A pathogen-encoded signaling receptor mediating host-like interactions through intrinsic disorder

Marta F. M. Vieira[a], Guillem Hernandez[a], Tiago Veloso[a], Miguel Arbesú[b], Andreas Zanzoni[c] Tiago N. Cordeiro[a]

[a]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157 Oeiras Portugal

[b]Department of NMR-supported Structural Biology, Leibniz-Forschungsinstitut für Molekulare Pharmakologie, Robert-Rössle-strasse 10, 13125 Berlin, Germany

[c]Aix-Marseille Univ, Inserm, TAGC, UMR_S1090, Marseille, France

Corresponding author: tiago.cordeiro@itqb.unl.pt

Abstract

The translocated intimin receptor (Tir) is a central effector of Attaching and Effacing (A/E) pathogens responsible for worldwide foodborne disease cases. Upon delivery into host cells, Tir acts as a cell-signaling receptor, rewiring host cellular processes to assist infection. We found that this bacterial-encoded transmembrane protein comprises highly disordered intracellular domains bearing host-like motifs that bind host proteins. This unexpected trait was found prevalent in several other effectors secreted by A/E bacteria. We assessed Tir's intrinsic disorder by an integrative structural biophysics approach, unveiling that its intracellular side comprises a partially structured N-terminal dimer (N-Tir) and a disordered C-terminal tail (C-Tir). NMR analysis revealed that C-Tir has pre-existing transient structures at phosphorylation sites, including host-like immunoreceptor tyrosine-based inhibitory motifs (ITIMs). These ITIM-like sequences were found to bind lipid bilayers as previously observed for host T-cell receptor cytoplasmic disordered domains. C-Tir's membrane affinity is residue-specific and modulated by lipid composition, suggesting a regulation layer based on membrane composition. Using NMR, we also observed that the disordered C-Tir displays multisite tyrosine-phosphorylation sites that mediate

promiscuous binding to the C-terminal SH2 domain of host SHP-1 in dynamic equilibrium. Together, these novel insights provide an updated picture of Tir's structural features and highlight Tir-mediated mimicry of host disordered membrane receptors as a molecular strategy for host cell subversion.

Keywords

Intrinsically disordered proteins | bacterial effectors | EPEC/EHEC | molecular mimicry and multivalency | Host-pathogen interactions

Summary

Tir is a cellular receptor secreted by life-threatening pathogens. Upon delivery into host cells, Tir inserts the host plasma membrane providing a means for these extracellular pathogens to control host intracellular processes. To prevent pathogens from relying on Tir, it is essential to understand its intracellular mechanics. This paper provides a coherent picture of the intracellular side of Tir, highlighting its ability to copycat the interactions of disordered intracellular domains of host immune receptors. This copycatting allows the bacterial pathogens to modulate critical host processes, allowing infection to spread further without triggering the immune system response. This work proposes that other bacterial secreted pathogenic proteins exploit intrinsic disorder to hijack human cells, suggesting a widespread host subversion mechanism.

Introduction

Intrinsically disordered proteins and regions (IDPs and IDRs, respectively) are widespread across all kingdoms of life (1). They do not adopt well-defined structures but exist as dynamic conformational ensembles that display unique properties complementary to globular proteins (2, 3). IDPs/IDRs play essential roles in virtually all signaling pathways in the cell (4), including cell cycle (5), circadian circuits (6), post-transcriptional regulation (7), and protein degradation (8). Given their ubiquitous relevance in cellular signaling and regulation, the onset of several human diseases, including cancer and neurodegeneration, are linked to dysfunctional disordered proteins (9). A common hallmark of IDPs/IDRs is the high occurrence of short linear motifs, also known as eukaryotic linear motifs (ELMs) (10), which are stretches of 3-10 contiguous residues mediating transient protein-protein interactions (e.g., post-translational modifications sites, targeting signals) (11). For instance, phosphorylation sites are often found in protein IDRs, where they can

64 modify charge and hydrophobicity and modulate interactions with partners (7). The prevalence
65 of such functional elements in IDRs provides versatility to cell interaction hubs by enabling
66 flexibility and adaptability to multiple interaction interfaces (12, 13). Importantly, IDR sequence
67 composition is biased towards low complexity and charged residues favoring electrostatic
68 interactions at lipid bilayer surfaces (14), with lipid-binding proteins accounting for 15% of all
69 disordered proteins (15). As a result, intrinsic disorder is also common in membrane receptors,
70 particularly in their intracellular domains (16, 17).
71
72 Disordered proteins make up 30-50% of eukaryotes' proteins (18, 19). In contrast, they are
73 remarkably less abundant in prokaryotes. However, increasing experimental evidence in the
74 literature shows that several effector proteins secreted by pathogenic bacteria contain functionally
75 relevant IDRs (20). A well-known case is the oncogenic *Helicobacter pylori* effector CagA, which
76 promiscuously recruits host proteins to potentiate oncogenic signaling via a long IDR (22).
77 Another illustrative example is the effector CyaA of *Bordetella pertussis*, the causative agent of
78 whooping cough that interacts with host calmodulin via a disordered stretch of 75 amino acid
79 residues (21). Other less-ordered effectors are also secreted by Enteropathogenic and
80 Enterohemorrhagic *Escherichia coli* strains (EPEC and EHEC, respectively), such as EspF(U)/TccP
81 (22) and EspB (23).
82
83 EPEC is a leading cause of child deaths worldwide (24), and EHEC causes hemolytic uremic
84 syndrome (HUS), defined by acute kidney failure, low platelet count, and destruction of red blood
85 cells (25). These bacteria attach to the gut mucosa, causing a histopathological condition known
86 as the 'attaching and effacing' (A/E) lesions, characterized by actin-rich pedestal formation and
87 microvilli degeneration. To establish a replicative niche on host cells, EPEC/EHEC delivers a vast
88 repertoire of effectors (26, 27). In light of this, we ran a structural disorder analysis on the
89 collection of effectors secreted by A/E-bacteria and searched for ELMs evidence. Among A/E-
90 bacteria disordered effectors, we found the prominent translocated intimin receptor (Tir) to be a
91 disordered membrane receptor with a relatively high density of host-like ELMs, particularly in its
92 intracellular domains. Tir is the first effector secreted during infection (28). Once inside the host,
93 it migrates to the plasma membrane to act as a receptor for intimin presented at the bacterial
94 surface (29), anchoring A/E pathogens to targeted host cells. By mediating attachment, Tir also
95 facilitates translocation of additional T3SS secreted effectors (30). Moreover, Tir promotes actin

96 polymerization (31), suppresses autophagy (32) and immune response (33) to allow bacterial
97 survival on the host cell's surface. It also triggers pyroptotic cell death (34) (35). As a membrane
98 receptor, Tir adopts a hairpin topology with the external intimin-binding domain (IBD) (36)
99 connecting two transmembrane domains, with both N- and C-terminus in the host cytoplasm,
100 interacting with multiple intracellular eukaryotic proteins (37, 38). Those intracellular regions
101 enable Tir to interact with at least 25 host intracellular targets (38), such as host tyrosine
102 phosphatases SHP-1/2 to suppress pro-inflammatory cytokines signaling (33, 39) and TAK1-
103 mediated immune response (40).
104
105 Despite all these findings, the structural mechanisms underlying Tir intracellular regions' function
106 during infection are poorly understood. To gain further insights into Tir structural dynamics, we
107 devised an integrative structure-based biophysical study to experimentally assess their intrinsic
108 disorder and multivalency. In particular, we focus our attention on self-assembly, accessibility to
109 multi-site phosphorylation, host protein- and membrane-binding of Tir intracellular regions. We
110 found that N-Tir is a flexible dimer with a hybrid architecture of ordered and disordered regions,
111 opening the question of whether host signaling activation involves structural changes of
112 preformed dimers. Moreover, the present work shows that C-Tir can bind to lipid bilayers through
113 short host-like motifs that, when phosphorylated, interact with the C-terminal SH2 domain (C-
114 SH2) of host SHP-1 in dynamic equilibrium. Together these findings highlight a molecular
115 mechanism through which Tir can interact with host components through mimicry and
116 multivalency.
117

118 Results and Discussion

119

120 Structural disorder and short linear motifs are common features of
121 A/E pathogen effectors

122 To quantify the prevalence of disorder propensity among A/E pathogens, we predicted the
123 disorder content of A/E effectors by using DISOPRED3 (41) (**Fig. 1**) and IUPred 1.0 (42, 43))
124 (**Fig. S1**) structural disorder predictors. We observed that A/E effectors have a higher disorder
125 propensity than their proteome counterparts in EHEC O157:H7, the closely related
126 Enteropathogenic *E. coli* H127:H6 (EPEC), and *Citrobacter rodentium* (CR) — the standard small

127 animal A/E pathogen model (**Fig. 1*A*, Table S1**). Subsequently, we classified them into five
128 structural categories, based on disorder content (see **Material and Methods for details**). Our
129 analysis shows that A/E pathogens have a structurally diverse repertoire of effectors, ranging
130 from fully unstructured to ordered proteins (**Fig. 1*B***). While most proteins are folded in all
131 prokaryotic collections as a whole and individually (**Fig. S2*A***), fully disordered effectors are
132 predicted to be two to four-fold more frequent than the whole bacterial proteome (1.9% vs. 7.7
133 % on average, far from the 32.7 % of the human proteome) (**Fig. S1*B***). Partially Disordered
134 Proteins (PDR) with long disordered regions occur in EPEC effectors similar to the human
135 proteome. For instance, we classified as PDR the EspB effector, which is involved in cytoskeleton
136 rearrangement and previously identified as an inherently less-ordered effector (23). An additional
137 example of PDR is the NleH effector, a protein kinase with a disordered N-terminal domain that
138 binds the ribosomal protein 3 (RPS3) to manipulate the NF-κB pathway (44, 45). In the order-to-
139 disorder continuum, the effector EspF is akin to EspF(U) as a fully disordered protein (**Fig.1*C***).
140 Both effectors contain multiple consecutive repeats of linear motif pairs able to bind critical host
141 components for actin assembly, with EspF having three to five repetitions depending on the strain
142 (46), and EspFu, only secreted by EHEC, five-and-a-half repeats (47) (**Fig. S2*B***).
143
144 Our analysis highlights Tir as a disordered effector (**Fig. 1*C***, **Fig. S2*B***). Except for the
145 transmembrane domains, Tir displays a high disorder propensity within its IBD and intracellular
146 regions (**Fig. S1*D*)**. Increasing evidence shows that pathogens exploit host-like motifs as a
147 strategy to interact with host components to subvert host cellular functions (48–50). On this
148 basis, we searched the A/E effectors for eukaryotic linear motifs (ELMs) and found that aside
149 from EspF and EspF(U), other A/E effectors bear multiple putative ELM instances (**Fig. 1*C***). Tir
150 emerged again as having several predicted ELMs in their intracellular regions, including
151 experimentally verified SH3- and SH2-domain binding motifs (51). Structural disorder and high
152 ELM density support Tir's ability to interact with several host proteins (38). There is a correlation
153 between disorder content and ELM density, being Tir, EspF(U), and EspF clustered with the
154 highest disorder fraction and motif overall density (**Fig. 1*C***), similarly to eukaryotic IDPs (11, 52).
155 This correlation highlights ELMs within disordered segments of bacterial effectors as a strategy
156 to disrupt host networks. Disordered effectors, such as Tir, are an exception to the general trend
157 that prokaryotic proteins are less disordered. Given the ubiquitous presence of IDPs as
158 transcriptional factors (53) and, more generally, as hubs in host networks (54), bacterial effectors

159 with host-like disordered proteins features are an efficient way to subvert host eukaryotic systems
160 and promote infection. We selected Tir to study how disordered effectors control the host by
161 validating the predicted structural disorder and molecular mimicry within its intracellular domains.
162

163 Tir bears an intrinsically disordered C-terminal tail

164 Tir secreted by A/E pathogens share substantial sequence-conservation and high frequency of
165 disorder-promoting residues at the N- and C-terminal regions (N-Tir and C-Tir, respectively, both
166 localized to the host intracellular side (**Fig. S3**). We assessed experimentally the structural
167 disorder propensity of EPEC N-Tir and C-Tir by multiple biophysical methods, including small-
168 angle X-ray scattering (SAXS), circular dichroism (CD), and liquid-state nuclear magnetic
169 resonance (NMR). For this integrative study, we used the protein constructs encompassing the
170 residues 1-233 and 388-550 from the N-terminal and C-terminal cytosolic regions of EPEC Tir,
171 respectively (**Fig. S3**). Analytical size-exclusion chromatography (SEC) in combination with SAXS
172 indicated a monomeric state for C-Tir in solution with a radius of gyration (*Rg*) of 38.8 ± 0.2 Å
173 and a maximum distance (*Dmax*) of 128.0 ± 5.0 Å (**Table S4**). The *Rg*-values of IDPs
174 approximately followed a power law of $Rg = 2.54N^{0.522}$ (55, 56) as a function of sequence length
175 (*N*). Using this relationship, the predicted *Rg*-value is 37.4 Å for the 173-residue C-Tir construct
176 ending with a StrepTag (8 extra residues). The similarity between the experimental and predicted
177 values suggests that C-Tir might adopt IDP-like structures in solution. The SAXS-derived Kratky
178 plot of C-Tir also shows characteristics of disordered or unfolded proteins; monotonically
179 increasing without a well-defined maximum (**Fig. 2*A***). This feature is absent in globular proteins
180 and implies conformational heterogeneity and flexibility for C-Tir (57). Likewise, the asymmetric
181 pair distance distribution function, *P(r)*, obtained from the scattering data is compatible with a
182 highly flexible protein sampling pairwise distances far exceeding those expected for a globular
183 protein of the same molecular weight (56) (**Fig. S4**). The intrinsic disorder of C-Tir is also
184 reflected in its CD and [^{1}H-^{15}N]-HSQC NMR spectra. C-Tir's CD profile has negative ellipticity at
185 200 nm and a shallow band in the 210–230 nm range (**Fig. 2*B***), indicating a high content of
186 random coil with minimal ordered structural elements (58). The [^{1}H-^{15}N]-HSQC NMR spectrum for
187 this construct displays a typical IDP fingerprint with low chemical shift dispersion (59) (**Fig. 2*C***).
188 We observed a similar spectrum for the equivalent region of Tir from EHEC (**Fig. 2*C***). Although
189 all A/E pathogens produce Tir, actin polymerization by EHEC Tir C-terminus requires Esp(U)/TccP
190 (60), which is unneeded in EPEC to initiate actin pedestal formation in host cells (31). Besides the

191 different mechanisms, C-Tir is disordered in both strains, thus highlighting the conservation and
192 functional importance of structural disorder in this intracellular domain for A/E virulence.
193 Overall, the different biophysical measurements are consistent with a highly dynamic protein,
194 unambiguously showing that the C-terminal half of Tir is an IDR that lacks a well-folded structure
195 under native-like conditions. This inherent flexibility potentially enables Tir for multiple binding
196 while working as a pathogenic scaffold/hub recruiting host proteins to perturb cellular functions.
197

198 C-Tir displays non-random structural preferences at sites of 199 phosphorylation and host interactions

200 We employed NMR to dissect with further detail subtle structural features of C-Tir. The NMR
201 chemical shift is the most readily accessible observable in NMR and sensitive to, even partial,
202 secondary structural elements present in IDPs (59). To gain access to this information, we
203 assigned the backbone resonances of C-Tir nuclei using standard triple-resonance spectra at high-
204 field, together with reverse labeling (**Fig. S5**). Reverse labeling of selected amino-acids (61)
205 proved to be a cost-efficient approach to assign this disordered protein by reducing the spectral
206 ambiguity and should be readily applied to other IDPs (see SI **Methods and Materials**).
207 Provided with the assignment, we calculated the NMR chemical shift deviations (Δδ) to random coil
208 values of intrinsically disordered proteins (62). Although C-Tir has properties of a random coil,
209 these so-called NMR secondary shifts allowed us to identify two regions with non-random
210 structural preferences. In IDPs, transient secondary structural elements are often crucial in acting
211 as molecular recognition units, playing significant roles in binding (63, 64). In figure **Fig. 3*A***, the
212 consecutive positive ΔδCα–ΔδCβ values indicate a tendency for α-helical structure in two distal
213 segments of C-Tir: (I) D420-Q435; II) I507-A515. Both stretches include residues that are
214 phosphorylated by host kinases, S434 and Y511, respectively. By taking into account all the
215 assigned backbone resonances into a neighbor-corrected sequence structural propensity
216 calculator (nsSPC) score (65), these apparent local α-helical regions became even more evident.
217 New stretches also emerged with a putative α-helical propensity, including a sequence contiguous
218 to Y511 and downstream tyrosine 483 and a few C-terminal residues (**Fig. 3*A***).
219 Structural signatures near phosphorylation sites can affect kinase selectivity to disordered
220 substrates (66), and phosphorylation can modulate partially structured elements for regulation
221 (67). Phosphorylation of Tir on serines S434 and S463 by host protein kinase A (PKA) triggers
222 changes that affect its conformational and self-assembling properties (68, 69). So, the enhanced

α-helical propensity near S434 might play a role in PKA-mediated phosphorylation effects on Tir structure and function (32, 70). Moreover, tyrosines Y483 and Y511 are core residues of Tir's host-like Immunoreceptor tyrosine-based inhibitory motifs (ITIMs). The intracellular signaling associated with Tir ITIMs relies on the phosphorylation of the central tyrosine by Src family kinases (SFKs), creating a binding site for SH2–containing proteins, which become activated upon recruitment. Yet, we found that unphosphorylated C-Tir can recognize host SH2-domains with low-intermediate affinity. Using the C-terminal SH2 domain (C-SH2) of phosphatase SHP-1, we identified, by NMR, that residues surrounding the unphosphorylated Y511 ITIM-like motif are involved in binding SH2 domains (**Fig. 3*B-C***). Binding to C-Tir caused a selective loss of intensities for the HSQC peaks, mainly from residues $A_{512}LLA_{515}$. Visible NMR signals retained low-dispersion, indicating that the protein remains mostly disordered and flexible in the complex. We estimated an apparent K_D of *ca.* 68.5μM for this interaction (**Fig. 3*D***). Thus, besides encoding for α-helical structures, the sequence surrounding Y511 (i+4) has a chemical signature for SH2-domain binding pre-phosphorylation. This observation highlights that those residues adjacent to Y511 can define a binding motif for C-SH2, likely serving as a precursor to a tighter binding upon phosphorylation. The presence of a pre-existing transiently helix at this region suggests a possible role for this structure as molecular recognition elements for C-Tir.

Remarkably, Tir ITIMs share with cytoplasmic tyrosine-based sequence motifs of host receptors the ability to adopt α-helical conformations. Indeed, host ITIMs and Immunoreceptor tyrosine-based activation (ITAMs) can form dynamic/transient α-helices, whether in free (71, 72) or membrane-bound state (73). Their dynamic binding to membranes is postulated to regulate tyrosine sites' access to phosphorylation, providing a switch between functional and non-functional conformations (74). The helical conformation in ITIMs/ITAMs is stabilized *in vitro* in the presence of 2,2,2-trifluoroethanol (TFE) or through binding to detergent micelles and lipids (72, 73, 75). In the presence of membrane-mimetic solvent TFE, we found that C-Tir exhibited a substantial increase in secondary structure, with the CD signature of a partial helical protein (**Fig. S6*A***). This conformational change is followed by chemical shift differences (**Fig. S6*B-C)*** and a differential decrease in peak intensities of the C-Tir amide proton resonances in the presence of increasing amounts of TFE, mainly around Y511 and Y483 and not at S434 (**Fig 3*E***). This effect supports an enhanced α-helical conformation propensity around the two tyrosines, and a putative

254 role in membrane-binding, as postulated for some host tyrosine-based motifs that undergo TFE-
255 induced structural changes (72).
256 Our NMR results show that unphosphorylated C-Tir host-like ITIM motifs can adopt pre-formed
257 structures relevant to membrane binding and molecular recognition of SH2 domains of signaling
258 proteins.
259

260 **Disordered C-Tir binds lipids**

261 The ability to form α-helices stabilized by TFE prompts us to further investigate lipid binding by
262 C-Tir. To do so, we incubated ^{15}N-labelled C-Tir with different bicelles, which are small planar
263 bilayers of long-chain lipids closed by curved micelle-like walls of short-chain lipids. These disk-
264 like structures are membrane models extensively used to explore protein-membrane interactions,
265 including those involving disordered proteins (76). As long-chain lipids, we used 1,2-dimyristoyl-
266 *sn*-glycero-3-phosphocholine (DMPC) or the acidic lipid 1,2-dimyristoyl-*sn*-glycero-3-
267 phosphorylglycerol (DMPG) mixed with the short-chain lipid 1,2-dihexanoyl-*sn*-glycero-3-
268 phosphocholine (DHPC), offering the possibility to create bicelles with different lipid charge
269 density.
270 In the presence of bicelles containing DMPC lipids, we observed a significant residue-specific
271 decrease in the ratio of NMR signals around Y511; yet more subtle, also around Y483, as
272 previously induced with TFE (**Fig. 4**). This finding shows that C-Tir does interact with non-charged
273 lipid bilayers predominantly via its Y511-motif, including mostly hydrophobic residues, such as
274 $A_{512}LLA_{515}$. We further tested the influence of increasing negatively charged lipid head groups on
275 C-Tir membrane association with bicelles containing DMPG lipids. Our NMR data show a more
276 extended NMR signal attenuation in the presence of DMPG/DHPC bicelles, even including the first
277 stretch of residues of this construct, which display some positively charged residues (e.g., R388
278 and K389) (**Fig. 4**). In the context of the full-length protein, these residues define the membrane-
279 proximal region (**Fig. S3**). Given its proximity to the cell plasma membrane, it is reasonable to
280 assume its involvement in membrane binding. Aside from Y511 surrounding residues, the
281 sequence around Y483 is also notably affected by anionic lipid content, indicating that lipid-charge
282 content modulates C-Tir membrane binding modes. Contrary, the residues 395-475 did not show
283 any membrane interaction. This region has a net negative charge (pI~4.5), whereas the
284 sequences displaying membrane ability are positive (pI~9.9), thus in line with an electrostatic
285 model. In this residue range, C-Tir bears two experimental confirmed host protein binding sites:

286 a) NCK Src Homology 2 (SH2) domain binding motif (31); and b) NPY motif that interacts with
287 the I-BAR domain of host IRSp53/IRTKS (77). With both bicelles, the last C-terminal residues of
288 C-Tir also show NMR-signal attenuation indicative of membrane-binding (**Fig. 4*B***). These residues
289 (A_{541}PTPGPVRFV$_{550}$) emerge as part of a potential new lipid-binding region, also affected by TFE
290 (**Fig. 3*E, Fig. S6C***) and showing ncSPC values indicating α-helical propensity (**Fig. 3*A***). Our data
291 thus suggest that C-Tir can undergo multivalent electrostatic interactions with lipid bilayers that
292 might potentially fine-tune Tir's activity in the host cell.
293

294 Multi-phosphorylated C-Tir is a disordered multivalent cytosolic tail

295 In the host, Tir becomes phosphorylated by Src family protein tyrosine kinases (PTKs) at several
296 tyrosine phosphorylation sites targeting SH2 domain-containing host proteins. To fully assess its
297 ability to recruit SH2 domains, we reconstructed C-Tir's Tyr-phosphorylated state and evaluated
298 its binding to C-SH2. To this end, we incubated C-Tir with the Src family PTK Fyn, and
299 quantitatively monitored its phosphorylation by NMR (78). The sensitivity of chemical shifts to
300 changes caused by phosphorylation allowed us to identify four phosphorylated tyrosine sites along
301 the disordered C-Tir modified by Fyn (i.e., Y454, Y474, Y483, Y511). Phosphorylation caused
302 small but noticeable changes in the chemical shifts of the tyrosines and adjacent residues. The
303 [^{1}H-^{15}N]-HSQC spectrum displayed low amide proton dispersion, a diagnostic that the C-Tir
304 remains disordered upon multisite phosphorylation. Moreover, the secondary chemical shifts of
305 this 4-fold phosphorylated state (pC-Tir) did not reveal substantial changes in local structure
306 propensity due to phosphorylation (**Fig. S7**), as similarly reported for other disordered proteins
307 (79, 80).
308 To evaluate the interaction of phosphorylated C-Tir (pC-Tir) with an SH2 domain, we titrated
309 unlabeled C-SH2 into a ^{15}N-labeled pC-Tir solution and monitored the binding interaction
310 broadening of signals in NMR [^{1}H-^{15}N]-HSQC spectra (**Fig. 5*A,B***). With the resonance re-
311 assignment of pC-Tir, we identified that upon phosphorylation, all tyrosine sites interact with C-
312 SH2, and not exclusively the Y511-based motif as observed for unphosphorylated C-Tir. The
313 interaction caused a drop in peak intensities around each site (i.e., i= pY454, pY474, pY483,
314 pY511), including roughly residues i+4 and i-1, even at sub-stoichiometric conditions (**Fig. 5*B***).
315 This suggests a "fuzzy" multivalent binding of C-SH2 to four xYx(x)ϕ motifs, where ϕ is
316 hydrophobic (V/L/I) (**Fig. 5*C***), with multiple pYs on C-Tir interacting with C-SH2 in dynamic
317 equilibrium (79). Except for these residues, backbone chemical shift changes were relatively

small, with spectra displaying sharp line widths and low dispersion, both characteristics of a disordered protein. Thus, pC-Tir binds C-SH2 retaining its high level of disorder. The relative signal intensities at each phospho-tyrosine (pY) and surrounding residues approximately reflect the fraction of unbound sites, allowing us to estimate their apparent local binding affinity (**Fig. S8**). Phosphorylation enhanced the binding to Y511 ITIM by ~10 fold and enabled other tyrosine-based motifs to engage C-SH2. Among the four pYs, the resonances around pY454 were less broadened during the titration, suggesting lower local binding. The remaining pYs bind C-SH2 with similar strength (~3-9 μM). Overall, phosphorylation at multiple tyrosine sites provides various SH2 docking sites and reinforces the role of Tir as a scaffolding hub.

Tir's N-terminal intracellular tail is a dimer with order-disordered duality

In a similar fashion, we evaluated whether the N-terminal cytosolic region of Tir (N-Tir; **Fig S2**) was also intrinsically disordered. Yet, contrary to C-Tir, this region adopts a flexible but more defined structure. Based on SEC-SAXS data, the N-Tir is a non-globular protein with an *Rg* of 37.70 ± 0.10 Å and a *Dmax* of 140.0 ± 10.0 Å (**Fig. 6*A***) (**Table S4**), and a molecular weight compatible with a 52kDa dimer, in line with the observed gel filtration retention time. Like C-Tir, the respective *P(r)* curve is highly asymmetric but bimodal, with a peak at 21.9 Å and a prominent shoulder at 35.9 Å, suggesting that the dimer is elongated with spatially separated lobes. Its SAXS-derived Kratky plot is also not bell-shaped as expected for a non-globular and flexible arrangement (**Fig. 6*B***). The corresponding SAXS-driven *ab initio* reconstruction highlights an elongated S-shaped core for the dimeric N-Tir 1-233 polypeptide fused to a C-terminal Strep-Tag (**Fig. S4**). This extended dimer is largely disordered but contains stable secondary structural elements. Its far-UV CD profile reveals a partially folded protein, marked by a negative maximum at 205 nm due to high disorder content. The positive signal at 190 nm and the negative shoulder at 220 nm reflect ordered elements (**Fig. 6*C***). So, SAXS and CD data indicated that N-Tir is a partially folded dimer. To probe the order-disorder interplay within N-Tir at high-resolution, we have collected NMR data at different temperatures.

For globular proteins, the 52kDa dimer size is beyond the practical range amenable to traditional NMR spectroscopy in solution. The NMR signals of larger molecules relax faster, leading to line broadening, low spectral sensitivity, and eventually loss of NMR signals. This problem is even more significant at low temperatures. Nevertheless, we found that the 52kDa dimer has a [^{1}H-

350 ^{15}N]-HSQC NMR spectrum at 5 ºC characterized by intense signals and low chemical shift
351 dispersion akin to an IDP (**Fig. 6*D***). We assigned the resonances at 5 ºC mostly to the first 80
352 and last 35 residues of the N-Tir, meaning that those regions are flexible in solution. The missing
353 resonances of the central residues became visible with increasing temperature due to faster
354 tumbling. At 35 ºC, they adopted an NMR fingerprint of a folded-like protein (**Fig. 6*E***), reflecting
355 a more well-structured region. With NMR, we could establish that N-Tir's central sequence has
356 the broad NMR fingerprint of a well-folded protein, flanked by residues with poorly dispersed $^{1}H^{N}$
357 backbone resonances pointing disorder. A closer look at the structural disorder predictions
358 computed for N-Tir also suggests an order-disorder duality, with flanking residues more flexible
359 than those in the central region (**Fig. S9*A***). Such disordered segments harbor linear motifs that
360 were experimentally identified to interact with host proteins **(FIg. S2B**) (51, 81).
361 The partial deletion of these disordered regions reduced the random-coil content reported by CD
362 data, highlighting a more structured central part of N-Tir (N-Tir$_{60\text{-}200}$, hereafter NS-Tir). The
363 difference between CD spectra resulted in a pure random-coil CD signature, reinforcing that the
364 flanking regions are disordered, whereas the central part folded. NS-Tir's CD profile displays
365 positive values below 200 nm and two negative bands at 210 and 222 nm, commonly associated
366 with structured secondary conformations (**Fig. 6*C***).
367 We have employed SAXS to probe further the overall structure of NS-Tir. Our synchrotron SEC–SAXS data
368 confirmed that NS-Tir is still a stable dimer in solution. A slightly less broadened Kratky profile reflects the
369 absence of the disordered flanking regions (**Fig. 6*B***) but compatible with an elongated dimer (**Fig.
370 6*F***), retaining a similar overall conformation supported by NMR. We acquired [^{1}H-^{15}N]-HSQC data
371 for NS-Tir and superimposed them to N-Tir spectra (**Fig. S9*B***). Over 75% of the visible
372 resonances overlapped between the two constructs, with a low dispersion for the ^{1}H dimension
373 at 5 ºC and a broader [^{1}H-^{15}N]-fingerprint at 35 ºC, indicating that NS-Tir maintained a similar
374 folded isolated and in the full-length N-terminal region context. The reversible thermal
375 denaturation of NS-Tir monitored by far-UV CD shows the dimer unfolding in two discrete steps,
376 capturing both folding and dimerization events. It reveals a stable dimer that unfolds and
377 dissociates by increasing temperature (82) (**Fig. S10**).
378 Overall, our results support a model of Tir spanning the host plasma membrane with flexible N-
379 terminal intracellular domains assembled in a dimer form (**Fig 7**). In this model, N-Tir is anchored
380 to the membrane by flexible segments and with dangling linear motifs located at its disordered
381 N-termini.

382 # Concluding Remarks

383 We observed that A/E pathogens have a structurally diverse repertoire of protein effectors
384 enriched in disorder-prone residues. Some effectors are fully disordered proteins with a high
385 density of putative host-like motifs that can mediate the interaction with human proteins. Among
386 those highly disordered effectors emerged Tir, a 56 kDa membrane protein essential for virulence.
387 Using multiple biophysical methods, we found that Tir-ICDs (i.e., N-Tir and C-Tir) are highly
388 disordered with IDR features similar to those found in the cytoplasmic domains of host
389 transmembrane proteins. NMR analysis showed that C-Tir is not exclusively disordered but can
390 adopt partial secondary structural elements near phosphosites and motifs involved in versatile
391 interactions with the host (i.e., protein and lipid-binding). The C-Tir displays cytoplasmic tyrosine-
392 based motifs present in the disordered cytosolic tails of host receptors. The standard mode of
393 action associated with these motifs is the phosphorylation of the central tyrosine by Src family
394 PTKs, creating a binding site for SH2–containing proteins. In C-Tir, we found that one of those
395 motifs, involving Y511, displays a pre-formed transient helical structure and binds the C-terminal
396 SH2 domain of SHP-1 pre-phosphorylation, interacting as well with lipid bilayers (**Fig.3, Fig. 7**).
397 These observations highlight a functional diversity for C-Tir and a possible cooperation between
398 host proteins and the membrane in signaling processes inside the host (14). Some T-cell receptor
399 cytoplasmic disordered domains associate with the plasma membrane via tyrosine-based motifs
400 with a helical propensity, and such association regulates phosphorylation and downstream
401 signaling (74). Like the CD3ε cytoplasmic domain of the T-cell receptor (73), C-Tir also has two
402 tyrosines involved in lipid-binding and helical propensity. Our results suggest that the C-terminal
403 half Tir can mimic this regulatory stratagem to fine-tune its ability to interact with human cell
404 components, thereby interfering with normal cellular functions. Further studies are needed to
405 corroborate this initial observation. Nevertheless, we show that the C-Tir can interact with
406 membranes via ITMs and its last C-terminal residues. Importantly, our results established that
407 this membrane affinity is residue-specific and modulated by lipid composition in a quantitative
408 and site-resolved way, suggesting the existence of a regulation layer based on lipid composition.
409 The flexibility and conformational plasticity of C-Tir make it readily accessible for phosphorylation.
410 We found in C-Tir four *bona fide* tyrosine phosphorylation sites that render the ability for
411 multivalent/promiscuous binding, supporting its ability to act as a signaling hub. Even though we
412 only used protein fragments and explored one phosphorylation state, our results support the
413 presence of four phospho-tyrosines (pY) in the C-Tir sequence coexisting simultaneously. All can

414 bind the C-SH2 in the micromolar range as reported for other pY-containing motifs (84).
415 Interestingly, the pY454 site, which displays the lowest affinity, is part of a conserved NPY motif
416 (**Fig. S3**) that interacts with the I-BAR domain of host IRSp53/IRTKS, linking Tir to the actin
417 polymerization machinery (77) in a phosphorylation-independent manner. This suggests that
418 Y454 might act as a binding site for distinct and competing host proteins depending on its
419 phosphorylation state.
420 The superposition of interplaying functional elements observed in C-Tir and their possible
421 interplay mirror the mechanics of human intrinsically disordered domains. Transiently structured
422 lipid-binding regions might bury tyrosine residues rendering them inaccessible, while their
423 phosphorylation perturbs membrane anchoring (85). At the same time, these multiple
424 modification sites promiscuously participate in fuzzy binding of host partners, their respective
425 affinities being interdependent with the phosphorylation state. Future work using high-resolution
426 structural studies targeting the interaction with multiple partners and different C-Tir's
427 phosphorylation states will provide additional insights into Tir binding specificity and action mode.
428 We unraveled for the first time that the N-terminal region, yet partially disordered, forms a stable
429 dimer with potential implications on Tir self-assembly and signaling. Clustering of Tir triggers host
430 signaling events when binding to intimin. The extracellular intimin binding domain (IBD) of Tir
431 binds intimin as a dimer (**Fig. 7**) (36) in a reticulating model (83). Thus, Tir's intracellular self-
432 assembly opens the question of whether host signaling activation involves intimin-induced
433 dimerization or structural changes of preformed dimers.
434 Together, these novel structural insights provide an updated picture of Tir's intracellular side (**Fig.**
435 **7**), laying the path towards illustrating how Tir hijacks host signaling. Importantly, yet less
436 abundant in prokaryotes, the structural disorder of Tir and several other A/E effectors reinforce
437 the idea of a positive evolutionary selection towards disordered proteins in the pool of secreted
438 effectors by bacterial pathogens to target host cellular machinery (20).
439

440 # Methods and Materials

441 ## Disorder prediction and short linear motif analysis

442 We collected the sequence of A/E pathogen effectors and corresponding reference proteomes
443 from the UniprotKB database (86) (**Table S1**). Three effector sets were assembled for the
444 following strains: EHEC, EPEC, and CR. The representative set of 20,365 human proteins was

445 extracted from UniprotKB human reference proteome. To compute disorder propensity at the
446 residue level, we used DISOPRED 3 (41) and IUPred 1.0 ("long" mode prediction) (42), structural
447 disorder predictors. Disordered residues were defined as those with a propensity score equal to
448 or above 0.5. We used this metric to calculate the fraction of disordered residues for each protein.
449 The one-sided Mann-WhitneyU-test (87) was used to compare disorder fraction distribution of
450 the effector collections and their corresponding proteomes (**SI Materials and Methods**). Both
451 per-residue scores and aggregated disorder fractions were used to classify each protein according
452 to the structural categories adapted from (88) (IDP: Intrinsically disordered proteins; PDR:
453 Proteins with intrinsically disordered regions; FRAG: Proteins with fragmented-disorder; NDR: Not
454 disordered proteins; ORD: Ordered Proteins). See **SI Materials and Methods** and **Table S2**
455 for a detailed description of the criteria used to define the structural categories. The list of
456 eukaryotic short linear motifs (ELMs) was downloaded from the ELM database (89) (version 1.4,
457 May 2017). We searched for occurrences of putative ELMs in effector sequences using the
458 ANCHOR tool (90). For each effector sequence, we calculated the motif density as the fraction
459 between the amino acids belonging to putative ELMs predicted as disordered and the total number
460 of amino acids predicted as disordered.

461 Protein expression and purification

462 All constructs (C-Tir, C-SH2, N-Tir, and NS-Tir) were subcloned into the pHTP8 plasmid
463 (NZYTech), bearing a cleavable His6-tagged thioredoxin tag (TRX-His_6) for improved solubility
464 and folding. They were successfully expressed using *E. coli* BL21 Star (DE3) pLysS and purified
465 by affinity chromatography and Size-Exclusion gel filtration. More details in **SI Material and**
466 **Methods.**

467 NMR spectroscopy

468 All NMR experiments were recorded on an 800 MHz Bruker Avance II+ spectrometer equipped
469 with a cryoprobe. The backbone assignment was performed combinating standard triple-
470 resonance experiments and reverse labeling (61). NMR is well suited to study such dynamic and
471 transient interactions in solution at single-residue resolution. We extensively explored the
472 sensitivity of NMR to changes in the chemical environment to monitor phosphorylation, protein-
473 and lipid-binding. More details in **SI Material and Methods.**

Small-angle X-ray scattering (SAXS)

We employed synchrotron SAXS coupled with size exclusion chromatography (SEC) to probe the overall size and conformational properties of Tir intracellular regions. A comprehensive description of SAXS measurements and analysis is described in detail in **SI Material and Methods.**

Bicelles preparation

Bicelles with different charge density were prepared based on the protocols described in refs (76, 91). See **SI Material and Methods** for additional details.

Tyrosine phosphorylation

Fully phosphorylation of four tyrosines in C-Tir was achieved by incubating the protein with human recombinant Fyn Kinase overnight (71). Optimized experimental conditions are detailed on **SI Material and Methods.**

Data and software availability

The NMR chemical shifts of C-Tir and pC-Tir are available at the Biological Magnetic Resonance (BMR) data bank under the entry 50758, 50759, respectively. The SEC-SAXS data can be found at the Small Angle Scattering Biological Data Bank (SASBDB) under the project "SAXS studies on the intracellular region of the translocated intimin receptor". The accession codes are SASDKF8, SASDKG8 and SASDKH8. Proteome and effector collections, disorder predictions, derived aggregated data and figures are available for download https://osf.io/3mka9/. The associated code is available at https://osf.io/cxkjf/.

Acknowledgements

MFMV is funded by a MolBioS FCT PhD program fellowship (PD/BD/135482/2018). National funds financially support GH through the PT-NMR FCT PhD Program (PD/BD/147227/2019). TNC is recipient of the grant CEECIND/01443/2017. This work was funded by national funds through FCT– Fundação para a Ciência e a Tecnologia, I.P. (FCT), Project MOSTMICRO-ITQB with refs

UIDB/04612/2020 and UIDP/04612/2020 and FEDER Funds through COMPETE 2020 (0145-FEDER-007660). We acknowledge using the ESRF-BM29 (MX-2085-BAG) and DLS-B21 Bio-SAXS beamlines (MX20161-1, SM21035-177). We thank Miquel Pons, from UB, and Lígia Martins, and Adriano Henriques, from ITQB-NOVA, for many helpful discussions. We thank Isabel Pacheco, João Carita, Cristina Timoteo, and Teresa Silva (Research Facilities, ITQB-NOVA) for their technical assistance. MA thanks Prof. Hartmut Oschkinat and FMP-Berlin for the financial support. AZ acknowledges the financial support of the JPI HDHL-INTIMIC action cofunded by the Agence Nationale de la Recherche (ANR-17-HDIM-0001-03) and thanks Fabrice Lopez and Aurélie Bergon (TAGC, Marseille) for technical assistance. NMR data were acquired at CERMAX, ITQB-NOVA, Oeiras, Portugal, with equipment funded by FCT, project AAC 01/SAICT/2016.

Author contributions

Conceptualization: TNC

Methodology: MA, AZ, TNC

Investigation: MFMV, GH, TV, MA, AZ, TNC

Supervision: TNC

Writing—review & editing: MA, AZ, TNC

Competing interests

Authors declare that they have no competing interests.

References

1. Z. Peng, *et al.*, Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **72**, 137–151 (2015).
2. M. M. Babu, R. W. Kriwacki, R. V. Pappu, Structural biology. Versatility from protein disorder. *Science* **337**, 1460–1461 (2012).
3. V. N. Uversky, Intrinsically disordered proteins and their "mysterious" (meta)physics. *Front. Phys.* **7** (2019).
4. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
5. C. A. Galea, Y. Wang, S. G. Sivakolundu, R. W. Kriwacki, Regulation of cell division by intrinsically unstructured proteins: intrinsic flexibility, modularity, and signaling conduits.

532 *Biochemistry* **47**, 7598–7609 (2008).

533 6. J. F. Pelham, J. C. Dunlap, J. M. Hurley, Intrinsic disorder is an essential characteristic of
534 components in the conserved circadian circuit. *Cell Commun. Signal.* **18**, 181 (2020).

535 7. A. Bah, J. D. Forman-Kay, Modulation of Intrinsically Disordered Protein Function by Post-
536 translational Modifications. *J. Biol. Chem.* **291**, 6696–6705 (2016).

537 8. R. van der Lee, *et al.*, Intrinsically disordered segments affect protein half-life in the cell
538 and during evolution. *Cell Rep.* **8**, 1832–1844 (2014).

539 9. V. N. Uversky, Wrecked regulation of intrinsically disordered proteins in diseases:
540 pathogenicity of deregulated regulators. *Front. Mol. Biosci.* **1**, 6 (2014).

541 10. N. E. Davey, *et al.*, Attributes of short linear motifs. *Mol. Biosyst.* **8**, 268–281 (2012).

542 11. K. Van Roey, *et al.*, Short linear motifs: ubiquitous and functionally diverse protein
543 interaction modules directing cell regulation. *Chem. Rev.* **114**, 6733–6778 (2014).

544 12. V. Csizmok, A. V. Follis, R. W. Kriwacki, J. D. Forman-Kay, Dynamic protein interaction
545 networks and new structural paradigms in signaling. *Chem. Rev.* **116**, 6424–6462 (2016).

546 13. N. Jespersen, E. Barbar, Emerging Features of Linear Motif-Binding Hub Proteins. *Trends*
547 *Biochem. Sci.* **45**, 375–384 (2020).

548 14. J. Cornish, S. G. Chamberlain, D. Owen, H. R. Mott, Intrinsically disordered proteins and
549 membranes: a marriage of convenience for cell signalling? *Biochem. Soc. Trans.* **48**,
550 2669–2689 (2020).

551 15. E. Deryusheva, *et al.*, Does Intrinsic Disorder in Proteins Favor Their Interaction with
552 Lipids? *Proteomics* **19**, e1800098 (2019).

553 16. J. Bürgi, B. Xue, V. N. Uversky, F. G. van der Goot, Intrinsic disorder in transmembrane
554 proteins: roles in signaling and topology prediction. *PLoS ONE* **11**, e0158594 (2016).

555 17. M. Kjaergaard, B. B. Kragelund, Functions of intrinsic disorder in transmembrane proteins.
556 *Cell. Mol. Life Sci.* **74**, 3205–3224 (2017).

557 18. W. Basile, M. Salvatore, C. Bassot, A. Elofsson, Why do eukaryotic proteins contain more
558 intrinsically disordered regions? *PLoS Comput. Biol.* **15**, e1007186 (2019).

559 19. B. Xue, A. K. Dunker, V. N. Uversky, Orderly order in protein intrinsic disorder distribution:
560 disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct.*
561 *Dyn.* **30**, 137–149 (2012).

562 20. M. Marín, V. N. Uversky, T. Ott, Intrinsic disorder in pathogen effectors: protein flexibility
563 as an evolutionary hallmark in a molecular arms race. *Plant Cell* **25**, 3153–3157 (2013).

564 21. D. P. O'Brien, *et al.*, Calmodulin fishing with a structurally disordered bait triggers CyaA
565 catalysis. *PLoS Biol.* **15**, e2004486 (2017).

566 22. O. Aitio, *et al.*, Enterohaemorrhagic Escherichia coli exploits a tryptophan switch to hijack
567 host f-actin assembly. *Structure* **20**, 1692–1703 (2012).

568 23. D. Hamada, M. Hamaguchi, K. N. Suzuki, I. Sakata, I. Yanagihara, Cytoskeleton-
569 modulating effectors of enteropathogenic and enterohemorrhagic Escherichia coli: a case
570 for EspB as an intrinsically less-ordered effector. *FEBS J.* **277**, 2409–2415 (2010).

571 24. T. H. Hazen, *et al.*, Genomic diversity of EPEC associated with clinical presentations of
572 differing severity. *Nat. Microbiol.* **1**, 15014 (2016).

573 25. H. Pennington, Escherichia coli O157. *Lancet* **376**, 1428–1435 (2010).

574 26. W. Deng, *et al.*, Dissecting virulence: systematic and functional analyses of a
575 pathogenicity island. *Proc Natl Acad Sci USA* **101**, 3597–3602 (2004).

576 27. T. Tobe, *et al.*, An extensive repertoire of type III secretion effectors in Escherichia coli
577 O157 and the role of lambdoid phages in their dissemination. *Proc Natl Acad Sci USA*
578 **103**, 14941–14946 (2006).

579 28. E. Mills, K. Baruch, X. Charpentier, S. Kobi, I. Rosenshine, Real-time analysis of effector
580 translocation by the type III secretion system of enteropathogenic Escherichia coli. *Cell*
581 *Host Microbe* **3**, 104–113 (2008).

582 29. B. Kenny, *et al.*, Enteropathogenic E. coli (EPEC) transfers its receptor for intimate
583 adherence into mammalian cells. *Cell* **91**, 511–520 (1997).

584 30. S. E. Battle, M. J. Brady, S. K. Vanaja, J. M. Leong, G. A. Hecht, Actin pedestal formation
585 by enterohemorrhagic Escherichia coli enhances bacterial host cell attachment and
586 concomitant type III translocation. *Infect. Immun.* **82**, 3713–3722 (2014).

587 31. S. Gruenheid, *et al.*, Enteropathogenic E. coli Tir binds Nck to initiate actin pedestal
588 formation in host cells. *Nat. Cell Biol.* **3**, 856–859 (2001).

589 32. Y. Xue, M. Du, H. Sheng, C. J. Hovde, M.-J. Zhu, Escherichia coli O157:H7 suppresses
590 host autophagy and promotes epithelial adhesion via Tir-mediated and cAMP-independent
591 activation of protein kinase A. *Cell Death Discov.* **3**, 17055 (2017).

592 33. D. Yan, X. Wang, L. Luo, X. Cao, B. Ge, Inhibition of TLR signaling by a bacterial protein
593 containing immunoreceptor tyrosine-based inhibitory motifs. *Nat. Immunol.* **13**, 1063–
594 1071 (2012).

595 34. P. J. Goddard, *et al.*, Enteropathogenic Escherichia coli Stimulates Effector-Driven Rapid
596 Caspase-4 Activation in Human Macrophages. *Cell Rep.* **27**, 1008-1017.e6 (2019).

597 35. Q. Zhong, *et al.*, Clustering of Tir during enteropathogenic E. coli infection triggers
598 calcium influx-dependent pyroptosis in intestinal epithelial cells. *PLoS Biol.* **18**, e3000986
599 (2020).

600 36. Y. Luo, *et al.*, Crystal structure of enteropathogenic Escherichia coli intimin-receptor
601 complex. *Nature* **405**, 1073–1077 (2000).

602 37. A. R. C. Wong, *et al.*, Enteropathogenic and enterohaemorrhagic Escherichia coli: even
603 more subversive elements. *Mol. Microbiol.* **80**, 1420–1438 (2011).

604 38. S. Blasche, *et al.*, The EHEC-host interactome reveals novel targets for the translocated

605 intimin receptor. *Sci. Rep.* **4**, 7531 (2014).

606 39. D. Yan, *et al.*, Enteropathogenic Escherichia coli Tir recruits cellular SHP-2 through ITIM
607 motifs to suppress host immune response. *Cell. Signal.* **25**, 1887–1894 (2013).

608 40. R. Zhou, *et al.*, Enterohemorrhagic Escherichia coli Tir inhibits TAK1 activation and
609 mediates immune evasion. *Emerg. Microbes Infect.* **8**, 734–748 (2019).

610 41. D. T. Jones, D. Cozzetto, DISOPRED3: precise disordered region predictions with
611 annotated protein-binding activity. *Bioinformatics* **31**, 857–863 (2015).

612 42. Z. Dosztányi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of
613 intrinsically unstructured regions of proteins based on estimated energy content.
614 *Bioinformatics* **21**, 3433–3434 (2005).

615 43. B. Mészáros, G. Erdos, Z. Dosztányi, IUPred2A: context-dependent prediction of protein
616 disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–
617 W337 (2018).

618 44. A. M. Grishin, *et al.*, NleH defines a new family of bacterial effector kinases. *Structure* **22**,
619 250–259 (2014).

620 45. X. Gao, *et al.*, Bacterial effector binding to ribosomal protein s3 subverts NF-kappaB
621 function. *PLoS Pathog.* **5**, e1000708 (2009).

622 46. A. Holmes, S. Mühlen, A. J. Roe, P. Dean, The EspF effector, a bacterial pathogen's Swiss
623 army knife. *Infect. Immun.* **78**, 4445–4453 (2010).

624 47. N. A. Sallee, *et al.*, The pathogen protein EspF(U) hijacks actin polymerization using
625 mimicry and multivalency. *Nature* **454**, 1005–1008 (2008).

626 48. T. Hagai, A. Azia, M. M. Babu, R. Andino, Use of host-like peptide motifs in viral proteins
627 is a prevalent strategy in host-virus interactions. *Cell Rep.* **7**, 1729–1739 (2014).

628 49. A. Via, B. Uyar, C. Brun, A. Zanzoni, How pathogens use linear motifs to perturb host cell
629 networks. *Trends Biochem. Sci.* **40**, 36–48 (2015).

630 50. N. E. Davey, G. Travé, T. J. Gibson, How viruses hijack cell regulation. *Trends Biochem.*
631 *Sci.* **36**, 159–169 (2011).

632 51. B. Bommarius, *et al.*, Enteropathogenic Escherichia coli Tir is an SH2/3 ligand that recruits
633 and activates tyrosine kinases required for pedestal formation. *Mol. Microbiol.* **63**, 1748–
634 1768 (2007).

635 52. P. Tompa, N. E. Davey, T. J. Gibson, M. M. Babu, A million peptide motifs for the
636 molecular biologist. *Mol. Cell* **55**, 161–169 (2014).

637 53. L. Staby, *et al.*, Eukaryotic transcription factors: paradigms of protein intrinsic disorder.
638 *Biochem. J.* **474**, 2509–2532 (2017).

639 54. C. Haynes, *et al.*, Intrinsic disorder is a common feature of hub proteins from four
640 eukaryotic interactomes. *PLoS Comput. Biol.* **2**, e100 (2006).

641 55. P. Bernadó, M. Blackledge, A self-consistent description of the conformational behavior of
642 chemically denatured proteins from NMR and small angle scattering. *Biophys. J.* **97**,
643 2839–2845 (2009).

644 56. T. N. Cordeiro, *et al.*, Structural Characterization of Highly Flexible Proteins by Small-
645 Angle Scattering. *Adv. Exp. Med. Biol.* **1009**, 107–129 (2017).

646 57. A. G. Kikhney, D. I. Svergun, A practical guide to small angle X-ray scattering (SAXS) of
647 flexible and intrinsically disordered proteins. *FEBS Lett.* **589**, 2570–2577 (2015).

648 58. L. B. Chemes, L. G. Alonso, M. G. Noval, G. de Prat-Gay, Circular dichroism techniques for
649 the analysis of intrinsically disordered proteins and domains. *Methods Mol. Biol.* **895**,
650 387–404 (2012).

651 59. E. Delaforge, T. N. Cordeiro, P. Bernadó, N. Sibille, "Conformational characterization of
652 intrinsically disordered proteins and its biological significance" in *Modern Magnetic*
653 *Resonance*, G. A. Webb, Ed. (Springer International Publishing, 2018), pp. 381–399.

654 60. J. Garmendia, *et al.*, TccP is an enterohaemorrhagic Escherichia coli O157:H7 type III
655 effector protein that couples Tir to the actin-cytoskeleton. *Cell. Microbiol.* **6**, 1167–1183
656 (2004).

657 61. C. Prasanna, A. Dubey, H. S. Atreya, Amino acid selective unlabeling in protein NMR
658 spectroscopy. *Meth. Enzymol.* **565**, 167–189 (2015).

659 62. K. Tamiola, B. Acar, F. A. A. Mulder, Sequence-specific random coil chemical shifts of
660 intrinsically disordered proteins. *J. Am. Chem. Soc.* **132**, 18000–18003 (2010).

661 63. M. Fuxreiter, I. Simon, P. Friedrich, P. Tompa, Preformed structural elements feature in
662 partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* **338**, 1015–1026
663 (2004).

664 64. A. De Biasio, *et al.*, p15PAF is an intrinsically disordered protein with nonrandom
665 structural preferences at sites of interaction with other proteins. *Biophys. J.* **106**, 865–
666 874 (2014).

667 65. K. Tamiola, F. A. A. Mulder, Using NMR chemical shifts to calculate the propensity for
668 structural order and disorder in proteins. *Biochem. Soc. Trans.* **40**, 1014–1020 (2012).

669 66. M.-H. Cho, J. O. Wrabl, J. Taylor, V. J. Hilser, Hidden dynamic signatures drive substrate
670 selectivity in the disordered phosphoproteome. *Proc Natl Acad Sci USA* **117**, 23606–
671 23616 (2020).

672 67. N. Stanley, S. Esteban-Martín, G. De Fabritiis, Kinetic modulation of a disordered protein
673 domain by phosphorylation. *Nat. Commun.* **5**, 5272 (2014).

674 68. A. Hawrani, *et al.*, Effect of protein kinase A-mediated phosphorylation on the structure
675 and association properties of the enteropathogenic Escherichia coli Tir virulence protein.
676 *J. Biol. Chem.* **278**, 25839–25846 (2003).

677 69. P. R. Race, A. S. Solovyova, M. J. Banfield, Conformation of the EPEC Tir protein in

678 solution: investigating the impact of serine phosphorylation at positions 434/463. *Biophys.*
679 *J.* **93**, 586–596 (2007).

680 70. S. Backert, B. Kenny, R. Gerhard, N. Tegtmeyer, S. Brandt, PKA-mediated
681 phosphorylation of EPEC-Tir at serine residues 434 and 463: A novel pathway in
682 regulating Rac1 GTPase function. *Gut Microbes* **1**, 94–99 (2010).

683 71. J. Rosenlöw, L. Isaksson, M. Mayzel, J. Lengqvist, V. Y. Orekhov, Tyrosine
684 phosphorylation within the intrinsically disordered cytosolic domains of the B-cell receptor:
685 an NMR-based structural analysis. *PLoS ONE* **9**, e96199 (2014).

686 72. H. Cheng, *et al.*, Conformational Changes in the Cytoplasmic Region of KIR3DL1 upon
687 Interaction with SHP-2. *Structure* **27**, 639-650.e2 (2019).

688 73. C. Xu, *et al.*, Regulation of T cell receptor activation by dynamic membrane binding of the
689 CD3epsilon cytoplasmic tyrosine-based motif. *Cell* **135**, 702–713 (2008).

690 74. X. Xu, H. Li, C. Xu, Structural understanding of T cell receptor triggering. *Cell. Mol.*
691 *Immunol.* **17**, 193–202 (2020).

692 75. E. Duchardt, A. B. Sigalov, D. Aivazian, L. J. Stern, H. Schwalbe, Structure induction of
693 the T-cell receptor zeta-chain upon lipid binding investigated by NMR spectroscopy.
694 *Chembiochem* **8**, 820–827 (2007).

695 76. Y. Pérez, *et al.*, Lipid binding by the Unique and SH3 domains of c-Src suggests a new
696 regulatory mechanism. *Sci. Rep.* **3**, 1295 (2013).

697 77. J. C. de Groot, *et al.*, Structural basis for complex formation between human IRSp53 and
698 the translocated intimin receptor Tir of enterohemorrhagic E. coli. *Structure* **19**, 1294–
699 1306 (2011).

700 78. F.-X. Theillet, *et al.*, Site-specific NMR mapping and time-resolved monitoring of serine
701 and threonine phosphorylation in reconstituted kinase reactions and mammalian cell
702 extracts. *Nat. Protoc.* **8**, 1416–1432 (2013).

703 79. T. Mittag, *et al.*, Dynamic equilibrium engagement of a polyvalent ligand with a single-site
704 receptor. *Proc Natl Acad Sci USA* **105**, 17772–17777 (2008).

705 80. A. L. Turner, *et al.*, Highly disordered histone H1-DNA model complexes and their
706 condensates. *Proc Natl Acad Sci USA* **115**, 11964–11969 (2018).

707 81. A. Patel, *et al.*, Host protein interactions with enteropathogenic Escherichia coli (EPEC):
708 14-3-3tau binds Tir and has a role in EPEC-induced actin polymerization. *Cell. Microbiol.*
709 **8**, 55–71 (2006).

710 82. C. Rat, J. C. Heiby, J. P. Bunz, H. Neuweiler, Two-step self-assembly of a spider silk
711 molecular clamp. *Nat. Commun.* **9**, 4779 (2018).

712 83. T. Touzé, R. D. Hayward, J. Eswaran, J. M. Leong, V. Koronakis, Self-association of EPEC
713 intimin mediated by the beta-barrel-containing anchor domain: a role in clustering of the
714 Tir receptor. *Mol. Microbiol.* **51**, 73–87 (2004).

84. M. J. Wagner, M. M. Stacey, B. A. Liu, T. Pawson, Molecular mechanisms of SH2- and PTB-domain-containing proteins in receptor tyrosine kinase signaling. *Cold Spring Harb. Perspect. Biol.* **5**, a008987 (2013).

85. M. Arbesú, G. Iruela, H. Fuentes, J. M. C. Teixeira, M. Pons, Intramolecular fuzzy interactions involving intrinsically disordered domains. *Front. Mol. Biosci.* **5**, 39 (2018).

86. UniProt Consortium, UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).

87. H. B. Mann, D. R. Whitney, On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18**, 50–60 (1947).

88. A. Giansanti, A. Deiana, S. Forcelloni, A. Porrello, New classification of intrinsic disorder in the human proteome. *BioRxiv* (2018) https:/doi.org/10.1101/446351.

89. H. Dinkel, *et al.*, ELM 2016--data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.* **44**, D294-300 (2016).

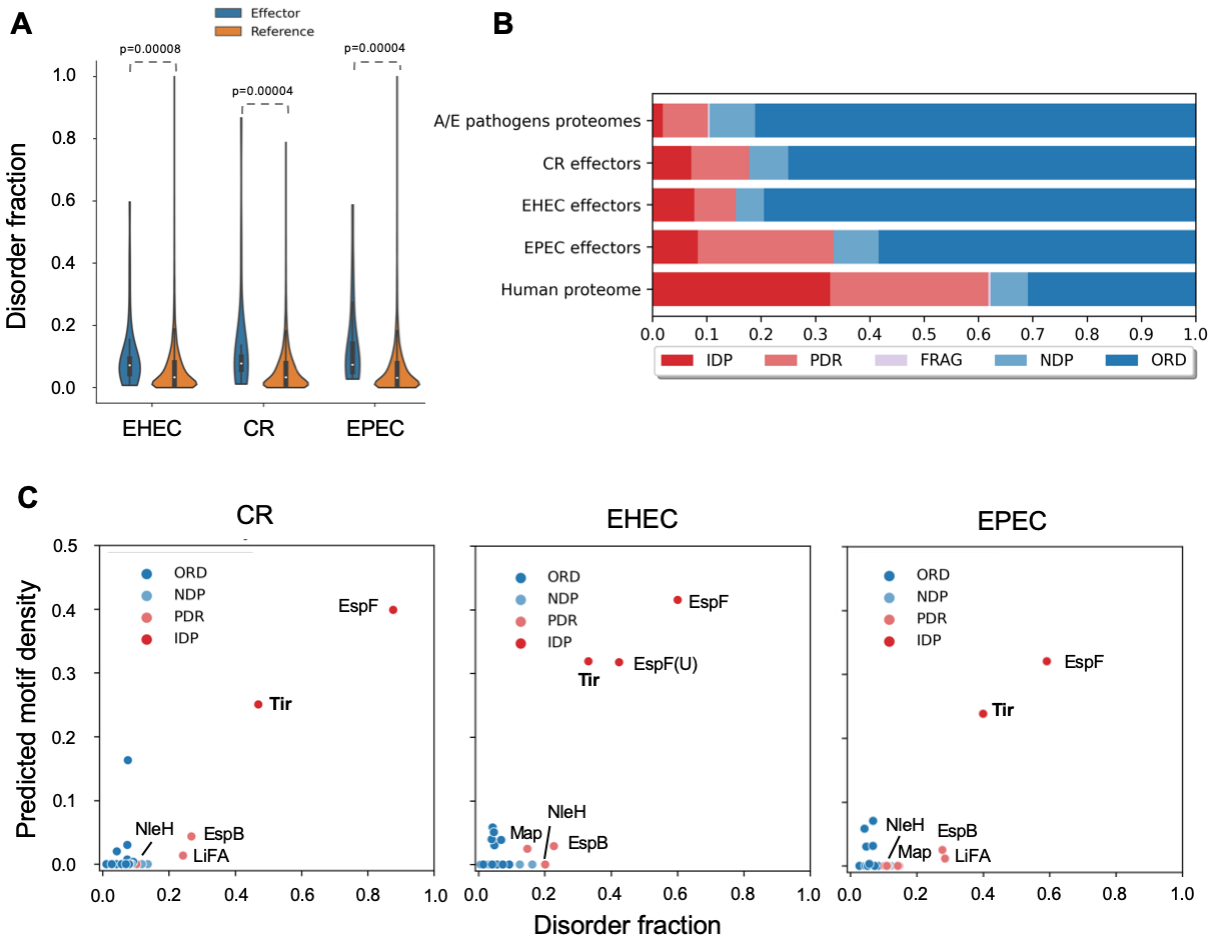
90. B. Mészáros, I. Simon, Z. Dosztányi, Prediction of protein binding regions in disordered proteins. *PLoS Comput. Biol.* **5**, e1000376 (2009).

91. M. Maffei, *et al.*, Lipid binding by disordered proteins. *Protoc. exch.* (2013) https:/doi.org/10.1038/protex.2013.094.

92. D. I. Svergun, M. V. Petoukhov, M. H. Koch, Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).

Figures

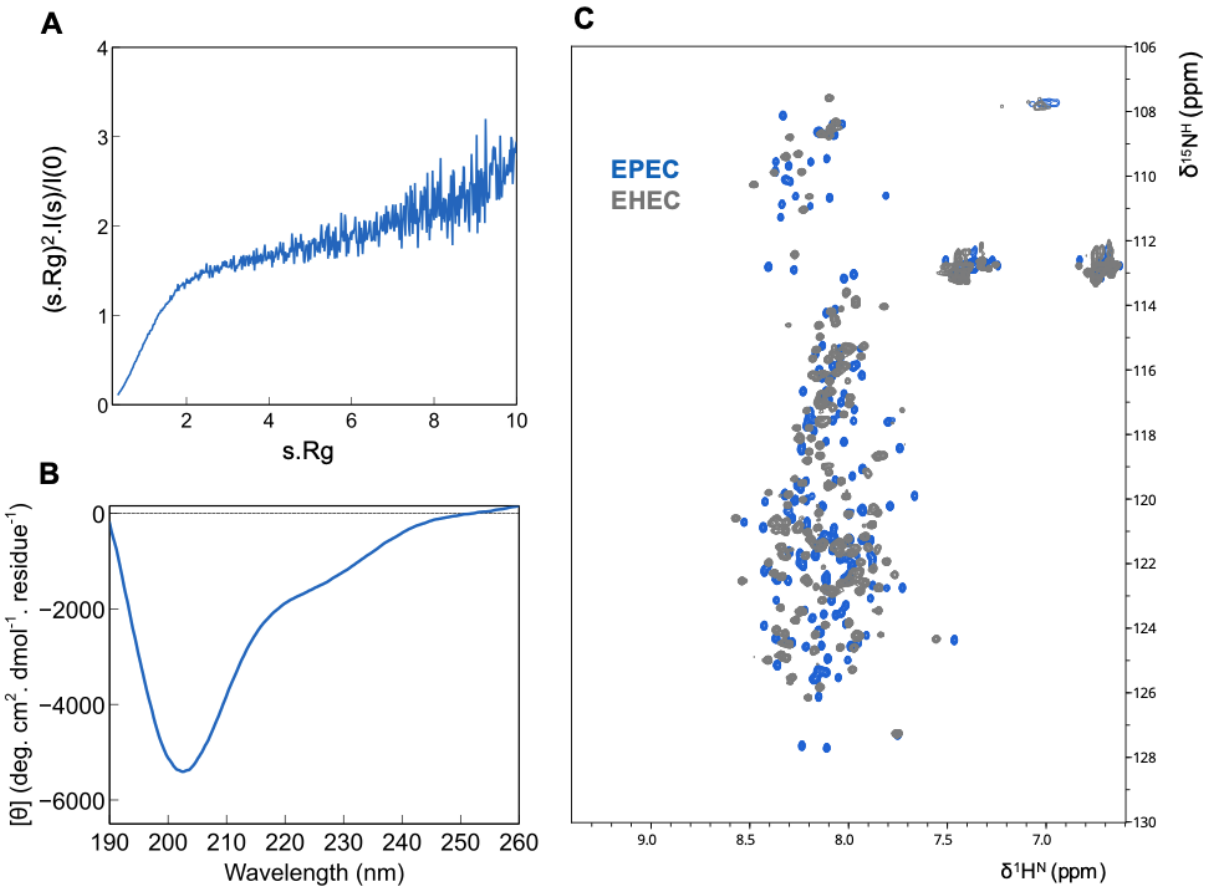
Fig. 1 - Predicted structural disorder in A/E pathogens.



(*A*) Distribution of disorder fraction displayed as violin plots for effectors (blue) and full-proteomes (red) of A/E pathogens. (*B*) Accumulated fractions of the structural categories in terms of structural disorder. IDP: Intrinsically disordered proteins; PDR: Proteins with intrinsically disordered regions; FRAG: Proteins with fragmented-disorder; NDR: Not disordered proteins; ORD: Ordered Proteins. See full definitions in **Table S2** and the individual A/E pathogen proteomes in **Fig. S2**. (*C*) Fraction of ELMs vs. disorder fraction in A/E effectors. Tir, EspF, and EspF(U) display a high motif content within disordered regions. IDP and PDR effectors are labeled. See all data in **Table S3.**

762 **Fig. 2 - C-Tir is an intrinsically disordered domain.**

763



764

765 (*A*) Kratky representation of the SAXS profile measured for C-Tir indicating its lack of compactness. (*B*) Far-
766 UV circular dichroism of C-Tir revealing the absence of high-populated structural secondary elements. (*C*)
767 C-Tir EPEC (light blue) and C-Tir EHEC (grey) [^{1}H-^{15}N]-HSQC spectra reveal the narrow ^{1}H chemical shift
768 dispersion characteristic of IDPs with ^{1}H amide backbone resonances clustering between 7.7 and 8.5 ppm.
769 All together provide a definite diagnostic of protein disorder.

770

771

772

773

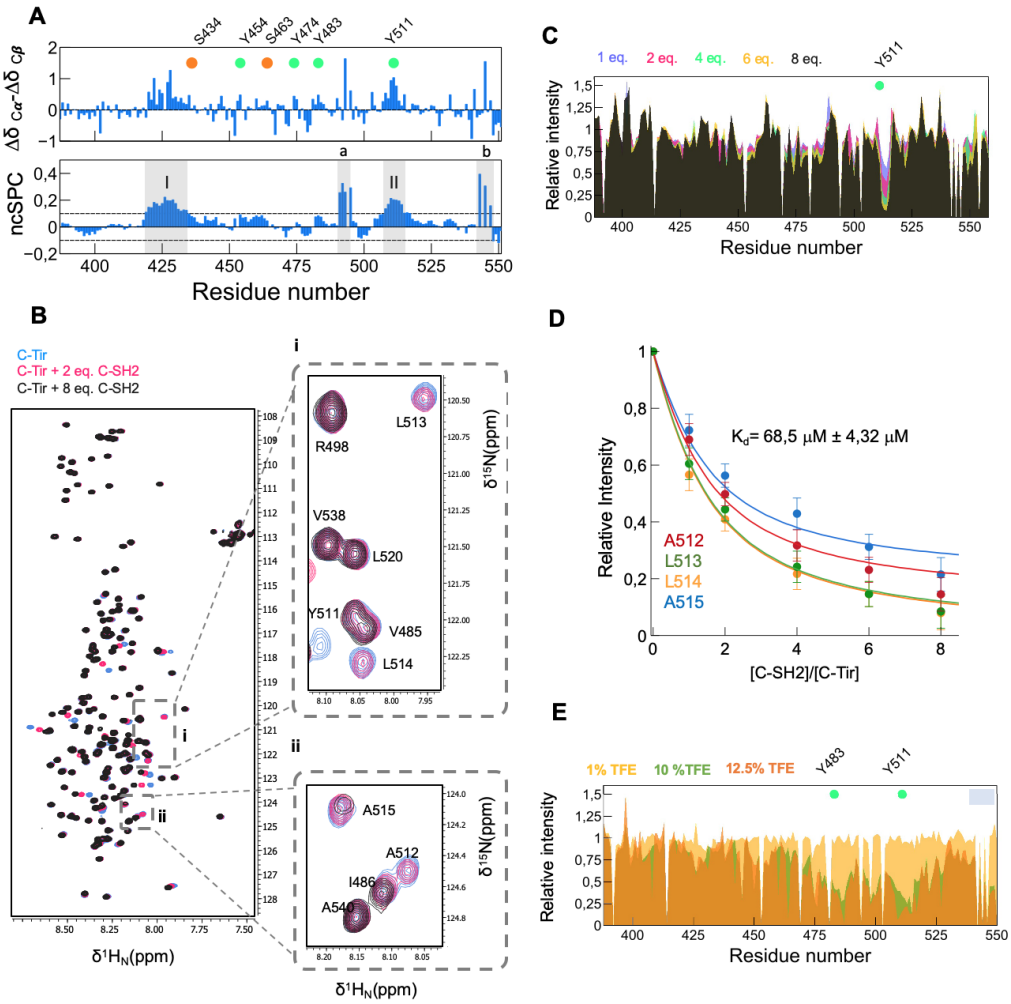
774

775

776

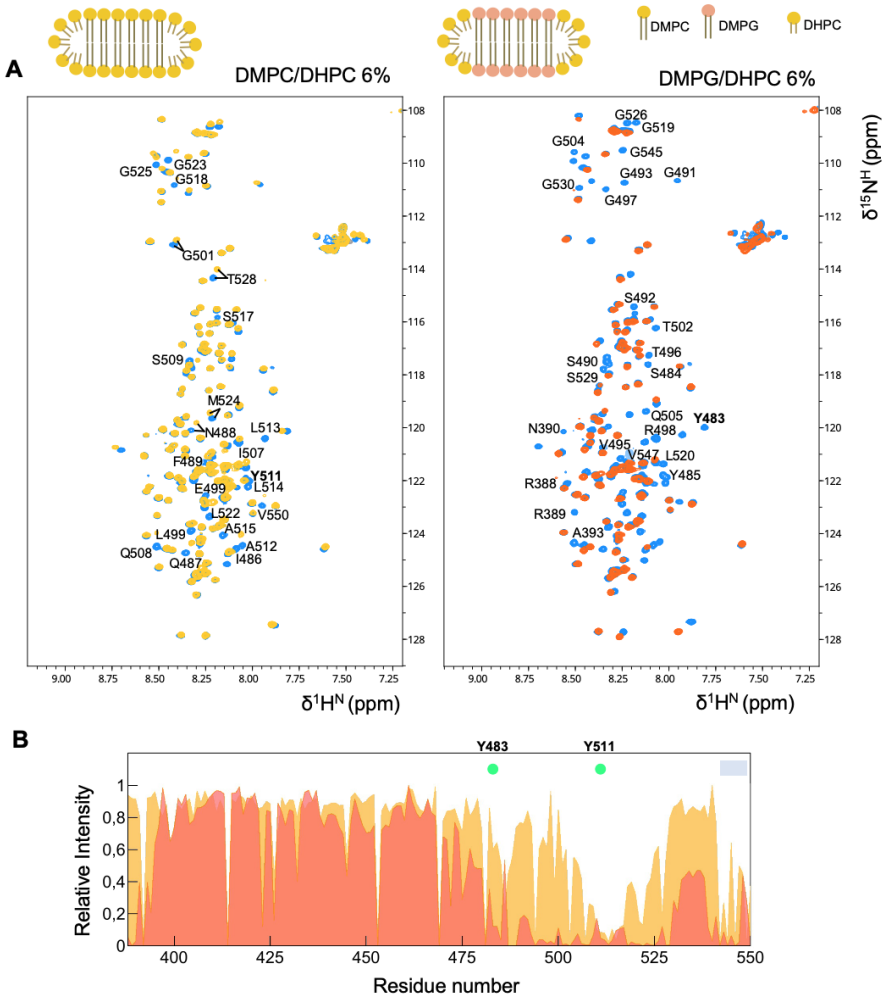
777

Fig. 3 - C-Tir is disordered with residual secondary structure at sites of host interaction.



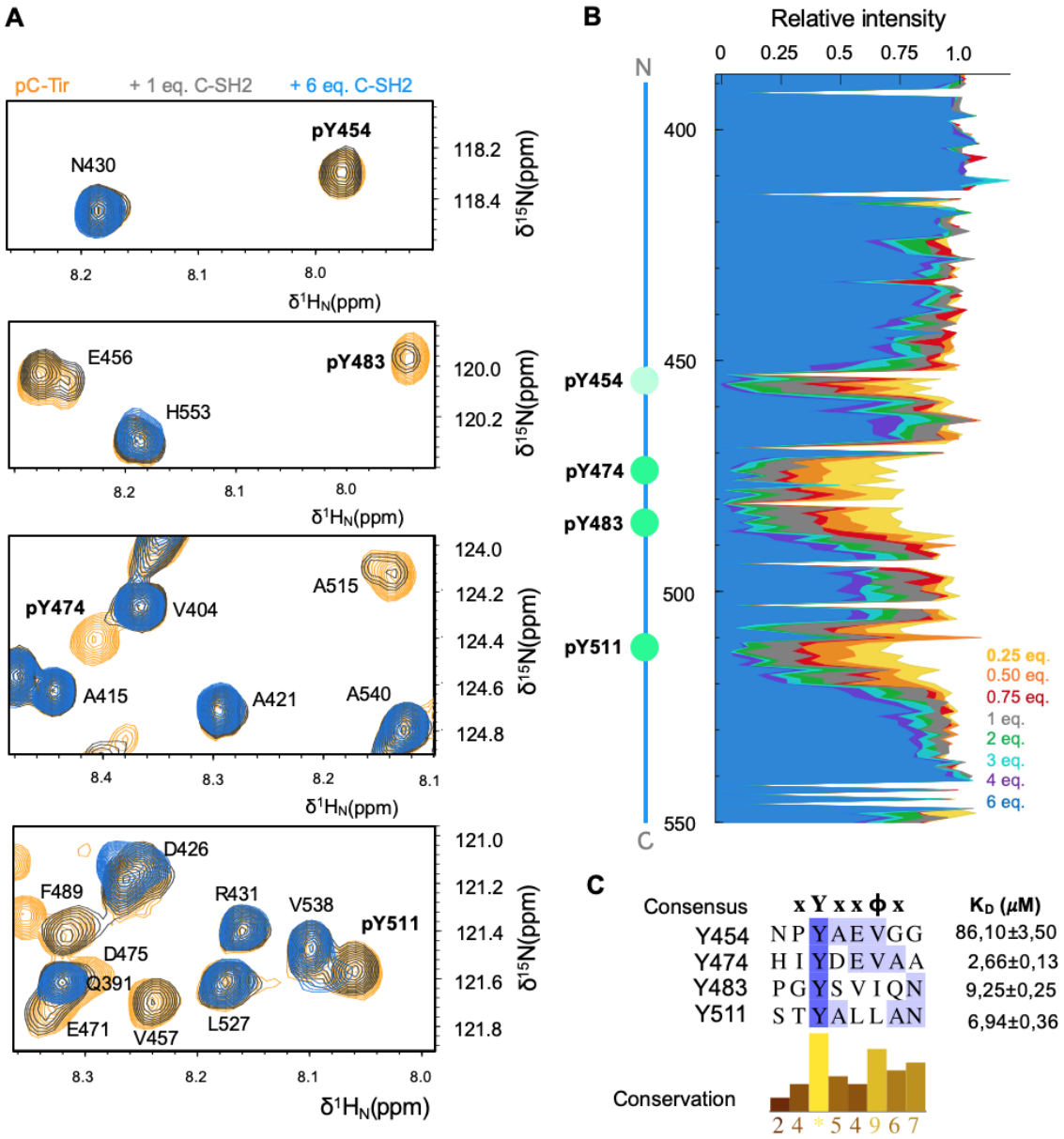
(*A*) Secondary chemical shifts (ΔδCα-ΔδCβ) (top) and structural propensity plot (bottom) for C-Tir where the dashed-lines depicts the random-coil threshold. Values above or below this threshold have α-helix or β-sheet propensities, respectively. Positive values (grey regions) show an increased α-helix tendency in regions D420-Q435 (I) and I507-A515 (II), as well as (a) F489-V495 and (b) T543-V547. Green and orange circles mark phosphorylation sites on C-Tir for tyrosine and serine residues, respectively. (*B*) Overlay of [^{1}H-^{15}N]-HSQC spectra of C-Tir in the absence (blue) and the presence of C-SH2 at 2.0 (pink) and 8.0 (dark grey) equivalents. Expansions (right panels) show the NMR signal of residues $A_{512}LLA_{515}$ with a significant intensity drop due to C-SH2 binding. (*C*) Relative [^{1}H-^{15}N]-peak intensities from the titration analysis. (*D*) Global fitting NMR quantification of the binding of C-SH2 to unphosphorylated C-Tir. (*E*) Residue-resolution mapping of the effect of TFE on C-Tir. TFE-induced NMR attenuation profiles, i.e., the ratio of peak intensity in the presence and absence of TFE, are plotted along the sequence. Green circles mark the position of Y483 and Y511. The grey bar highlights the C-terminal residues with NMR-signal attenuation.

Fig. 4 - Lipid-binding by disordered C-Tir.



(*A*) NMR [^{1}H-^{15}N]-HSQC spectra of C-Tir in the absence (blue) and the presence of DMPC/DHPC (6% w/v), yellow) or DMPG/DHPC (6% w/v), orange) bicelles. C-Tir's HSQC data in bicelles superimpose well to free C-Tir, with identical chemical shifts for most visible resonances, pointing to intrinsic disorder. Residues with NMR signal attenuation upon addition of DMPC/DHPC bicelles are indicated in the spectra (left). The additional residues affected by DMPG/DHPC are also shown (right). (*B*) Lipid-induced NMR attenuation profiles are plotted against C-Tir primary sequence. Residues comprising the Y483 and Y511 ITIM motifs became less intense in lipids' presence, reinforcing their inherent ability to bind lipids. The grey bar highlights the C-terminal residues affected by lipids.

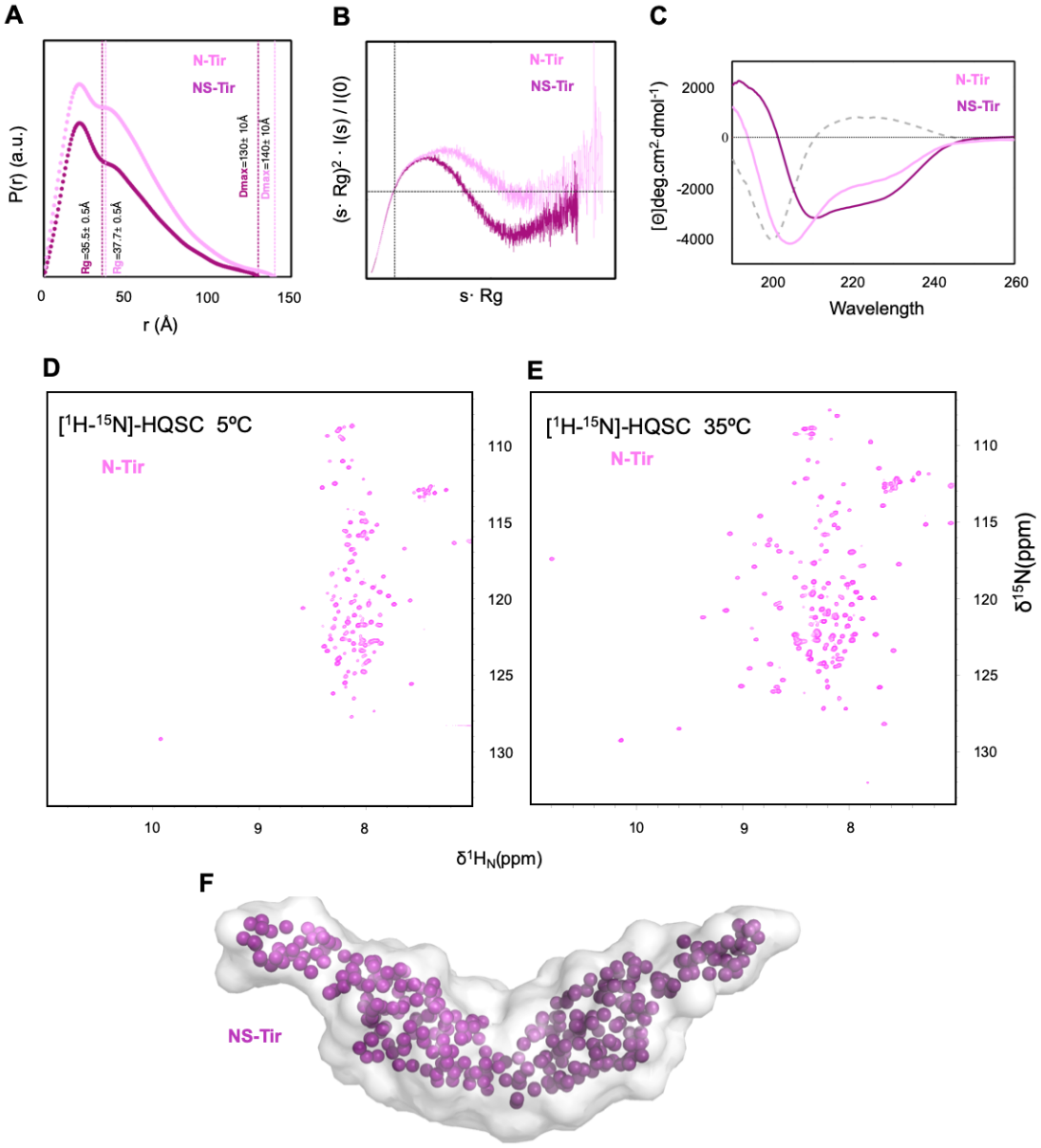
807 **Fig. 5 - C-Tir contains multiple phosphorylated docking sites.**



808

809 (*A*) NMR-detected binding of C-SH2 to phosphotyrosine pY454, pY474, pY483, pY511 before (orange) and
810 after adding 1.0 and 6.0 equivalents of C-SH2 (blue and grey, respectively). (*B*) Site-specific NMR signal
811 attenuation on pC-Tir due to C-SH2 binding, with local intensity drop around each phosphorylation site. (*C*)
812 Sequence alignment of xYx(x)ϕ motifs of C-Tir interacting with C-SH2 and their respective apparent K_D
813 values obtained by global fitting 1:1 model the local intensity drop around each phosphorylation site. Darker
814 shades of blue refer to higher sequence identity. The conservation of physicochemical properties in each
815 alignment position is reported in the corresponding barplot below the alignment. Green circles denote
816 phosphorylation tyrosine-sites with color intensity reflecting the apparent relative affinity.

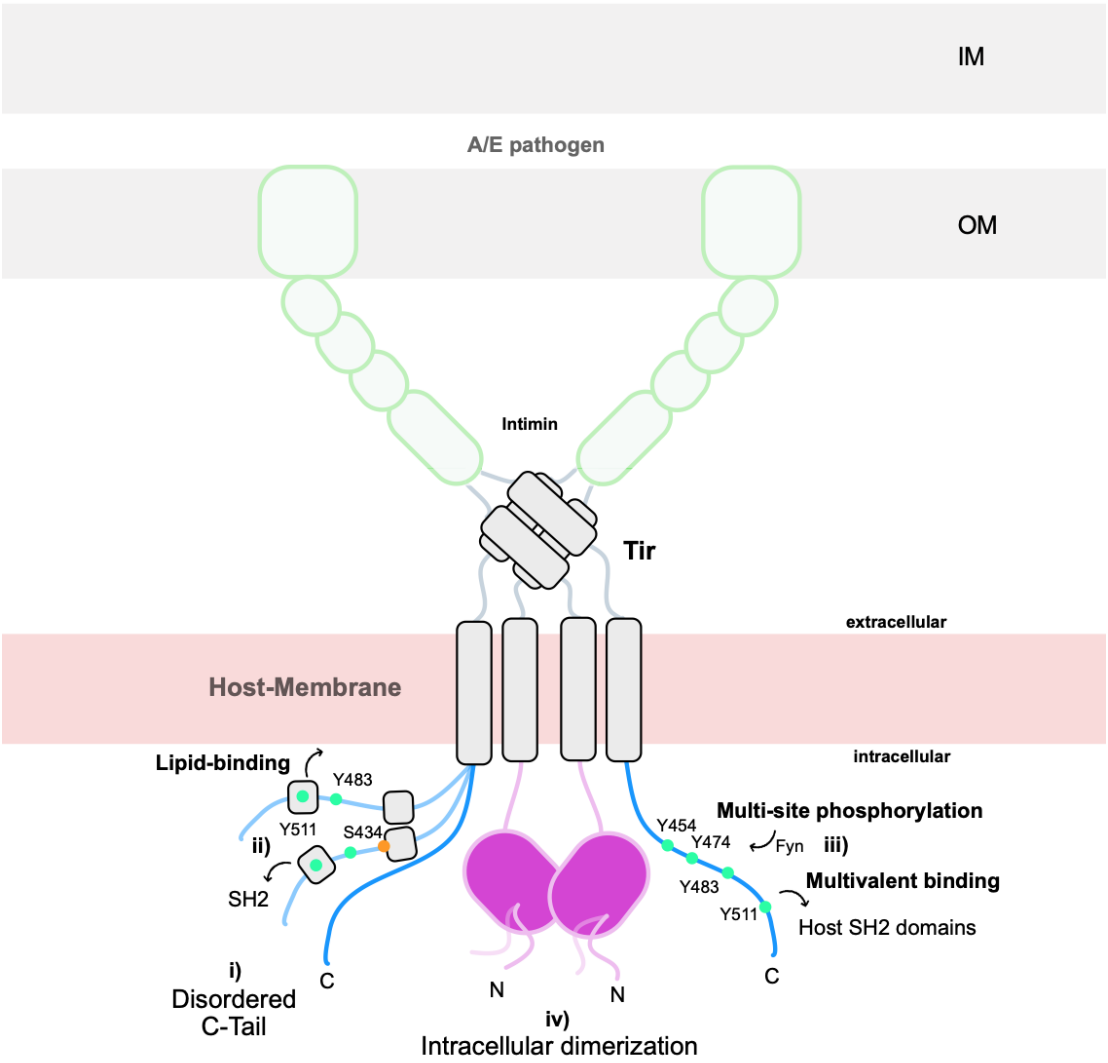
Fig. 6 - N-Tir is a partially disordered dimer.



(*A*) Normalized pairwise distance distribution, P(r), computed from experimental SAXS curves of N-Tir (pink) and NS-Tir (purple) consistent with an elongated dimer. Dashed lines indicate the derived Rg and Dmax values. (*B*) Kratky plots of N-Tir (pink) and NS-Tir (dark-magenta), highlighting a high conformationally flexibility for N-Tir. (*C*) Far-CD spectra of N-Tir (pink) and NS-Tir (dark-magenta) and their difference (grey dashed line). (*D*) [^{1}H-^{15}N]-HSQC spectra of N-Tir at 5 ºC and (*E*) 35 ºC revealing two dynamically different regions within the protein. (*F*) NS-tir dimer at low-resolution driven from SAXS with GASBOR (92).

827 **Fig. 7 - A proposed model of Tir structural organization.**

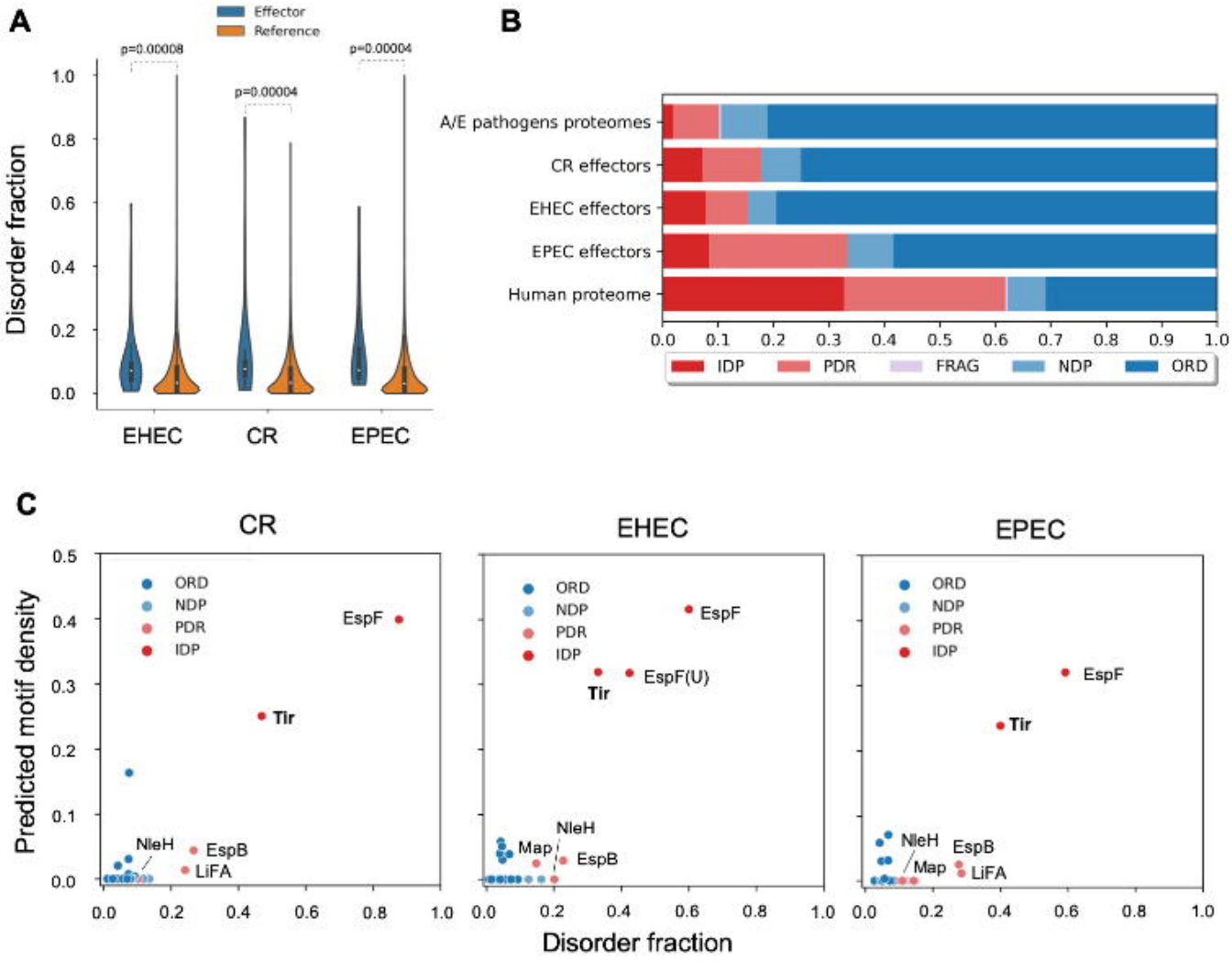
828

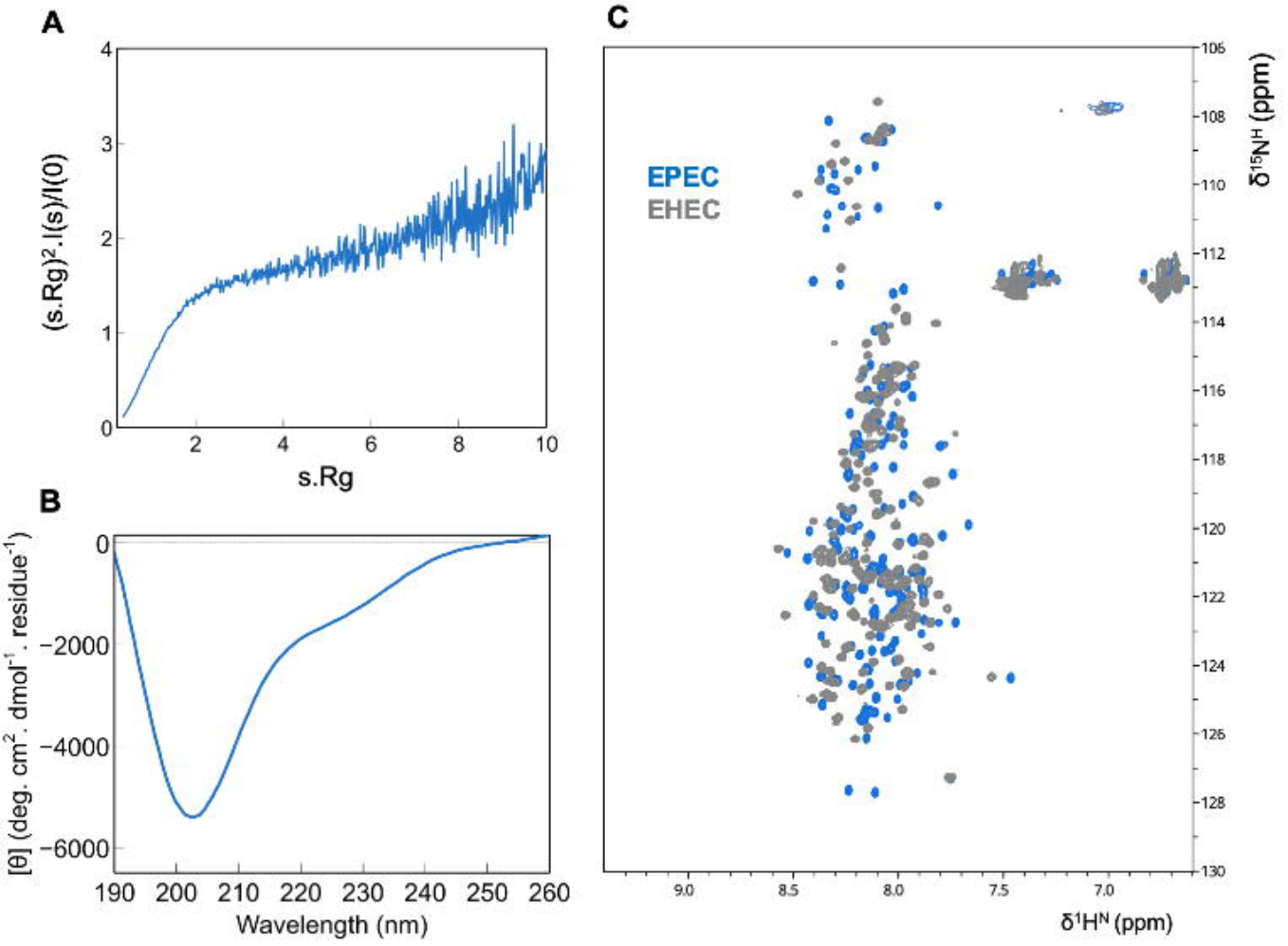


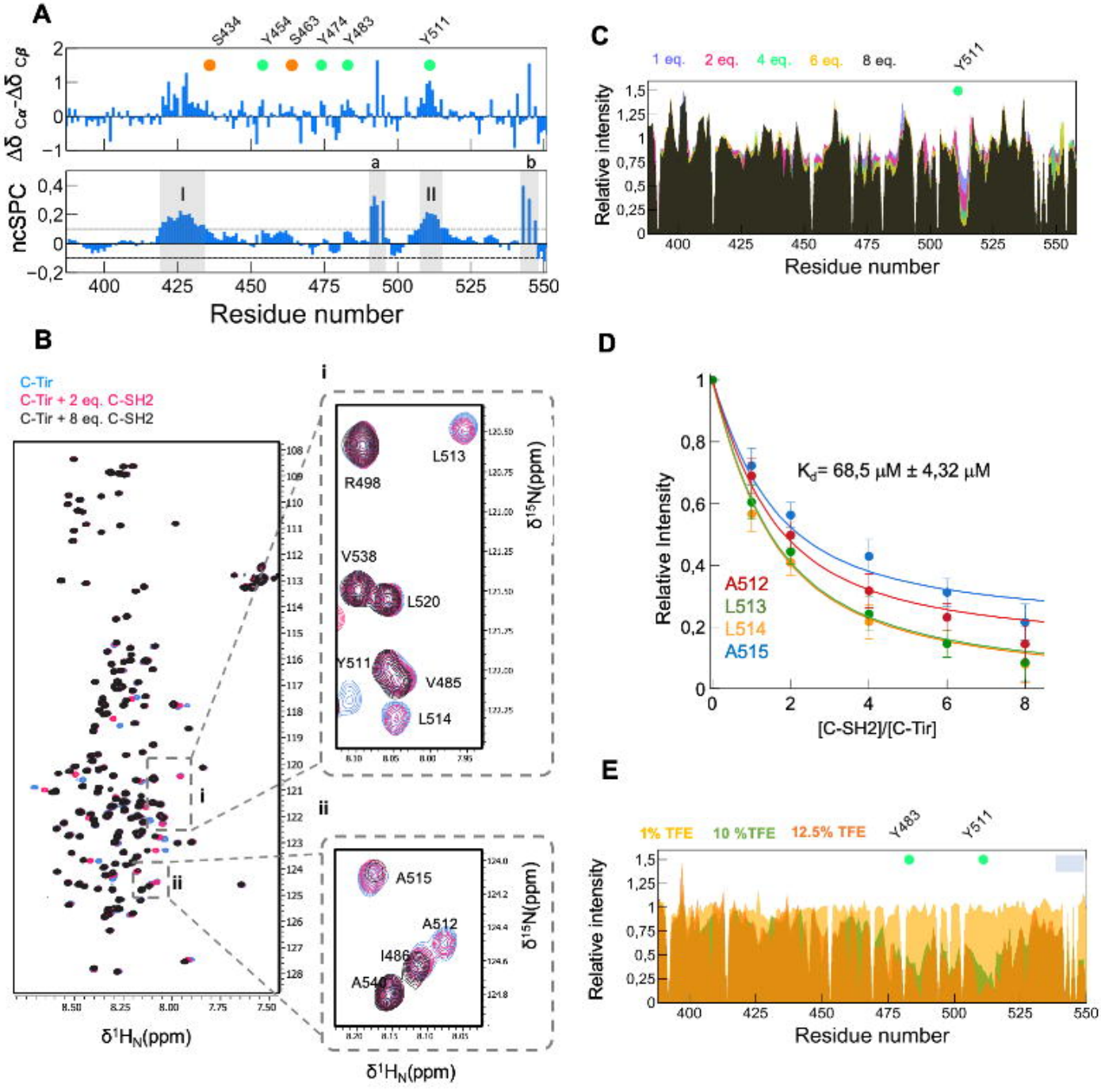
829

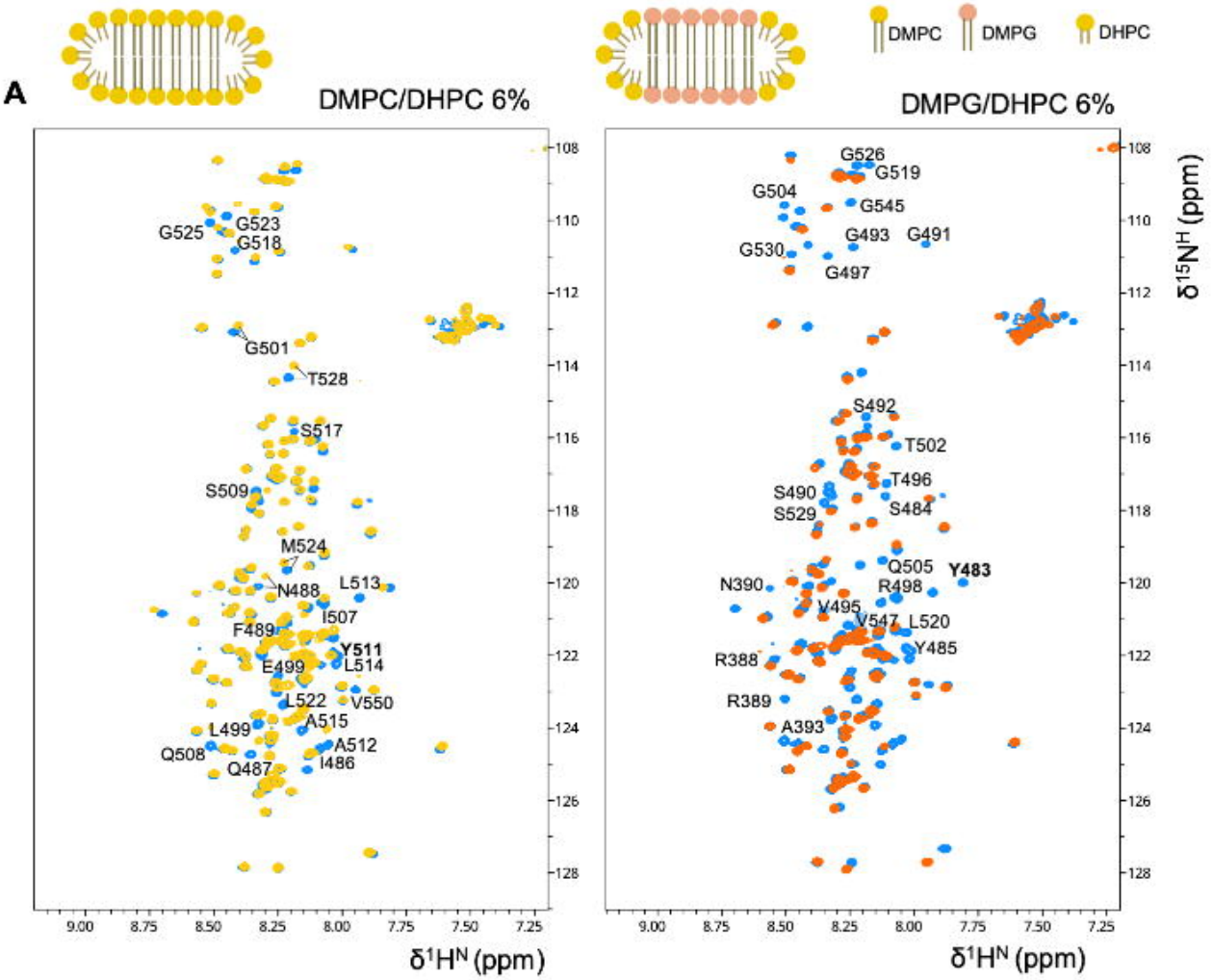
830

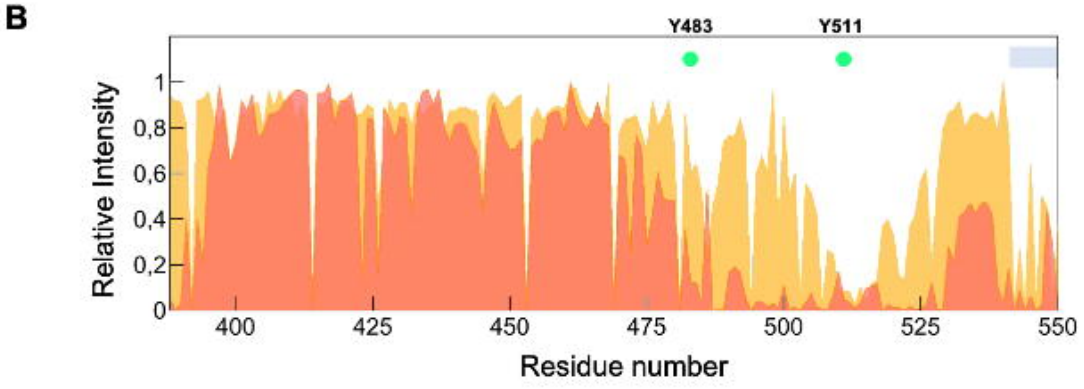
831 Tir in the host plasma membrane is depicted in a hairpin-topology while bound to intimin at the bacterial
832 outer membrane. Our data support that (i) the C-terminal intracellular region of Tir (C-Tir) membrane
833 receptor is disordered; (ii) with transient helical structural elements involved protein- and lipid-interactions;
834 (iii) and host-like multi-phosphorylation sites that serve for docking various host SH2-domains. (iv) The
835 intracellular N-terminal region of Tir (N-Tir) is partially disordered with a folded domain that self-assemble
836 into dimer flanked by disordered residues, opening the question of whether host signaling activation
837 involves intimin-induced dimerization or structural changes of preformed dimers. Green and orange circles
838 denoted tyrosine and serine phosphorylation sites, respectively.

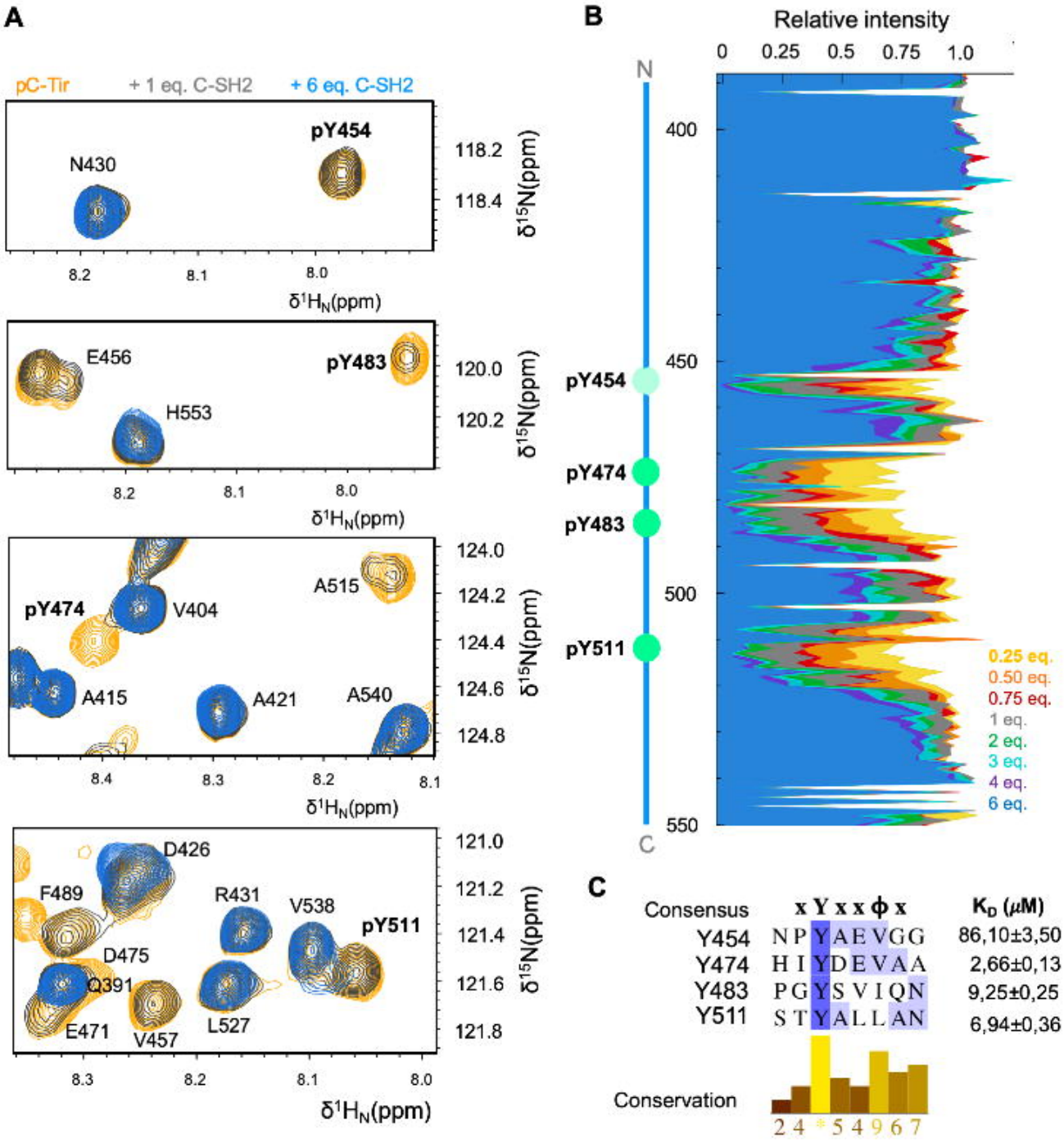

A
Effector
Reference
p=0.00008
p=0.00004
p=0.00004
Disorder fraction
1.0
0.8
0.6
0.4
0.2
0.0
EHEC
CR
EPEC
B
A/E pathogens proteomes
CR effectors
EHEC effectors
EPEC effectors
Human proteome
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
IDP
PDR
FRAG
NDP
ORD
C
CR
EHEC
EPEC
Predicted motif density
0.5
0.4
0.3
0.2
0.1
0.0
ORD
NDP
PDR
IDP
EspF
Tir
NleH
EspB
LiFA
EspF(U)
Map
Disorder fraction
0.0 0.2 0.4 0.6 0.8 1.0

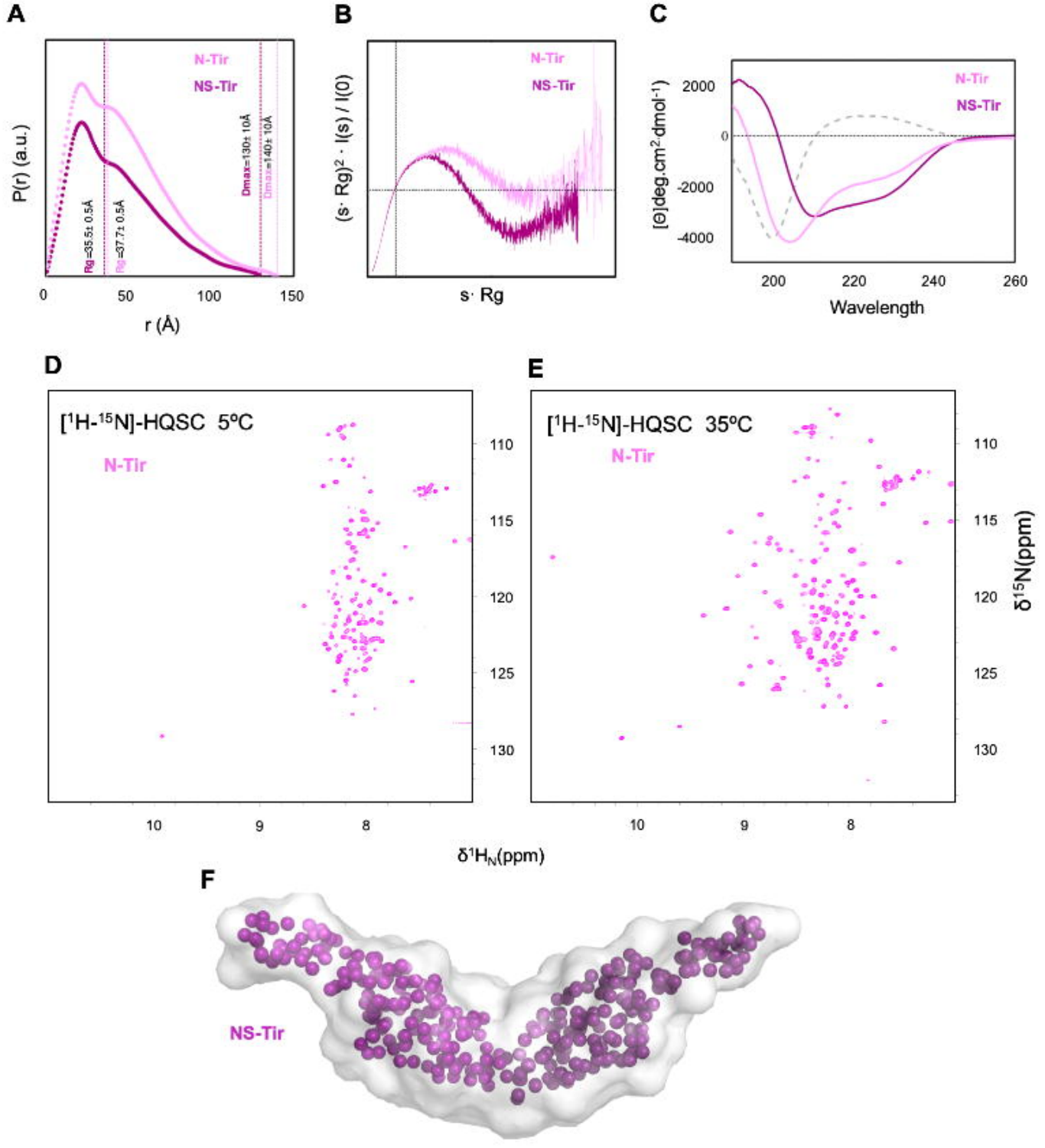

A
(s.Rg)².I(s)/I(0)
s.Rg
B
[θ] (deg. cm². dmol⁻¹. residue⁻¹)
Wavelength (nm)
C
EPEC
EHEC
δ¹⁵N^H (ppm)
δ¹H^N (ppm)

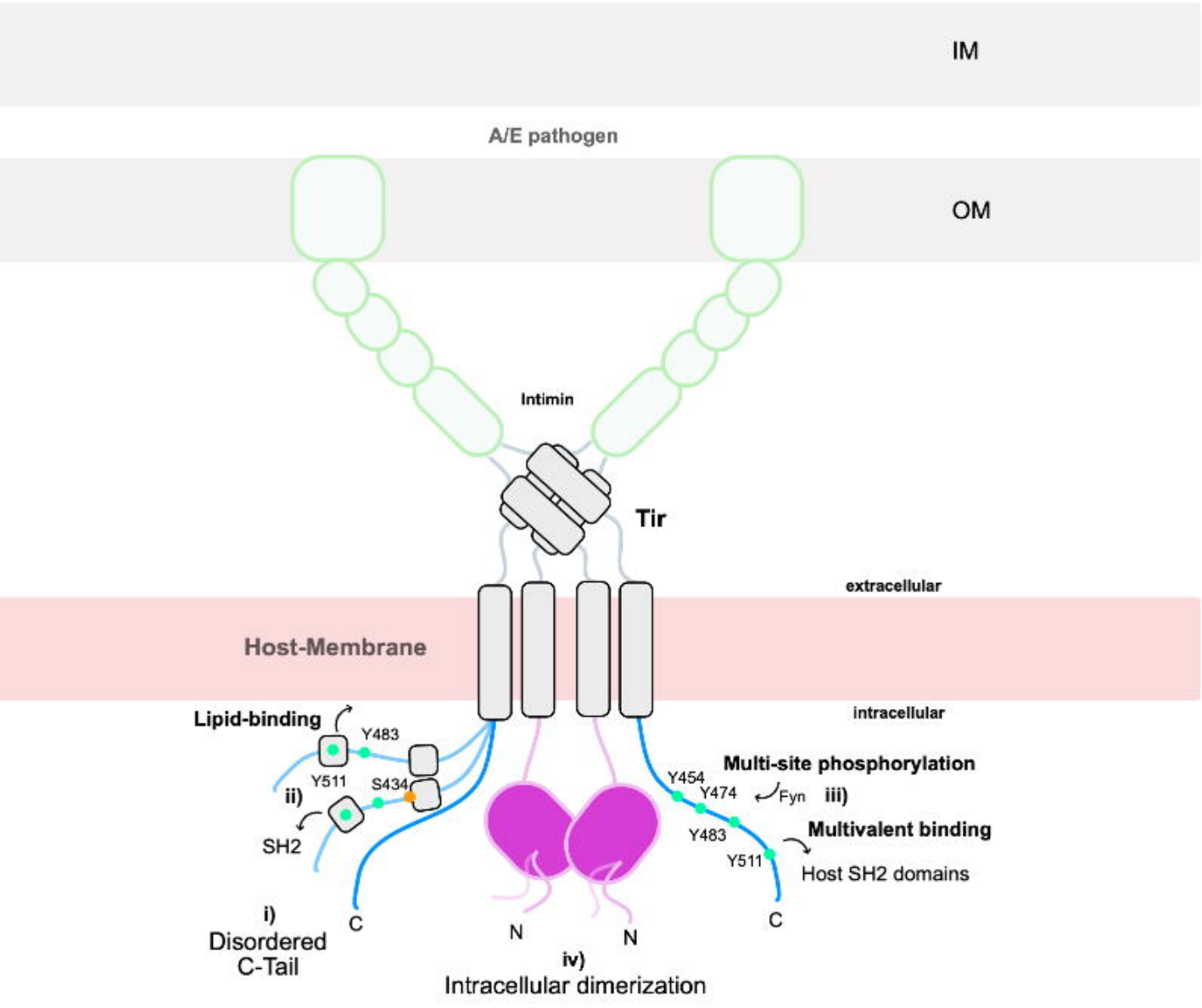

A
S434
Y454
S463
Y474
Y483
Y511
Δδ Cα-Δδ Cβ
ncSPC
a
b
I
II
Residue number
B
C-Tir
C-Tir + 2 eq. C-SH2
C-Tir + 8 eq. C-SH2
i
ii
R498
L513
V538
L520
Y511
V485
L514
A515
A512
I486
A540
δ15N(ppm)
δ1HN(ppm)
C
1 eq.
2 eq.
4 eq.
6 eq.
8 eq.
Y511
Relative intensity
Residue number
D
Kd= 68,5 μM ± 4,32 μM
Relative Intensity
A512
L513
L514
A515
[C-SH2]/[C-Tir]
E
1% TFE
10 %TFE
12.5% TFE
Y483
Y511
Relative intensity
Residue number


A
DMPC/DHPC 6%
DMPG/DHPC 6%
DMPC
DMPG
DHPC
G525
G523
G518
G501
T528
S517
S509
M524
N488
L513
I507
F489
Y511
E499
L514
L522
V550
A515
L499
A512
Q508
Q487
I486
G526
G519
G504
G545
G493
G491
G530
G497
S492
T502
T496
S490
S529
S484
Q505
Y483
N390
R498
V495
V547
L520
Y485
R388
R389
A393
δ1HN (ppm)
δ15NH (ppm)


B
Y483
Y511
Relative Intensity
Residue number


A
pC-Tir
+ 1 eq. C-SH2
+ 6 eq. C-SH2
N430
pY454
118.2
118.4
δ15N(ppm)
8.2
8.1
8.0
δ1HN(ppm)
E456
pY483
H553
120.0
120.2
pY474
V404
A515
A415
A421
A540
124.0
124.2
124.4
124.6
124.8
8.4
8.3
D426
F489
R431
V538
pY511
D475
Q391
E471
V457
L527
121.0
121.2
121.4
121.6
121.8
B
Relative intensity
0
0.25
0.5
0.75
1.0
N
C
400
450
500
550
0.25 eq.
0.50 eq.
0.75 eq.
1 eq.
2 eq.
3 eq.
4 eq.
6 eq.
C
Consensus
x Y x x ϕ x
KD (μM)
Y454 N P Y A E V G G 86,10±3,50
Y474 H I Y D E V A A 2,66±0,13
Y483 P G Y S V I Q N 9,25±0,25
Y511 S T Y A L L A N 6,94±0,36
Conservation
2 4 * 5 4 9 6 7


A
N-Tir
NS-Tir
P(r) (a.u.)
Rg=35.5± 0.5Å
Rg=37.7± 0.5Å
Dmax=130± 10Å
Dmax=140± 10Å
0
50
100
150
r (Å)
B
N-Tir
NS-Tir
(s· Rg)² · I(s) / I(0)
s· Rg
C
N-Tir
NS-Tir
[Θ]deg.cm²dmol⁻¹
2000
0
-2000
-4000
200
220
240
260
Wavelength
D
[¹H-¹⁵N]-HQSC 5ºC
N-Tir
110
115
120
125
130
10
9
8
E
[¹H-¹⁵N]-HQSC 35ºC
N-Tir
110
115
120
125
130
10
9
8
δ¹⁵N(ppm)
δ¹H_N(ppm)
F
NS-Tir


IM
A/E pathogen
OM
Intimin
Tir
extracellular
Host-Membrane
intracellular
Lipid-binding
Y483
Y511
S434
ii)
SH2
i)
Disordered
C-Tail
C
N
N
iv)
Intracellular dimerization
Multi-site phosphorylation
Y454
Y474
Fyn
iii)
Y483
Y511
Multivalent binding
Host SH2 domains
C