View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DSpace@MIT

1 **Spatial tethering of kinases to their substrates relaxes evolutionary**
2 **constraints on specificity**

3

4 Emily J. Capra[1], Barrett S. Perchuk[1], Orr Ashenberg[1], Charlotte A. Seid[1], Hana R. Snow[1],
5 Jeffrey M. Skerker[3], Michael T. Laub[1,2,4]

6

7

8 [1]Department of Biology
9 [2]Howard Hughes Medical Institute
10 Massachusetts Institute of Technology
11 Cambridge, MA 02139

12

13 [3]Department of Bioengineering
14 University of California, Berkeley
15 Berkeley, CA 94720

16

17 [4]corresponding author: laub@mit.edu, 617-324-0418

18

19

20

21

22 Keywords: histidine kinase, coevolution, cross-talk, protein evolution

23

24 Running Title: Specificity of hybrid histidine kinases

25

26 ## Summary

27 Signal transduction proteins are often multidomain proteins that arose through the fusion of
28 previously independent proteins. How such a change in the spatial arrangement of proteins
29 impacts their evolution and the selective pressures acting on individual residues is largely
30 unknown. We explored this problem in the context of bacterial two-component signaling
31 pathways, which typically involve a sensor histidine kinase that specifically phosphorylates a
32 single cognate response regulator. Although usually found as separate proteins, these proteins
33 are sometimes fused into a so-called hybrid histidine kinase. Here, we demonstrate that the
34 isolated kinase domains of hybrid kinases exhibit a dramatic reduction in phosphotransfer
35 specificity *in vitro* relative to canonical histidine kinases. However, hybrid kinases
36 phosphotransfer almost exclusively to their covalently attached response regulator domain,
37 whose effective concentration exceeds that of all soluble response regulators. These findings
38 indicate that the fused response regulator in a hybrid kinase normally prevents detrimental
39 cross-talk between pathways. More generally, our results shed light on how the spatial
40 properties of signaling pathways can significantly affect their evolution, with additional
41 implications for the design of synthetic signaling systems.

42

Introduction

Cells can sense and respond to a remarkable diversity of signals and stimuli. This sensory capability typically involves a limited number of signal transduction protein families that have expanded through gene duplication. Although the relative ease of duplication and divergence has enabled cells to dramatically expand their signaling repertoires, the use of highly related signaling proteins has a significant cost, or risk. Cells must avoid detrimental cross-talk and ensure the fidelity of information flow through different signaling pathways. How the specificity of each signaling pathway is determined and how it evolves following gene duplication events are important problems that remain incompletely understood.

In bacteria, the dominant form of signal transduction is known as two-component signaling and typically involves a sensor histidine kinase that can autophosphorylate and then transfer its phosphoryl group to a cognate response regulator, which effects changes in cellular physiology or behavior (Stock *et al.*, 2000) (Fig. 1A). Two-component signaling genes have undergone extensive duplication and horizontal transfer, such that most species possess tens or hundreds of these pathways (Galperin, 2005). Previous work has shown that the interaction between a histidine kinase and its cognate response regulator is highly specific with limited cross-talk between pathways *in vivo* (Capra *et al.*, 2012, Fisher *et al.*, 1996, Grimshaw *et al.*, 1998, Laub & Goulian, 2007, Skerker *et al.*, 2005). This specificity is determined predominantly at the level of molecular recognition rather than relying on cellular factors such as scaffolds. Consequently, a histidine kinase preferentially phosphorylates its cognate response regulator *in vitro*, relative to all other response regulators (Skerker et al., 2005).

Canonical histidine kinases harbor two highly-conserved domains, a dimerization and histidine phosphotransfer (DHp) domain and a catalytic and ATP binding (CA) domain. The DHp domain promotes homodimerization and harbors the histidine that is autophosphorylated by the CA domain. Response regulators also typically have two domains, a receiver domain and an output domain. The receiver domain contains a conserved aspartate that receives a phosphoryl group from the autophosphorylated kinase while the output domains are variable, but are often DNA-binding domains.

71 Phosphotransfer relies primarily on an interaction between the DHp domain of the kinase and
72 the receiver domain of the regulator (Casino *et al.*, 2009). The residues that determine the
73 specificity of this interaction were identified through analyses of amino acid coevolution in
74 large sets of cognate kinase-regulator pairs (Capra *et al.*, 2010, Skerker *et al.*, 2008). These
75 studies pinpointed a small set of strongly coevolving residues that determine the specificity of
76 two-component signaling proteins and that enable the rational rewiring of both the kinase and
77 the regulator (Bell *et al.*, 2010, Capra et al., 2010, Skerker et al., 2008).

78 The coevolution of specificity-determining residues in two-component signaling proteins is
79 driven by negative selection against pathway cross-talk following gene duplication (Capra et
80 al., 2012). The insulation of recently duplicated two-component proteins requires changes in
81 the residues that govern molecular recognition, such that each cognate pair of signaling
82 proteins continues interacting while avoiding cross-talk with the other pathway. In some
83 cases, changes in the specificity residues of other two-component signaling proteins, that were
84 not recently duplicated, are also necessary to achieve a system-wide insulation of all pathways
85 in a given cell (Capra et al., 2012).

86 A common variant of two-component signaling involves hybrid histidine kinases, in which a
87 conventional histidine kinase is fused to a receiver domain similar to those found in soluble
88 response regulators (Fig. 1B). Hybrid kinases autophosphorylate and are thought to transfer
89 the phosphoryl group intramolecularly to their receiver domains. The phosphoryl group can
90 then be transferred to a histidine phosphotransferase and finally to a soluble response
91 regulator, completing a phosphorelay. Hybrid histidine kinases are found in over 50% of all
92 bacterial genomes and nearly 25% of all bacterial histidine kinases are hybrids (Wuichet *et*
93 *al.*, 2010). These hybrid kinases likely arise through the fusion of canonical, co-operonic
94 histidine kinases and response regulators, and may further expand through gene duplication
95 (Whitworth & Cock, 2009, Zhang & Shi, 2005).

96 Despite their prevalence, the phosphotransfer properties and specificity of hybrid kinases are
97 poorly characterized relative to canonical histidine kinases. Here, we investigated the global
98 phosphotransfer specificity of hybrid histidine kinases. We find that these hybrid kinases
99 exhibit significantly reduced phosphotransfer specificity when liberated from their receiver
100 domains. The covalently attached receiver domain thus normally serves as an intramolecular

101 phosphoacceptor and helps prevent unwanted cross-talk inside cells. Our data further indicate
102 that, following the duplication of a hybrid kinase, there is reduced selective pressure to
103 diversify the residues responsible for binding its attached response regulator domain, in stark
104 contrast to canonical histidine kinases. In sum, we propose that the spatial arrangement of
105 domains in hybrid histidine kinases strongly influences the evolution of these proteins with
106 implications for understanding the evolution of multi-domain signaling proteins throughout
107 biology and for designing synthetic circuits.

108

109 ## Results

110 ***Hybrid kinases show reduced amino acid coevolution between kinase and receiver domains***

111 Analyses of amino acid coevolution using mutual information as a metric have helped
112 pinpoint the residues that govern protein-protein interaction specificity in two-component
113 signal transduction systems (Capra et al., 2010, Skerker et al., 2008). These analyses
114 identified a small set of residues that map to the molecular interface formed during
115 phosphotransfer (Casino et al., 2009), and were used to guide the rational rewiring of
116 substrate specificity for the model histidine kinase EnvZ, validating their role in dictating
117 specificity (Skerker et al., 2008). To assess whether the same residues coevolve in hybrid
118 histidine kinases, we examined amino acid coevolution in a large set of hybrid kinases. This
119 analysis was performed on a multiple sequence alignment containing 2681 hybrid histidine
120 kinases, drawn from a wide phylogenetic range of organisms. This sequence alignment
121 contained the DHp and CA domains of each hybrid kinase as well as its receiver domain, but
122 omitted sensory domains. To measure coevolution we used a mutual information-based
123 algorithm that helps adjust for phylogenetic and sampling biases in sequence alignments
124 (Martin *et al.*, 2005). Adjusted MI values were calculated for all possible pairs of positions
125 within the sequence alignment (Fig. 1C, S1A-D). A similar analysis for canonical kinase-
126 regulator pairs was used for comparison (Capra et al., 2010). The two alignments have similar
127 entropy at each position, facilitating a comparison of mutual information scores (Fig S1E-F).

128 We focused primarily on residue pairs in which one position corresponds to a site within the
129 DHp or CA domains and the other to a site within the receiver domain. The overall shape of
130 the distribution of adjusted MI values was similar for the canonical kinase-regulator pairs and
131 the hybrid kinase-receiver domain pairs (Fig. S1C-D). However, the hybrid kinase
132 distribution did not contain the same long tail seen in the canonical distribution. There are 12
133 pairs of amino acids in the canonical kinase-regulator alignment that have adjusted MI values
134 greater than 3.5, which indicates significant coevolution. In contrast, in the hybrid kinase-
135 receiver domain alignment, no residue pair had an MI value greater than 3.5, and only one
136 pair had a value greater than 3.0 (Fig. 1C). The scores for residue pairs in the hybrid kinase
137 alignment were not simply reduced relative to those from the canonical alignment. Of the 12
138 top-scoring residue pairs from the canonical kinase-regulator alignment, only 5 were included

139 in the top 12 scoring pairs from the hybrid kinase alignment. The other 7 had substantially
140 reduced scores, falling throughout the distribution, although each had a positive score (Fig.
141 1D). This analysis suggests that hybrid kinases do not exhibit the same extensive amino acid
142 coevolution between DHp and receiver domains as canonical kinase-regulator pairs.

143 ***Hybrid kinases exhibit limited phosphotransfer specificity***

144 To determine whether the reduced coevolution in hybrid kinases translates into a difference in
145 kinase specificity, we performed phosphotransfer profiling (Skerker et al., 2005). In this
146 approach, a histidine kinase is autophosphorylated using [γ-^{32}P]ATP and then systematically
147 tested for phosphotransfer to a large panel of full-length response regulators or receiver
148 domains, using SDS-PAGE and phosphorimaging. Robust phosphotransfer typically
149 manifests both with a band corresponding to a phosphorylated response regulator and,
150 sometimes, with depletion of the radiolabeled kinase band.

151 We profiled 10 different hybrid kinases from the α-proteobacterium *C. crescentus*. In each
152 case we purified an epitope-tagged construct harboring the DHp and CA domains, but not the
153 receiver domain. We first profiled each kinase against the entire set of receiver domains from
154 the 27 annotated *C. crescentus* hybrid kinases, using incubation times of 15 minutes (Fig. 2A-
155 B, S2). Strikingly, most of the kinases phosphorylated several of the hybrid kinase receiver
156 domains. In fact, some kinases phosphorylated the majority of the receiver domains. These
157 profiles stand in sharp contrast to our results with canonical histidine kinases in which the
158 phosphotransfer profiles were typically extremely sparse, with kinases phosphorylating a
159 single cognate response regulator (Skerker et al., 2008, Skerker et al., 2005).

160 Interestingly, not all of the hybrid histidine kinases phosphorylated their own receiver
161 domains. For example, the kinase CC0723 phosphorylated the receiver domains of CC3075
162 and CC2670, but not its own, even though other hybrid kinases were able to phosphorylate the
163 CC0723 receiver domain. There were also several cases in which a hybrid kinase
164 phosphorylated its own receiver domain, but did so more weakly than other receiver domains.
165 For example, CC3191 phosphorylated the CC0921 receiver domain to a greater extent than its
166 own (Fig. 2A, S4B). Thus, unlike canonical kinases for which the cognate response regulator

167 is usually the kinetically preferred target, hybrid kinases display a variety of behaviors, and
168 often harbor substantially less specificity.

169 Next, we profiled each of the 10 hybrid kinases against the entire set of 44 canonical, soluble
170 response regulators encoded in the *C. crescentus* genome (Fig. 2C, S3). Although these
171 profiles were sparser than those performed against the hybrid kinase receiver domains, there
172 were significant interactions observed with several of response regulators. For instance, the
173 kinase domain of CC2501 showed significant phosphotransfer to the regulators CheYIV,
174 DivK, and CC3015. There were also several response regulators that were phosphorylated by
175 multiple hybrid kinases, including CC0630, CC2576, CC3015, and CC3286. Finally, we
176 noted that two hybrid kinases, CC0723 and CC2324, showed stronger phosphotransfer to
177 CC0630 than to any of the hybrid kinase receiver domains, including their own. These
178 profiles reinforce the conclusion that hybrid kinases exhibit relaxed phosphotransfer
179 specificity and are fundamentally different in this respect from canonical histidine kinases.

180 ***Physical attachment of a receiver domain reduces signaling cross-talk***

181 Although our data demonstrated a reduced specificity of hybrid kinases, these profiles were
182 performed using kinases that had been physically separated from their receiver domains. The
183 kinetic preference and phosphotransfer behavior of these liberated kinase domains likely
184 differ substantially from those of full-length hybrid kinases. For example, although the kinase
185 domain for CC0138 (ShkA) phosphorylated 16 receiver domains and 3 full-length response
186 regulators, previous studies have indicated that ShkA exclusively phosphorylates its own
187 receiver domain *in vivo* (Biondi *et al.*, 2006b). Similarly, although the kinase domain of
188 CC1078 (CckA) showed apparent promiscuity *in vitro* and phosphorylated the response
189 regulator PetR, there is no evidence of cross-talk to this regulator *in vivo* and CckA does not
190 activate PetR-dependent genes *in vivo* (Biondi *et al.*, 2006a). Thus, we propose that the high
191 local concentration of a covalently attached receiver domain normally allows this domain to
192 outcompete other response regulators for access to an autophosphorylated kinase domain.

193 To further probe the effect of covalently attaching a receiver domain to a histidine kinase, we
194 focused on the hybrid kinase CC3191. We first compared the phosphotransfer behavior of the
195 CC3191 construct used in Fig. 2 that harbors the DHp and CA domains to a construct that

196 also contains the C-terminal receiver domain of CC3191. The kinase-only construct for
197 CC3191 phosphorylated its own receiver domain *in vitro*, although it also phosphorylated the
198 soluble response regulator CheYV at a similar rate (Fig. 2A, S4B). In contrast, the longer
199 construct containing the C-terminal receiver domain no longer detectably phosphotransferred
200 to CheYV (Fig. 3A, S4C). This result demonstrates that the receiver domain in a hybrid
201 kinase normally prevents cross-talk between the kinase domain and other, soluble response
202 regulators.

203 The suppression of cross-talk provided by a receiver domain could arise through steric
204 hindrance or because the kinase domain is engaged in intramolecular phosphotransfer. To
205 determine whether productive phosphotransfer contributes, we first generated a full-length
206 CC3191 construct in which the phosphoaccepting aspartate (D563) in the receiver domain
207 was mutated to alanine. This construct exhibited significantly more phosphotransfer to soluble
208 CheYV than the wild-type CC3191 construct, indicating that engagement of the kinase
209 domain in intramolecular phosphotransfer contributes to the suppression of cross-talk (Fig.
210 3B), although the receiver domain may also prevent cross-talk, in part, by occluding the
211 binding of other regulators.

212 To further understand the contribution of a receiver domain to the prevention of cross-talk, we
213 created chimeric hybrid kinases, fusing the kinase domain of CC3191 to a receiver domain
214 from CheYIV or CC1182 (soluble response regulators) or from CC0026 or CC2670 (hybrid
215 kinases). In our profiling studies, the liberated kinase domain of CC3191 had not detectably
216 phosphorylated CheYIV, and had only weakly phosphorylated CC1182 and the receiver
217 domain of CC2670, but it had strongly phosphorylated the receiver domain of CC0026 (Fig.
218 2C). To test whether these four chimeras could phosphotransfer intramolecularly from the
219 CC3191 kinase domain to the heterologous receiver domain attached, we autophosphorylated
220 each in buffer, acid, or base (Fig. 3A). Histidyl-phosphate bonds are sensitive to acid and
221 aspartyl-phosphate bonds are sensitive to base (Fig. S4A). The phosphorylation of CC3191
222 was decreased in the presence of either acid or base, indicating that it was phosphorylated on
223 both the histidine and aspartate. In contrast, the phosphorylation of CC3191(D563A) was
224 primarily acid sensitive. Together, these patterns of acid/base sensitivity indicate that CC3191
225 normally autophosphorylates and transfers its phosphoryl group intramolecularly to its

226 receiver domain. We observed a similar pattern, consistent with intramolecular
227 phosphotransfer, for the chimera CC3191-CC0026 and, to a lesser extent, CC3191-CC2670,
228 but not CC3191-CheYIV or CC3191-1182. These findings are consistent with our results
229 indicating that the CC3191 kinase domain alone can phosphorylate its own receiver domain
230 and the receiver domains of CC0026 and CC2670, but not CC1182 or CheYIV (Fig. 2). These
231 results also indicate that tethering non-cognate receiver domains to a histidine kinase is not
232 always sufficient to promote phosphotransfer.

233 Next, we tested whether the four chimeras would phosphorylate, or cross-talk to, soluble
234 CheYV. All four chimeras showed reduced phosphotransfer to CheYV compared to the
235 CC3191 kinase-only construct (Fig. 3B, S4C), with the strongest suppression of cross-talk
236 occurring with CC3191-CC2670 and CC3191-CC0026, the two chimeras that also
237 demonstrated the most significant intramolecular phosphotransfer. Only the CC3191-CC0026
238 chimera, whose kinase and receiver domains displayed an interaction similar to that of
239 CC3191-CC3191, both in isolation and when fused, completely prevented cross talk. Taken
240 together, our results indicate that the receiver domain of a hybrid histidine kinase plays an
241 important role in reducing, or eliminating, cross-talk with other response regulators by
242 interacting with, and receiving phosphoryl groups from, the linked kinase domain.

243 ***Hybrid kinases lacking their receiver domains likely cross-talk to other response regulators***
244 **in vivo**

245 Previous work has shown that, with only a few exceptions, canonical histidine kinase-
246 response regulator pairs are insulated from each other *in vivo* (Laub & Goulian, 2007, Skerker
247 et al., 2005) and, importantly, that cross-talk between non-cognate pairs can be severely
248 detrimental to an organism's fitness (Capra et al., 2012). We have shown here that many of
249 the hybrid kinases, when separated from their receiver domains, interact readily with
250 noncognate response regulators *in vitro*. Thus, we hypothesized that expressing only the
251 kinase domain of a hybrid histidine kinase might induce cross-talk *in vivo* and affect the
252 growth or fitness of cells.

253 We tested this hypothesis by inducing expression of CC3191 lacking its C-terminal receiver
254 domain in *C. crescentus* and assessing cellular growth in swarm plates. Wild-type *C.*

255 *crescentus* cells can swim through low-percentage agar, creating a large circular colony, or
256 swarm; defects in motility, chemotaxis, cell growth, or cell division can affect swarm size,
257 making this a convenient assay for assessing gross cellular phenotype (Skerker et al., 2005).
258 We found that cells producing the kinase-only portion of CC3191 produced a small swarm
259 relative to the wild type without affecting growth or morphology. This observation is
260 consistent with the notion that a kinase-only version of CC3191 inappropriately
261 phosphotransfers to CheYV *in vivo*, as it does *in vitro* (Fig. 2C). In contrast, cells synthesizing
262 either a full-length construct that contains the receiver domain or the receiver domain alone
263 did not exhibit significant swarm phenotypes (Fig. 3C-D). The phenotype seen with cells
264 expressing the kinase portion of CC3191 was dependent on autophosphorylation, as cells
265 overexpressing a construct in which the conserved histidine was mutated to an alanine no
266 longer exhibited a severe swarm phenotype.

267 We then tested the effects of overexpressing three other hybrid histidine kinases that we
268 profiled above: CC0026, CC0138, and CC2670. Like CC3191, these kinases do not contain
269 transmembrane domains. As with CC3191, overproducing the N-terminal and kinase domains
270 of CC0138 and CC2670 led to a small swarm phenotype, whereas constructs containing both
271 the kinase and receiver domains, or the receiver domain alone, did not (Fig. 3D, S4D). For the
272 kinase-only constructs of CC0138 and CC2670, the phenotype was suppressed by substituting
273 the phosphorylatable histidine with an alanine suggesting that autokinase activity is required
274 for the small swarm phenotype. Unlike CC0138 and CC2670, cells synthesizing the kinase-
275 only version of CC0026 did not exhibit a significant swarm phenotype. Notably, however, the
276 kinase domain of CC0026 had not significantly phosphorylated any non-hybrid receiver
277 domains *in vitro* (Fig. 2C). Taken together, these data are consistent with the idea that some
278 hybrid kinases are promiscuous, but that their attached receiver domains normally help to
279 prevent cross-talk with other response regulators *in vivo*.

280 ***Hybrid histidine kinases are under reduced selective pressure to diversify***

281 Collectively, our results indicate that hybrid histidine kinases are subject to different selective
282 pressures than canonical histidine kinases. We previously found that canonical histidine
283 kinases and response regulators are under strong selective pressure to diversify their
284 specificity residues following gene duplication, but are otherwise relatively static (Capra et

285 al., 2012). This diversification of specificity residues post-duplication is critical to preventing
286 cross-talk and ultimately ensures the system-wide optimization of phosphotransfer specificity
287 (Capra & Laub, 2012, Capra et al., 2012). Consistently, inspection of the six key specificity
288 residues (those from α-helix 1 in the DHp domain) in genome-wide sets of canonical histidine
289 kinases indicates fewer than three identities at these six positions in most pairwise
290 comparisons (Fig. S5).

291 We extracted the corresponding six residues from each of 24 hybrid histidine kinases in *C.*
292 *crescentus* (Fig. S5). Although there are 27 annotated hybrid kinases that contain CA and
293 receiver domains, 3 did not have intact DHp domains. Strikingly, many of the 24 hybrid
294 kinases share four, five, or even six identities at these positions with other hybrid kinases.
295 This similarity does not arise simply because the hybrid kinases duplicated recently, as
296 pairwise comparisons of the entire DHp and CA domains demonstrated extensive variability
297 at other sites (Fig. S1E-F), resulting in significant separation in a neighbor-joining tree built
298 from those domains (Fig. 4A).

299 The lack of variability at the sites corresponding to the six key specificity residues in
300 canonical kinases was also evident in sequence logos for the 24 hybrid and 21 canonical
301 kinases from *C. crescentus* (Fig. 4B). The logo for canonical kinases indicated relatively low
302 conservation at each specificity position except the first, which may be constrained due to
303 involvement in autophosphorylation (Capra et al., 2010, Casino *et al.*, 2010). In contrast, the
304 logo for hybrid kinases indicated higher conservation at each site.

305 The kinase domains of hybrid histidine kinases are likely under less selective pressure than
306 canonical kinases to diversify following gene duplication. The effective concentration of the
307 attached receiver domain is high enough to ensure that a hybrid kinase will transfer its
308 phosphoryl group intramolecularly and not to another regulator or receiver domain. Hence,
309 after duplication of a hybrid kinase, the residues that bind to the receiver domain do not need
310 to change to insulate the new proteins from one another, as occurs in canonical kinases (Fig.
311 5). Consistent with this hypothesis, many of the hybrid histidine kinases in *C. crescentus,*
312 which were likely derived from a common ancestral gene through duplication and divergence,
313 had similar specificity residues and exhibited similar phosphotransfer profiles when liberated
314 from their receiver domains (Fig. 2B). One exception to this trend was CC1078 (CckA),

315 which had a distinct set of specificity residues relative to the other hybrid kinases and,
316 consequently, had a significantly different phosphotransfer profile. Notably, CckA did not
317 group with the other hybrid kinases in a tree of *Caulobacter* kinases (Fig. 4A) suggesting that
318 CckA may be relatively ancient and not derived from a recent duplication.

319

320 **Discussion**

321 The expansion of existing signaling protein families has enabled cells to rapidly evolve the
322 ability to sense and response to a wide range of stimuli. In bacteria, two-component signaling
323 proteins have expanded dramatically, such that most species encode dozens, and sometimes
324 hundreds, of these proteins. For canonical pathways involving a single histidine kinase and
325 response regulator, these pathways are exquisitely specific and a cognate response regulator
326 can outcompete all other non-cognate regulators to receive phosphoryl groups from a given
327 histidine kinase. Consequently, phosphotransfer profiles of canonical kinases have
328 demonstrated that each possesses a strong kinetic preference for its cognate substrate (Skerker
329 et al., 2005). This preference is determined by a small number of specificity-determining
330 residues in both the kinase and regulator. These residues must coevolve to maintain a tight,
331 specific interaction between cognate partners, particularly after a gene duplication event as a
332 means of insulating the new pathways from one another (Fig. 5) (Capra et al., 2012).

333 In contrast to the canonical systems, we demonstrated here that kinase domains of hybrid
334 kinases typically exhibit relaxed substrate specificity, often phosphorylating soluble response
335 regulators or other receiver domains as well or better than they phosphorylate their own
336 receiver domains. A similar observation was made previously in *Myxococcus xanthus* with a
337 limited set of response regulators. In that case, the kinase domain of RodK was shown to
338 preferentially phosphorylate the soluble regulator RokA relative to its own receiver domain,
339 RodK-R3 even though the latter is the *in vivo* target of RodK (Wegener-Feldbrugge &
340 Sogaard-Andersen, 2009).

341 Although hybrid kinases are more promiscuous on their own, our data indicate that the
342 covalently attached receiver domain helps to prevent cross-talk with other cytoplasmic
343 response regulators. The local concentration of an attached receiver domain likely exceeds the
344 concentration of all soluble response regulators quite significantly. Consequently,
345 intramolecular phosphotransfer from the kinase domain to the attached receiver domain will
346 be strongly favored, thereby ensuring minimal cross-talk to other pathways.

347 The enforcement of intramolecular phosphotransfer specificity through spatial tethering of
348 domains likely eliminates selective pressure to diversify the residues in a hybrid kinase that

349 mediate docking to the receiver domain. Hence, after a hybrid kinase duplicates, these
350 residues either will not change or will change more rarely through processes such as genetic
351 drift (Fig. 4B). The net result of the reduced rate of change is that for hybrid kinases in extant
352 organisms, the interfacial residues show substantially reduced variability compared to the
353 same set of residues in canonical histidine kinases.

354 The enforcement of phosphotransfer within hybrid kinases has also likely reduced the need
355 for their kinase and receiver domains to coevolve (Fig. 1). Mutations that reduce or weaken
356 the interaction of these domains are probably more easily tolerated because the domains are
357 spatially tethered. By contrast, with canonical two-component pathways, the cognate proteins
358 are under strong pressure to coevolve, as a means of maintaining their interaction and
359 preventing interaction with non-cognate proteins. However, merely increasing the effective
360 concentration of a receiver domain was not always sufficient to induce phosphotransfer from
361 a kinase domain (Fig. 3A) indicating some requirement for molecular recognition and a
362 proper pairing of interfacial residues. It may be that the fusion of domains in a hybrid kinase
363 serves primarily to prevent cross talk, rather than driving phosphotransfer.

364 Why some two-component pathways involve hybrid histidine kinases instead of canonical
365 kinases is not clear. Hybrid kinases are often involved in phosphorelays, and the additional
366 number of components in a phosphorelay may create additional points for integrating signals
367 (Burbulys *et al.*, 1991). However, not all hybrid kinases necessarily participate in
368 phosphorelays. Recent work with the hybrid kinase VirA from *Agrobacterium tumefaciens*
369 suggests that the receiver domain binds the response regulator VirG, somehow stimulating its
370 activity as a transcriptional activator (Wise *et al.*, 2010). There are also hybrid kinases in
371 some Gram-positive bacteria, such as *Bacteroides thetaiotaomicron*, that have DNA-binding
372 domains C-terminal to their receiver domains, suggesting that these kinases may directly
373 regulate transcription (Raghavan & Groisman, 2010). In short, although nearly a quarter of all
374 kinases are of the hybrid variety, our understanding of their functions, properties, and
375 advantages remains limited.

376 The notion that spatial proximity can overcome relaxed specificity of signaling proteins is
377 relevant in all cells. Multi-domain signaling proteins are quite common, particularly in
378 eukaryotes. Additionally, some signal transduction proteins are spatially constrained through

379 the action of scaffolds. For example, in the *S. cerevisiae* pheromone pathway, the scaffold
380 Ste5 enforces the proximity of three separate MAP kinases, helping to prevent them from
381 inappropriately phosphorylating other substrates (Choi *et al.*, 1994). This spatial
382 colocalization may, in turn, have relaxed evolutionary constraints on these MAP kinases.

383 Finally, our results suggest that information flow through two-component pathways could be
384 rationally engineered by fusing together non-cognate kinases and regulators. Such an
385 arrangement can also prevent unwanted cross-talk with other pathways. Indeed, we showed
386 here that fusing heterologous receiver domains to a hybrid kinase was, in some cases,
387 sufficient to allow phosphotransfer and prevent cross-talk with a soluble regulator. Synthetic
388 scaffolds that bring non-cognate two-component signaling proteins in close proximity may
389 also be used to promote phosphotransfer or prevent cross-talk. A similar approach of
390 artificially colocalizing proteins has been applied in metabolic engineering studies, where
391 enzymes have been tethered together to enhance the synthesis and yield of desired compounds
392 (Dueber *et al.*, 2009).

393 In sum, our work has revealed new aspects of signaling protein evolution in bacteria that will
394 likely inform similar evolutionary studies in other organisms and help guide efforts to
395 construct synthetic signaling circuits.

396

397 # Experimental Procedures

398 ***Sequence analyses***

399 Histidine kinase and response regulator receiver domains were identified, aligned, and filtered
400 as described previously (Capra et al., 2010). Hybrid kinases were defined as those proteins
401 that had a single match to each of the three Pfam models: HisKA, HATPase_C, and
402 Response_reg. The final alignment included 2681 hybrid kinases. Shannon entropy values
403 were calculated for each position in the alignment. Mutual information for every pair of
404 columns in the sequence alignment was calculated as previously reported (9). Raw and
405 adjusted MI values are provided in Tables S1 and S2 and multiple sequence alignments are
406 provided as Supporting Information. Sequence logos were built using WebLogo
407 (weblogo.berkeley.edu). Neighbor-joining trees were built using the PHYLIP package and
408 multiple sequence alignments built from the DHp domain of each canonical and hybrid
409 histidine kinase in the *C. crescentus* genome.

410 ***Strain construction and growth conditions***

411 *E. coli* and *C. crescentus* strains were grown as described previously (Skerker et al., 2005).
412 Primers used are listed in Table S3. Full-length hybrid kinases and the kinase domains of
413 hybrid kinases were amplified from genomic CB15N DNA and ligated into the Gateway
414 pENTR vector (Invitrogen). Chimeric hybrid kinases were cloned by separately amplifying
415 the kinase domain from CC3191 and the specified receiver domain, amplifying the chimeric
416 sequence using splicing with overlap extension PCR and ligating the resulting product into
417 pENTR. pENTR clones were moved into pDEST-His_6-MBP or pDEST-TRX-His_6 vectors for
418 purification, or the pDEST-P_{xyl}-M2 vector derived from pJS71 for overexpression studies.
419 Overexpression vectors were introduced into wild-type CB15N via electroporation.

420 ***Protein purification and phosphotransfer assays***

421 Expression, protein purification, and phosphotransfer profiling experiments were carried out
422 as described previously (Capra et al., 2012, Skerker et al., 2008, Skerker et al., 2005, Biondi
423 et al., 2006a). All reactions used 500 μM ATP, and 0.5 μCi/μL [γ–^{32}P]ATP. For
424 phosphotransfer experiments in Fig. 3A, CC3191-HK was autophosphorylated under the same

425 conditions as the phosphotransfer profiles and then incubated with the given receiver domain
426 in a 1:1 ratio for the time indicated. For phosphotransfer experiments in Fig. 3C, 2.5 μM of
427 the specified kinase was mixed with 2.5 μM CheYV before ATP was added the reaction
428 allowed to proceed for the indicated time before being stopped with the addition of 4X
429 loading buffer. To test acid or base stability of phosphoryl groups, 5 μM of kinase was
430 autophosphorylated at room temperature for 15 minutes. The reaction was then stopped by the
431 addition of 4X loading buffer, and then buffer, 1 M HCl or 0.5 M NaOH was added. After 20
432 minutes, reactions were neutralized. All phosphotransfer experiments were analyzed by SDS-
433 PAGE and phosphorimaging.

434

435 **Acknowledgements**

436 We thank Anna Podgornaia for helpful comments on the experiments and on the manuscript.
437 This work was supported by an NSF CAREER award to MTL and an NSF GRFP award to
438 EJC. MTL is an Early Career Investigator at the Howard Hughes Medical Institute.

439

440 ## References

441 Bell, C. H., S. L. Porter, A. Strawson, D. I. Stuart & J. P. Armitage, (2010) Using structural
442 information to change the phosphotransfer specificity of a two-component chemotaxis
443 signalling complex. *PLoS Biol* **8**: e1000306.
444 Biondi, E. G., S. J. Reisinger, J. M. Skerker, M. Arif, B. S. Perchuk, K. R. Ryan & M. T.
445 Laub, (2006a) Regulation of the bacterial cell cycle by an integrated genetic circuit.
446 *Nature* **444**: 899-904.
447 Biondi, E. G., J. M. Skerker, M. Arif, M. S. Prasol, B. S. Perchuk & M. T. Laub, (2006b) A
448 phosphorelay system controls stalk biogenesis during cell cycle progression in
449 Caulobacter crescentus. *Mol Microbiol* **59**: 386-401.
450 Burbulys, D., K. A. Trach & J. A. Hoch, (1991) Initiation of sporulation in B. subtilis is
451 controlled by a multicomponent phosphorelay. *Cell* **64**: 545-552.
452 Capra, E. J. & M. T. Laub, (2012) Evolution of Two-Component Signal Transduction
453 Systems. *Annu Rev Microbiol*.
454 Capra, E. J., B. S. Perchuk, E. A. Lubin, O. Ashenberg, J. M. Skerker & M. T. Laub, (2010)
455 Systematic dissection and trajectory-scanning mutagenesis of the molecular interface
456 that ensures specificity of two-component signaling pathways. *PLoS Genet* **6**:
457 e1001220.
458 Capra, E. J., B. S. Perchuk, J. M. Skerker & M. T. Laub, (2012) Adaptive Mutations that
459 Prevent Crosstalk Enable the Expansion of Paralogous Signaling Protein Families.
460 *Cell* **150**: 222-232.
461 Casino, P., V. Rubio & A. Marina, (2009) Structural insight into partner specificity and
462 phosphoryl transfer in two-component signal transduction. *Cell* **139**: 325-336.
463 Casino, P., V. Rubio & A. Marina, (2010) The mechanism of signal transduction by two-
464 component systems. *Curr Opin Struct Biol* **20**: 763-771.
465 Choi, K. Y., B. Satterberg, D. M. Lyons & E. A. Elion, (1994) Ste5 tethers multiple protein
466 kinases in the MAP kinase cascade required for mating in S. cerevisiae. *Cell* **78**: 499-
467 512.
468 Dueber, J. E., G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V. Ullal, K. L.
469 Prather & J. D. Keasling, (2009) Synthetic protein scaffolds provide modular control
470 over metabolic flux. *Nat Biotechnol* **27**: 753-759.
471 Fisher, S. L., S. K. Kim, B. L. Wanner & C. T. Walsh, (1996) Kinetic comparison of the
472 specificity of the vancomycin resistance VanSfor two response regulators, VanR and
473 PhoB. *Biochemistry* **35**: 4732-4740.
474 Galperin, M. Y., (2005) A census of membrane-bound and intracellular signal transduction
475 proteins in bacteria: bacterial IQ, extroverts and introverts. *BMC Microbiol* **5**: 35.
476 Grimshaw, C. E., S. Huang, C. G. Hanstein, M. A. Strauch, D. Burbulys, L. Wang, J. A. Hoch
477 & J. M. Whiteley, (1998) Synergistic kinetic interactions between components of the
478 phosphorelay controlling sporulation in Bacillus subtilis. *Biochemistry* **37**: 1365-1375.
479 Laub, M. T. & M. Goulian, (2007) Specificity in two-component signal transduction
480 pathways. *Annu Rev Genet* **41**: 121-145.
481 Martin, L. C., G. B. Gloor, S. D. Dunn & L. M. Wahl, (2005) Using information theory to
482 search for co-evolving residues in proteins. *Bioinformatics* **21**: 4116-4124.
483 Raghavan, V. & E. A. Groisman, (2010) Orphan and hybrid two-component system proteins
484 in health and disease. *Curr Opin Microbiol* **13**: 226-231.

485 Skerker, J. M., B. S. Perchuk, A. Siryaporn, E. A. Lubin, O. Ashenberg, M. Goulian & M. T.
486 Laub, (2008) Rewiring the specificity of two-component signal transduction systems.
487 *Cell* **133**: 1043-1054.
488 Skerker, J. M., M. S. Prasol, B. S. Perchuk, E. G. Biondi & M. T. Laub, (2005) Two-
489 component signal transduction pathways regulating growth and cell cycle progression
490 in a bacterium: a system-level analysis. *PLoS Biol* **3**: e334.
491 Stock, A. M., V. L. Robinson & P. N. Goudreau, (2000) Two-component signal transduction.
492 *Annu Rev Biochem* **69**: 183-215.
493 Wegener-Feldbrugge, S. & L. Sogaard-Andersen, (2009) The atypical hybrid histidine protein
494 kinase RodK in Myxococcus xanthus: spatial proximity supersedes kinetic preference
495 in phosphotransfer reactions. *J Bacteriol* **191**: 1765-1776.
496 Whitworth, D. E. & P. J. Cock, (2009) Evolution of prokaryotic two-component systems:
497 insights from comparative genomics. *Amino Acids* **37**: 459-466.
498 Wise, A. A., F. Fang, Y. H. Lin, F. He, D. G. Lynn & A. N. Binns, (2010) The receiver
499 domain of hybrid histidine kinase VirA: an enhancing factor for vir gene expression in
500 Agrobacterium tumefaciens. *J Bacteriol* **192**: 1534-1542.
501 Wuichet, K., B. J. Cantwell & I. B. Zhulin, (2010) Evolution and phyletic distribution of two-
502 component signal transduction systems. *Curr Opin Microbiol* **13**: 219-225.
503 Zhang, W. & L. Shi, (2005) Distribution and evolution of multiple-step phosphorelay in
504 prokaryotes: lateral domain recruitment involved in the formation of hybrid-type
505 histidine kinases. *Microbiology* **151**: 2159-2173.
506
507

508

509 ## Figure Legends

510 **Figure 1. Amino acid coevolution analysis of hybrid histidine kinases. (**A) Diagram of
511 canonical two-component signaling pathways and (B) phosphorelays, indicating the
512 conserved domains in each protein. (C) Coevolving residues in cognate pairs of canonical
513 histidine kinases and response regulators. Residue pairs with adjusted mutual information
514 scores greater than 3.5 are listed, connected by lines (left), and shown in spacefilling on a
515 structure of the *T. maritima* HK853-RR468 complex (right). The only pair in the hybrid
516 kinase alignment with a score greater than 3.0 is highlighted. For clarity, only the DHp
517 domain of HK853 is shown. Residue numbers correspond to positions within EnvZ and
518 OmpR (see Fig. S1A-B). (D) Histogram of adjusted mutual information scores for all residue
519 pairs in the hybrid histidine kinase alignment. Arrows indicate the residue pairs scoring higher
520 than 3.5 in the analysis of canonical two-component proteins, with scores for these pairs in
521 each alignment listed in the table.

522 **Figure 2. Hybrid histidine kinases show reduced phosphotransfer specificity *in vitro*.** (A)
523 Phosphotransfer profiles for kinase domains from three *C. crescentus* hybrid histidine kinases
524 against all 27 receiver domains from hybrid kinases. (B) Quantification of phosphotransfer
525 profiles for 10 hybrid kinases against the 27 hybrid kinase receiver domains; for raw profile
526 data, see Fig. S2. (C) Quantification of phosphotransfer profiles for 10 hybrid kinases against
527 the 44 soluble *C. crescentus* response regulators; for raw profile data, see Fig. S3. For panels
528 B-C, the ratio of receiver domain or response regulator band intensity to the
529 autophosphorylated kinase band intensity was calculated and converted to color based on the
530 legend shown. All phosphotransfer reactions were incubated 15 minutes.

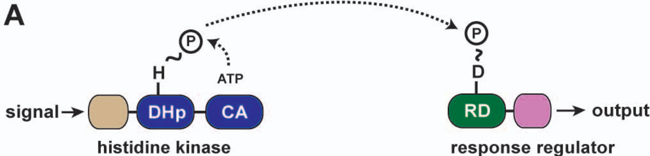
531 **Figure 3. Hybrid kinases lacking their receiver domains exhibit cross-talk.** (A) Chimeric
532 hybrid kinases were autophosphorylated in the presence of buffer, HCl, or NaOH to assess
533 whether phosphoryl groups resided on the conserved histidine, aspartate, or both. (B)
534 Chimeric hybrid kinases were autophosphorylated and then tested for phosphotransfer to
535 soluble CheYV at the time points indicated. Error bars represent standard deviation from three
536 independent replicates. Raw gel images are shown in Fig. S4C. The identity of domains in
537 each chimeric kinase are listed. (C) Swarm plate assay for strains expressing each of the
538 CC3191 constructs listed or vector alone. (D) Quantification of swarm sizes for strains

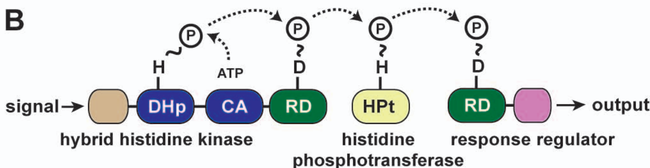
539 expressing various constructs for each of the four hybrid histidine kinases indicated. Swarm
540 areas were measured and plotted relative to the empty vector control. Error bars represent
541 standard deviations from three replicates. Swarm plate images are shown in Fig. S4D.

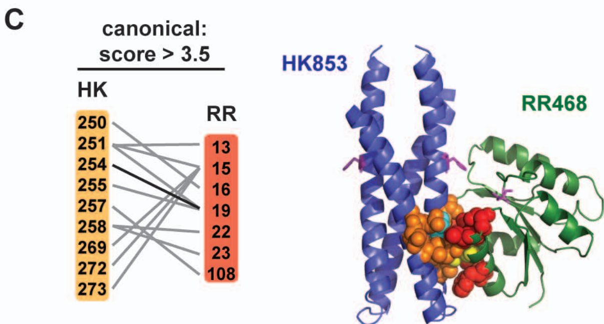
542 **Figure 4. Specificity residues are conserved among hybrid histidine kinases.** (A) An
543 unrooted neighbor-joining tree of the *C. crescentus* kinases was built from an alignment of the
544 DHp domains of all 24 hybrid and 21 canonical histidine kinases from *C. crescentus*. Hybrid
545 kinases are labeled in red. (B) Sequence logos for the residues that dictate phosphotransfer
546 specificity in canonical kinase-regulator pairs. Logos were built from an alignment of the 21
547 canonical histidine kinases and 44 soluble response regulators (top), and from an alignment of
548 the 24 hybrid histidine kinases in *C. crescentus* (bottom).

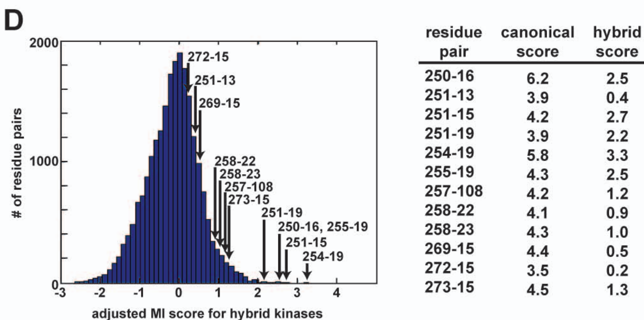
549 **Figure 5. Model for changes in specificity residues following duplication of canonical**
550 **and hybrid histidine kinases.** Ovals represent niches within sequence space, or the set of
551 response regulators recognized by a given histidine kinase as determined by its specificity
552 residues. Post-duplication, canonical kinases separate in sequence space to insulate the two
553 pathways and prevent cross-talk. In contrast, hybrid kinases do not separate, as the tethered
554 receiver domain effectively insulates the duplicated kinases against cross-talk.

555

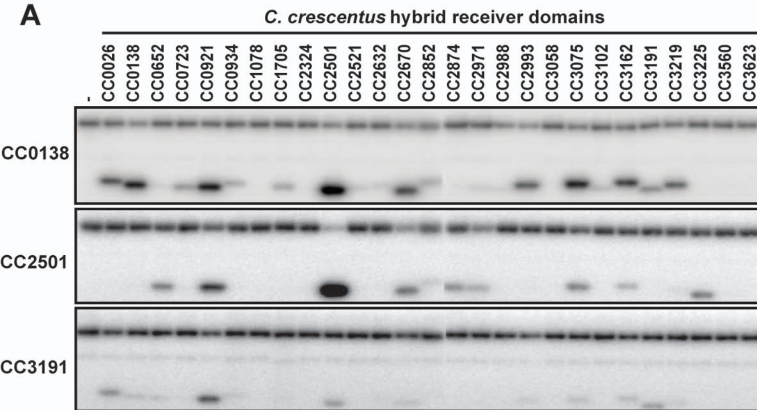


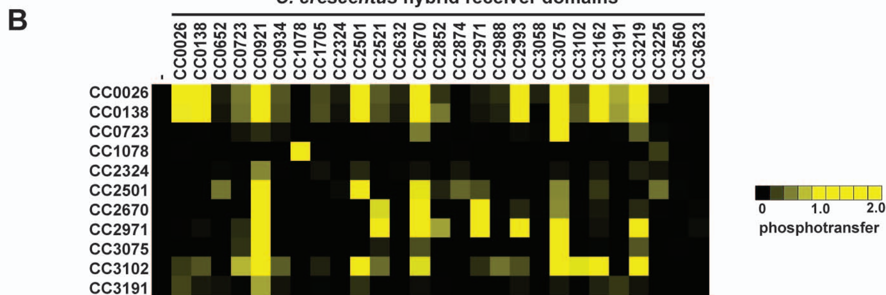


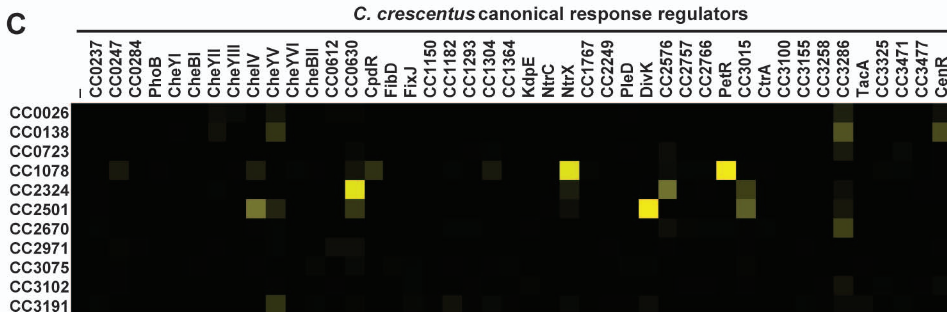


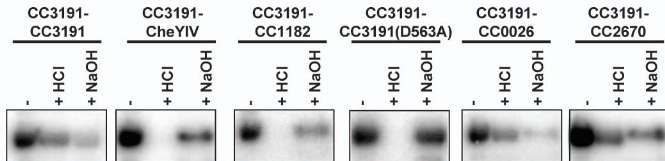


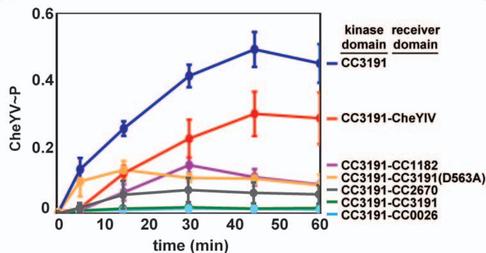
residue pair	canonical score	hybrid score
250-16	6.2	2.5
251-13	3.9	0.4
251-15	4.2	2.7
251-19	3.9	2.2
254-19	5.8	3.3
255-19	4.3	2.5
257-108	4.2	1.2
258-22	4.1	0.9
258-23	4.3	1.0
269-15	4.4	0.5
272-15	3.5	0.2
273-15	4.5	1.3

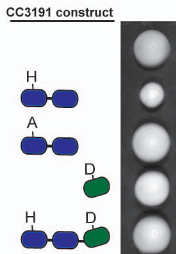

A
C. crescentus hybrid receiver domains
-
CC0026
CC0138
CC0652
CC0723
CC0921
CC0934
CC1078
CC1705
CC2324
CC2501
CC2521
CC2632
CC2670
CC2852
CC2874
CC2971
CC2988
CC2993
CC3058
CC3075
CC3102
CC3162
CC3191
CC3219
CC3225
CC3560
CC3623
CC0138
CC2501
CC3191

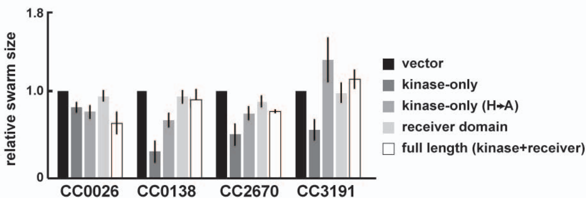

B
C. crescentus hybrid receiver domains
-
CC0026
CC0138
CC0652
CC0723
CC0921
CC0934
CC1078
CC1705
CC2324
CC2501
CC2521
CC2632
CC2670
CC2852
CC2874
CC2971
CC2988
CC2993
CC3058
CC3075
CC3102
CC3162
CC3191
CC3219
CC3225
CC3560
CC3623
CC0026
CC0138
CC0723
CC1078
CC2324
CC2501
CC2670
CC2971
CC3075
CC3102
CC3191
0
1.0
2.0
phosphotransfer


C
C. crescentus canonical response regulators
-
CC0237
CC0247
CC0284
PhoB
CheYI
CheBI
CheYII
CheYIII
CheIV
CheYV
CheYVI
CheBII
CC0612
CC0630
CpdR
FibD
FixJ
CC1150
CC1182
CC1293
CC1304
CC1364
KdpE
NtrC
NtrX
CC1767
CC2249
PleD
DivK
CC2576
CC2757
CC2766
PetR
CC3015
CtrA
CC3100
CC3155
CC3258
CC3286
TacA
CC3325
CC3471
CC3477
CenR
CC0026
CC0138
CC0723
CC1078
CC2324
CC2501
CC2670
CC2971
CC3075
CC3102
CC3191


A
CC3191-CC3191
CC3191-CheYIV
CC3191-CC1182
CC3191-CC3191(D563A)
CC3191-CC0026
CC3191-CC2670
-
+ HCl
+ NaOH


B
CheYV~P
time (min)
kinase domain
receiver domain
CC3191
CC3191-CheYIV
CC3191-CC1182
CC3191-CC3191(D563A)
CC3191-CC2670
CC3191-CC3191
CC3191-CC0026


C
CC3191 construct
H
A
D
H
D


D
relative swarm size
CC0026
CC0138
CC2670
CC3191
vector
kinase-only
kinase-only (H→A)
receiver domain
full length (kinase+receiver)

A

CC1594 CC2884 CC3327 CC0238 CC0248 CC1294 CC0289 CC1740 CC2482 CC0759 CC1062 CC1078 CC1063 CC2874 CC2632 CC3225 CC3191 CC0138 CC0026 CC1742 CC2501 CC0586 CC0530 CC0652 CC2324 CC2755 CC0934 CC2521 CC1305 CC2852 CC2971 CC2670 CC3219 CC0921 CC3102 CC2988 CC0723 CC2932 CC2765 CC2993 CC3075 CC3623 CC1705 CC3484 CC1181

0.25

B

histidine kinases

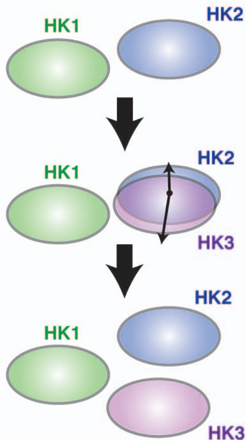
response regulators

canonical

hybrid

bits

specificity residues


canonical histidine kinases
HK1
HK2
HK1
HK2
HK3
HK1
HK2
HK3
ancestral state,
pre-duplication
gene duplication
derived state,
post-duplication
hybrid histidine kinases
HK1
HY1
HK1
HY1
HY2
HK1
HY1
HY2