A fuzzy complex in the negative arm regulates circadian clock robustness.

Jennifer Hurley (✉ hurlej2@rpi.edu)
Rensselaer Polytechnic Institute

Meaghan Jankowski
Rensselaer Polytechnic Institute

Daniel Griffith
Washington University School of Medicine   https://orcid.org/0000-0002-9633-9601

Divya Shastry
Rensselaer Polytechnic Institute

Jacqueline Pelham
Rensselaer Polytechnic Institute

Garrett Ginell
Washington University School of Medicine   https://orcid.org/0000-0001-6511-5480

Joshua Thomas
Rensselaer Polytechnic Institute

Pankaj Karande
Rensselaer Polytechnic Institute

Alex Holehouse
Washington University in St. Louis   https://orcid.org/0000-0002-4155-5729

Article

A fuzzy complex in the negative arm regulates circadian clock robustness.

Meaghan S. Jankowski[1], Daniel Griffith[2], Divya G. Shastry[1], Jacqueline F. Pelham[1], Garrett M. Ginell[2], Joshua Thomas[1], Pankaj Karande[3,4], Alex S. Holehouse[2,5], and Jennifer M. Hurley[1,4,*]

[1] Department of Biological Sciences, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA.

[2] Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO, 63110, USA.

[3] Department of Chemical and Biological Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA.

[4] Center for Biotechnology and Interdisciplinary Sciences, Rensselaer Polytechnic Institute, Troy, NY, 12180, USA.

[5] Center for Science and Engineering of Living Systems, Washington University in St. Louis, St. Louis, MO, 63110, USA.

*Correspondence: hurlej2@rpi.edu

Summary

Organismal physiology is widely regulated by the circadian clock, a molecular circuit composed of a Transcription-Translation Feedback Loop [1,2]. Protein components of the molecular clock are enriched in intrinsically disordered regions, inherently flexible regions that interact with other proteins via short linear binding motifs (SLiMs) [3–5]. SLiM-

24   driven interactions contribute to circadian timing and the circadian regulation of the cell.

25   However, the mechanism that allows the formation of dynamic clock complexes remains

26   unclear as structural analysis of these protein-protein interactions has been limited due

27   to inherent protein disorder. Here, we apply a synthetic peptide microarray approach to

28   demonstrate that the core clock forms a fuzzy complex to support circadian robustness

29   [6,7]. We found positively charged islands on the clock protein FREQUENCY (FRQ) drove

30   a multi-valent interaction between FRQ and its partner FRQ-interacting RNA Helicase

31   (FRH) that enabled clock robustness rather than the previously-reported feedback [8]. We

32   found these positively charged islands were a conserved molecular feature throughout

33   clocks in fungi, insects, and mammals, and may enable the formation of fuzzy complexes.

34   This study constitutes the first mechanistic reason for the uniquely-broad conservation of

35   intrinsic disorder in circadian negative-arm proteins and will aid in the development of the

36   molecular model of clock protein interactions. Furthermore, we anticipate the application

37   of synthetic peptide microarrays to study disordered clock proteins and will be useful in

38   characterizing sites of interaction for clock-specific drug discovery [9].

39

43

44   **Introduction**

45   Circadian clocks have evolved as an adaptive mechanism to anticipate daily

46   environmental changes and are recognized as an important regulator of the cellular

47   environment amongst eukaryotes [1]. The underlying architecture of the circadian clock in

48   higher eukaryotes (including fungi, insects, and mammals) is a molecular oscillator

49   composed of a Transcription-Translation Feedback Loop (TTFL), made up of a positive

50   transcriptional-activating protein complex and a negative repressing protein complex

51   (Extended Data Fig. 1A) [2]. Despite an understanding of clock regulation effected through

52   transcriptional activation by positive arm proteins, questions remain regarding the

53   fundamental biophysical mechanisms of feedback and circadian post-transcriptional

54   regulation [10].

55   Clues to sources of circadian regulation have come from studies into the

56   ontologically-diverse and dynamic macromolecular protein complexes that coalesce

57   around negative arm clock proteins [11]. Negative-arm centered interactomes are likely

58   enabled by the enrichment of Intrinsically Disordered Regions (IDRs) within negative arm

59   proteins, regions characterized by a lack of a fixed tertiary structure that are also known

60   to form interaction hubs [4,12]. Genetically-based studies of negative arm clock proteins

61   have focused on conserved domains, ignoring that poorly-conserved IDRs commonly

62   engage in molecular interactions via short (5-15 residue), degenerate regions termed

63   <u>S</u>hort <u>Li</u>near <u>M</u>otifs (SLiMs) [5,13]. This approach, likely due to technical limitations,

64   hampered biophysical studies of the interactions that occur within negative-arm IDRs,

65   constraining investigations into this potentially important circadian regulatory mechanism.

66   Recently, synthetic peptide microarrays have been used to identify SLiMs within a single

67   IDR for a more structured domain, suggesting a protein microarray approach could be

68   extended to investigate binding behavior within the IDRs of clock negative-arm proteins

69   [7].

70   Here we report the use of the L̲inear m̲ot̲if disc̲overy using r̲at̲ional de̲sign

71   (LOCATE) approach, which employed rationally-designed printed synthetic peptides to

72   explore interaction domains in negative-arm clock protein complexes [14]. LOCATE

73   characterized the binding relationship between the negative arm clock protein

74   FREQUENCY (FRQ) from the clock model *Neurospora crassa* (*N. crassa*) and its partner

75   protein FRQ-interacting RNA Helicase (FRH). This analysis revealed the interaction

76   between FRQ and FRH occurred predominantly along conserved, positively charged,

77   "islands" within FRQ [8,15]. These positively charged islands were found to be a conserved

78   feature in negative-arm clock proteins in higher eukaryotes (e.g., the functionally

79   orthologous PER protein family), suggesting that distributed electrostatic interactions,

80   characteristic of fuzzy protein complexes, are a vital element in the formation of negative-

81   arm protein complexes. "Hotspot" residues within electrostatic charge islands facilitated

82   the conversion of an hourglass-like timer into a robust circadian timekeeper, highlighting

83   electrostatic interactions as a principal tenet of circadian timing.

84

85   **Results**

86   *LOCATE identifies novel interaction sites*

87   To explore interactions between negative arm clock proteins, we selected FRQ as

88   our model as it is largely intrinsically disordered, binds many different partners, and

89   contains many putative SLiMs (Extended Data Figs. 1b, 1c and 3d) [3,16]. We constructed

90   a 15-mer "peptide map" of sequentially overlapping peptides that scanned through the

91   primary sequence of FRQ (NCU02265), shifting by 3 a.a. (Extended Data Fig. 2a,

92   Supplementary Data 1) (see methods) [14]. Regions of interest were investigated by

93   rationally designed peptides (e.g., scrambles, single-point mutations, and truncations) [14].

94   To validate the LOCATE approach, we examined binding between the library and the

95   FRQ Nanny protein, FRH (NCU03363), using *E. coli* expressed FRH (a.a. 110-1106, with

96   a 6x His-tag) (Fig. 1a and Extended Data Fig. 2b) [15,17]. The LOCATE approach recognized

97   the previous genetically-identified FRQ-FRH domain (FFD, a.a. 774-782) (Extended Data

98   Fig. 2c, areas highlighted in orange) [8]. However, beyond this domain, there were many

99   instances of valid interaction, suggesting the FRQ/FRH interaction was more complex

100   and involved regions beyond the FFD (Extended Data Fig. 2c).

101

102   *Negative arm interaction occurs via charged islands*

103   In characterizing overall binding behavior, we found that the top 10% of FRH-

104   binding peptides were enriched in basic residues and FRH binding was skewed towards

105   more positively charged peptides (Fig. 1b and c, Extended Data Fig. 2d, and

106   Supplementary Data 1). This correlated with previously-solved FRH crystal structures that

107   show the surface electrostatic potential of FRH is mainly negative, with the exception of

108   the RNA-binding groove (Extended Data Fig. 2e) [17]. Notably, the relative position of

109   oppositely-charged residues, known as "charge patterning", has emerged as an important

110   sequence feature for highly disordered proteins [18,19]. To evaluate the role of charge

111   patterning in FRQ/FRH binding, we surveyed the linear distribution of Net-Charge Per

112   Residue (NCPR) across FRQ. Positive and negative NCPR scores were clustered on

113   FRQ, forming charge "islands", with FRH binding correlated strongly with positive NCPR

114   islands (Fig. 2a).

115   To quantify positive charge clustering, we used Inverse-Weighted Distance (IWD),

116   which analyzes composition and sequence distribution using single residues (Eq. 2 in

117   methods) [20]. Comparing the IWD of positive residues within FRQ to the IWD calculated

118   from 10,000 randomly shuffled FRQ sequences, we found that positively charged

119   residues in FRQ were significantly clustered (Fig. 2b) [20–22]. Clustering was also seen for

120   negative, but not aromatic, residues (Extended Data Figs. 3a and b). Analysis of IWD

121   across different orthologs, and conversion of IWD into Z-scores, allowed for direct

122   comparison between orthologs [21,22]. Most fungal FRQ orthologs and the functional

123   PERIOD (PER) ortholog proteins in *D. melanogaster*, *M. musculus*, and *H. sapiens* all

124   demonstrated significant positive charge clustering (Fig. 2c and Extended Data Fig. 4a),

125   suggesting that positive charge clustering may be the conserved molecular feature

126   amongst disordered negative arm clock proteins in higher eukaryotes [23].

127   To determine if the temporal phosphorylation of FRQ modulated these charge

128   islands, we calculated NCPR across FRQ using known time-specific phosphorylations [16].

129   Positive charge clusters were in general maintained, even when all detected

130   phosphosites were considered (Extended Data Fig. 3c). Markedly, predicted SLiMs for

131   verified FRQ interactors occurred near negatively charged islands, suggesting a

132   relationship between FRQ/FRH binding and the formation of FRQ macromolecular

133   complexes (Extended Data Figs. 3c and d) [3].

134

135   *Charged islands support the circadian feedback loop*

136   Given the distribution of positively charged islands across FRQ, we hypothesized

137   modulating positively charged residues within an island might impact clock function. To

138 test this hypothesis, we mutated residues in a positively charged island (KKK, a.a. 315-

139 317) that is adjacent to the genetically identified CK-1A interaction site known as FCD-1

140 (a.a. 319-326) [24]. We targeted VHF-tagged (V5, 10-His, 3-Flag) alleles of Wild-type

141 (FRQ^{VHF}) and KKK315AAA (FRQ$^{KKK/AAA}$) substitutions to the cyclosporin (*csr-1*) locus of

142 an *frq* KO strain with a banding and *frq* promoter luciferase-reporter background [25,26].

143 Banding mutants (bd+) lay a conidial "band" once per clock cycle, allowing for the direct

144 analysis of the FRQ$^{KKK/AAA}$ mutation on overt clock rhythms [27]. Conidial band formation

145 demonstrated that, while the FRQ^{VHF} strain maintained a clock, the FRQ$^{KKK/AAA}$ strain had

146 no overt clock rhythms (Fig. 2d and Extended Data Fig. 4b).

147 Banding is a measurement of clock output but is not always representative of the

148 activity of the core clock. Therefore, we next assayed *frq* promoter activity using the

149 luciferase reporter present in the strains as a proxy for the TTFL, as the activation of the

150 *frq* promoter leads to the transcription/translation of luciferase [28]. Compared to the

151 FRQ^{VHF} strain, which showed a typical ~21 hr. oscillation in *frq* promoter activity, the

152 FRQ$^{KKK/AAA}$ strain displayed a non-circadian oscillation (~ 48 hrs) (Fig. 2e and Extended

153 Data Fig. 4c). The loss of clock function due to the KKK to AAA substitution supports a

154 model in which positively charged residues can steer conserved hydrophobic binding

155 motifs on FRQ by generating an appropriate sequence context. To further explore this

156 idea, we turned to a known binding motif in FRQ, the FFD.

157

158 *Hotspot residues are critical for FRH binding*

159 While the overall binding of FRH to FRQ appeared to be related to electrostatics,

160 LOCATE also highlighted FRH binding to FRQ peptides with a net neutral charge (see

161   arrows in Fig. 1c). These peptides mapped to the FFD, a region composed of multiple

162   hydrophobic residues flanked by charged residues (a.a. 774-782) (Extended Data Figs.

163   1b and 5a) [8]. While the FFD SLiM is not strictly conserved, similar physicochemical

164   residues are retained, suggesting an evolutionary chemical signature (Extended Data

165   Figs. 5a and b). When we analyzed the binding of FRH to scrambled FFD peptides that

166   maintained the same amino acid composition and neutral net charge, we found scrambles

167   with diminished binding (Extended Data Fig. 5c). A comparable analysis of the

168   electrostatically-driven interaction near the FCD-1 region (described above, peptide

169   MTDKE**KKK**LVVRRLE, predicted net charge +3) showed minimal changes in binding

170   intensity upon shuffling (Extended Data Fig. 5d). Therefore, the FFD SLiM is distinct from

171   other FRH-binding regions in that it has some degree of binding specificity.

172   SLiM-based protein interactions are often driven by specific residues within the

173   motif, termed "hotspots" [29]. In our LOCATE analysis, we noted a sharp increase in binding

174   when two arginines C-terminal of the canonical FFD motif entered the peptide window

175   (a.a. 783-784, Fig. 3a). Rational mutations of this region showed these arginines were

176   critical for FRQ/FRH binding (Figs. 3b and c, Supplementary Data 1). This is consistent

177   with arginines being a common amino acid in protein interaction regions, and suggested

178   that the double arginines and the FFD region are vital in the FRH SLiM [29,30].

179   To verify the importance of the RR hotspot *in vivo*, we again targeted VHF-tagged

180   (V5, 10-His, 3-Flag) alleles of FRQ to the *csr-1* locus of an *frq* KO strain (Extended Data

181   Fig. 5e) [25,26]. We created a previously published alanine substitution of the FFD (FRQ[FFD2],

182   VMLVTT to AAAAAA), a double arginine to double alanine (RR783AA, FRQ[RR/AA]), and a

183   double arginine to double histidine (RR783HH, FRQ[RR/HH]) mutant strain (Extended Data

184 Fig. 5e) [8]. We tested the ability of these FRQ isoforms to interact with FRH *in vivo* using

185 co-immunoprecipitation (Fig. 4a and Extended Data Fig. 6a). While FRQ[VHF] pulled down

186 ample FRH, as did FRQ[RR/HH], as predicted by LOCATE and previous publications, both

187 FRQ[FFD2] and FRQ[RR/AA] were unable to co-immunoprecipitate FRH (Fig. 4a and Extended

188 Data Fig. 6a) [8].

189   To gain mechanistic insight into how the double arginine region affected the

190 binding of FRQ to FRH, we performed all-atom simulations of a 50 a.a. section of FRQ

191 that centered on the FFD region (Extended Data Figs. 6b-e, Supplementary Movie 1).

192 Simulations of wildtype (WT), RR783AA, RR783HH with a neutral histidine, and RR783HH

193 with a positively charged histidine, were performed. RR to AA and RR to HH variants had

194 a limited impact on the structural ensemble of the region outside of the specific position

195 where the mutations occur (Extended Data Figs. 6b-d). Given the limited impact on the

196 local conformation and that local context of a motif is important for the specificity of

197 interactions, our data suggests that the arginine residues contribute an electrostatic

198 component that works in combination with the hydrophobic residues to facilitate FRH

199 binding to the FFD [31,32]. This implied that FRQ binding to FRH was driven by distributed,

200 multivalent binding sites in which both charged and hydrophobic residues can contribute

201 in ways that depend on both the sequence order and the sequence chemistry.

202

203 *Hotspot residues are essential for clock robustness*

204   The functional consequences of deleting the hydrophobic residues in the FFD are

205 well-established, with a reduction in FRQ/FRH interaction correlated with a loss of FRQ

206 stability [8,15,33]. This phenomenon has been attributed to FRH's function as a Nanny

207   protein, preventing the constitutive degradation of FRQ [15,34,35]. However, we wondered if

208   disrupting electrostatically-mediated interactions in the FFD would lead to distinct

209   outcomes due to the orthogonal alteration of IDR chemistry. We therefore analyzed FRQ

210   stability over time using cycloheximide (CHX) [15,34]. As expected, FRQ^{FFD2} showed a

211   decrease in FRQ stability compared to FRQ^{VHF} (Fig. 4a and b and Extended Data Fig.

212   7a) [8,15]. Conversely, FRQ$^{RR/AA}$ had a longer half-life than FRQ^{VHF} despite the loss of FRH

213   binding (Figs. 4a and b). In addition, we noted an increase in the overall levels of the

214   positive transcriptional activator White-Collar 1 (WC-1, NCU02356) (Fig. 4c and Extended

215   Data Fig. 7b), unusual as WC-1 levels are typically lower in strains where *frq* does not

216   mediate White Collar Complex (WCC) phosphorylation to close the feedback loop [36].

217   Together, these data implied that FRQ can close the negative feedback loop without

218   binding FRH.

219   To determine the ability of FRQ$^{RR/AA}$ to close the TTFL independently of FRH, we

220   followed the banding phenotype of the above strains using a race tube assay. We found

221   the FRQ^{VHF} and FRQ$^{RR/HH}$ strains maintained a typical clock period, whereas the FRQ^{FFD2}

222   and FRQ$^{RR/AA}$ strains were arrhythmic (Fig. 4d and Extended Data Fig. 8a) [8]. However,

223   when we tracked clock functionality at the molecular level using luciferase reporter strains

224   as described above, we found that while the FRQ^{VHF}, FRQ$^{RR/HH}$, and FRQ^{FFD2} strains

225   mirrored their overt phenotypes, the FRQ$^{RR/AA}$ strain showed a re-activation of *frq*

226   promoter activity within the first 24 hrs that dampened on the second day (Fig. 4e and

227   Extended Data Fig. 7c) [25,28]. This is consistent with FRQ$^{RR/AA}$ closing the feedback loop

228   but not robustly reactivating the TTFL. To corroborate this finding, we assessed

229   conidiation in race tubes in a 12hr light:12hr dark (12L:12D) entrainment regime, which

230　will only oscillate if FRQ is able to close the feedback loop. While FRQ^{VHF}, FRQ$^{RR/HH}$ and

231　FRQ^{FFD2} strains again mirrored the overt clock phenotypes, the FRQ$^{RR/AA}$ strain showed

232　banding rhythms, consistent with the ability for FRQ$^{RR/AA}$ to repress the WCC without FRH

233　(Fig. 4f and Extended Data Fig. 8b).

234

235　**Discussion**

236　　　A mechanistic understanding of how the circadian TTFL regulates biology has

237　been hampered by the highly disordered nature of the core clock proteins. Our application

238　of the LOCATE approach allowed us to overcome this hurdle by granting high amino acid

239　resolution analysis of the interaction between FRQ and FRH. This analysis identified a

240　hotspot that generates an electrostatically favorable context for FRH binding and once

241　mutated broke binding with FRH without decreasing FRQ stability (Figs. 4a and b). Stable

242　FRQ allowed us to investigate the role of FRQ in the core clock independently of FRH,

243　demonstrating FRQ alone was able to exert negative feedback onto the WCC to create

244　an hourglass circuit (Extended Data Fig. 9b). In support of this data, FRQ can enter the

245　nucleus without FRH and also interact with the WCC and its principle kinase Casein

246　Kinase 1a (CK-1A) independently of FRH [8,37,38]. In total, our evidence supports a model

247　where FRQ can carry out WCC repression without FRH (Extended Data Fig. 9b). In higher

248　eukaryotes, CRYPTOCHROME (CRY) similarly contributes to clock robustness rather

249　than feedback [39]. The mechanism by which this robustness is imparted will be a focus of

250　future work.

251　　　Beyond the role of FRH in robustness, LOCATE demonstrated that the interaction

252　between FRQ and FRH was principally electrostatically driven by positively charged

253 islands (Fig. 2a). These positively charged islands were a conserved molecular feature

254 across FRQ orthologs in fungi and higher eukaryotes (Fig. 2c and Extended Data Fig. 4a)

255 [16,23]. This parallels work illustrating the importance of electrostatic interactions and

256 multivalency in other complexes involving disordered proteins, including mammalian

257 clock proteins [40–45]. The pattern of multivalent electrostatic islands supports a fuzzy

258 interaction model, where clock proteins interact as a heterogeneous ensemble rather than

259 a single fixed structure and the further investigation of the importance of conformational

260 dynamics in this context is warranted (Extended Data Fig. 9a) [3,6,46,47]. Given the insights

261 garnered from our investigation of clock protein-protein interactions, we propose that the

262 LOCATE method could be employed broadly to investigate interactions between highly

263 disordered proteins, perhaps focusing on the effect of post-translational modifications

264 [16,48,49,3,12]. As clock protein IDRs, and IDRs in general, are potentially druggable targets,

265 the LOCATE method will also be relevant to identify candidate SLiMs within IDRs for drug

266 development [5,7,9].

267

268

269

270

271

272

273

274

275

References

1. Partch, C. L. Orchestration of Circadian Timing by Macromolecular Protein Assemblies. *Journal of Molecular Biology* **432**, 3426–3448 (2020).

2. Hurley, J. M., Loros, J. J. & Dunlap, J. C. Circadian Oscillators: Around the Transcription–Translation Feedback Loop and on to Output. *Trends in Biochemical Sciences* **41**, 834–846 (2016).

3. Pelham, J. F. *et al.* Conformational Changes in the Negative Arm of the Circadian Clock Correlate with Dynamic Interactomes Involved in Post-transcriptionally Regulated Processes. *bioRxiv* 2021.11.20.469315 (2021) doi:10.1101/2021.11.20.469315.

4. Cumberworth, A., Lamour, G., Babu, M. M. & Gsponer, J. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem J* **454**, 361–369 (2013).

5. Van Roey, K. *et al.* Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem. Rev.* **114**, 6733–6778 (2014).

6. Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends in Biochemical Sciences* **33**, 2–8 (2008).

7. Katz, C. *et al.* Studying protein-protein interactions using peptide arrays. *Chem Soc Rev* **40**, 2131–2145 (2011).

8. Guo, J., Cheng, P. & Liu, Y. Functional Significance of FRH in Regulating the Phosphorylation and Stability of Neurospora Circadian Clock Protein FRQ. *The Journal of Biological Chemistry* **285**, 11508–11515 (2010).

298   9. Wójcik, S., Birol, M., Rhoades, E., Miranker, A. D. & Levine, Z. A. Targeting the
299   Intrinsically Disordered Proteome Using Small-Molecule Ligands. *Methods Enzymol*
300   **611**, 703–734 (2018).

301   10. Green, C. B. Circadian Posttranscriptional Regulatory Mechanisms in Mammals. *Cold*
302   *Spring Harb Perspect Biol* **10**, (2018).

303   11. Mosier, A. E. & Hurley, J. M. Circadian Interactomics: How Research Into Protein-
304   Protein Interactions Beyond the Core Clock Has Influenced the Model of Circadian
305   Timekeeping.   *J   Biol   Rhythms*   07487304211014622   (2021)
306   doi:10.1177/07487304211014622.

307   12. Pelham, J. F., Dunlap, J. C. & Hurley, J. M. Intrinsic disorder is an essential
308   characteristic of components in the conserved circadian circuit. *Cell Communication*
309   *and Signaling* **18**, 181 (2020).

310   13. Akashi, M. *et al.* A Positive Role for PERIOD in Mammalian Circadian Gene
311   Expression. *Cell Reports* **7**, 1056–1064 (2014).

312   14. Shastry, D. G. & Karande, P. Microarrays for the screening and identification of
313   carbohydrate-binding peptides. *Analyst* **144**, 7378–7389 (2019).

314   15. Hurley, J. M., Larrondo, L. F., Loros, J. J. & Dunlap, J. C. Conserved RNA helicase
315   FRH acts nonenzymatically to support the intrinsically disordered Neurospora clock
316   protein FRQ. *Molecular cell* **52**, 832–843 (2013).

317   16. Baker, C. L., Kettenbach, A. N., Loros, J. J., Gerber, S. A. & Dunlap, J. C. Quantitative
318   Proteomics Reveals a Dynamic Interactome and Phase-Specific Phosphorylation in
319   the Neurospora Circadian Clock. *Molecular Cell* **34**, 354–363 (2009).

320 17. Conrad, K. S. *et al.* Structure of the frequency - interacting RNA helicase : a protein
321   interaction hub for the circadian clock. *The EMBO Journal* **35**, 4–7 (2016).

322 18. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are
323   influenced by linear sequence distributions of oppositely charged residues. *Proc Natl*
324   *Acad Sci USA* **110**, 13392 (2013).

325 19. Sawle, L. & Ghosh, K. A theoretical method to compute sequence dependent
326   configurational properties in charged polymers and proteins. *J Chem Phys* **143**,
327   085101 (2015).

328 20. Holehouse, A. S., Ginell, G. M., Griffith, D. & Böke, E. Clustering of Aromatic Residues
329   in Prion-like Domains Can Tune the Formation, State, and Organization of
330   Biomolecular Condensates. *Biochemistry* (2021) doi:10.1021/acs.biochem.1c00465.

331 21. Cohan, M. C., Shinn, M. K., Lalmansingh, J. M. & Pappu, R. V. Uncovering non-
332   random sequence patterns within intrinsically disordered proteins. *bioRxiv*
333   2021.08.19.456831 (2021) doi:10.1101/2021.08.19.456831.

334 22. Martin, E. W. & Holehouse, A. S. Intrinsically disordered protein regions and phase
335   separation: sequence determinants of assembly or lack thereof. *Emerg Top Life Sci*
336   **4**, 307–329 (2020).

337 23. Zarin, T. *et al.* Proteome-wide signatures of function in highly diverged intrinsically
338   disordered regions. *eLife* **8**, e46883 (2019).

339 24. Querfurth, C. *et al.* Circadian Conformational Change of the Neurospora Clock Protein
340   FREQUENCY Triggered by Clustered Hyperphosphorylation of a Basic Domain.
341   *Molecular Cell* **43**, 713–722 (2011).

342 25. Bardiya, N. & Shiu, P. K. T. Cyclosporin A-resistance based gene placement system

343   for Neurospora crassa. *Fungal Genetics and Biology* **44**, 307–314 (2007).

344 26. Wang, B., Kettenbach, A. N., Gerber, S. A., Loros, J. J. & Dunlap, J. C. Neurospora

345   WC-1 Recruits SWI/SNF to Remodel frequency and Initiate a Circadian Cycle. *PLOS*

346   *Genetics* **10**, e1004599–e1004599 (2014).

347 27. Sargent, M. L. & Woodward, D. O. Genetic determinants of circadian rhythmicity in

348   Neurospora. *J Bacteriol* **97**, 861–866 (1969).

349 28. Gooch, V. D. *et al.* A kinetic study of the effects of light on circadian rhythmicity of the

350   frq promoter of Neurospora crassa. *Journal of biological rhythms* **29**, 38–48 (2014).

351 29. Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**,

352   1–9 (1998).

353 30. Mendoza, V. L. & Vachet, R. W. Probing protein structure by amino acid-specific

354   covalent labeling and mass spectrometry. *Mass Spectrom Rev* **28**, 785–815 (2009).

355 31. Bugge, K. *et al.* Interactions by Disorder – A Matter of Context. *Frontiers in Molecular*

356   *Biosciences* **7**, 110 (2020).

357 32. Ivarsson, Y. & Jemth, P. Affinity and specificity of motif-based protein–protein

358   interactions. *Current Opinion in Structural Biology* **54**, 26–33 (2019).

359 33. Cheng, P., He, Q., He, Q., Wang, L. & Liu, Y. Regulation of the Neurospora circadian

360   clock by an RNA helicase. *Genes & Development* 234–241 (2005)

361   doi:10.1101/gad.1266805.

362 34. He, Q. *et al.* FWD1-mediated degradation of FREQUENCY in Neurospora establishes

363   a conserved mechanism for circadian clock regulation. *The EMBO Journal* **22**, 4421–

364   4430 (2003).

35. Suskiewicz, M. J., Sussman, J. L., Silman, I. & Shaul, Y. Context-dependent resistance to proteolysis of intrinsically disordered proteins. *Protein Sci* **20**, 1285–1297 (2011).

36. Wang, B., Kettenbach, A. N., Zhou, X., Loros, J. J. & Dunlap, J. C. The Phospho-Code Determining Circadian Feedback Loop Closure and Output in Neurospora. *Molecular Cell* **74**, 771-784.e3 (2019).

37. Cha, J., Yuan, H. & Liu, Y. Regulation of the Activity and Cellular Localization of the Circadian Clock Protein FRQ. *The Journal of Biological Chemistry* **286**, 11469–11478 (2011).

38. Froehlich, A. C., Loros, J. J. & Dunlap, J. C. Rhythmic binding of a WHITE COLLAR-containing complex to the frequency promoter is inhibited by FREQUENCY. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 5914–9 (2003).

39. Putker, M. *et al.* CRYPTOCHROMES confer robustness, not rhythmicity, to circadian timekeeping. *The EMBO Journal* 15 (2021).

40. Sherry, K. P., Das, R. K., Pappu, R. V. & Barrick, D. Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. *Proc Natl Acad Sci USA* **114**, E9243 (2017).

41. Czarna, A. *et al.* Quantitative Analyses of Cryptochrome-mBMAL1 Interactions: MECHANISTIC INSIGHTS INTO THE TRANSCRIPTIONAL REGULATION OF THE MAMMALIAN CIRCADIAN CLOCK. *Journal of Biological Chemistry* **286**, 22414–22425 (2011).

387   42. Holehouse, A. S. Chapter 7 - IDPs and IDRs in biomolecular condensates. in

388   *Intrinsically Disordered Proteins* 209–255 (Elsevier, 2019). doi:10.1016/B978-0-12-

389   816348-1.00007-7.

390   43. Ozber, N. *et al.* Identification of two amino acids in the C-terminal domain of mouse

391   CRY2 essential for PER2 interaction. *BMC Mol Biol* **11**, 69 (2010).

392   44. Xu, H. *et al.* Cryptochrome 1 regulates the circadian clock through dynamic

393   interactions with the BMAL1 C terminus. *Nat Struct Mol Biol* **22**, 476–484 (2015).

394   45. Zhao, J. *et al.* Intrinsically disordered N-terminal domain (NTD) of p53 interacts with

395   mitochondrial PTP regulator Cyclophilin D. *bioRxiv* 2021.07.23.453429 (2021)

396   doi:10.1101/2021.07.23.453429.

397   46. Sharma, R., Raduly, Z., Miskei, M. & Fuxreiter, M. Fuzzy complexes: Specific binding

398   without complete folding. *FEBS Letters* **589**, 2533–2542 (2015).

399   47. Wiggers, F. *et al.* Diffusion of a disordered protein on its folded ligand. *Proc Natl Acad*

400   *Sci USA* **118**, e2106690118 (2021).

401   48. Dittmar, G. *et al.* PRISMA: Protein Interaction Screen on Peptide Matrix Reveals

402   Interaction Footprints and Modifications-Dependent Interactome of Intrinsically

403   Disordered C/EBPβ. *iScience* **13**, 351–370 (2019).

404   49. Tapia, V. *et al.* Evaluating the coupling efficiency of phosphorylated amino acids for

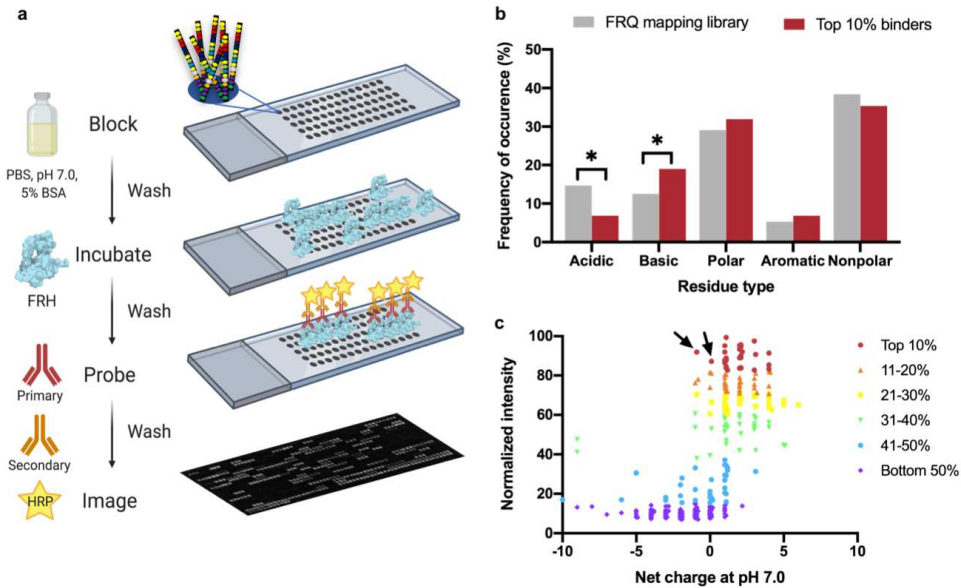405   SPOT synthesis. *J Pept Sci* **14**, 1309–1314 (2008).

406

407

408

409

Figures and Figure Legends

Figure 1. LOCATE reveals for positively charged islands are essential for negative-arm protein-protein interaction. a, The experimental workflow for the LOCATE method; blocking, incubating with FRH, probing using an FRH-specific antibody, and chemiluminescent detection. **b,** Comparison of the amino acid types in the overall 3-mer shift mapping library and the amino acid types in the top 10% of binding peptides (two-tailed z test for population proportions; * = p<0.05). Acidic = D and E; basic = R and K; polar = H, C, N, Q, and S; aromatic = Y, F, and W; and nonpolar = G, A, V, I, L, M, and P. **c,** Ranking of peptides by normalized binding-intensity and estimated net charge at pH 7.0. Peptides coloured by their binding decile as indicated. Arrows denote two peptides containing the FFD with an overall neutral charge. Note that (**b**) and (**c**) are based on

424   average normalized intensities from anti-His and anti-FRH experiments using ~10 nM

425   FRH. Related to Extended Data Figs. 1, 2 and 9.
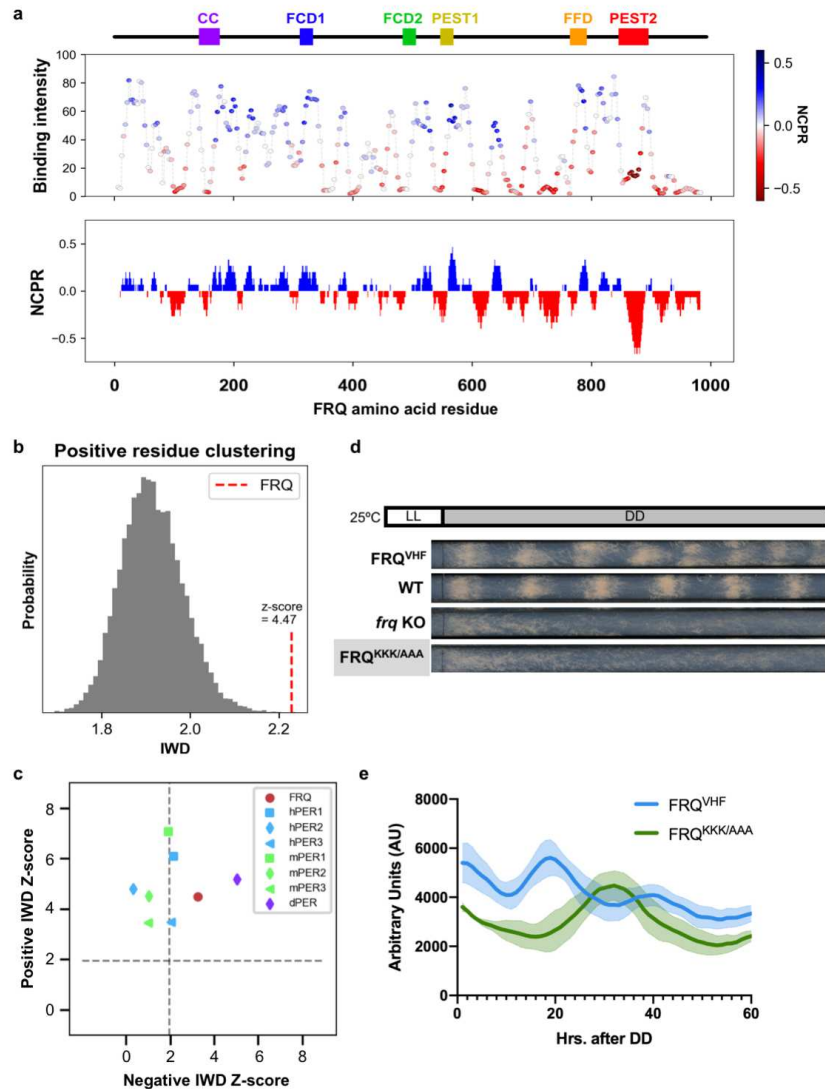
426

427

428

429

430

431

432

433

434

Figure 2. FRQ/FRH electrostatically-driven binding within positively charged islands is essential for clock function. a, Normalized average FRH-binding intensity (~10 nM FRH, anti-His), with peptides colour-coded by net charge per residue (NCPR) as shown in the legend, and local NCPR (using a 15 residue sliding window) plotted against the sequence and known domains of FRQ (abbreviations as in Supplemental Fig. 1). **b,** The Inverse Weighted Distance (IWD) parameter to measure positive residue clustering in FRQ (red) compared to the IWD null distribution (gray) based on 10,000 FRQ sequence shuffles. **c,** The normalized Z-score for the calculated IWD parameter for

444   positive and negative residues for FRQ (red circle) compared to the IWD Z-scores of PER

445   orthologs (see legend). Values above the dashed lines are significant (p<0.05). **d,**

446   Representative race tubes of FRQ^{VHF}, Wild-type FRQ (WT), *frq* KO and FRQ$^{KKK/AAA}$

447   strains grown in DD. **e,** Average and standard deviation values of n = 3 FRQ^{VHF} (tagged

448   wild-type) vs. FRQ$^{KKK/AAA}$ strains with a luciferase reporter of *frq* promoter activity (*Pfrq(c-*

449   *box)::luc*) background grown in constant darkness (DD) in a 96 well format in the presence

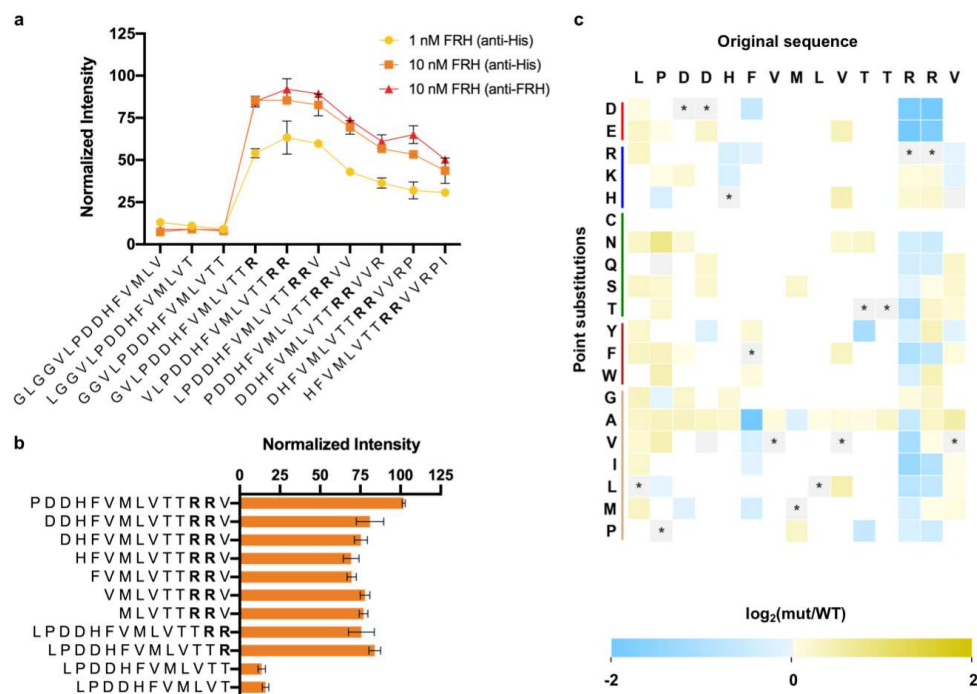450   of luciferin. Related to Extended Data Figs. 3, 4, and 9.

451

452

453

454

455

456

457

458

459

Figure 3. A LOCATE-identified hotspot is vital for FRQ/FRH interaction. a, The average normalized intensity of FRH binding to the stated FRQ-based peptides from Library I. **b,** Average normalized microarray intensity of FRH-binding to different FFD-based peptide truncations. **c,** Point substitution analysis of the FFD SLiM with the original peptide listed along the top and the tested amino acid point substitution along the left-hand side. Grayed boxes with stars denote the wild-type residue at that position, while yellow denotes an increase in FRH associated with a given substitution, while blue denotes a decrease in FRH binding intensity. Note that the scale is $\log_2$ of mutant peptide (mut)/wild type peptide (WT). White boxes correspond to amino acids substitutions that were not tested. Unless otherwise noted, results in this figure are based on Library II peptides, incubated with ~100 nM FRH and visualized with anti-His. Error bars denote standard deviation (n = 2-3 peptide spots). Related to Extended Data Figs. 5 and 9.
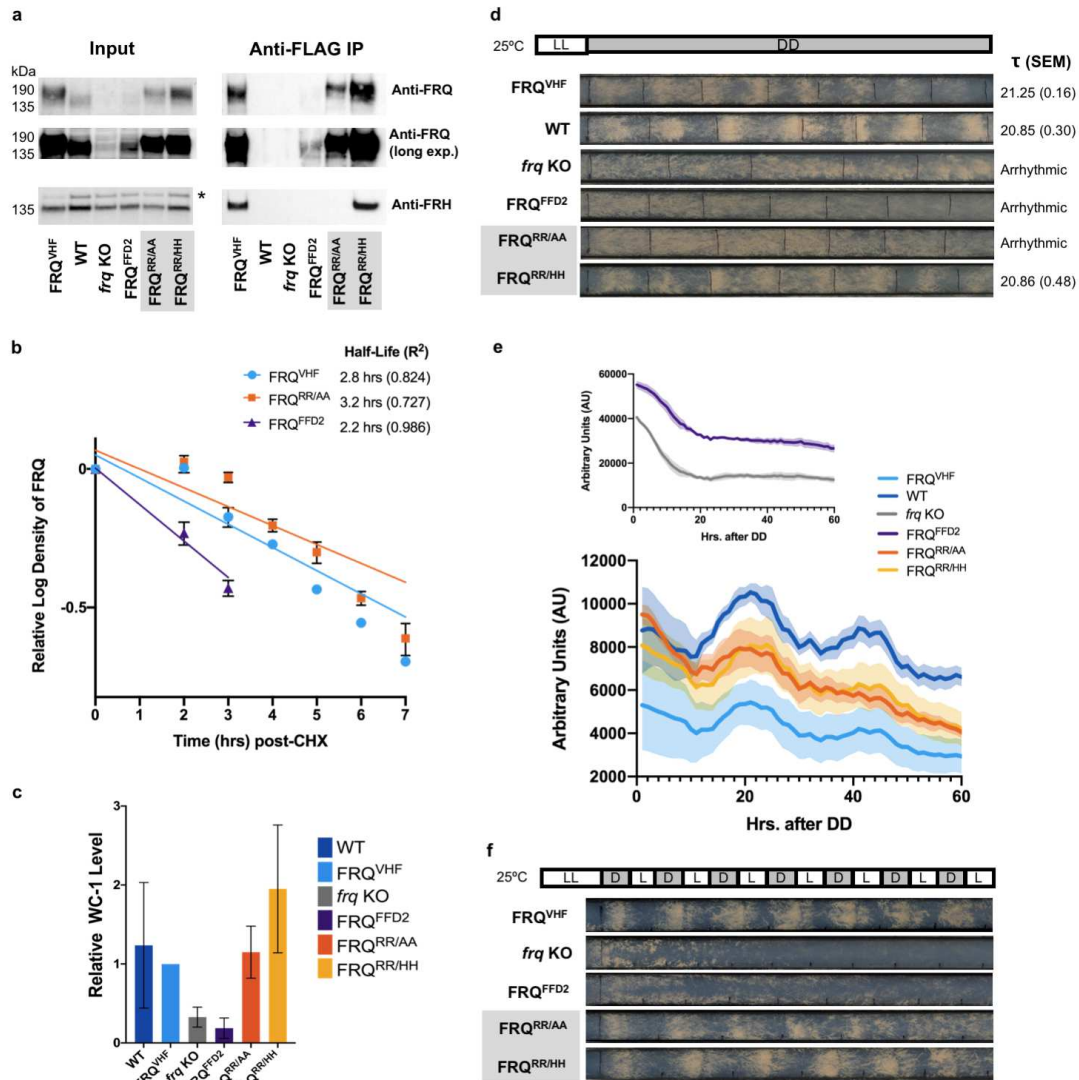
472

473

474 **Figure 4. *In vivo* mutation of the LOCATE-identified hotspot leads to a loss of FRQ-**

475 **FRH interaction and clock robustness. a,** Western blot of the anti-Flag co-

476 immunoprecipitation of the described VHF-tagged FRQ strains. **b,** Semi-quantitative FRQ

477 half-life analysis based on a cycloheximide assay assuming one-phase decay. **c,**

478 Densitometry analysis of WC-1 lysate levels in the described VHF-tagged FRQ strains.

479 Error bars in (**b**) and (**c**) denote standard deviation, n = 3 biological replicates. **d,**

480 Representative race tubes of the VHF-tagged FRQ strains grown in DD. Average period

481   (τ) in hrs. (with SEM) derived from n = 5-6 race tubes. **e,** Average and standard deviation

482   of luciferase levels over 2.5 circadian days of the VHF-tagged FRQ strains with a

483   luciferase reporter for *frq* promoter activity (*Pfrq(c-box)::luc*) (n = 3) grown in DD in a 96

484   well format in the presence of luciferin. Inset to separate scales of expression. **f,**

485   Representative race tubes of FRQ^{VHF}, *frq* KO, FRQ^{FFD2}, FRQ$^{RR/AA}$, and $^{RR/HH}$ strains grown

486   under a 12L:12D lighting regime (n = 5-6). Related to Extended Data Figs. 5-9.

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

Methods

<u>Peptide library design, synthesis, and microarray printing.</u>

504

505

506   Two peptide libraries were designed, with Library I primarily made up of a linear

507   peptide "mapping" of the primary sequence of FRQ (NCU02265) and Library II containing

508   further rationally designed peptides to investigate the specificity of FRH binding to original

509   "parent" peptides (Supplementary Data 1). The FRQ "mapping" peptides in Library I were

510   15 amino acids (a.a.) in length, beginning at the N-terminus of FRQ's sequence, with

511   consecutive peptides shifting by 3 a.a.. Peptide mapping for some regions of interest (eg.

512   FFD region) were repeated at a finer scale of 15 a.a., shifting by 1 a.a. between

513   consecutive peptides. Any parent peptides of interest were used as the basis of further

514   rationally designed peptides, such as scrambled sequences (using the Genscript random

515   library tool; https://www.genscript.com/random_library.html) or truncation series, to

516   identify sequence-specific interactions and the minimal SLiM. Parent peptides containing

517   a candidate SLiM were further singly or doubly mutated to other residues of interest (not

518   all combinatorial possibilities were investigated) to identify permissive or prohibited

519   residues at different positions within the SLiM.

520   Peptides were synthesized using standard Fluoroenylmethyloxycarbonyl (Fmoc)

521   chemistry in an automated peptide synthesizer (Multipep RS, INTAVIS Bioanalytical

522   Instruments AG, Germany), as done previously [1]. Specifically, parallel peptide synthesis

523   proceeded from C- to N-terminus on solid cellulose discs, with all peptides N-terminally

524   acetylated to better mimic the charge of a peptide segment within the parent protein's

525   sequence. Note that the first peptide based on the parent protein's N-terminus was not

526   N-terminally acetylated. See Supplementary Data 1 for more details on the two peptide

527   libraries. Once synthesized, peptide-cellulose discs were reconstituted in 250 μL dimethyl

528   sulfoxide (DMSO) following Intavis' standard work-up procedure. The resulting peptide

529   stock solution were used 1:1 for spotting in triplicate in a microarray format on

530   nitrocellulose-coated glass microscope slides using a slide-spotting robot (Intavis

531   Bioanalytical Instruments AG). Peptide microarrays were then air-dried for 2 hrs. at 65°C.

532

533   <u>FRH Protein Expression and Purification.</u>

534   The FRHΔ100 plasmid [2], consisting of a pET28a vector, FRH (100-1106 a.a.), and

535   an N-terminal 6x His tag, was transformed into BL21 (DE3) Competent *E. coli* cells (New

536   England Biolabs, C2527) and plated on selective LB plates containing 30 μg/ml

537   kanamycin (AMRESCO, 0408-10G). Positive colonies were grown in 3L of liquid LB with

538   30 μg/ml kanamycin at 37°C, 225 rpm, until Abs_{600} was ~0.45. After cooling the culture

539   on ice, protein expression of the FRHΔ100 plasmid was induced with 0.2 mM IPTG

540   (Biotium, 10021), at 18°C, 185 rpm, for 16 hrs. Cells were pelleted by centrifugation at

541   4°C, 4800 rpm, for 15 minutes, and kept on ice for immediate protein extraction. Protein

542   extraction buffers contained 50mM HEPES (Sigma Aldrich, H0887-100 ml) and 150 mM

543   NaCl, pH 7.0, and varied amounts of imidazole (Amresco, 0527-100G). Pelleted cells

544   were resuspended in 30 ml of Lysis Buffer that contained 10 mM imidazole. Cells were

545   lysed by three rounds of French Pressing at 1000 psi. The cell lysate was clarified by

546   centrifugation at 4°C, 14,000 rpm for 20 min. The soluble fraction was split between two

547   columns each with 2 ml Ni-NTA agarose beads (Qiagen, 30210) pre-equilibrated with

548   Lysis Buffer, and nutated for 1 hr. at 4°C. The columns were washed with two rounds of

549   15 ml Wash Buffer containing 30 mM Imidazole. Proteins were eluted in 500 μL fractions

550 using an Elution Buffer with 200 mM imidazole. Each elution was applied to a 40 kDa cut-

551 off Zeba column (ThermoScientific, 87769) to de-salt and exchange the buffer to PBS

552 (100 mM NaCl, 10 mM Potassium Phosphate, pH 7.0). A BSA-based Bradford assay

553 (BioRad, 5000006) was used to quantify resulting amounts of FRH protein, and the

554 percentage of FRH in the final product was visually estimated based on the percentage

555 of the FRH band relative to other bands (~30% FRH) when the eluted protein was

556 visualized on a Coomassie-stained gel (Amresco, 0472-25G). Expression and purification

557 of FRHΔ100 was verified by SDS-PAGE analysis (ThermoScientific, WG1602BOX), and

558 Western blotting with primary anti-His at 1:1000 (Sigma Aldrich, SAB1305538),

559 secondary anti-Mouse at 1:10000 (ThermoFisher, 31430) and Pico (ThermoScientific,

560 34577). Further purification was not considered as the non-specific protein products acted

561 beneficially as a built-in competition assay to decrease false positives for FRH binding.

562

563 Microarray incubation and data analysis.

564   Basic microarray screening protocol was carried out at room temperature on a

565 rocker using 5 mL of each buffer per slide, as in [1]. FRQ-based microarrays were first

566 blocked for 3 hrs. in PBS (10 mM phosphate, 100 mM NaCl, pH 7.0) with 5% w/v BSA,

567 followed by 3 x 10min washes in PBS. Incubation with different approximate

568 concentrations (1, 10, or 100 nM) of the purified and buffer-exchanged FRH occurred for

569 3 hrs followed by another 3x10 min washes and 1 hr. incubation with antibodies (in 2.5%

570 w/v BSA in PBS) with 3x10 min washes between each step, either anti-His at 1:1,000

571 then anti-Mouse at 1:500,000 (see above), or anti-FRH at 1:12,500 [3] followed by anti-

572 Rabbit at 1:500,000 (Invitrogen, 31460), then SuperSignal West FEMTO

573   (ThermoScientific, 34094). Non-specific antibody binding was assessed by incubating

574   with antibodies and FEMTO, but without first incubating with FRH. Chemiluminescence

575   imaging was performed with a ChemiDoc XRS+ System (Bio-Rad) and Image Lab 4.0

576   software, using signal accumulation mode (SAM) with the high-resolution option with 2x2

577   binning.

578   Microarray images were normalized for each library within ImageLab by

579   standardizing the intensity range to allow better comparison amongst replicates, and

580   exported to Fiji (ImageJ v2.0.0, NIH) to convert to 8-bit grayscale, inverted, and

581   background subtracted (rolling ball radius = 50 pixels) [4]. TIGR Spotfinder (Release 2009-

582   08-21) was used to quantify spot intensities using the Otsu segmentation method with

583   local background subtraction (Saeed et al., 2003) (Supplementary Data 1). Basic plots of

584   sequential normalized microarray intensity, truncations, and scrambled peptides were all

585   plotted in PRISM 9.0.2. Compositional analysis of the peptide residues was also carried

586   out as in [1], and plotted in PRISM 9.0.2.. The frequency of occurrence of residue types

587   (acidic, basic, polar, aromatic, or nonpolar) in the 3-mer shift peptide mapping portion of

588   the peptide library were compared to the frequency of occurrence in the top 10% of

589   peptides, i.e., the peptides with the highest normalized binding intensities. The statistical

590   significance of the changes in residue occurrence amongst populations was determined

591   using the test statistic z:

592

593
$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \quad (1)$$

594

595   where $\widehat{p}$ is the percentage of a residue or residue type occurring in the top 10% of

596   peptides; p_0 is the percentage of the same residue or residue type occurring in the

597   peptide mapping portion of the library; and n is the total number of residues in the partial

598   peptide mapping library. For net charge at pH 7.0 calculations, residues D and E were

599   each considered -1 charge, R and K as +1 charge and H as +0.091, while all other residues

600   were considered charge neutral. Resulting peptide charges were plotted in PRISM 9.0.2.

601   Analysis for the FFD region was calculated as a fraction of the WT peptide

602   LPDDHFVMLVTTRRV value of 64, an average of the Library II peptides 215 and 235

603   (Supplementary Data 1), that was then transformed to the $\log_2$ scale.

604

605   Electrostatic potential calculation for FRH.

606   The solved crystal structure of FRH was downloaded from the Protein Data Bank

607   (PDB 5E02; Conrad et al., 2016). Using PyMol (v2.4.0), the file was converted to PQR

608   format and the APBS tool used to calculate the solvent-accessible electrostatic potential

609   according to the Poisson-Boltzmann equation [6]. The electrostatic map was visualized in

610   PyMol, using a gradient of -3 kT/e (red) to +3 kT/e (blue).

611

612   Bioinformatics for BLAST Multiple Sequence Alignment and Cladogram.

613   The *N. crassa* FREQUENCY (NCU02265) protein sequence was run in NCBI

614   BLAST [7]. Homologous protein sequences with over 50% identities or 50% positives with

615   FRQ were chosen for further analysis, yielding a list of ten other homologous proteins.

616   Fungal protein sequences of interest were exported from FungiDB, release 47 [8]. Next the

617   protein sequences were aligned using the UniProt alignment tool and exported in

618   Stockholm format [9]. The multi-sequence alignment was further visualized in SnapGene

619   (version 5.1.1) and a logo was made using the interactive tool Skylign (accessed July

620   2020). Parameters used in Skylign were logos based on the full alignments, letter heights

621   determined by information content above background, and their weighted counts method,

622   where weights are applied to account for highly similar sequences before calculating a

623   maximum-likelihood estimate for each column in a multi-sequence alignment [10]. The

624   relationship amongst the species included in our FRQ homology analysis were

625   represented in a rectangular cladogram, created using the Interactive Tree of Life online

626   tool, ver. 4 [11].

627

628   Generation of FRQ[VHF] and FRQ-FFD mutant strains.

629   An FRQ[VHF] cassette was developed for insertion into the cyclophilin locus of *N.*

630   *crassa* as described [12]. Starting at the 5' end, we fused 1000 bp from the upstream portion

631   of the cyclophilin locus (*csr-1,* NCU00726) from FungiDB to the 3000 bp upstream

632   promoter of the *frq* ORF and the complete wildtype *frq* Open-Reading Frame (ORF),

633   followed by a 10x glycine linker and V5-10His-3Flag (VHF) tag [13], followed by a new stop

634   codon, followed by 1000 bp downstream of the *frq* ORF, and finally 1000 bp from the

635   downstream portion of the target csr-1 locus [8]. Template genomic DNA was harvested

636   from a wildtype *N. crassa* strain 87-3 (bd+, mat a) using the Gentra Puregene Tissue kit

637   (Qiagen, 158622) and following the manufacturer's instructions. PCR reactions using

638   primers found in Table 1 were carried out using Phusion Flash High-Fidelity PCR Master

639   Mix (ThermoFisher, F548S), following the manufacturer's instructions and using an

640   Eppendorf Mastercycler Nexus Thermal Cycler with the following program: 98°C 10 sec,

641 (98°C 5 sec, 67°C 5 sec, 72°C 4 min) x35, 72°C 5 min, 4°C 5 min. Appropriate cDNA

642 product sizes were confirmed using a 0.8% agarose gel in 1X TAE buffer, relative to a 1

643 kb DNA ladder (New England Bio, N3232L). The cDNA pieces with primer overhangs

644 were ligated along with a gapped selective yeast vector (pRS426, containing *URA3* and

645 ampicillin resistance; Hurley et al., 2013) using the homologous recombination system

646 endogenously found in yeast (strain FY834 from the Fungal Genetics Stock Center), by

647 carrying out a Lithium Acetate/ PEG transformation [15,16]. After isolating all plasmid DNA

648 using a "Smash and Grab" protocol [17], the harvested plasmids were transformed into *E.*

649 *coli* (DH5alpha derivative; New England Bio, C2989K) using the manufacturer's

650 instructions (BTX Harvard Apparatus, 45-2001) and plated onto LB agar plates with 100

651 µg/ml ampicillin to grow overnight at 37°C. Colonies were picked and grown in liquid LB

652 broth with 100 µg/ml ampicillin overnight at 37°C and shaken at 225 rpm. After

653 centrifugation of the culture, plasmids were isolated and purified using the QIAprep Spin

654 Miniprep kit (Qiagen, 27106), and PCR of the final linear cassette was done as stated

655 above except using primers MSJ001F and MSJ006R. The cassette was purified using

656 the QIAquick PCR Purification Kit (Qiagen, 28106). Cassette sequences were verified

657 through Sanger sequencing using sequencing primers spaced ~700-800 nt apart

658 (Genewiz and Eurofins Operon, LLC). Verified cassettes were transformed into the *csr-1*

659 locus of an *frq* KO *N. crassa* strain, 122 (bd+, delta-frq::hph+, mat a), using methods

660 previously described [12]. Transformants recovered after electroporation in a liquid VM

661 medium with additional yeast extract and were then plated on an agar plate containing a

662 growth-restrictive mixture of fructose-glucose-sorbose (FGS) and 5 µg/ml of cyclosporin

663 A [12,18]. After 3-4 days of growth at 25°C, isolated colonies were picked and placed onto

664   selective slants (small test tubes with agar-based VM media) and 5 µg/ml cyclosporin A

665   (Sigma, 20024).

666   We designed our FRQ mutations around the concept of maintaining the relative

667   adaptiveness of codons used, since previous studies have shown that FRQ codon usage

668   has effects on its structure and stability [19]. By accessing the codon table for *N. crassa* [20],

669   we calculated the relative adaptiveness of each codon, with 100% being the highest

670   frequency amino acid codon, and then relatively scaled the frequency of the remaining

671   codons [21]. We used this Relative Adaptiveness measure of codons to rank from most

672   adaptive to least adaptive and substituted an alanine or histidine codon of similar rank for

673   our mutations. For our FRQ-FFD mutants, the program PrimerX

674   (http://www.bioinformatics.org/primerx/cgi-bin/DNA_1.cgi) was used to design the

675   needed primers, using the following specifications: Melting temp. 50-85°C, GC content

676   40-60%, length 40-60 bp, 5' flanking region 20-30 bp, 3' flanking region 20-30 bp,

677   terminates in G or C, mutation site at center, and complementary primer pair [22]. Designed

678   primers are found in Table 1.

679

680   Table 1. Primers used in this study synthesized from Integrated DNA Technologies, Inc.

681   IDT. Underlined bases denote the substitutions made for the different FRQ mutations.

Purpose	Primer Name	Forward or Reverse Complement?	Primer Sequence (mutation underlined)
Joining plasmid	MSJ001F	Forward	GGGTTTTCCCAGTCACGACG*GGGTCTGCAG CTGTACCGGG*

pRS426 to *5' csr-1 locus*			
FRQ^{VHF}: *5' csr-1* to *frq* locus	MSJ002F	Forward	*GAACCGTGCTTAATCAGGTA*CGGAAGAGGTTGTTGCGAACAAAG
FRQ^{VHF}: *5' csr-1* to *frq* locus	MSJ002R	Reverse Complement	CTTTGTTCGCAACAACCTCTTCC*GTACCTGATTAAGCACGGTTC*
FRQ^{VHF}: *frq* ORF to *VHF tag*	MSJ003F	Forward	GATGGAGGACGTCTCATCCTCG*GGCGGAGGCGGCGGAGGCGG*
FRQ^{VHF}: *frq* ORF to *VHF tag*	MSJ003R	Reverse Complement	*CCGCCTCCGCCGCCTCCGCC*CGAGGATGAGACGTCCTCCATC
VHF tag template (*Glycine linker*,3Flag/10His/V5 tag).	MSJTag001	Forward	*GGCGGAGGCGGCGGAGGCGGAGGCGGAGGCGG*GTAAGCCTATCCCTAACCCTCTCCTCGGTCTCGATTCTACG/CATCATCACCATCACCATCATCACCACCAC/GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGACTACAAGGATGACGATGACAAG*TAG*
VHF tag to 3' end of *frq* locus	MSJ004F	Forward	*CAAGGATGACGATGACAAGTAG*GACCTGAGTGGGTATTTTTC

VHF tag to 3' end of *frq* locus	MSJ004R	Reverse Complement	GAAAAATACCCACTCAGGTC*CTACTTGTCAT CGTCATCCTTG*
frq locus to 3' *csr-1 locus*	MSJ005F	Forward	GGGCGGGCTACACAGACAGT*CAACGCCTAG ATGAAACCAAATTAC*
frq locus to 3' *csr-1 locus*	MSJ005R	Reverse Complement	*GTAATTTGGTTTCATCTAGGCGTTG*ACTGTC TGTGTAGCCCGCCC
Joining plasmid pRS426 to *3' csr-1 locus*	MSJ006R	Reverse Complement	CAATTTCACACAGGAAACAGC*GCCGACTCG CTTATGAAGCATTG*
FRQ Mutation: VMLVTT777 AAAAAA	MSJ011F	Forward	CTTCCTGACGACCATTTT<u>GCTGCCGCCGCC GCCGCG</u>CGCCGCGTCGTCAGACCTATC
FRQ Mutation: VMLVTT777 AAAAAA	MSJ011R	Reverse Complement	GATAGGTCTGACGACGCGGCG<u>CGCGGCGG CGGCGGCAGC</u>AAAATGGTCGTCAGGAAG
FRQ Mutation: RR783AA	MSJ012F	Forward	GTGATGCTCGTCACCACT<u>GCCGCC</u>GTCGTC AGACCTATCCTG
FRQ Mutation: RR783AA	MSJ012R	Reverse Complement	CAGGATAGGTCTGACGAC<u>GGCGGC</u>AGTGG TGACGAGCATCAC

FRQ Mutation: RR783HH	MSJ013F	Forward	GTGATGCTCGTCACCACT<u>CACCAC</u>GTCGTCAGACCTATCCTG
FRQ Mutation: RR783HH	MSJ013R	Reverse Complement	CAGGATAGGTCTGACGAC<u>GTGGTG</u>AGTGGTGACGAGCATCAC

682

683   For luciferase reporter assays, we created a new strain (uber #6) which was the

684   result of a cross between the X200-3 strain (bd+, his3::pLL26, mat A) where a minimal

685   frq promoter fused to luciferase (Pfrqmin1::luc), and the 122 strain (*frq* KO), to create an

686   *frq* KO strain with a luciferase reporter attached to a minimal frq promoter in the

687   background (Delta-frq::hph+, bd+, his3+::Pfrqmin::luc). For the race tube and Co-IP

688   assays, we used WT strain 328-4 (bd+, mat A).

689

690   <u>Race tube assays and period determination.</u>

691   Race tube assays were carried out using custom glass tubes filled with 14 mL of

692   race tube medium (1x Vogel's salts, 0.05% glucose, 0.1% arginine, 50 ng/ml biotin and

693   1.5% bacto-agar). See fgsc.net/neurosporaportocols/How to choose and prepare

694   media.pdf for more details on Vogel's salts or other Neurospora media. Prepared race

695   tubes were inoculated with 10 µL of a conidial suspension from the noted strain and grown

696   for ~24 hrs at 25ºC in constant light, before being synchronized with a transition to

697   constant dark, 25ºC, and marked each 24hrs until strains reached the end of the race

698   tube. Race tubes were scanned using an EPSON GT-1500 scanner, and images were

699   cropped and converted to black and white. Clock period was analyzed using ChronOSX

700   (v1.1.0) [23], using the Period Analysis option and including 6 days of constant dark

701   densitometry data from each race tube to calculate the mean and standard deviation for

702   each tube, followed by a mean and standard error of the mean for each set of replicates

703   per strain.

704

705   <u>CCD array trials and analysis.</u>

706   Low Nitrogen-CCD media (LN-CCD; 0.03% glucose, 0.05% arginine, 50 ng/ml

707   biotin, 1x Vogel's salts, 1.5% bacto-agar, 25 µM luciferin) with 0.001 M Quinic Acid (pH

708   4.75) was used for all luciferase reporter assays. 185 µL of this media was used per well

709   in a black 96-well plate (Eppendorf, 951040196), inoculated with 10 µL of the relevant

710   conidial suspension. Plates were sealed with a breathable membrane (BreatheEasy, USA

711   Scientific, 9123-6100), and incubated at 25C in constant Light for ~48 hrs before being

712   placed at 25C constant darkness in an incubator with a PIXIS CCD array (Princeton

713   Instruments, 1024B) with a 35 mm Nikon DX lens (AF-S NIKKOR, 1:1 8G), run by the

714   program Lightfield (version 5.2, Princeton Instruments). Images were acquired for 15 min

715   every hour and final image stacks were imported into FIJI (ImageJ v2.0.0, NIH) to adjust

716   brightness, and denoised using the default "Remove outliers" tool option. A custom image

717   analysis plugin was used called "Toolset Image Analysis Larrondo's Lab 1.0" (courtesy of

718   Luis Larrondo) using the 96-well plate quantifying tool. Data were smoothed using a

719   moving average of three timepoints, before plotting in PRISM 9.0.2.

720

721   <u>Co-immunoprecipitation and western blotting.</u>

722   Tissue from each *N. crassa* strain was grown in triplicate, by making a conidial

723   suspension using 1mL of Liquid Culture Media (LCM; 2% Glucose, 0.5% Arginine, 1x

724   Vogel's Salts, 50ng/ml Biotin) to resuspend conidia from a ~1 week old Vogel's minimal

725   media slant. The centrifuged and washed conidia were then inoculated into a 125 mL

726   flask containing 50 mL of LCM and grown at 25°C, 125 rpm in constant light for 48 hrs

727   before harvesting. Harvesting was carried out using vacuum filtration and flash freezing

728   in Liquid Nitrogen before storing at -80°C. Tissue was ground in a mortar and pestle along

729   with Liquid Nitrogen, before extracting tissue lysate using a similar volume of chilled

730   Protein Extraction Buffer was added (pH 7.4, 50 mM HEPES, 137 mM NaCl, 10%

731   Glycerol, 0.4% NP-40 alternative) with 1x HALT protease and phosphatase inhibitor,

732   EDTA-free (87785, ThermoScientific). Co-Immunoprecipitation employed 40 μL of

733   resuspended anti-FLAG magnetic beads (M8823, Sigma) or anti-V5 agarose beads

734   (A7345, Sigma) that were prepared according to manufacturer's instructions for each 3.5

735   - 4.8 mg of total lysate, as measured by Bradford assay. Final volumes were brought up

736   to 1 mL total with further Protein Extraction Buffer with 1x HALT. Samples were incubated

737   overnight at 4°C while nutating. Following placement on a magnetic rack, the flow through

738   was removed and the beads were washed three times with 1 μL of Protein Extraction

739   Buffer. Eluted proteins were retrieved by adding 40 μL of 2x LDS Buffer (Cat #) and boiling

740   for 15 min. at 65°C. Samples were removed to a fresh Eppendorf tube and boiled at 100°C

741   for 5 min. with the addition of 3%v/v ß-mercaptoethanol before freezing at -20°C. Thawed

742   samples were run on precast NuPAGE 3-8% Tris-Acetate gels (Invitrogen, WG1602BOX)

743   following manufacturer's protocol, and transferred to PVDF membrane using a BioRad

744   Trans-Blot Turbo Transfer System (BioRad, 1704150). Membrane was blocked using 5%

745   Milk in PBS buffer with 0.2% Tween-20, Primary antibodies were custom antibodies,

746   courtesy of the Dunlap-Loros Labs at Dartmouth, used at 1:5000 for anti-FRQ [24], 1:12000

747   for anti-FRH [3] or 1:5000 for anti-WC1 [25], in 1% Milk with PBS and 0.2% Tween-20.

748   Secondary antibody was 1:5000 (Goat anti-Rabbit; Invitrogen, 31460) in PBS and 0.2%

749   Tween-20. Western blots were incubated with SuperSignal West FEMTO

750   (ThermoScientific, 34094) or ATTO (ThermoScientific, A38554) for anti-WC-1 blots, and

751   imaged using a Bio-Rad GelDoc imager.

752

753   <u>Cycloheximide assay and semi-quantitative western blotting.</u>

754   The cycloheximide assay was adapted from Hurley et al. [14]. Briefly, conidial

755   suspensions from the designated *N. crassa* strain were inoculated into a petri dish filled

756   with LCM (2% glucose) and grown in constant light at 25°C. After 24-36 hours (dependent

757   upon the growth rate of the strain) plugs were cut from the resulting mycelial mat and

758   placed into individual Erlenmeyer flasks with ~50 ml of LCM (2% glucose). After plugs

759   grew in constant light at 25°C (125 rpm) for a further 24 hours, 40 µg/uL of Cycloheximide

760   (94271, VWR) was added to each culture and this was designated time 0. Samples

761   continued to grow at LL, 25°C, 125 rpm until they were harvested using vacuum filtration

762   at different timepoints (0, 2, 3, 4, 5, 6, 7 hrs post-cycloheximide addition) via flash freezing

763   in liquid nitrogen before storing at -80°C. Protein extraction and standardization was

764   executed as described above. SDS-PAGE and Western blot protocols were carried out

765   as above, except using precast NuPAGE 4-12% Bis-Tris gels (Invitrogen, WG1402BOX),

766   1:5000 anti-V5 primary (Invitrogen, 46-1157) and 1:25000 Goat anti-Mouse (Invitrogen,

767   313430) secondary antibodies, and SuperSignal West FEMTO (ThermoScientific, 34094)

768 or ATTO (ThermoScientific, A38554) in the case of FRQ[FFD2]. Image Lab software (version

769 6.0.1) was used for relative quantification and Amido Black staining was used as a total

770 protein loading control for normalization. Pixel density was quantified from hour 0 to 7 or

771 0 to 3 hours within each western blot (dependent upon the stability of FRQ) and

772 normalized by dividing by the pixel density of a matching area of the Amido Black stained

773 membrane. Normalized data was plotted as a ratio relative to timepoint 0 in PRISM 9.0.2.

774 An exponential fit was made to the biological triplicate timepoints for each strain, and a

775 half-life calculated using first-order decay kinetics [14,26]. For the WC-1 relative lysate levels,

776 pixel density after normalization by the matching area of the Amido Black stained

777 membrane, was plotted as a ratio relative to FRQ[VHF] WC-1 levels, again plotted in PRISM

778 9.0.2.

779

780 <u>Sequence analysis.</u>

781   Overall predictions of FRQ sequence disorder were carried out using the package

782 metapredict, a machine learning-based method that predicts whether a residue is in a

783 disordered region by predicting that residue's consensus score (smoothing using a 5

784 residue window) across multiple disorder predictors [27]. Metapredict considers residues

785 with a predicted consensus value of >0.3 as disordered (meaning that more than 30% of

786 the predictors agreed that the residue was in a disordered region), while residues <0.3

787 are considered ordered. To calculate sequence conservation, we extracted 86

788 orthologous *N. crassa* FRQ sequences from the eggNOG (v5.0) [28], aligned these

789 sequences, then calculated per residue conservation scores by using the approach

790 described by [29], grouping amino acids by residue properties. Using the primary amino

791 acid sequence of FRQ (NCU02265), we calculated the net charge per residue (NCPR)

792 using a sliding window of 15 residues and average Kyte-Doolittle hydropathy using the

793 localCIDER program (v0.1.18) with a sliding window size of 5, meaning the value at each

794 position is based on that residue and the two residues to each side [30]. NCPR was also

795 recalculated at different timepoints including the phosphorylations reported in Baker et al.

796 (2009) as negative charges, and plotted using PRISM 9.0.2. We measured clustering of

797 particular residue classes (*e.g.*, positive, negative, and aromatic residues) within FRQ's

798 sequence by calculating the average inverse weighted distance (IWD). The IWD is

799 defined as:

800
$$IWD = <\frac{1}{d}> = \frac{1}{N_{pairs}} \sum_{i=1}^{N_S-1} \sum_{j=i+1}^{N_S} \frac{1}{S_j - S_i} \quad (2)$$

801 where S is the set of target residue positions, S_i is the i-th element of S, N_S is the

802 number of items in S, and N_{pairs} is the number of pairwise combinations between elements

803 of S [13,2731,32]. The IWD was then compared to the IWD calculated from 10,000 randomly

804 shuffled FRQ sequences to assess significance as IWD values above the 95 percentile,

805 and following the approach presented in [33] we calculated a residue class specific Z-score

806 to allow comparisons across orthologs that differ in length and amino acid composition.

807

808 <u>All-atom simulations.</u>

809 All-atom simulations were performed using the CAMPARI simulation engine with

810 the ABSINTH implicit solvent model [34,35] (http://campari.sourceforge.net/). A 50-residue

811 fragment from FRQ (residues 754 – 803) was simulated in a spherical droplet with a

812 radius of 94 Å. Simulations were performed at 340 K and 15 mM NaCl as has been done

813   previously [36–38]. Ten independent simulations were run for each of the different constructs,

814   which include a wildtype construct, an RR to AA construct, an RR to HH construct with

815   neutral histidine, and an RR to HH construct with a positively charged histidine. Each

816   simulation was run for 126 x 10^6 Monte Carlo steps, with the first 6 x 10^6 discarded as

817   equilibration. Trajectory information was saved every 80,000 steps, such that the final

818   ensembles consist of 15,000 distinct confirmations.

819   Simulations were analyzed using MDTraj and SOURSOP v0.2.0

820   (https://soursop.readthedocs.io/). Secondary structure was calculated using the DSSP

821   algorithm [39]. The solvent accessible surface area was calculated for the sidechains only,

822   using a probe radius of 1.4 Å. To compare changes in the solvent accessible surface area

823   along the sequence necessitates correcting for the intrinsic differences in sidechain

824   volume. To account for this, we performed simulations in which all attractive non-bonded

825   interactions are turned off. In these simulations only repulsive component of the Lennard-

826   Jones potential determines the energetically accessible ensemble. This excluded volume

827   (EV) ensemble allows for the intrinsic SASA of each reside sidechain in the appropriate

828   sequence context to be computed, as done previously [40]. The normalized SASA is then

829   calculated as the ratio of the per-residue SASA from the full simulation divided by the

830   SASA from the EV ensemble. Finally, the change in normalized SASA was computed by

831   calculating the difference between the normalized SASA in the wildtype simulation and

832   each of the variants.

833

834   AlphaFold2 structural modelling.

835   The modeled 50 a.a. portion of FRQ (centered on the FFD region) shown in Figure

836   7A was done using a Google Colab notebook based on a simplified version of AlphaFold

837   v2.1.0   [41],   made   available   at   the   following   website:

838   https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/Alph

839   aFold.ipynb.

840

841   **Data Availability Statement**

842   The authors declare that the data supporting the findings of this study are available within

843   the paper and its extended data and supplementary data files.

844

845   **Code Availability Statement**

846   Data from simulation analysis, subsampled trajectories information, and simulation

847   input   information   can   be   found   at   https://github.com/holehouse-

848   lab/supportingdata/tree/master/2022/jankowski_2022.

849

850   **Methods References**

851   1. Shastry, D. G. & Karande, P. Microarrays for the screening and identification of

852   carbohydrate-binding peptides. *Analyst* **144**, 7378–7389 (2019).

853   2. Conrad, K. S. *et al.* Structure of the frequency - interacting RNA helicase : a protein

854   interaction hub for the circadian clock. *The EMBO Journal* **35**, 4–7 (2016).

855   3. Shi, M., Collett, M., Loros, J. J. & Dunlap, J. C. FRQ-interacting RNA helicase

856   mediates negative and positive feedback in the Neurospora circadian clock. *Genetics*

857   **184**, 351–361 (2010).

858 4. Rueden, C. T. *et al.* ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics* **18**, 529 (2017).

860 5. Saeed, A. I. *et al.* TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–378 (2003).

862 6. Fogolari, F., Brigo, A. & Molinari, H. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J Mol Recognit* **15**, 377–392 (2002).

864 7. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **49**, D10–D17 (2021).

866 8. Basenko, E. Y. *et al.* FungiDB: An Integrated Bioinformatic Resource for Fungi and Oomycetes. *Journal of Fungi* **4**, 39–39 (2018).

868 9. The UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**, D506–D515 (2019).

870 10. Wheeler, T. J., Clements, J. & Finn, R. D. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. *BMC Bioinformatics* **15**, 7 (2014).

873 11. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**, W256–W259 (2019).

875 12. Bardiya, N. & Shiu, P. K. T. Cyclosporin A-resistance based gene placement system for Neurospora crassa. *Fungal Genetics and Biology* **44**, 307–314 (2007).

877 13. Emerson, J. M. *et al.* period-1 encodes an ATP-dependent RNA helicase that influences nutritional compensation of the Neurospora circadian clock. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 15707–15712 (2015).

881 14. Hurley, J. M., Larrondo, L. F., Loros, J. J. & Dunlap, J. C. Conserved RNA helicase

882   FRH acts nonenzymatically to support the intrinsically disordered Neurospora clock

883   protein FRQ. *Molecular cell* **52**, 832–843 (2013).

884 15. Collopy, P. D. *et al.* High-throughput construction of gene deletion cassettes for

885   generation of Neurospora crassa knockout strains. *Methods Mol Biol* **638**, 33–40

886   (2010).

887 16. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS

888   carrier DNA/PEG method. *Nature Protocols* **2**, 31–34 (2007).

889 17. Hoffman, C. S. Preparation of yeast DNA. *Curr Protoc Mol Biol* **Chapter 13**, Unit13.11

890   (2001).

891 18. Park, G. *et al.* High-throughput production of gene replacement mutants in

892   Neurospora crassa. *Methods Mol Biol* **722**, 179–189 (2011).

893 19. Zhou, M. *et al.* Non-optimal codon usage affects expression, structure and function of

894   clock protein FRQ. *Nature* **494**, 111–115 (2013).

895 20. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international

896   DNA sequence databases: status for the year 2000. *Nucleic Acids Res* **28**, 292

897   (2000).

898 21. Fuhrmann, M. *et al.* Monitoring dynamic expression of nuclear genes in

899   Chlamydomonas reinhardtii by using a synthetic luciferase reporter gene. *Plant

900   Molecular Biology* **55**, 869–881 (2004).

901 22. Lapid, C. & Gao, Y. *PrimerX - Automated design of mutagenic primers for site-directed

902   mutagenesis.* (2006).

903   23. Roenneberg, T. & Taylor, W. Automated recordings of bioluminescence with special

904   reference to the analysis of circadian rhythms. *Methods Enzymol* **305**, 104–119

905   (2000).

906   24. Garceau, N. Y., Liu, Y., Loros, J. J. & Dunlap, J. C. Alternative Initiation of Translation

907   and Time-Specific Phosphorylation Yield Multiple Forms of the Essential Clock

908   Protein FREQUENCY. *Cell* **89**, 469–476 (1997).

909   25. Lee, K., Loros, J. J. & Dunlap, J. C. Interconnected feedback loops in the Neurospora

910   circadian system. *Science* **289**, 107–110 (2000).

911   26. Hong, C. I., Ruoff, P., Loros, J. J. & Dunlap, J. C. Closing the circadian negative

912   feedback loop: FRQ-dependent clearance of WC-1 from the nucleus. *Genes &*

913   *Development* **22**, 3196–3204 (2008).

914   27. Emenecker, R. J., Griffith, D. & Holehouse, A. S. Metapredict: a fast, accurate, and

915   easy-to-use predictor of consensus disorder and structure. *Biophysical Journal* **120**,

916   4312–4319 (2021).

917   28. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically

918   annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic*

919   *Acids Research* **47**, D309–D314 (2019).

920   29. Capra, J. A. & Singh, M. Predicting functionally important residues from sequence

921   conservation. *Bioinformatics* **23**, 1875–1882 (2007).

922   30. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. & Pappu, R. V.

923   CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically

924   Disordered Proteins. *Biophys J* **112**, 16–21 (2017).

925 31. Schueler-Furman, O. & Baker, D. Conserved residue clustering and protein structure

926   prediction. *Proteins: Structure, Function, and Bioinformatics* **52**, 225–235 (2003).

927 32. Yang, Z., Deng, X., Liu, Y., Gong, W. & Li, C. Analyses on clustering of the conserved

928   residues at protein-RNA interfaces and its application in binding site identification.

929   *BMC Bioinformatics* **21**, 57 (2020).

930 33. Cohan, M. C., Shinn, M. K., Lalmansingh, J. M. & Pappu, R. V. Uncovering non-

931   random sequence patterns within intrinsically disordered proteins. *bioRxiv*

932   2021.08.19.456831 (2021) doi:10.1101/2021.08.19.456831.

933 34. Vitalis, A. & Pappu, R. V. Chapter 3 Methods for Monte Carlo Simulations of

934   Biomacromolecules. in *Annual Reports in Computational Chemistry* (ed. Wheeler, R.

935   A.) vol. 5 49–76 (Elsevier, 2009).

936 35. Vitalis, A. & Pappu, R. V. ABSINTH: A new continuum solvation model for simulations

937   of polypeptides in aqueous solutions. *Journal of Computational Chemistry* **30**, 673–

938   699 (2009).

939 36. Cubuk, J. *et al.* The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and

940   phase separates with RNA. *Nature Communications* **12**, 1936 (2021).

941 37. Martin, E. W. *et al.* Valence and patterning of aromatic residues determine the phase

942   behavior of prion-like domains. *Science* **367**, 694–699 (2020).

943 38. Sherry, K. P., Das, R. K., Pappu, R. V. & Barrick, D. Control of transcriptional activity

944   by design of charge patterning in the intrinsically disordered RAM region of the Notch

945   receptor. *Proc Natl Acad Sci USA* **114**, E9243 (2017).

946   39. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern
947   recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–
948   2637 (1983).

949   40. Holehouse, A. S., Garai, K., Lyle, N., Vitalis, A. & Pappu, R. V. Quantitative
950   assessments of the distinct contributions of polypeptide backbone amides versus side
951   chain groups to chain expansion via chemical denaturation. *J Am Chem Soc* **137**,
952   2984–2995 (2015).

953   41. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature*
954   **596**, 583–589 (2021).

955   42. Cheng, P., Yang, Y., Heintzen, C. & Liu, Y. Coiled-coil domain-mediated FRQ–FRQ
956   interaction is essential for its circadian clock function in Neurospora. *The EMBO*
957   *Journal* **20**, 101–108 (2001).

958   43. Querfurth, C. *et al.* Circadian Conformational Change of the Neurospora Clock Protein
959   FREQUENCY Triggered by Clustered Hyperphosphorylation of a Basic Domain.
960   *Molecular Cell* **43**, 713–722 (2011).

961   44. Görl, M. *et al.* A PEST-like element in FREQUENCY determines the length of the
962   circadian period in Neurospora crassa. *The EMBO Journal* **20**, 7074–7084 (2001).

963   45. Guo, J., Cheng, P. & Liu, Y. Functional Significance of FRH in Regulating the
964   Phosphorylation and Stability of Neurospora Circadian Clock Protein FRQ. *The*
965   *Journal of Biological Chemistry* **285**, 11508–11515 (2010).

966   46. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of
967   a protein. *J Mol Biol* **157**, 105–132 (1982).

968 47. Baker, C. L., Kettenbach, A. N., Loros, J. J., Gerber, S. A. & Dunlap, J. C. Quantitative

969   Proteomics Reveals a Dynamic Interactome and Phase-Specific Phosphorylation in

970   the Neurospora Circadian Clock. *Molecular Cell* **34**, 354–363 (2009).

971 48. Gouw, M. *et al.* The eukaryotic linear motif resource – 2018 update. *Nucleic Acids*

972   *Research* **46**, D428–D434 (2017).

973 49. Pelham, J. F. *et al.* Conformational Changes in the Negative Arm of the Circadian

974   Clock Correlate with Dynamic Interactomes Involved in Post-transcriptionally

975   Regulated   Processes.   *bioRxiv*   2021.11.20.469315   (2021)

976   doi:10.1101/2021.11.20.469315.

977 50. Ren, J. *et al.* DOG 1.0: illustrator of protein domain structures. *Cell Res* **19**, 271–273

978   (2009).

979

1002

1003
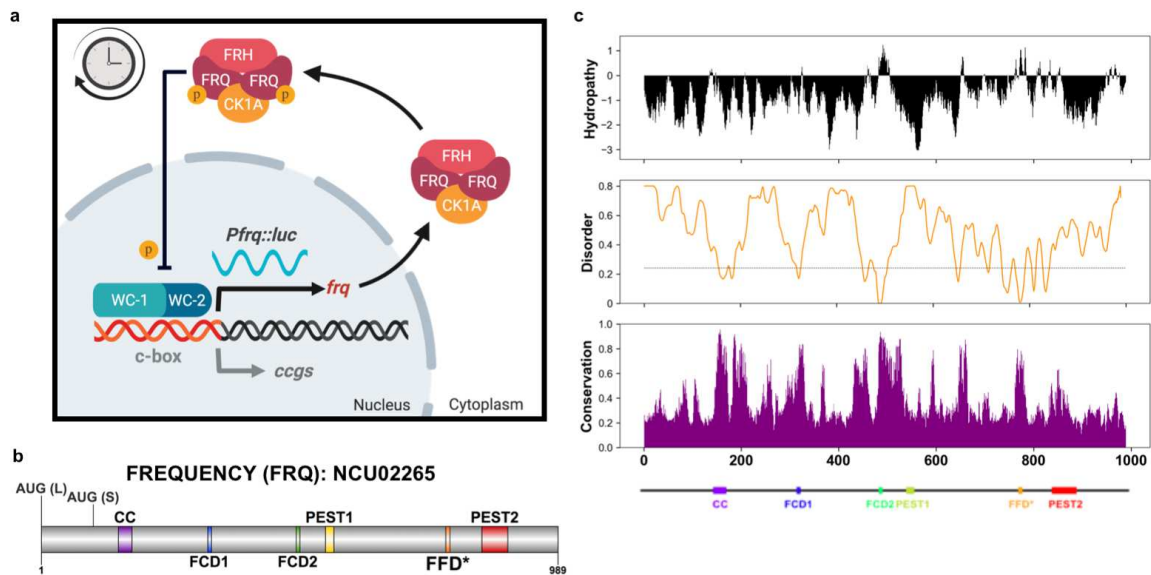
1004

1005

1006

1007

1008

1009

1010

1011

1012

1013 **Extended Data Figures and Legends**

1014

1015

1016 **Extended Data Figure 1. FRQ is a highly disordered protein in the negative arm of**

1017 **the *N. crassa* clock. a,** Schematic of the molecular clock in *N. crassa.* The transcriptional

1018 activators WC-1 and WC-2 activate *frq* transcription. Once translated, FRQ binds CK-1A

1019 and FRH leading to repression of WC-1/WC-2 activity at the *frq* promoter, presumably by

1020 phosphorylation, closing the feedback loop. Robust oscillation of the clock is represented

1021 by the continuous oscillation of luciferase expressed from the *frq* promoter in our reporter

1022 strain. FRQ = FREQUENCY, CK1a = CASEIN KINASE 1a, FRH = FRQ-INTERACTING

1023 RNA HELICASE, *Pfrq* = *frq* promoter, *luc* = *luciferase*, c-box = clock box, *ccgs* = clock-

1024 controlled genes, p = phosphorylation. **b,** FRQ protein topology with known binding

1025 domains highlighted, AUG (L) = Long FRQ start codon, AUG (S) = Short FRQ start codon

1026 [24], CC = Coiled-coil region [42], FCD1/2 = FRQ-CK-1A interaction Domain 1/2 [43], PEST 1/2

1027 = Proline-Glutamic Acid-Serine-Threonine rich region 1/2 [44], FFD = FRQ-FRH interacting

1028 Domain, * = region as defined by [45]). **c,** Linear sequence analysis of FRQ. Top: Linear

1029   hydropathy profile reveals per-residue hydrophobicity based on the Kyte-Doolittle scale

1030   [46]. The majority of the sequence is highly depleted for hydrophobic residues. Middle:

1031   Linear disorder profile calculated using Metrapredict, using with a five-residue smoothing

1032   window [27]. FRQ is predicted to be almost entirely disordered. Bottom: Per-residue

1033   conservation calculated across 83 FRQ orthologs taken from eggNOG. Regions of high

1034   conservation coincide with hydrophobic regions with lower disorder tendencies. Related

1035   to Fig. 1.

1036

1037

1038

1039

1040
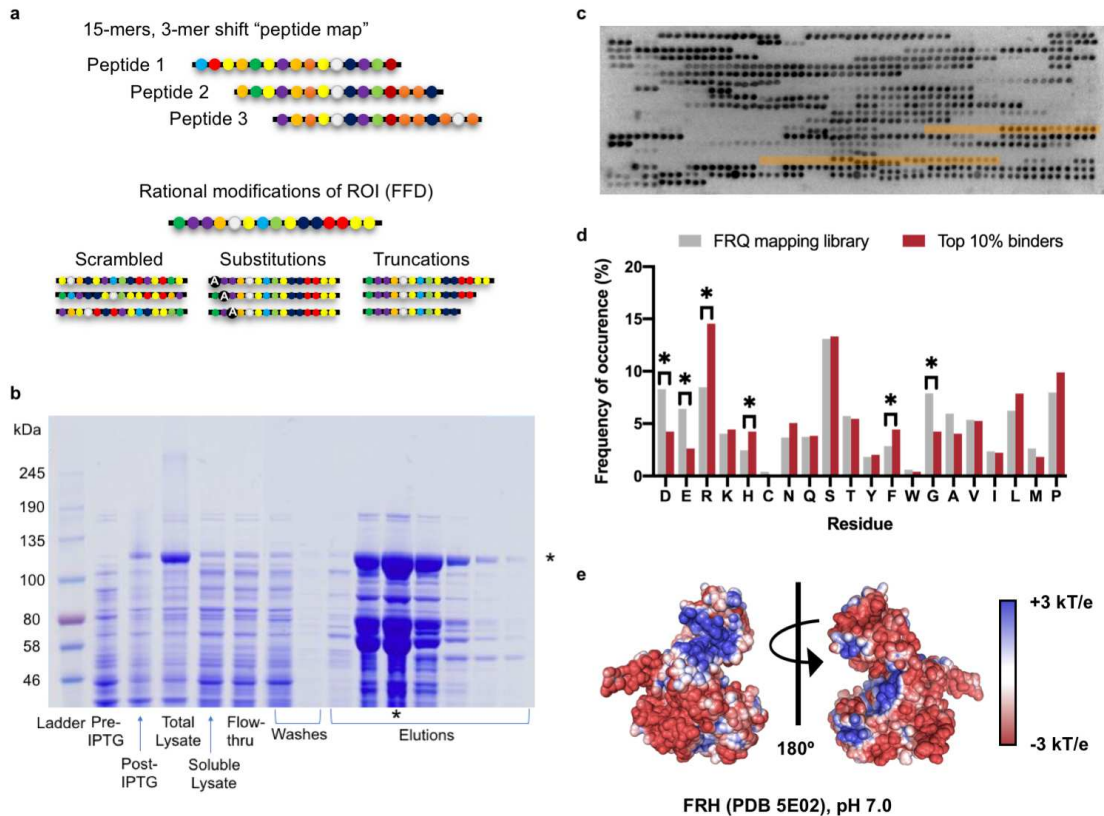
1041

1042

1043 **Extended Data Figure 2. The LOCATE method uncovers charged residues on FRQ**

1044 **are preferred for FRH binding. a,** The design of the FRQ LOCATE peptide library based

1045 on a 15-mer "peptide map" scan, shifting by 3 a.a., through the primary sequence of FRQ.

1046 Rational modifications were designed for Regions of Interest (ROI) using scrambled

1047 sequences, mutations, and truncations. **b,** Coomassie-stained gel showing all steps of

1048 the FRHΔ100 induction (pre- and post-IPTG addition), as well as total versus soluble

1049 lysate after French press and nickel column purification steps (flow-thru, washes and final

1050 elutions). The asterisk (*) denotes the elution used in subsequent microarray assays as

1051 well as the band that corresponds to FRHΔ100. **c,** Example of a FRQ microarray

1052 challenged with ~10 nM FRH and probed with anti-His antibodies. Image inverted for

1053 density analysis of FRH binding. Regions highlighted in orange are peptides including the

1054 FFD region (3-mer shift vs. 1-mer shift). **d,** Comparison of the amino acids in the overall

1055 3-mer shift mapping library and the amino acids in the top 10% of binding peptides (two-

1056 tailed **z** test for population proportions; * = p<0.05). Statistical significance of frequency

1057 changes of residues C, Y, W and M was not determined due to low residue occurrence in

1058 the sample population. **e,** The calculated surface electrostatic potential map for FRH. Blue

1059 denotes positive charge while red denotes negative charge, ranging from +3 kT/e to -3
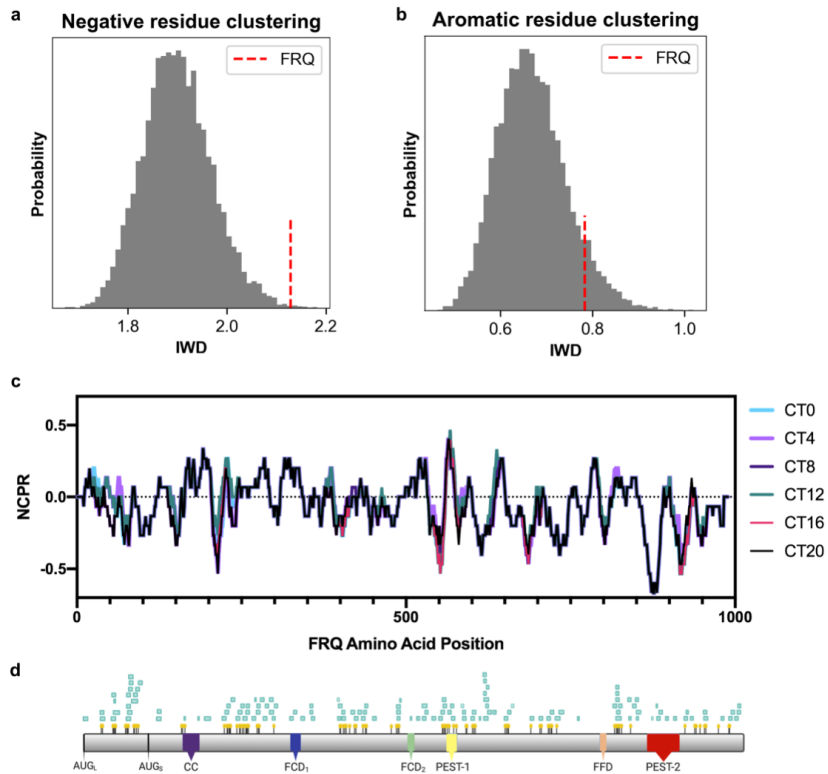
1060 kT/e. Related to Fig. 1.

1061

1062

1063

1064

1065

Extended Data Figure 3. FRQ positive charge islands are anticorrelated with predicted interaction motifs. a, The statistical significance of negatively charged clusters (islands) of residues is assessed by calculating the expected null distribution for negatively charged residue clustering from randomly shuffled of sequences that match the composition and length of FRQ. Clustering is calculated using the Inverse Weighted Distance (IWD) (see methods). The real FRQ sequence (red dashed line) has negatively charged residues that are more well clustered than almost all randomly shuffled sequences, suggesting that at least for FRQ from *N. cassa,* negatively residues (like positively charged residues) are well-clustered. **b,** The statistical significance of aromatic clusters of residues is assessed by calculating the expected null distribution for aromatic residue clustering from randomly shuffled of sequences that match the composition and

1079   length of FRQ. Clustering is calculated using the Inverse Weighted Distance (IWD) (see

1080   methods). The real FRQ sequence (red dashed line) has aromatic residues that are not

1081   more clustered than almost all randomly shuffled sequences. **c,** Net Charge Per Residue

1082   (NCPR) plot of FRQ incorporating extra negative charges at known phosphorylated

1083   residues at the different Circadian Times (CT), as published in [47]. Note that CT = 0 is

1084   relative dawn and CT = 12 is relative dusk. **d,** Predicted SLiMs (green boxes) from the

1085   ELM database for verified interactors of FRQ, along with detected phosphosites denoted

1086   with yellow pins [47–49]. See Extended Data Fig. 1 for more details about the known FRQ

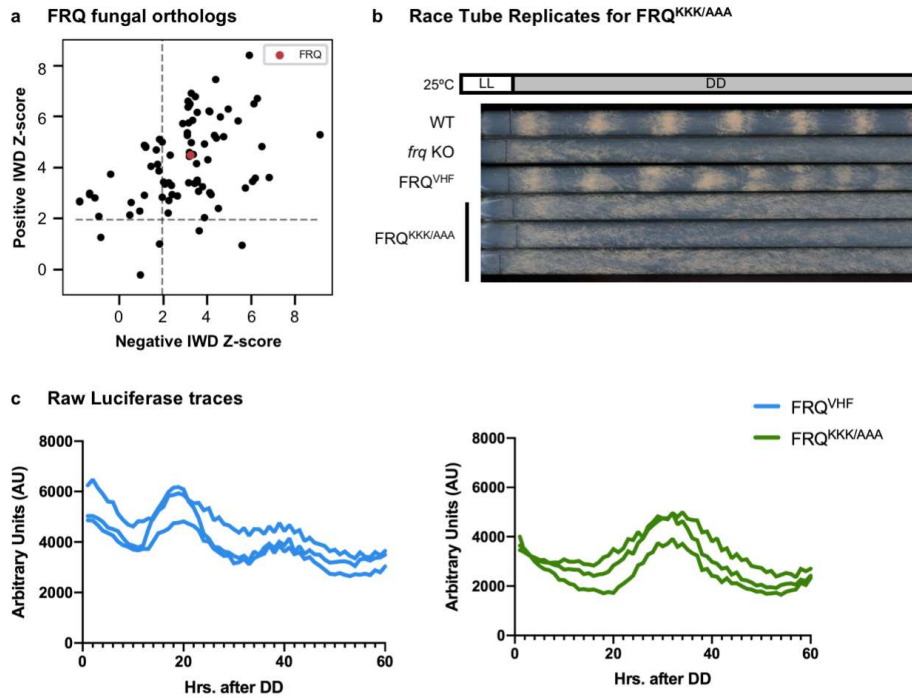1087   domains highlighted in color. Related to Fig. 2.

1088

1089

1090

1091

1092

a FRQ fungal orthologs

b Race Tube Replicates for FRQ^{KKK/AAA}

c Raw Luciferase traces

1093

1094

1095 **Extended Data Figure 4. Positively charged islands are a conserved feature across**

1096 **FRQ fungal orthologs and affect clock output. a,** The Z-score of positive residue

1097 Inverse Weighted Distance (IWD) versus negative residue IWD of FRQ (red circle)

1098 compared to 86 FRQ orthologs (black circles). Dotted lines denote significance (p < 0.05).

1099 Most sequences (64/86) show significant clustering of both positively-charged and

1100 negatively-charged residues (top right quadrant). A subset of sequences (17/86) only

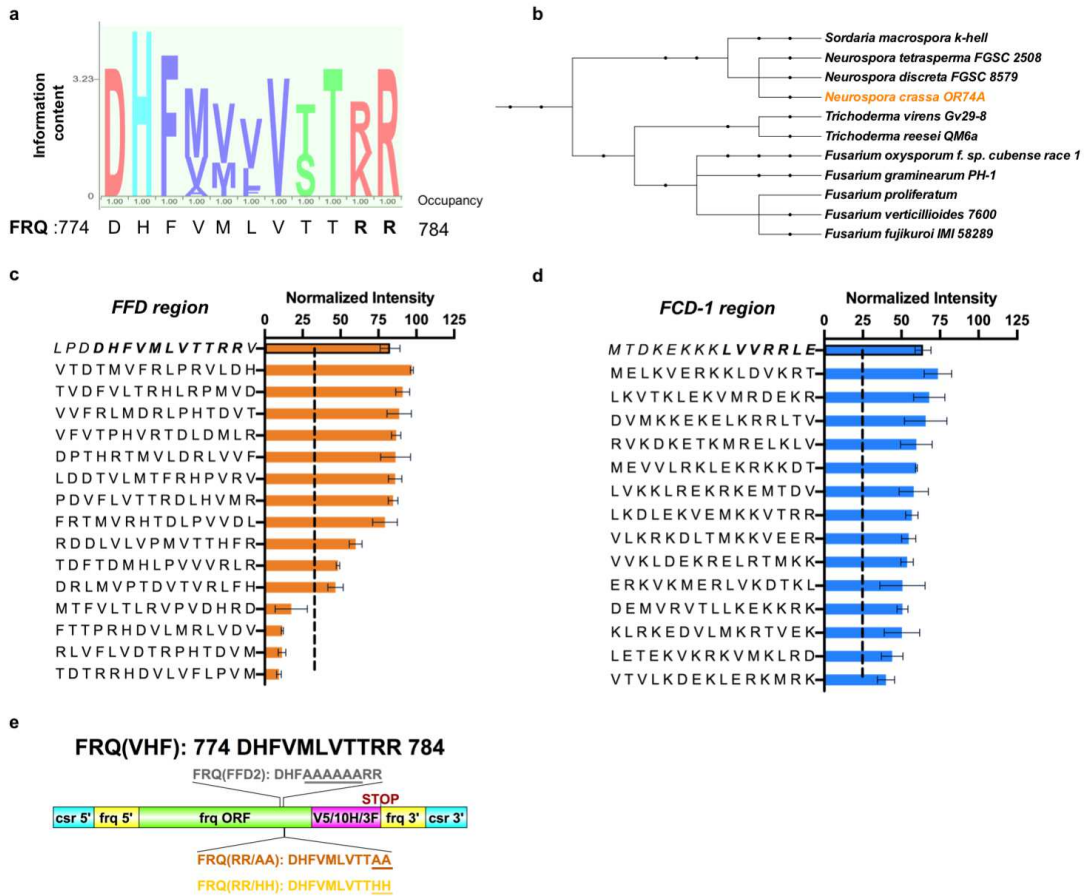1101 show clustering of positively charged residues (top left quadrant). A much smaller subset

1102 show either non-significant clustering (3/86) or only clustering of negatively charged

1103 residues (2/86). **b,** Race tube replicates of the FRQ^KKK/AAA mutant strain. **c,** Raw luciferase

1104 traces (n = 3) from the FRQ^KKK/AAA mutant strain 96-well plate experiment. Related to Fig.

1105 2.

1106

a

Information content

3.23

0

FRQ :774 D H F V M L V T T **R R** 784

Occupancy 1.00

b

Sordaria macrospora k-hell
Neurospora tetrasperma FGSC 2508
Neurospora discreta FGSC 8579
Neurospora crassa OR74A
Trichoderma virens Gv29-8
Trichoderma reesei QM6a
Fusarium oxysporum f. sp. cubense race 1
Fusarium graminearum PH-1
Fusarium proliferatum
Fusarium verticillioides 7600
Fusarium fujikuroi IMI 58289

c

FFD region

Normalized Intensity
0 25 50 75 100 125

LPD **DHFVMLVTTRR** V-
VTDTMVFRLPRVLDH-
TVDFVLTRHLRPMVD-
VVFRLMDRLPHTDVT-
VFVTPHVRTDLDMLR-
DPTHRTMVLDRLVVF-
LDDTVLMTFRHPVRV-
PDVFLVTTRDLHVMR-
FRTMVRHTDLPVVDL-
RDDLVLVPMVTTHFR-
TDFTDMHLPVVVRLR-
DRLMVPTDVTVRLFH-
MTFVLTLRVPVDHRD-
FTTPRHDVLMRLVDV-
RLVFLVDTRPHTDVM-
TDTRRHDVLVFLPVM-

d

FCD-1 region

Normalized Intensity
0 25 50 75 100 125

MTDKEKKK **LVVRRLE**-
MELKVERKKLDVKRT-
LKVTKLEKVMRDEKR-
DVMKKEKELKRRLTV-
RVKDKETKMRELKLV-
MEVVLRKLEKRKKDT-
LVKKLREKRKEMTDV-
LKDLEKVEMKKVTRR-
VLKRKDLTMKKVEER-
VVKLDEKRELRTMKK-
ERKVKMERLVKDTKL-
DEMVRVTLLKEKKRK-
KLRKEDVLMKRTVEK-
LETEKVKRKVMKLRD-
VTVLKDEKLERKMRK-

e

FRQ(VHF): 774 DHFVMLVTTRR 784

FRQ(FFD2): DHFAAAAAARR

STOP

| csr 5' | frq 5' | frq ORF | V5/10H/3F | frq 3' | csr 3' |

FRQ(RR/AA): DHFVMLVTT**AA**
FRQ(RR/HH): DHFVMLVTT**HH**

1107

1108

1109 **Extended Data Figure 5. Sequence composition contributes to the interaction**

1110 **between the FRQ-FFD and FRH. a,** Weighted observed residues at each position for the

1111 FFD region of FRQ over 10 FRQ homologues, with *N. crassa* FRQ residues below. Colour

1112 denotes residue type (pink is charged, blue is histidine or tyrosine, purple is

1113 small/hydrophobic, and green is a hydroxyl/amine following ClustalX scheme).

1114 Occupancy refers to the percent the position was filled. **b,** A phylogenetic tree of the FRQ

1115 orthologs considered in the FFD region alignment shown in (**a**). **c,** Normalized binding

1116 intensity of the native and scrambled peptides of the FRQ-FFD region to FRH, based on

1117 peptides from Library I, using ~10 nM FRH and visualized using anti-His. Peptide

1118 sequences are reported along the y-axis with the native FFD in bold. Normalized binding

1119   intensity values are reported for each peptide along the x-axis and error bars report the

1120   standard deviation (n = 3). The dashed line represents one Standard Deviation above

1121   background. **d,** Normalized binding intensity of the native and scrambled peptides of the

1122   FRQ-FCD-1 region to FRH, based on peptides from Library II, using ~100 nM FRH and

1123   visualized using anti-His. Peptide sequences are reported along the y-axis with the native

1124   FCD-1 in bold. Binding intensity values are reported for each peptide along the x-axis,

1125   and error bars report the standard deviation (n = 3). The dashed line represents one

1126   standard deviation above background-subtracted zero, for the whole library. **e,** Schematic

1127   of the genetic mutations made within the FRQ-FFD region, plotted using DOG (v2.0) [50].

1128   Related to Figs. 3 and 4.

a Co-Immunoprecipitation replicates
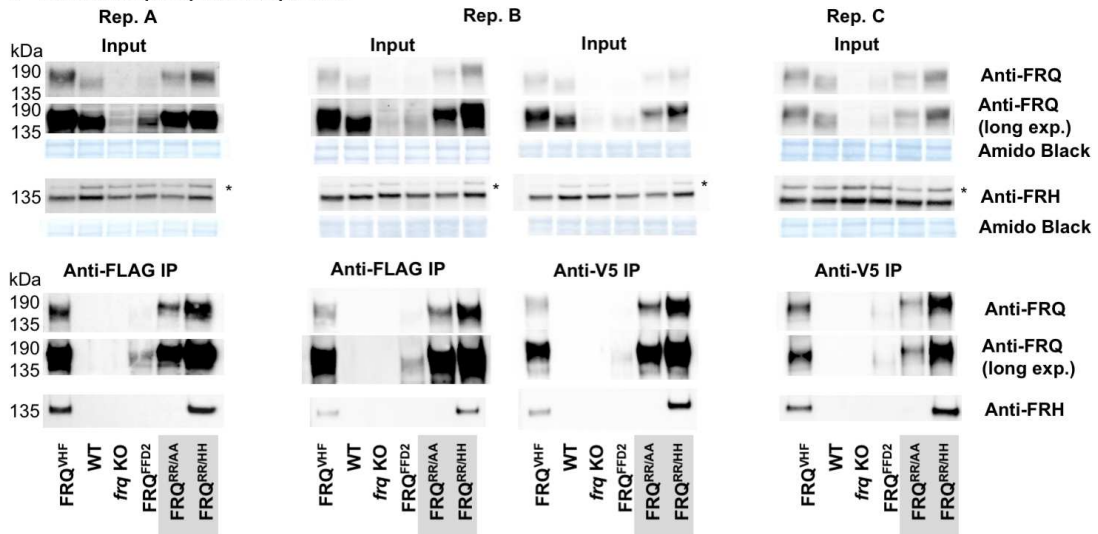
b Solvent Accessible Surface Area (SASA)

Site of Arg mutations

Canonical FFD (774-782)
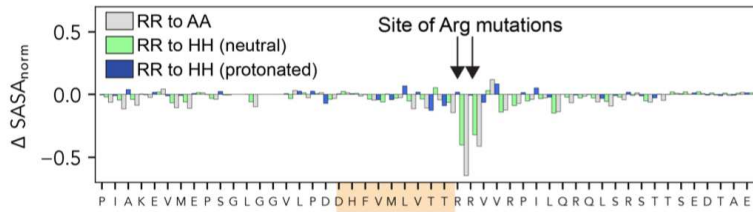
c Transient Helicity
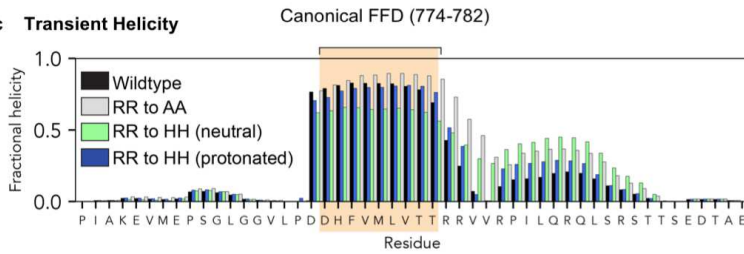
d Global Dimensions
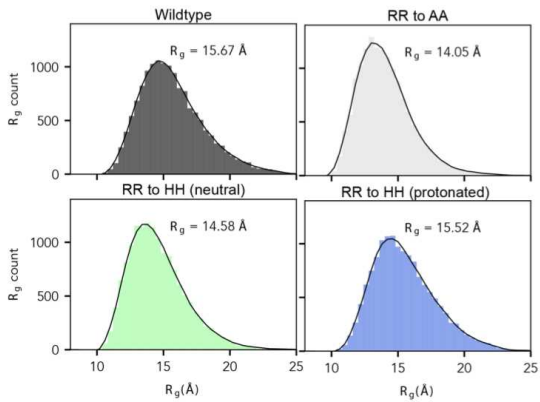
e Simulation Snapshots
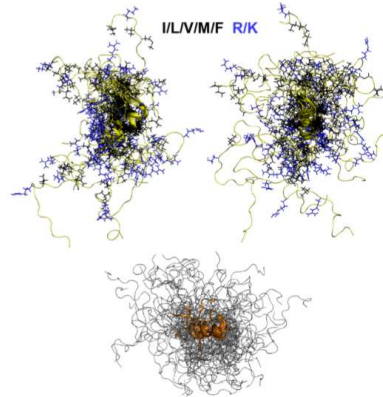
1129

1130

1131

1132

Extended Data Figure 6. Loss of a hotspot leads to loss of FRH interaction without a change in the FRQ FFD microenvironment. a, Biological replicates of the Input and anti-FLAG immunoprecipitation (IP) or anti-V5 IP for the noted strains. Note that Replicate B shows a technical replicate from carrying out anti-FLAG and anti-V5 IPs on the same harvested tissue. Asterisk (*) denotes a non-specific band in the lysate lanes when using anti-FRH. **b,** The change in normalized solvent accessible surface area (SASA) for each residue in a 50-residue FFD region assessed by all-atom Monte Carlo simulations (see methods). Other than the arginine residues themselves, no significant changes in solvent accessibility are observed upon changing the arginine hotspot residues to alanine, neutral histidine, or protonated histidine. The canonical FFD region is highlight in orange. **c,** Transient helicity reveals the canonical FFD lies along a region that is predominantly in a transient helix, a common binding mode for intrinsically disordered regions. RR783 mutations do not alter helicity profiles, with the exception of RR783AA which becomes slightly more helical. **d,** Global dimensions of the four ensembles are assessed based on histograms of the radius of gyration (R_g), a measure of overall ensemble size. All four sequences have almost identical global dimensions. **e,** Snapshots of superimposed structures from a su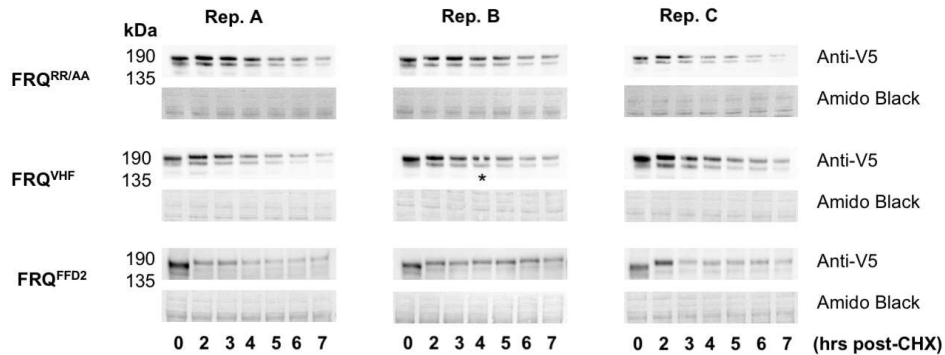bset of the simulations also shown in Supplementary Movie 1. Positively charged residues are shown in blue, hydrophobic residues in black, and the chain backbone in yello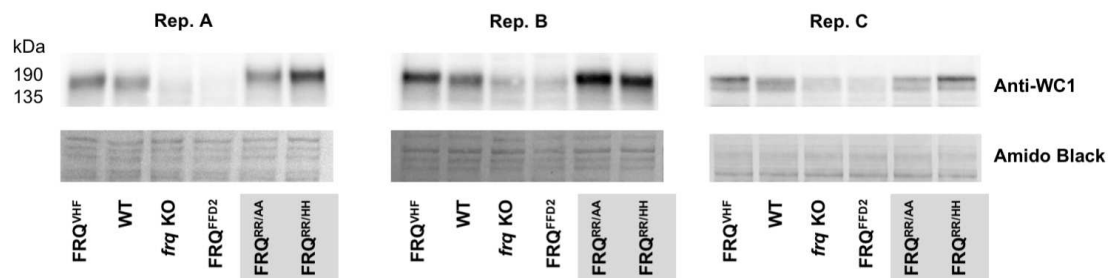w. On the bottom, helical conformations are shown in orange, illustrating the transient and highly disordered nature of the ensemble. Related to Fig. 4.
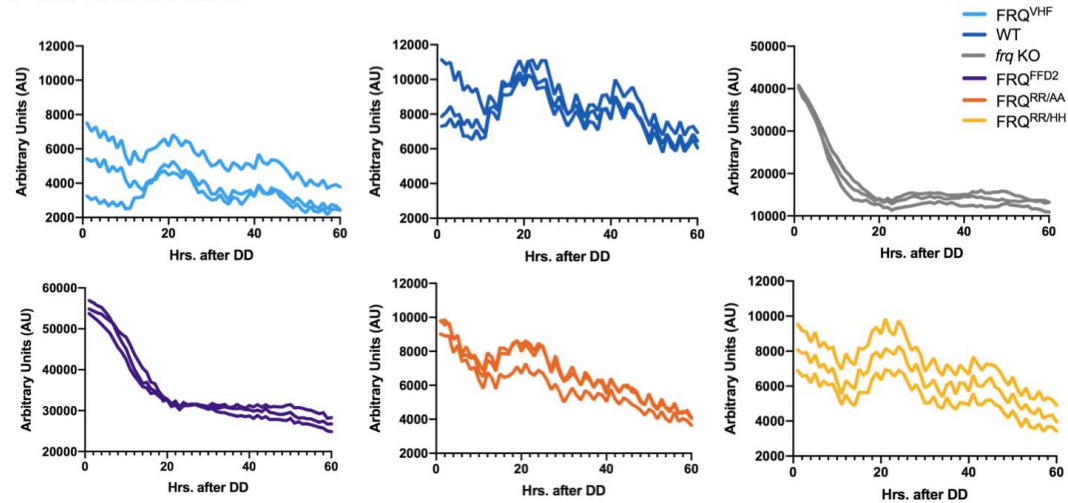
1153

a Cycloheximide assay replicates

b Anti-WC1 lysate replicates

c Raw Luciferase Traces

1154

1155

Extended Data Figure 7. Stable FRQ^RR/AA supports wild-type levels of WC-1, suggesting the closure of the TTFL independently of FRH binding. a, Biological replicate blots and Amido B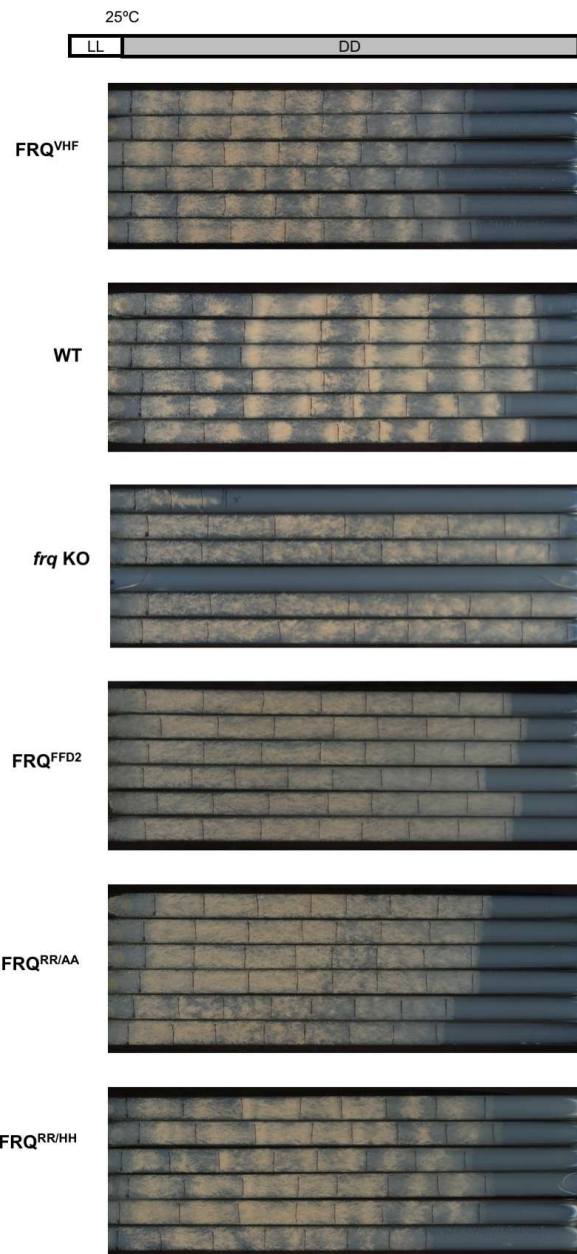lack stained membranes for the cycloheximide assay. Asterisk (*) notes that this lane was omitted from analysis due to an air bubble. **b,** Replicates of

1160   anti-WC-1 lysate levels amongst the different strains, and matching Amido Black stained

1161   membranes showing even loading. Note that replicates A and B were run on Tris-Acetate

1162   gels (3-8%) while replicate C was run on a Bis-Tris gel (4-12%). **c,** Raw Luciferase Traces

1163   (n = 3) for each of the indicated strains, measured in Arbitrary Units (AU) over 15 min.,

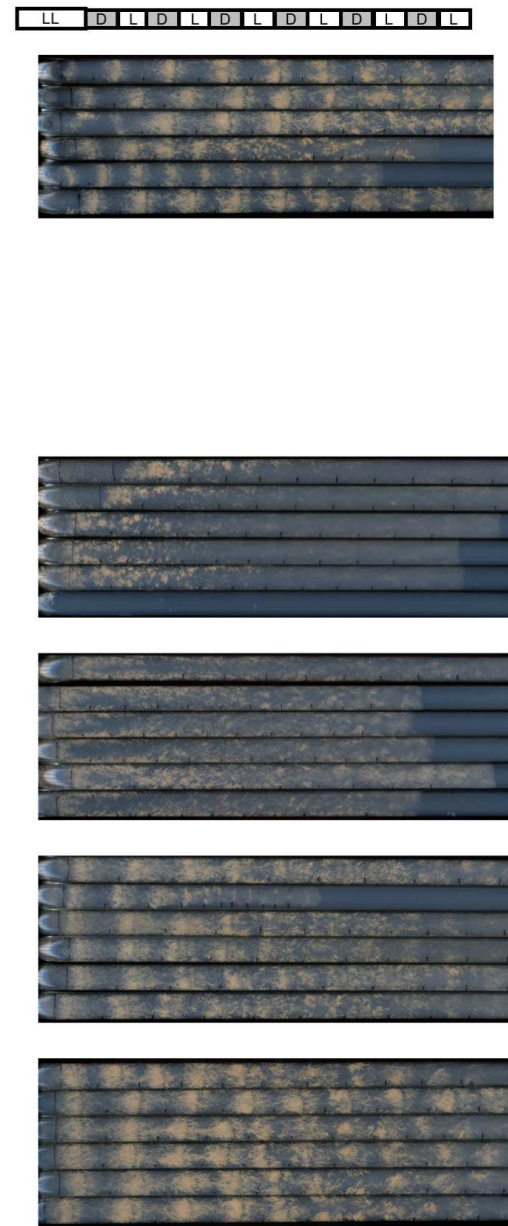1164   repeated every 1 hr for 60 hrs total. Related to Fig. 4.

1165

a   Race tube trials for period determination   **b**   Race Tubes under 12L:12D lighting regime

1166

1167

1168   **Extended Data Figure 8. FRQ$^{RR/AA}$ is overtly arrhythmic in constant conditions yet**

1169   **rhythmic in a 12L:12D lighting regime. a,** Race tubes grown in constant light (LL) before

1170   being allowed to free run in constant dark (DD) for each of the described *N. crassa* strains.

1171   **b,** Race tubes grown in a 12 hr. light: 12 hr. dark (12L:12D) lighting regime for each of the

1172   described *N. crassa* strains. Daily marks are denoted with black lines. Related to Fig. 4.
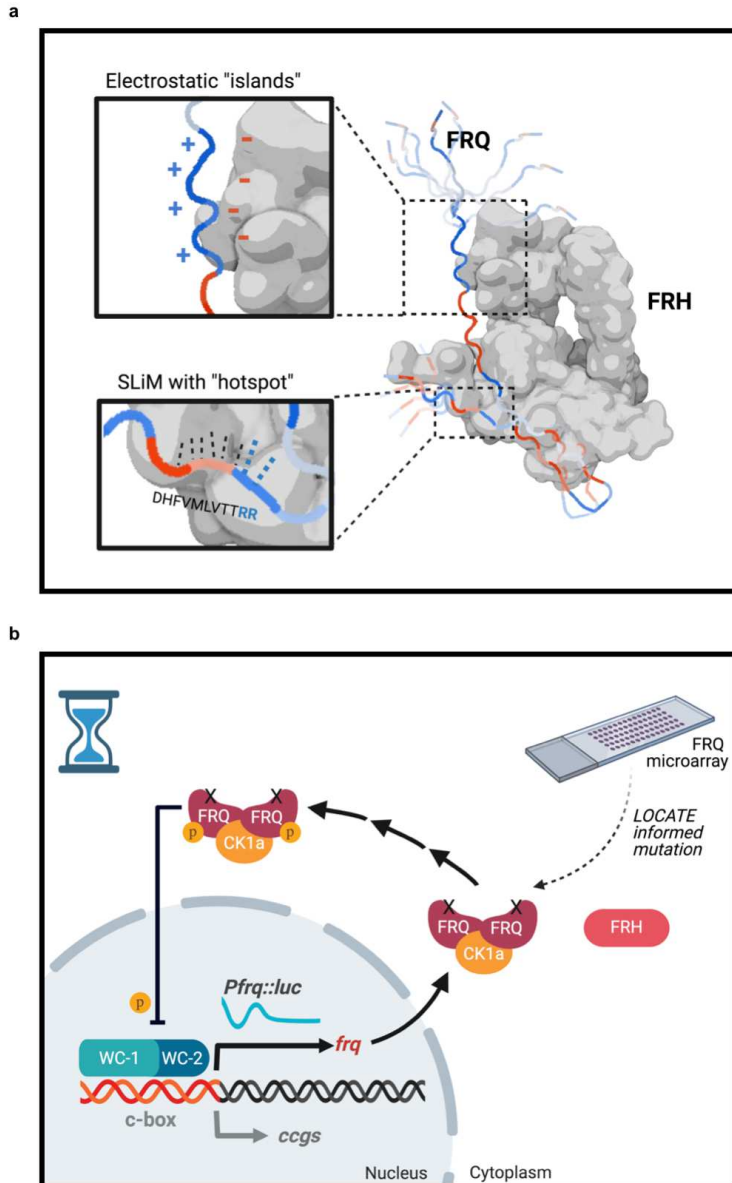
1173

1174

1175

1176

1177

1178

1179

Extended Data Figure 9. Proteins in the negative arm of the clock form a multivalent complex aided by electrostatic islands and SLiM hotspots to support clock robustness. a, Illustration of the fuzzy interaction model using a space-filling model of a known crystal structure of FRH (PDB 4XGT; grey protein) and a portion of FRQ (linear protein) modeled using AlphaFold2 [41], not shown to scale. Blue regions represent positively charged islands and red regions represent negatively charged islands. The top inset shows expected electrostatic interactions between FRQ and the overall negative

1187   outer surface of FRH. The bottom inset shows the specific FFD motif, with the double

1188   arginine hotspot highlighted in blue. **b,** Model based on our LOCATE-informed mutation

1189   FRQ$^{RR/AA}$. In this strain, FRQ is not able to interact with FRH yet can feed back on WC-

1190   1/2 to close the circuit in an hourglass-like manner. The loss of robustness in the clock is

1191   demonstrated by the damping oscillation of luciferase as expressed from the *frq* promoter.

1192   Abbreviations as in Extended Data Fig. 1. Related to Figs. 1-4.

1193

1194   **Supplementary Movie 1. Movie of Monte Carlo all-atom simulation showing the**

1195   **overall disordered nature, albeit with transient helices, that characterizes the FFD**

1196   **region of FRQ.** Residues 754-803 of Wild-type FRQ were simulated using the programs

1197   CAMPARI and ABSINTH (see methods for complete details). Residue and helices

1198   colored as in Extended Data Fig. 6e.

1199

1200

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- movies1.mp4