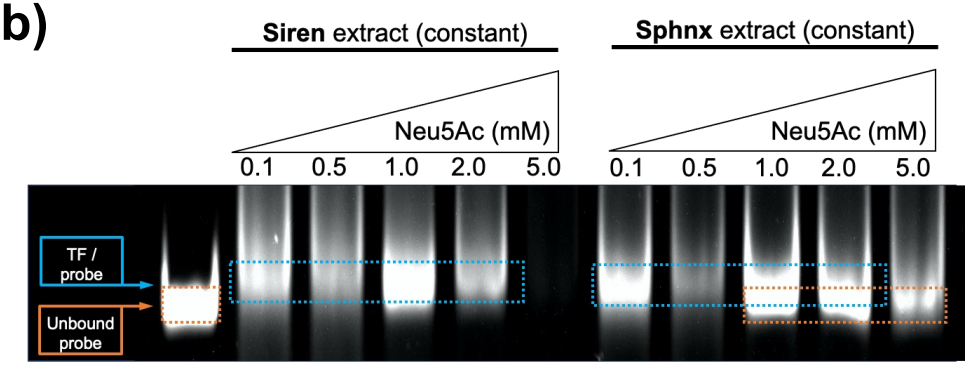
Figure 1. Screening strategy to find functional chimeric transcription factors. a) Biosensor genetic circuit for the *in vivo* testing of the functionality of cloned chimeric TFs. MG1655 (pCKT-*Chimera*, pHC_DYOLacI-R) cells express a chimeric TF under the control of TetR. The chimeric product can repress 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 fluorescence 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* (*solid red line*), displays background fluorescence by a strain that does not carry *gfp*. *Basal* (*dark blue line*), *Repressed* (*dashed red line*), *De-repressed* (*light 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.

Figure 2

a)


Log2 Fold
Time
Condition
Induced
Repressed

b)


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

c)

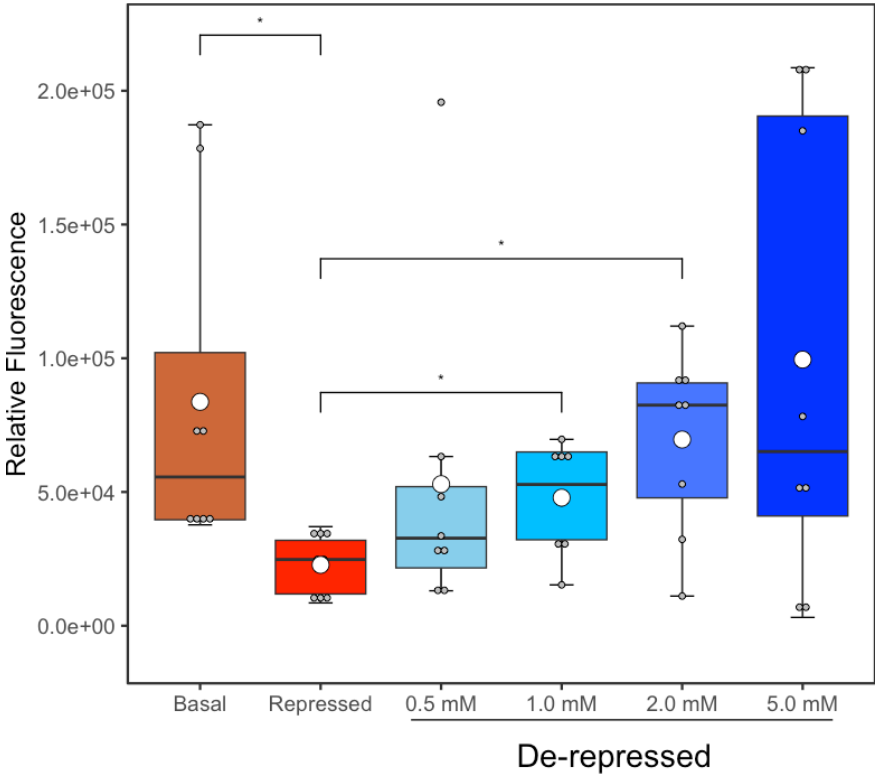

Relative Fluorescence
2.0e+05
1.5e+05
1.0e+05
5.0e+04
0.0e+00
Basal
Repressed
0.5 mM
1.0 mM
2.0 mM
5.0 mM
De-repressed

Figure 2. *V*alidation 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 over time (hours). 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). 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).

Figure 3

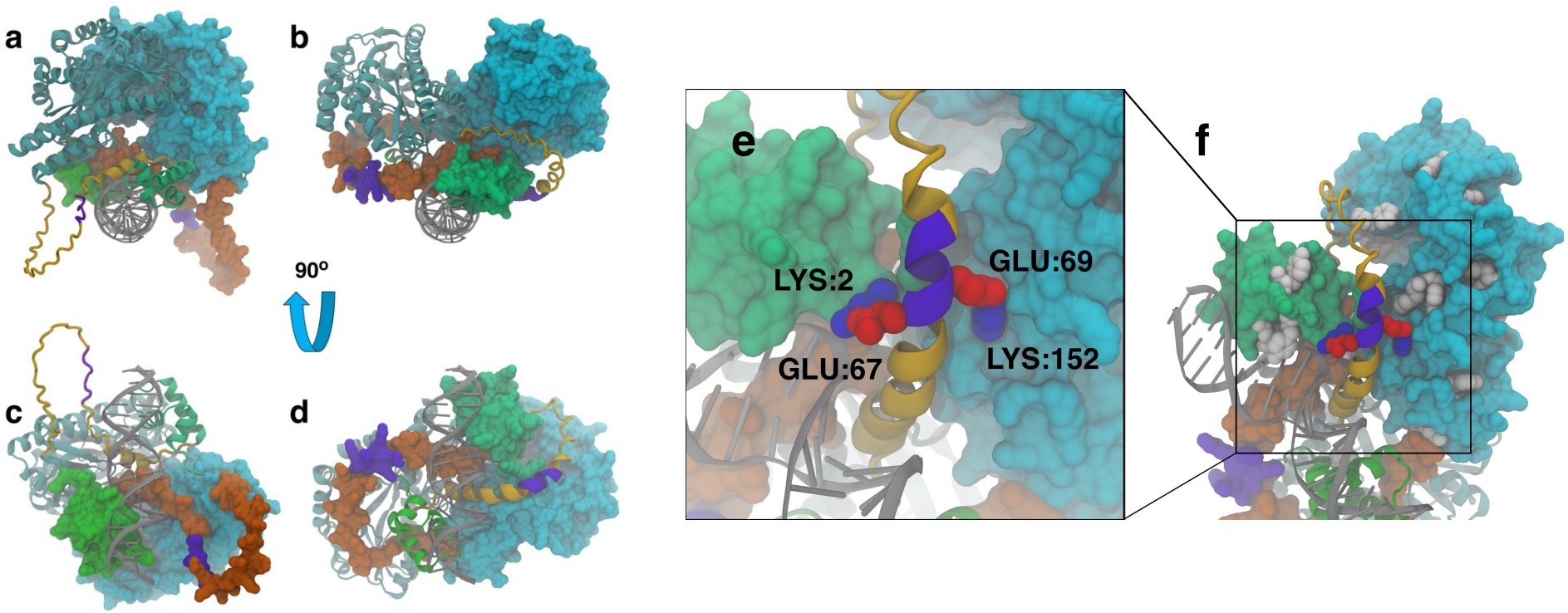


Figure 3. Modeled structures of chimeric protein dimers with double helix DNA obtained with Alphafold 3. For each structure, one monomer is shown using the New Cartoon representation and the other using the space-filling Surf representation. The DBD is shown in *green*, the PBP in *cyan*, and the connecting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

556 **BIBLIOGRAPHY**

557 1. Pancrazio, J. J., Whelan, J. P., Borkholder, D. A., Ma, W. & Stenger, D. A. Development
558 and application of cell-based biosensors. *Ann. Biomed. Eng.* **27**, 697–711 (1999).
559 2. Chang, H.-J., Voyvodic, P. L., Zúñiga, A. & Bonnet, J. Microbially derived biosensors for
560 diagnosis, monitoring and epidemiology. *Microb. Biotechnol.* **10**, 1031–1035 (2017).
561 3. Wu, Y., Wang, C.-W., Wang, D. & Wei, N. A Whole-Cell Biosensor for Point-of-Care
562 Detection of Waterborne Bacterial Pathogens. *ACS Synth. Biol.* **10**, 333–344 (2021).
563 4. Zhang, F. & Keasling, J. Biosensors and their applications in microbial metabolic
564 engineering. *Trends Microbiol.* **19**, 323–329 (2011).
565 5. He, W., Yuan, S., Zhong, W.-H., Siddikee, M. A. & Dai, C.-C. Application of genetically
566 engineered microbial whole-cell biosensors for combined chemosensing. *Appl. Microbiol.*
567 *Biotechnol.* **100**, 1109–1119 (2016).
568 6. Saltepe, B., Kehribar, E. Ş., Su Yirmibeşoğlu, S. S. & Şafak Şeker, U. Ö. Cellular
569 Biosensors with Engineered Genetic Circuits. *ACS Sens* **3**, 13–26 (2018).
570 7. Berepiki, A., Kent, R., Machado, L. F. M. & Dixon, N. Development of High-Performance
571 Whole Cell Biosensors Aided by Statistical Modeling. *ACS Synth. Biol.* **9**, 576–589 (2020).
572 8. Schleif, R. DNA binding by proteins. *Science* **241**, 1182–1187 (1988).
573 9. Rogers, J. K. *et al.* Synthetic biosensors for precise gene control and real-time monitoring of
574 metabolites. *Nucleic Acids Res.* **43**, 7648–7660 (2015).
575 10. Juárez, J. F., Lecube-Azpeitia, B., Brown, S. L., Johnston, C. D. & Church, G. M. Biosensor
576 libraries harness large classes of binding domains for construction of allosteric
577 transcriptional regulators. *Nat. Commun.* **9**, 3101 (2018).
578 11. Koch, M., Pandi, A., Borkowski, O., Batista, A. C. & Faulon, J.-L. Custom-made

579 transcriptional biosensors for metabolic engineering. *Curr. Opin. Biotechnol.* **59**, 78–84
580 (2019).
581 12. Wren, M. E., Shirtcliff, E. A. & Drury, S. S. Not all biofluids are created equal: chewing
582 over salivary diagnostics and the epigenome. *Clin. Ther.* **37**, 529–539 (2015).
583 13. Bandodkar, A. J. & Wang, J. Non-invasive wearable electrochemical sensors: a review.
584 *Trends Biotechnol.* **32**, 363–371 (2014).
585 14. Campuzano, S., Pedrero, M., Torrente-Rodríguez, R. M. & Pingarrón, J. M. Affinity-based
586 wearable electrochemical biosensors: Natural versus biomimetic receptors. *Anal. Sens.*
587 (2022) doi:10.1002/anse.202200087.
588 15. Omar, E. Current concepts and future of noninvasive procedures for diagnosing oral
589 squamous cell carcinoma--a systematic review. *Head Face Med.* **11**, 6 (2015).
590 16. Wang, S., Yang, M., Li, R. & Bai, J. Current advances in noninvasive methods for the
591 diagnosis of oral squamous cell carcinoma: a review. *Eur. J. Med. Res.* **28**, 53 (2023).
592 17. Haines-Menges, B. L., Whitaker, W. B., Lubin, J. B. & Boyd, E. F. Host Sialic Acids: A
593 Delicacy for the Pathogen with Discerning Taste. *Microbiol Spectr* **3**, (2015).
594 18. Blix, G. Concerning the carbohydrate groups of submaxillary mucin. *Hoppe Seylers Z.*
595 *Physiol. Chem.* **240**, 43–54.
596 19. Lundblad, A. Gunnar Blix and his discovery of sialic acids. Fascinating molecules in
597 glycobiology. *Ups. J. Med. Sci.* **120**, 104–112 (2015).
598 20. Röhrig, C. H., Choi, S. S. H. & Baldwin, N. The nutritional role of free sialic acid, a human
599 milk monosaccharide, and its application as a functional food ingredient. *Crit. Rev. Food*
600 *Sci. Nutr.* **57**, 1017–1038 (2017).
601 21. Vajaria, B. N. *et al.* Salivary glyco-sialylation changes monitors oral carcinogenesis.

602 *Glycoconj. J.* **31**, 649–659 (2014).
603 22. Phansopa, C. *et al.* Characterization of a sialate-O-acetylesterase (NanS) from the oral
604 pathogen Tannerella forsythia that enhances sialic acid release by NanH, its cognate
605 sialidase. *Biochem. J* **472**, 157–167 (2015).
606 23. Arthisri, A. S., Sathiyamoorthy, A., Meenakshi, B. & Chandran, C. R. Ratio of Salivary
607 Sialic Acid to Fucose as Tumor Markers in Potentially Malignant Disorders and Oral
608 Cancer. *Contemp. Clin. Dent.* **11**, 131–135 (2020).
609 24. Chaudhari, V., Pradeep, G. L., Prakash, N. & Mahajan, A. M. Estimation of salivary sialic
610 acid in oral premalignancy and oral squamous cell carcinoma. *Contemp. Clin. Dent.* **7**, 451–
611 456 (2016).
612 25. Markopoulos, A. K. Current aspects on oral squamous cell carcinoma. *Open Dent. J.* **6**,
613 126–130 (2012).
614 26. Coletta, R. D., Yeudall, W. A. & Salo, T. Grand Challenges in Oral Cancers. *Front Oral*
615 *Health* **1**, 3 (2020).
616 27. Warnakulasuriya, S. & Greenspan, J. S. *Textbook of Oral Cancer: Prevention, Diagnosis*
617 *and Management*. (Springer Nature, 2020).
618 28. Robson, B. Bioinformatics studies on a function of the SARS-CoV-2 spike glycoprotein as
619 the binding of host sialic acid glycans. *Comput. Biol. Med.* **122**, 103849 (2020).
620 29. Stanton, B. C. *et al.* Systematic transfer of prokaryotic sensors and circuits to mammalian
621 cells. *ACS Synth. Biol.* **3**, 880–891 (2014).
622 30. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol.*
623 *Biol.* **3**, 318–356 (1961).
624 31. Meinhardt, S. *et al.* Novel insights from hybrid LacI/GalR proteins: family-wide functional

625 attributes and biologically significant variation in transcription repression. *Nucleic Acids*
626 *Res.* **40**, 11139–11154 (2012).
627 32. Fukami-Kobayashi, K., Tateno, Y. & Nishikawa, K. Parallel evolution of ligand specificity
628 between LacI/GalR family repressors and periplasmic sugar-binding proteins. *Mol. Biol.*
629 *Evol.* **20**, 267–277 (2003).
630 33. Sakaguchi-Mikami, A., Taniguchi, A., Sode, K. & Yamazaki, T. Construction of a novel
631 glucose-sensing molecule based on a substrate-binding protein for intracellular sensing.
632 *Biotechnol. Bioeng.* **108**, 725–733 (2011).
633 34. Giuliani, S. E. *et al.* Environment sensing and response mediated by ABC transporters.
634 *BMC Genomics* **12 Suppl 1**, S8 (2011).
635 35. Locher, K. P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat.*
636 *Struct. Mol. Biol.* **23**, 487–493 (2016).
637 36. Post, D. M. B., Mungur, R., Gibson, B. W. & Munson, R. S., Jr. Identification of a novel
638 sialic acid transporter in Haemophilus ducreyi. *Infect. Immun.* **73**, 6727–6735 (2005).
639 37. Gruteser, N., Marin, K., Krämer, R. & Thomas, G. H. Sialic acid utilization by the soil
640 bacterium Corynebacterium glutamicum. *FEMS Microbiol. Lett.* **336**, 131–138 (2012).
641 38. Müller, A. *et al.* Conservation of structure and mechanism in primary and secondary
642 transporters exemplified by SiaP, a sialic acid binding virulence factor from Haemophilus
643 influenzae. *J. Biol. Chem.* **281**, 22212–22222 (2006).
644 39. Severi, E. *et al.* Sialic acid transport in Haemophilus influenzae is essential for
645 lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite
646 ATP-independent periplasmic transporter. *Mol. Microbiol.* **58**, 1173–1185 (2005).
647 40. Gangi Setty, T., Cho, C., Govindappa, S., Apicella, M. A. & Ramaswamy, S. Bacterial

648 periplasmic sialic acid-binding proteins exhibit a conserved binding site. *Acta Crystallogr.*
649 *D Biol. Crystallogr.* **70**, 1801–1811 (2014).
650 41. Gangi Setty, T. *et al.* Molecular characterization of the interaction of sialic acid with the
651 periplasmic binding protein from. *J. Biol. Chem.* **293**, 20073–20084 (2018).
652 42. Oertel-Buchheit, P., Reinbolt, J., John, M., Granger-Schnarr, M. & Schnarr, M. A LexA
653 mutant repressor with a relaxed inter-domain linker. *Protein Sci.* **7**, 512–515 (1998).
654 43. Tsirigotaki, A., De Geyter, J., Šoštaric, N., Economou, A. & Karamanou, S. Protein export
655 through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **15**, 21–36 (2017).
656 44. Glick, B. R. Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13**, 247–
657 261 (1995).
658 45. Liu, Q., Schumacher, J., Wan, X., Lou, C. & Wang, B. Orthogonality and Burdens of
659 Heterologous AND Gate Gene Circuits in E. coli. *ACS Synth. Biol.* **7**, 553–564 (2018).
660 46. Logic Gates. *The Department of Mathematics and Computer Science at Oxford College of*
661 *Emory University* https://mathcenter.oxford.emory.edu/site/cs170/gates/.
662 47. Lewis, M. The lac repressor. *C. R. Biol.* **328**, 521–548 (2005).
663 48. Deans, T. L., Cantor, C. R. & Collins, J. J. A tunable genetic switch based on RNAi and
664 repressor proteins for regulating gene expression in mammalian cells. *Cell* **130**, 363–372
665 (2007).
666 49. Langdon, R. C., Burr, T., Pagan-Westphal, S. & Hochschild, A. A chimeric activator of
667 transcription that uses two DNA-binding domains to make simultaneous contact with pairs
668 of recognition sites. *Mol. Microbiol.* **41**, 885–896 (2001).
669 50. Durante-Rodríguez, G. *et al.* Identification of a missing link in the evolution of an enzyme
670 into a transcriptional regulator. *PLoS One* **8**, e57518 (2013).

671 51. Durante-Rodríguez, G., Mancheño, J. M., Díaz, E. & Carmona, M. Refactoring the λ phage
672 lytic/lysogenic decision with a synthetic regulator. *Microbiologyopen* **5**, 575–581 (2016).
673 52. Oktay, S. *et al.* Serum and saliva sialic acid in periodontitis patients with and without
674 cardiovascular disease. *Pathophysiol. Haemost. Thromb.* **37**, 67–71 (2010).
675 53. Snoek, T. *et al.* Evolution-guided engineering of small-molecule biosensors. *Nucleic Acids*
676 *Res.* **48**, e3 (2020).
677 54. Hebisch, E., Knebel, J., Landsberg, J., Frey, E. & Leisner, M. High variation of
678 fluorescence protein maturation times in closely related Escherichia coli strains. *PLoS One*
679 **8**, e75991 (2013).
680 55. Chen, X., Zaro, J. L. & Shen, W.-C. Fusion protein linkers: property, design and
681 functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–1369 (2013).
682 56. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with
683 AlphaFold 3. *Nature* (2024) doi:10.1038/s41586-024-07487-w.
684 57. Meuzelaar, H., Vreede, J. & Woutersen, S. Influence of Glu/Arg, Asp/Arg, and Glu/Lys
685 Salt Bridges on α-Helical Stability and Folding Kinetics. *Biophys. J.* **110**, 2328–2341
686 (2016).
687 58. Raman, S., Taylor, N., Genuth, N., Fields, S. & Church, G. M. Engineering allostery.
688 *Trends Genet.* **30**, 521–528 (2014).
689 59. Levine, M. J. Development of artificial salivas. *Crit. Rev. Oral Biol. Med.* **4**, 279–286
690 (1993).
691 60. Persico, M. *et al.* From Protein Communication to Drug Discovery. *Curr. Top. Med. Chem.*
692 **15**, 2019–2031 (2015).
693 61. Bertani, G. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *J.*

694 *Bacteriol.* **186**, 595–600 (2004).

695 62. Miller, J. H. *Experiments in Molecular Genetics*. (1972).

696 63. Sambrook, J. & Russell, D. W. *Molecular Cloning: A Laboratory Manual*. (CSHL Press,

697 2001).

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

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

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

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

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

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

704 (1985).

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

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

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

708 **14**, 33–8, 27–8 (1996).

709