# Nematode Extracellular Protein Interactome Expands Connections between Signaling Pathways

Wioletta I. Nawrocka,[1,2,3,*] Shouqiang Cheng,[1,2,3,*] Bingjie Hao,[4,*] Matthew C. Rosen,[1,2,3] Elena Cortés,[1,2,3] Elana E. Baltrusaitis,[1,2,3] Zainab Aziz,[1,2,3] István A. Kovács,[4,5,6] Engin Özkan[1,2,3,†]

[1] Department of Biochemistry and Molecular Biology, [2] Institute for Neuroscience, and [3] Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637, USA.

[4] Department of Physics and Astronomy, [5] Northwestern Institute on Complex Systems, and [6] Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208, USA.

* equal contribution

[†] contact: eozkan@uchicago.edu

## ABSTRACT

Multicellularity was accompanied by the emergence of new classes of cell surface and secreted proteins. The nematode *C. elegans* is a favorable model to study cell surface interactomes, given its well-defined and stereotyped cell types and intercellular contacts. Here we report our *C. elegans* extracellular interactome dataset, the largest yet for an invertebrate. Most of these interactions were unknown, despite recent datasets for flies and humans, as our collection contains a larger selection of protein families. We uncover new interactions for all four major axon guidance pathways, including ectodomain interactions between three of the pathways. We demonstrate that a protein family known to maintain axon locations are secreted receptors for insulins. We reveal novel interactions of cystine-knot proteins with putative signaling receptors, which may extend the study of neurotrophins and growth-factor-mediated functions to nematodes. Finally, our dataset provides insights into human disease mechanisms and how extracellular interactions may help establish connectomes.

**Keywords:** Protein-protein interactions; Extracellular space; Cell surface receptor; Secreted protein; High-throughput interaction assay; *Caenorhabditis elegans*

## INTRODUCTION

Cell surface receptors and secreted ligands build structures connecting cells, mediate communications between cells, sense and respond to extracellular cues, and function as molecular tags to identify cellular populations. These proteins include cell adhesion molecules, signaling receptors, cytokines, growth factors, and secreted cues for cellular migration among others, which are essential for most multicellular life.[1] However, much of the extracellular proteome remains without known binding partners, which is a roadblock for revealing the various functional roles of cell surface and secreted proteins in multicellular systems. Since membrane proteins and cell surface receptors constitute the majority of targets for current therapies and are more readily accessible to new therapies,[2] extracellular interactions are important for understanding the mode-of-action of existing therapies, as well as developing new ones.

Recent expansions of genomic sequence data and bioinformatics tools to recognize cell surface and secreted proteins have opened the study of extracellular interactions to specialized high-throughput methods.[3–5] One of the most effective approaches has been secreted bait/prey capture strategies that use recombinantly expressed ectodomain libraries where bait and prey have been oligomerized, resulting in significant increase in effective affinity for the bait-prey pair. Such avidity-enhanced assays have been instrumental in the discovery of interactions controlling cell-cell recognition for synapse targeting and avoidance,[4,5] neuronal repulsion,[6] and regulatory signaling networks in plant growth and immunity,[7] among others.

Despite these improvements, several issues with extracellular interactome studies remain: First, the existing studies have focused on an important but limited set of protein families, including the immunoglobulin (IG) superfamily[3,4,8–10] and proteins with leucine-rich repeat (LRR) domains,[4,7,11] while most of the known fold space in the extracellular milieu has been ignored. Second, there is a need for throughput improvements to allow for expansion of datasets towards the full interactome of cell surface and secreted proteomes. Finally, the application of the methodology to important model organisms lags. Animal genomes contain groups of conserved cell surface receptors and ligands with related functions throughout many taxa, as well as specific classes of proteins to mediate functions within few taxa or proteins with different binding partners in different taxa. The challenge of predicting whether a given set of molecular interactions are preserved

across many species is a significant barrier to incorporating knowledge generated across model organisms.

The nematode *Caenorhabditis elegans* has been a pioneering model organism for studying embryogenesis, nervous system development, behavior, aging and many other aspects of biology.[12] The *C. elegans* genome contains a smaller number of paralogs for cell surface protein families shared across bilaterians,[13] which makes it easier to cover a larger group of receptor and ligand families relevant for human physiology. In this study, we increased the throughput of our previous avidity-based interaction screen strategy and applied it to a collection of 379 *C. elegans* ectodomains, for a putative interaction space of 72,010 possible interactions, covering 12 domain families in full and 74 domain families in part, nearly all with representation in the human genome (**Table S1**). Using a statistical method we developed for data analysis, we report 185 interactions at our intermediate confidence level, including 159 (86%) previously unknown or unpredicted by homology to known complexes. Here, we highlight and validate novel interactions between axon guidance cues and ligands from separate pathways, insulin interactions with a class of secreted IG domain proteins, and differences in interactions between mammalian and nematode orthologs. We further reveal complexes of secreted proteins of growth factor folds with their signaling receptors, and compare them to mammalian counterparts, including a novel pair with connections to human disease. Finally, we discuss interactions likely to be important in the synaptic wiring of the nervous system.

## RESULTS

### Ectodomain Library

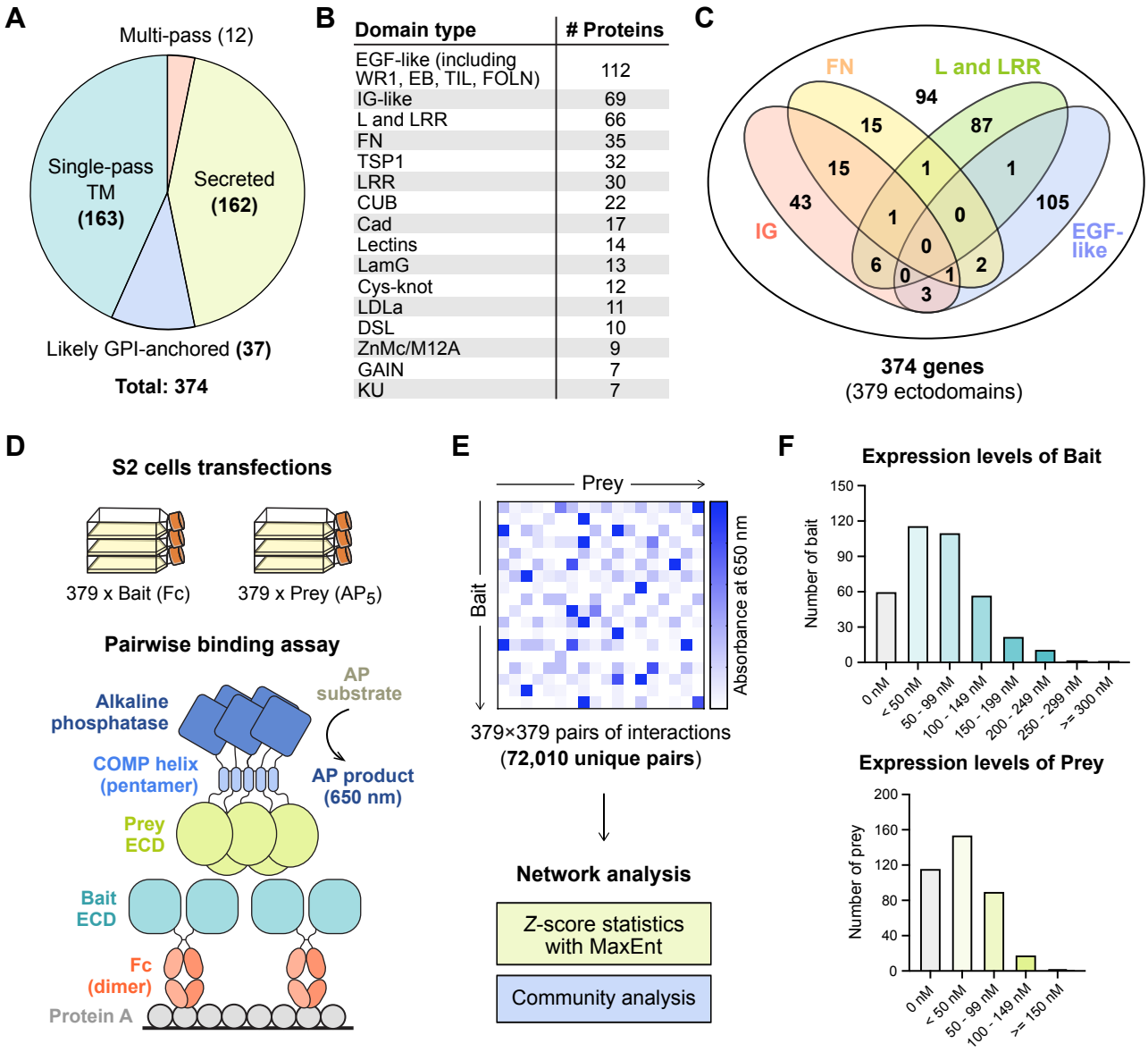
While classical high-throughput interaction discovery technologies have proven to be highly effective for many protein classes, these strategies have remained ineffective for studying membrane proteins, cell surface receptors and secreted proteins,[14] which make up a large fraction of animal proteomes (**Figure S1A**). Published extracellular interactome studies have focused on IG, FN (Fibronectin-type III) and LRR domains, which are some of the largest protein families in the human proteome,[3,4,7–9,11] but have ignored the rest. To address the knowledge gap in the extracellular interactomes, we chose to study the *C. elegans* proteome (**Figure S1A**), which have

86 smaller IG, FN, and LRR families compared to mammalian proteomes (e.g., ~600 IG proteins in
87 humans vs. ~70 in C. elegans),[15,16] allowing us to include other domain families, covering a more
88 diverse selection of protein fold and function space with our high-throughput assay. Here, we
89 subcloned nearly all ectodomains containing IG, FN, LRR, Cadherin (Cad), Epidermal growth
90 factor (EGF), CUB, Laminin-N (LamN), Thrombospondin type-I (TSP1), Integrin, GAIN, Cystine-
91 knot cytokine, and ADAM-type Cys-rich domains, as well as domain families found in known axon
92 guidance cues and receptors (**Figures 1A-1C, Table S1**). Our ectodomain collection also
93 includes some, but not all, ectodomains with Furin, Insulin, Kazal, Kunitz, LDLa, LY, SEA,
94 Sushi/CCP, von Willebrand Factor type A, C and D (VWA, VWC, VWD), and WAP domains
95 among others. Our diverse collection comprises expression constructs of 379 unique ectodomain
96 variants, from 374 cell surface receptor and secreted protein genes with ~86 types of domains or
97 folds, cloned in both bait and prey expression plasmids for interaction screening, ranging from 57
98 to 3,220 amino acids (average: 672 amino acids), not including the tags in bait and prey
99 expression constructs. 43% of the ectodomain library are from secreted proteins, while the rest
100 are predicted to be cell surface receptors anchored by transmembrane helices or
101 glycosylphosphatidylinositol (GPI) anchors (**Figure 1A**).

102 **Assay Development**

103 The Extracellular Interactome Assay (ECIA) uses Fc-tagged secreted ectodomains as bait
104 captured on Protein A-coated plates, and Alkaline phosphatase-tagged ectodomains as prey for
105 the detection of binding to the immobilized bait (**Figures 1D and E**). Essential to the sensitivity of
106 the assay is the pentameric coiled coil included in the prey constructs, which increases effective
107 affinity up to 10,000-fold through avidity,[4] as previously implemented in similar strategies.[3,5]

108 The first generation of the ECIA methodology used the inducible metallothionein promoter for
109 protein expression in the *Drosophila* S2 cell line. For this study of the *C. elegans* interactome, we
110 continued to use the well-established S2 line, as insects are phylogenetically closer to nematodes
111 when compared to other sources of established protein expression lines. However, we
112 implemented several changes to simplify our protocols and improve throughput. First, we modified
113 our expression plasmids to use the constitutively active Actin 5C promoter, which removed the
114 induction step during expression, while improving expression levels as we observed (**Figures**




| Domain type | # Proteins |
|---|---|
| EGF-like (including WR1, EB, TIL, FOLN) | 112 |
| IG-like | 69 |
| L and LRR | 66 |
| FN | 35 |
| TSP1 | 32 |
| LRR | 30 |
| CUB | 22 |
| Cad | 17 |
| Lectins | 14 |
| LamG | 13 |
| Cys-knot | 12 |
| LDLa | 11 |
| DSL | 10 |
| ZnMc/M12A | 9 |
| GAIN | 7 |
| KU | 7 |

**Figure 1.** The *C. elegans* ectodomain collection and high-throughput interaction assay design.
**A.** The distribution of secreted and membrane-anchored proteins in our *C. elegans* ectodomain collection.
**B, C.** The distribution of protein domains in the ectodomain collection.
**D, E.** The design of the ECIA pipeline and data analysis.
**F.** The expression levels of bait and prey proteins in the S2 cell culture media.

115 **S1B-E**) and as previously reported for other proteins in low-throughput studies.[17,18] We also
116 implemented the use of 384-well plates and robotics to decrease transfection volumes, speed up
117 the interaction assay, and decrease cost. Lastly, the advent of Gibson assembly protocols for
118 subcloning allowed us to simplify our previous cloning strategy of Topo TA cloning followed by
119 Gateway recombination. We cloned 63 ectodomain open-reading frames using existing cDNAs in
120 a *C. elegans* ORF Clone Collection (GE Healthcare), 16 from our laboratories' collections, 9 using
121 RT-PCR from a mixed *C. elegans* mRNA library, and corrected any mutation(s) and/or intron(s)
122 likely introduced during RT-PCR. We had the remaining 288 ORFs synthesized, choosing the
123 longest splice variant whenever possible and practical. As we kept the Gateway recombination
124 sequences intact in our plasmids, our ectodomain cDNA collection can be easily re-purposed for
125 other formats and assays. Despite the improvements, 23% of our constructs did not yield
126 detectable expression and secretion, as judged by western blotting of conditioned media (**Figure**
127 **1F**). The median concentrations in the media were estimated to be 57 nM for bait and 17 nM for
128 prey constructs.

129 **Analysis of ECIA Results**

130 ECIA reports interactions by measuring absorbance from the product of the alkaline phosphatase
131 reaction. To detect significant interactions, we have previously calculated *z*-scores using trimmed
132 mean and standard deviation values for each bait, and then used these normalized values to
133 calculate *z*-scores across each prey.[4] To improve our detection of significant hits here, we utilized
134 a maximum entropy network ensemble-based technique[19] to construct a statistical model that
135 captures the absorbance background. Significant interactions are identified by comparing the
136 observations with the modeled background (see Methods and **Figure S2** for details) using *z*-
137 scores, where the standard deviation quantifies systematic errors. As true PPIs are likely to be
138 detectable regardless of whether a protein is the bait or prey, we symmetrize the z-scores as $z =$
139 $\frac{z_{A\to B}+z_{B\to A}}{\sqrt{2}}$ (**Table S2**). This approach yields one continuous-valued weight/score for the interaction
140 between each pair of proteins. Using absorbance values collected at 2 hours following application
141 of the AP substrate, an intermediate score threshold of $z_{min} = 8.4$ yields a protein-protein
142 interaction graph with 185 interactions ($z > z_{min}$) (**Figures S2D, S2E** and **Table S3**). These
143 interactions align strongly with those identified using our previously employed scoring approach

(**Table S3**), but with a more rigorous standard in selecting the threshold. Lower values of $z_{min}$ correspond with more liberal criteria for determining which interactions to include, and thus networks with more edges.

**Community Analysis of ECIA Interaction Networks**

Although protein-protein interaction networks are built from pairwise measurements, they often also reflect higher-order, functionally-relevant relationships between groups of proteins. One particularly successful approach for identifying higher-order relationships is community detection,[20] which seeks to identify 'insular' subgroups of proteins – where connections between pairs of proteins within the group are stronger than those made with proteins outside the group. The interactions within these groups can reflect the sharing of latent relationships even between pairs of proteins that do not themselves directly interact. We applied a modularity optimization method to the ECIA interaction network to investigate its community structure.[21] As expected, communities tend to group proteins with known functions (**Figure 2**). For example, proteins known to mediate axon guidance, those belonging to Robo, Slit, Ephrin and Eph, and one of the Semaphorins and Plexins, belong to the same community (community 9 in Fig. 2), while the remaining Semaphorins and Plexin are in the same community with an RPTP (PTP-4), VER-1/VEGFR, and Calsyntenin (CASY-1), proteins with known functions in regulating axon guidance (community 11 in Fig. 2).[22,23] To capture functional groupings among a larger subset of the network (beyond the intermediate thresholds described above), we conducted the same analysis across 20-fold variation in $z_{min}$, from conservative ($z_{min} = 40$, $N = 60$ interactions) to liberal ($z_{min} = 2$, $N = 654$ interactions). Using the communities calculated over this range of thresholds, we found nodes that repeatedly belonged to the same community. For every protein *P* for which we observed an interaction, we report (a) the full set of proteins partitioned into the same community as *P* for any tested $z_{min}$, and (b) *P*'s "canonical neighbors", the subset of proteins that belong to the same community as *P* across different choices for $z_{min}$ (**Table S4**).

**Higher-order axon guidance complexes**

One of the highly conserved processes in nervous system development is the control of the direction of axonal growth by interactions of guidance cues with their neuronal receptors. Our

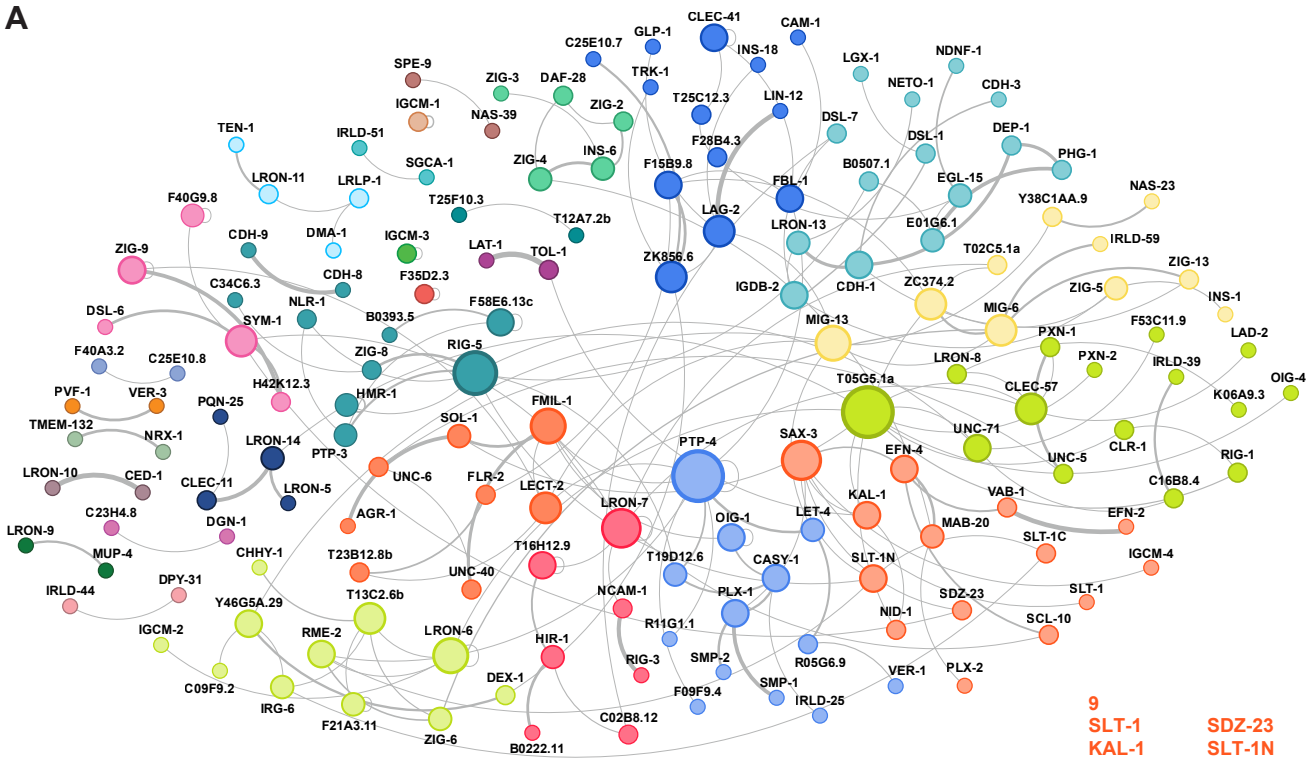



**Figure 2.** Interactions and the community structure of the extracellular interaction.
**A.** Network of moderate-confidence ($z_{min}$ > 8.4) protein-protein interactions ($N$ = 185) identified via the maximum-entropy method; proteins are colored by community grouping.
**B.** Network of community-community interactions. Only *connected communities* (communities with at least one protein making at least one interaction outside the community; $N$ = 14) are shown.
**C.** List of proteins in each connected community.

172 study included the known cue-receptor pairs from the four classical guidance systems: Slit and
173 Robo, Ephrins and Ephs, Semaphorins and Plexins, and Netrin (UNC-6) and its receptors, UNC-
174 40/DCC and UNC-5.[24] To our surprise, we observed novel interactions with high confidence that
175 connected the three axes in the extracellular space (**Figure 3A and Table S3**). SAX-3/Robo
176 interacts with its classical ligand SLT-1 and its N-terminal processed fragment (SLT-1N), but also
177 with the Ephrin EFN-4 and Plexin PLX-1. This connects the Robo signaling axis to receptors of
178 both the Ephrin and Plexin axes. In addition, EFN-4 interacts with the Semaphorin MAB-20,
179 physically connecting Ephrin receptor with a Plexin ligand in the extracellular space. We
180 successfully replicated these results and others involving guidance receptors and cues using
181 ECIA (**Figures 3B and S3**), resulting in a connected network of interactions where only the Netrin
182 axis remained unconnected in the extracellular space (**Figure 3A**). It should be noted that the
183 DCC class of Netrin receptors were shown previously to interact with the Robo receptor
184 intracellularly.[25] We also identified two novel ligands for Netrin receptors: PXN-1, a peroxidasin
185 associated with neuronal phenotypes,[26] interacting with UNC-5; and FMIL-1, an adhesion GPCR,
186 interacting with UNC-40/DCC.

187 Given the central importance of axonal guidance pathways to neuronal wiring and their
188 involvement in neurodevelopmental disorders, we chose to further validate these interactions with
189 orthogonal methods. First, we performed ECIA experiments with protein domains to learn about
190 the architectures of these new complexes. We showed that the IG domains of SAX-3/Robo bind
191 PLX-1 IPT domains 3 to 6, and EFN-4 domain 2 (**Figures 3C and 3D**). To confirm these data, we
192 produced SAX-3 IG1-4, EFN-4 D2 and PLX-1 IPT3-6 domain constructs using the baculoviral
193 expression system in lepidopteran cells, and purified them to homogeneity. Purified SAX-3 IG1-4
194 and EFN-4 D2 can form a stable complex observed via size-exclusion chromatography (SEC)
195 (**Figure 3E**). We also validated the interaction of the Semaphorin MAB-20 with the Ephrin EFN-
196 4, using purified MAB-20 ectodomain with the first domain of EFN-4 with both ECIA and SEC
197 (**Figures 3F and 3G**). Our results implicate the first domain of Ephrin/EFN-4 in interactions with
198 its classical receptors (EPHs) and Semaphorins, and the second domain in its interactions with
199 Robos. Interestingly, we did not observe Semaphorin or Robo interactions with the other *C.*
200 *elegans* Ephrins, which suggests specialization of EFN-4 to integrate multiple cell surface signals.

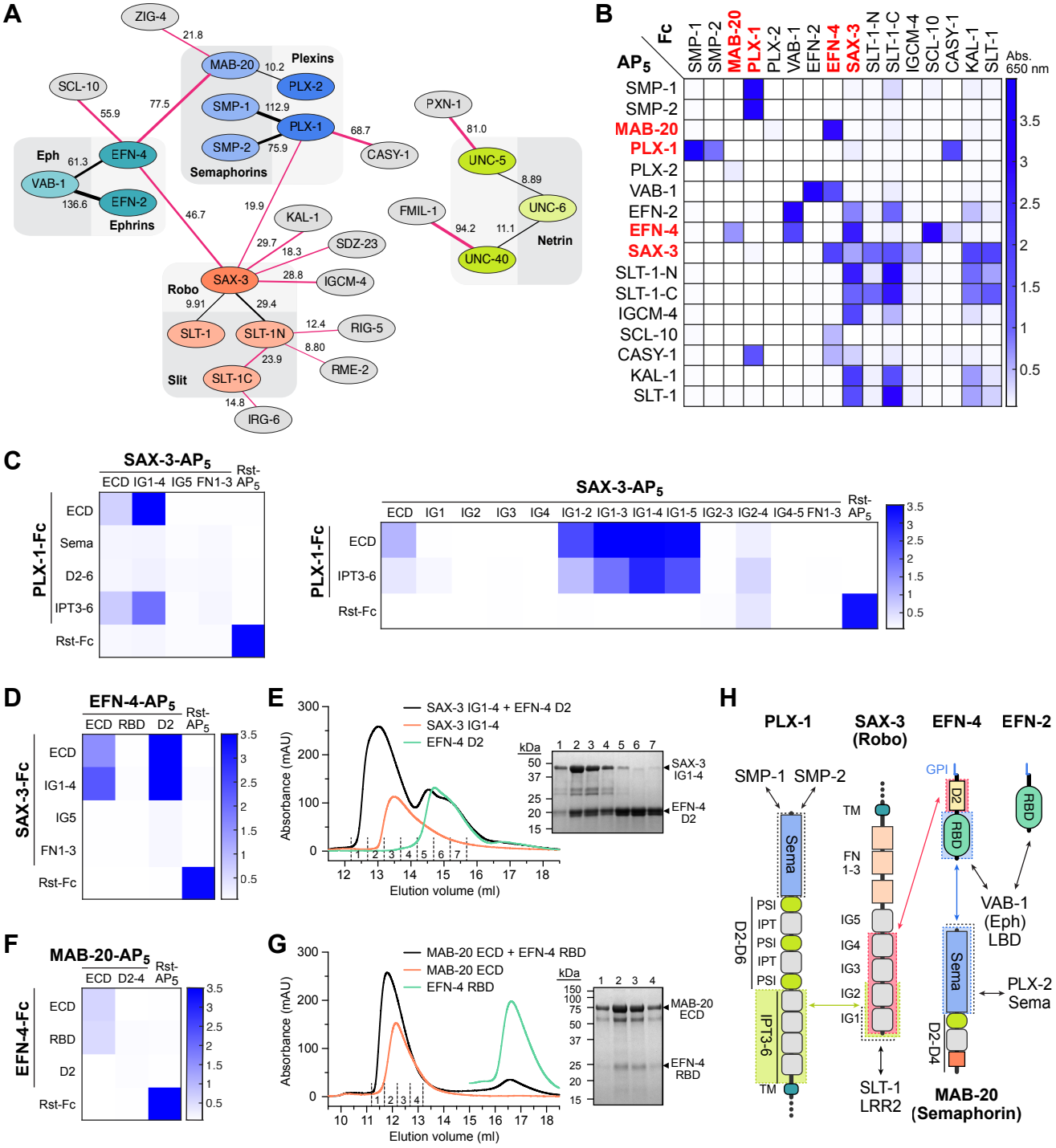



**Figure 3.** Axon guidance receptors and cues interact with each other outside the known cue-receptor axes.
**A.** Schematic of interactions from the high-throughput assay, where line thickness is scaled to symmetrized MaxEnt *z*-scores (Figure 2). Interactions that were not previously known are shown with red lines. The numbers next to the lines indicate *z*-scores.
**B.** Interactions observed in the high-throughput assay are reproduced (see Figure S3 for more). As expected, interactions observed with a bait-prey (Fc-AP) pair are also observed in the reciprocal orientation, resulting in diagonally symmetrical assay results.
**C-G.** ECIA can be used to identify domains required for novel interactions: The first two IG domains of SAX-3 interact with the IPT domains 3 to 6 (C). SAX-3 IG domains also interact with the second domain of EFN-4, as observed by ECIA (D) and size-exclusion chromatography (SEC) (E). The interaction of MAB-20 with EFN-4 is mediated by the first domain (RBD) of EFN-4, as observed by ECIA (F) and SEC (G). SDS-polyacrylamide gels show the presence of both ectodomains in the complex fractions of SEC runs (E, G).
**H.** Summary of interactions between the domains of the axon guidance receptors and cues. Black arrows refer to interactions we observed but also previously characterized in other taxa.

201 *C. elegans* EFN-4 is also unique among Ephrins as we could not identify an EFN-4 D2-like
202 sequence in other Ephrins outside nematodes.

203 **A family of insulin-receptor complexes revealed in nematodes**

204 Insulin and related peptides are conserved peptide hormones that regulate metabolism, cell
205 proliferation, aging and longevity across animals through a conserved set of downstream
206 signaling molecules, starting at the cell membrane through insulin receptors (IR) and the related
207 Insulin-like growth factor receptors (IGFR).[27–29] In nematodes, the insulin family has expanded to
208 a set of 40 insulin-like peptides (ILPs: INS-1 to -39 and DAF-28),[30,31] compared to only eight in *D.*
209 *melanogaster* and ten in humans. Most nematode ILPs are expressed in neurons,[32] and they are
210 proposed to act through the only insulin receptor, DAF-2, in *C. elegans*.[27] Depending on the
211 context, ILPs have been observed to be agonists or antagonists of DAF-2,[33] which presents a
212 mechanistic conundrum in the absence of co-ligands or co-receptors that may modulate DAF-2
213 signaling in response to insulin-like peptides.

214 To advance our understanding of insulin signaling in *C. elegans*, we included four insulins (INS-
215 1, INS-6, INS-18 and DAF-28) in our interaction screen. We observed that two insulins, DAF-28
216 and INS-6, interact with ZIG-2 and ZIG-4, respectively (**Figure 2 and Table S3**). We also
217 observed that ZIG-3 interacts with INS-6 and ZIG-5 interacts with INS-1. Multiple lines of evidence
218 led us to classify ZIG-2 to -5 as a novel class of insulin-binding proteins distinct from other ZIGs:
219 (1) Our sequence alignments of ZIG molecules, previously classified as neuronal surface or
220 secreted proteins containing two-immunoglobulin domains, showed that ZIG-2 to -5 are closely
221 related to each other and not to other ZIGs (**Figure 4A**); (2) ZIG-2 to -5 share sequence features
222 uncommon to other IgSF proteins, especially at the linker connecting the two IG domains,
223 including a disulfide bond (**Figure S4A**); and (3) ZIG-2 to -5 are all secreted, where most other
224 ZIGs appear to have transmembrane helices or GPI membrane anchors (**Table S1**). Finally, ZIG-
225 2 to -5 are co-expressed in the PVT neuron, along with ZIG-1 and -8, and have been implicated
226 in the maintenance of axon position in the ventral nerve chord, suggesting a functional connection
227 among these proteins.[34,35]

228 Next, we wanted to answer if other nematode insulins interact with ZIGs. To answer this question,

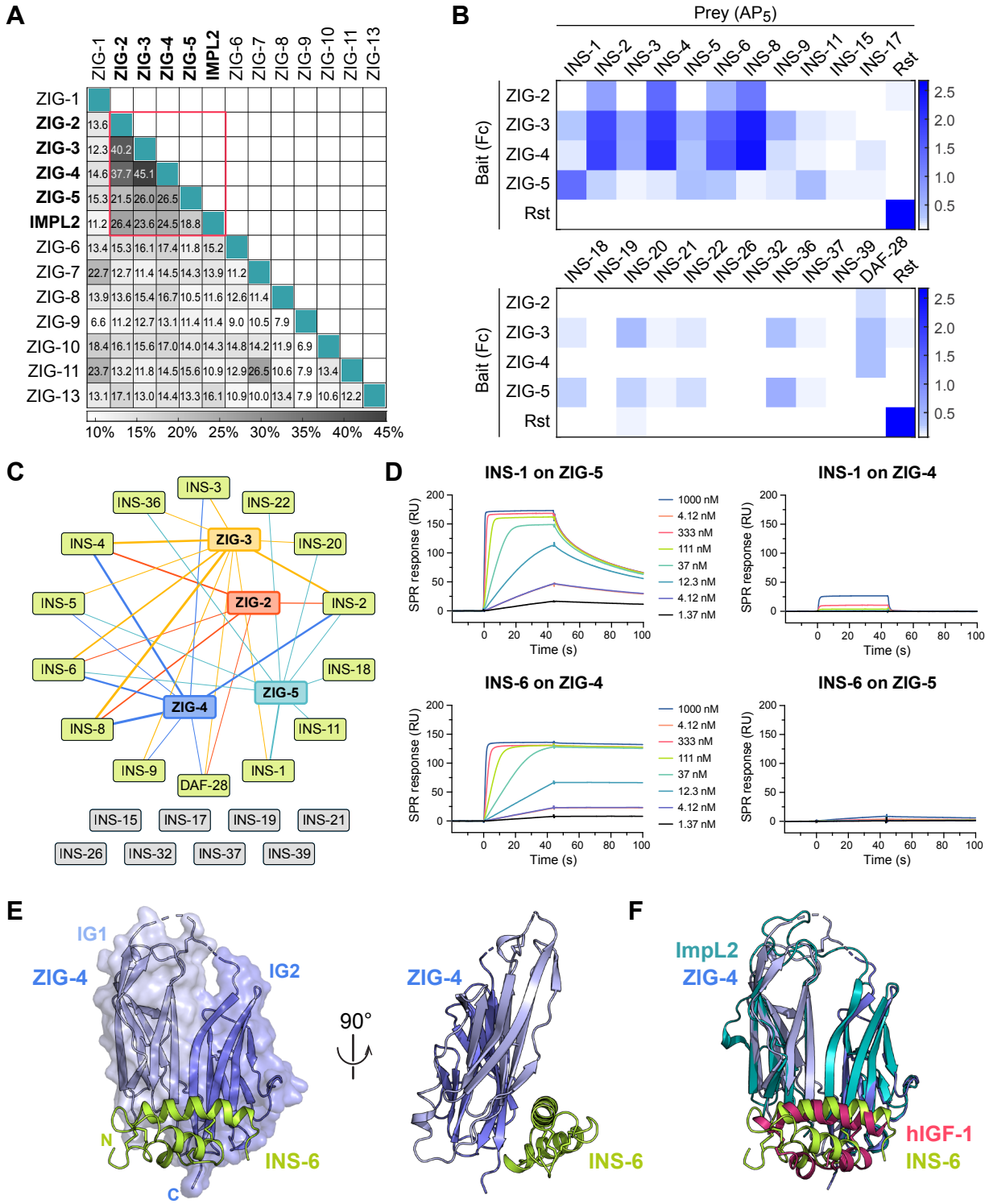



**Figure 4.** ZIG-2, -3, -4 and -5 make up a family of insulin-binding IgSF proteins.
**A.** Pairwise sequence identities show that ZIG-2 to -5 are more closely related to each other and to *Drosophila* ImpL2 than the rest of the ZIG family.
**B.** ECIA for 22 insulin family members against ZIG-2 to -5. For expression levels of ZIGs and insulins, see Figures S4C, S4D.
**C.** ZIG-insulin interaction network inferred from the ECIA shown in B. The thickness of the connecting lines reflects the absorbance values in B for each ZIG-insulin interaction.
**D.** SPR sensorgrams for INS-1 and INS-6 binding to ZIG-5 and ZIG-4 immobilized on SPR chips. Kinetic fits with estimated on- and off-rates and equilibrium constants are shown in Figure S4E.
**E.** The structure of ZIG-4 (purple) bound to INS-6 (green) as observed in our tetragonal crystals.
**F.** The structure of the ZIG-4–INS-6 complex strongly resembles that of Drosophila ImpL2 bound to human IGF-1 (PDB ID: 6FF3).

229 we created ECIA expression constructs for 18 other diverse insulins via RT-PCR, covering all
230 three classes of *C. elegans* insulins.[32] When we repeated ECIA with these four ZIGs against the
231 expanded set of 22 insulins, we observed binding between the four ZIGs and several other
232 Insulins (**Figures 4B, 4C and S4B**). Since the ECIA signal depends on expression levels of bait
233 and prey, which are different among the ZIGs and among the insulins (**Figures S4C and S4D**),
234 we cannot compare affinities between various ZIG–Insulin interactions. However, we observed a
235 trend where the β-class of insulins (INS-1 to -10 and DAF-28) gave higher signals of binding
236 compared to other classes. Different binding specificities and affinities of insulins against the ZIG
237 proteins may provide one means of differentiating their activities, even though all insulins likely
238 act through the same receptor (DAF-2).

239 To validate these results, we performed SPR experiments for two pairs of insulins and ZIGs. ECIA
240 results suggested that INS-6 binds most strongly to ZIG-4, while INS-1 binds most strongly to
241 ZIG-5. SPR results confirmed these findings (**Figures 4D and S4E**), yielding dissociation
242 constants ($K_D$) for INS-1 binding to ZIG-5 and INS-6 binding to ZIG-4 were 3 nM and 56 pM, while
243 the cross pairs had $K_D$ > 1 μM. The very strong affinities observed suggest that insulins are
244 unlikely to be free of ZIGs in contexts where ZIG-2 to -5 are expressed and secreted.

245 For more insights into the function of ZIG-insulin complexes, we crystallized the ZIG-4–INS-6
246 complex and determined its crystal structure in three different crystal forms (**Figure 4E**). All three
247 structures reveal the same complex (**Figure S4F**) and show that the two immunoglobulin domains
248 of ZIG-4 create a continuous sheet made from the *ABE* strands of IG1 and the *ACFG* strands of
249 IG2, to which INS-6 binds (**Figure 4E**). The ZIG-4-INS-6 structure is related to recently determined
250 structures of a *Drosophila* IgSF protein ImpL2 bound to *Drosophila* DILP5 and to human IGF-1,[36]
251 the atypical features of the two IG domains of ImpL2 are preserved in ZIG-4, and the overall root-
252 mean-square displacement between the two structures is 1.4 Å over 158 (out of 202) Cα atoms
253 (**Figure 4F**). Indeed, ZIG-2 to -5 are more closely related to ImpL2 than to other *C. elegans* ZIGs
254 (**Figure 4A**). These similarities imply that the ancestral ecdysozoan had a ZIG-2–5/ImpL2-like
255 protein able to interact with insulin(s). We could not identify any ZIG-4 or ImpL2 orthologs in
256 vertebrates; however, we found ZIG-2 to -5 and ImpL2-like sequences across protostome
257 genomes. *Drosophila* ImpL2 has been proposed as a molecule aiding the bioavailability of insulins,

258 similar to vertebrate IGFBPs, despite sharing no ancestry, structural or sequence similarities with
259 IGFBPs.[36] Therefore, it is possible that ZIG-2 to -5 may be functionally related to IGFBPs in
260 vertebrates, while not being related in sequence or structure.

261 As *Drosophila* and nematode insulins share the unexpected ability to act as both agonists and
262 antagonists of insulin/IGF receptor,[31] we analyzed how ZIG-binding to insulins could control
263 insulin activation of the insulin/IGF receptor. We first overlaid the ZIG-4-INS-6 structure on the
264 crystal structure of human Insulin bound to a minimal insulin receptor (IR) fragment, including the
265 L1 domain, the Cys-rich domain, the L2 domain and the C-terminal α-helix (αCT).[37] We saw that
266 a ternary complex of INS-6, ZIG-4 and the insulin receptor was possible, where ZIG-4 was
267 positioned to further interact with the IR, and not clash with it (**Figure S4G**). However, when the
268 structure of the INS-6–ZIG-4 complex was aligned with any of the structural models of the dimeric
269 IR bound with up to four insulins, presumably in partial or fully active states of the complex (e.g.,
270 **Figure S4H**),[38] we observed that ZIG-4 severely clashes with one of the IR protomers forming the
271 dimeric form of the IR complex, while aligning INS-6 with the site 1 insulin (**Figure S4I**). Similarly,
272 the second insulin-binding site on IR overlaps with the ZIG-4-binding site, which would prevent
273 the T-shaped $IR_2$-$insulin_4$ from forming (**Figure S4J**). Therefore, when ZIGs are present, insulin
274 receptor may be able to bind insulins, but will be sequestered in an inactive form, unable to make
275 interactions with both protomers to form an active-state IR dimer. This presents an attractive and
276 testable mechanism for how nematode and arthropod insulins may act as antagonists of DAF-
277 2/insulin receptor.

278 **Cystine-knot proteins, putative neurotrophins, growth factors and their receptors in *C.***
279 ***elegans***

280 As we curated *C. elegans* proteins that are on the cell surface or are secreted, we came across
281 several proteins sharing common growth factor and cytokine-like folds. These include cystine-
282 knot proteins, as well as FGF-like growth factors, and other molecules with growth factor-like
283 sequences. We included several of these molecules in our ectodomain collection (annotated in
284 **Table S1**) and identified binding partners for them.

285 First, we observed that two related cystine-knot family proteins, ZK856.6 and B0416.2, are the

286 strongest candidates for being ligands for TRK-1 (at $z = 22.7$ and 6.3, respectively), the
287 designated *C. elegans* ortholog of the vertebrate high-affinity neurotrophin receptors, the Trk
288 family of receptor tyrosine kinases (RTK) (**Figure 5A**). These two secreted proteins are, therefore,
289 putative neurotrophins, a class of growth factors necessary for neuronal growth, survival and
290 regeneration.[39] Vertebrate neurotrophins (NGF, BNDF, NT3 and NT4) are similarly Cys-knot
291 proteins.[40] Searches for homologs of the two putative neurotrophins using BLAST returned no
292 human, mouse or fly proteins, and only each other when searched for *C. elegans* paralogs. This
293 demonstrates that interaction screening is more effective in identifying Cys-knot ligand-receptor
294 pairs than using sequence similarity to known pairs. To confirm our findings, we expressed and
295 purified the TRK-1 ectodomain and one of the putative neurotrophins ZK856.6, and performed
296 SPR experiments, which validated the interaction (**Figure 5B**). Last, structural modeling using
297 AlphaFold-multimer showed ZK856.6 and B0416.2 dimers interacting with the FN domains of
298 TRK-1 in a manner reminiscent of vertebrate Neurotrophin-Trk Receptor complexes (**Figure S5**),
299 supporting our claim that TRK-1 and its interactors may signal and function similarly to
300 neurotrophin-receptor systems in vertebrates. Interestingly, one of the putative neurotrophins,
301 ZK856.6, is a hub also interacting with C25E10.7, F15B9.8, LAG-2 (a Jagged/Delta homolog) and
302 DEX-1.

303 Among the family of Cys-knot family of secreted proteins, we observed that B0222.11, C02B8.12
304 and T16H12.9 interact with one receptor, HIR-1. This receptor was recently identified to direct
305 hypoxia response and hypoxia-associated modeling of the extracellular matrix.[41] HIR-1 is also a
306 tyrosine kinase (RTK) with a cytoplasmic domain that resembles vertebrate RET and Fibroblast
307 growth factor (FGF) receptors (FGFR). Based on this sequence similarity, the FGF homolog LET-
308 756, which is a growth factor but not of the Cys-knot fold, was proposed as a potential HIR-1
309 ligand.[41] Instead, we recommend that Cys-knot family proteins, including the three HIR-1 ligands,
310 should be studied for hypoxia response in *C. elegans*. To validate our findings, we tested HIR-1
311 ectodomain against B0222.22 and LET-756/FGF with SPR using purified proteins (**Figure 5C**).
312 We observed a high-affinity complex of HIR-1 with B0222.22 ($K_D$ = 0.50 nM) (**Figure S5E**), while
313 LET-756/FGF showed no binding to HIR-1.

314 We had included the two FGFs (EGL-17 and LET-756) and the FGFR ortholog (EGL-15) in our

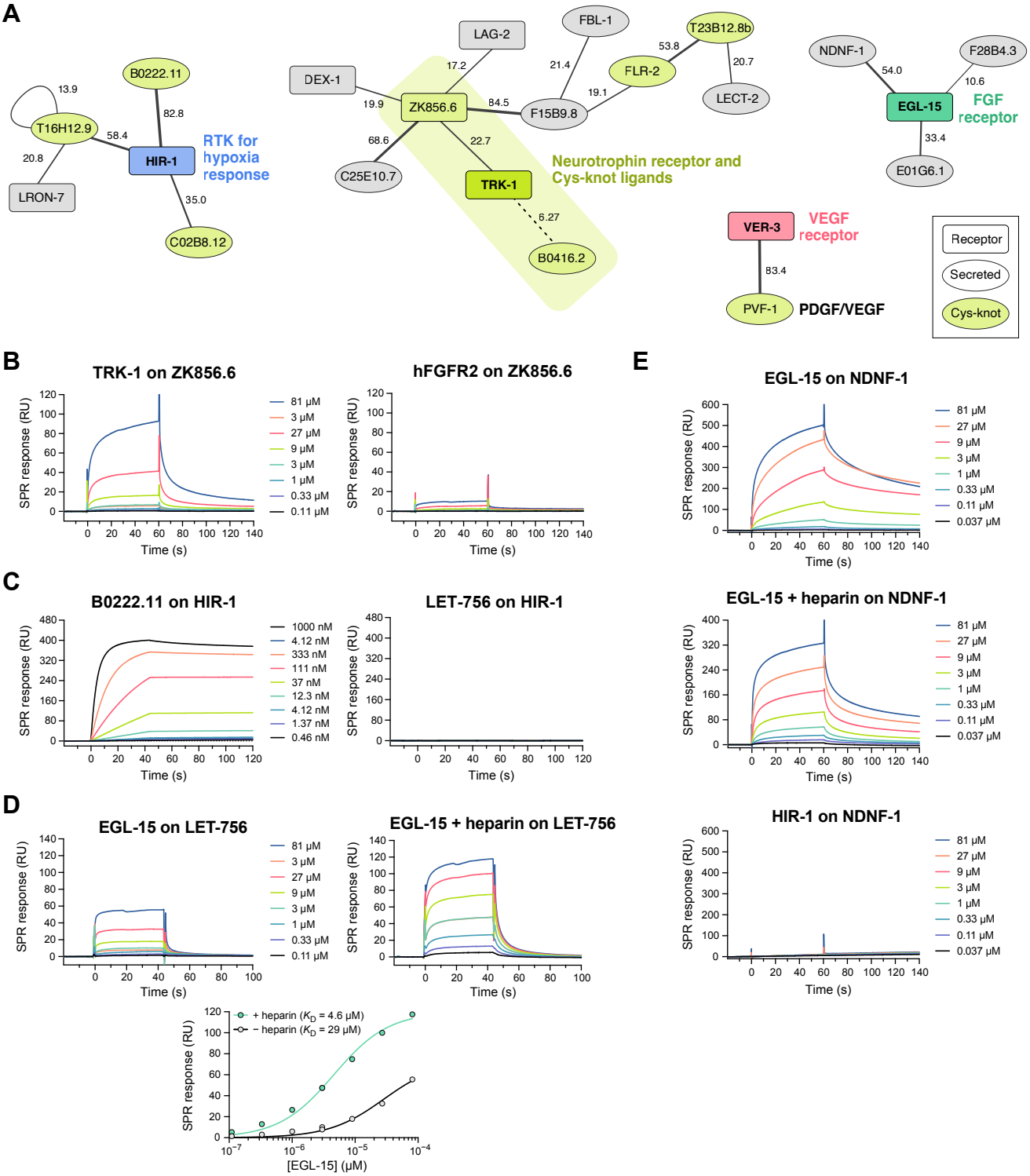



**Figure 5.** Networks of interactions with growth factor-like molecules and receptors.
**A.** Network of interactions between cytokine- and growth factor-like molecules and receptors as observed in our screen (Figure 2). Line thickness corresponds to the symmetrized *z*-scores for each interaction. The numbers next to the lines indicate the *z*-score values.
**B.** SPR sensorgrams for the interaction of TRK-1 with immobilized ZK856.6 and negative control (hFGFR2 against ZK856.6).
**C.** SPR sensorgrams for the interaction of B0222.11 with immobilized HIR-1 and negative control (LET-756 against HIR-1). Kinetic fits and parameters are shown in Figure S5E.
**D.** SPR sensorgrams and binding isotherms for the interaction of EGL-15 with immobilized LET-756 in the presence or absence of 50 μg/mL (3.1 μM) heparin.
**E.** SPR sensorgrams for the interaction of EGL-15 with immobilized Fc-tagged NDNF-1 and negative control (HIR-1 against NDNF-1-Fc).

315 ectodomain collection. Surprisingly, we did not observe an interaction between these putative
316 FGFs and the FGF receptor in our screen, likely as a result of complete lack of expression of the
317 FGF ligands using our bait and prey plasmids in S2 cells (**Table S1**). To confirm this, we
318 expressed and purified EGL-15/FGFR ectodomain and LET-756/FGF using lepidopteran cells
319 and performed SPR experiments, which demonstrated the FGF-FGFR interaction in *C. elegans*
320 (**Figure 5D**). This confirms that the lack of an FGF-FGFR hit in our screen is a false negative and
321 was due to lack of protein expression. We observed stronger binding in the presence of heparin
322 (**Figure 5D**), as previously observed for vertebrate FGFR-FGF interactions.[42]

323 In our screen, we observed NDNF-1, the nematode ortholog of the human neuron-derived
324 neurotrophic factor, interacting with EGL-15/FGFR. NDNF and FGF are not homologous and do
325 not share structural similarities, and an NDNF-FGFR interaction was not previously suspected or
326 reported. However, NDNF overexpression is known to inhibit FGF signaling in cultured vertebrate
327 cells, and FGFR-1 and NDNF are both implicated in congenital hypogonadotropic hypogonadism
328 (CHH) and Kallmann syndrome (KS).[43,44] Our discovery of an NDNF-FGFR complex, may provide
329 the missing mechanistic link that connects NDNF with FGFR signaling, especially in disease
330 states.

331 To validate the interaction of NDNF-1 with EGL-15/FGFR, we set out to express NDNF, but failed
332 to produce stable protein with common expression systems. Therefore, we used Protein A-
333 coupled SPR chips to capture Fc-tagged NDNF-1, originally produced as bait for ECIA, and
334 measured binding to EGL-15 as an SPR analyte. We observed strong binding in the absence and
335 presence of heparin (**Figure 5E**), demonstrating the validity of the NDNF-1–EGL-15/FGFR
336 interaction.

337 Finally, *C. elegans* has a single platelet-derived growth factor (PDGF) homolog, PVF-1. In our
338 screen, we observed that PVF-1 binds VER-3, the predicted PDGF/VEGF receptor ortholog,
339 suggesting that the PDGF/VEGF-mediated biology is conserved from nematodes to mammals.

340 ***C. elegans* Wirins: ZIG-8 and RIG-5 and their interactions**

341 An important result of our previous interactome study was the discovery of Dpr and DIP protein
342 families (32 total members) in the fruit fly.[4] Dprs and DIPs interact with each other, and have been

343 strongly associated with synaptic specificity, axon guidance and fasciculation, cell fate
344 determination and survival, and animal behavior.[45] We recently identified the nematode
345 homologues of Dprs and DIPs, ZIG-8 and RIG-5, respectively, as part of a conserved family
346 named Wirins,[45] and our interactome dataset reports a high-confidence ZIG-8–RIG-5 interaction
347 as expected (**Table S3**). While homophilic and heterophilic interactions between Dprs and DIPs,
348 and among their vertebrate orthologs (IgLONs) have been well established by others and us, it is
349 unclear how Dprs and DIPs signal since they are GPI anchored proteins with no intracellular
350 domains.[46] No extracellular cell surface receptors have been reported as binding partners for Dprs
351 and DIPs to-date.

352 In our dataset, we observed several interactions with ZIG-8/Dpr and RIG-5/DIP (**Figure 6A**).
353 These include ZIG-8 binding to C34C6.3, and RIG-5 interactions with B0507.1, PTP-3, NLR-1,
354 T19D12.6 and HMR-1. Among the binding partners, NLR-1, T19D12.6 and HMR-1 belong to the
355 LamG+EGF family of proteins, most prominently represented by the synaptic Neurexin proteins.
356 NLR-1 is the *C. elegans* ortholog of Contactin-associated proteins (CNTNAP), known as Nrx-IV
357 in flies. Since CNTNAPs are known to take part in the formation of various types of cell junctions
358 (such as *C. elegans* gap junctions, *Drosophila* septate junctions and vertebrate axo-glial
359 junctions),[47–50] and have known neuronal and synaptic developmental functions,[51,52] they are
360 plausible candidates for mediating Dpr/DIP signaling. The other strong candidate interaction for
361 mediating Dpr/DIP signaling is the RIG-5–PTP-3 interaction: PTP-3 is a receptor protein tyrosine
362 phosphatase (RPTP) and the worm ortholog for LAR. LAR has been implicated in *Drosophila* to
363 control synaptic development and signaling in the photoreceptor cells and neuromuscular
364 junctions,[53,54] where Dpr-DIP was previously shown to control synaptic signaling.[55] HMR-1 has
365 been shown to be required for axon fasciculation and dendrite extension in *C. elegans*,[56–58] and
366 its fly ortholog, CadN, functions in synaptic targeting in the developing optic lobe;[59,60] Dprs and
367 DIPs are similarly known to control synaptic targeting of photoreceptor neurons.[55] These strong
368 functional links support the validity of these interactions. To further confirm these interactions, we
369 performed SPR with purified RIG-5 ectodomain against PTP-3, NLR-1 and HMR-1 ectodomains,
370 and validated the interaction hits (**Figures 6B-E**). These interactions proved to be lower affinity,
371 as would be expected for co-receptors residing on the same cell, which may also require the

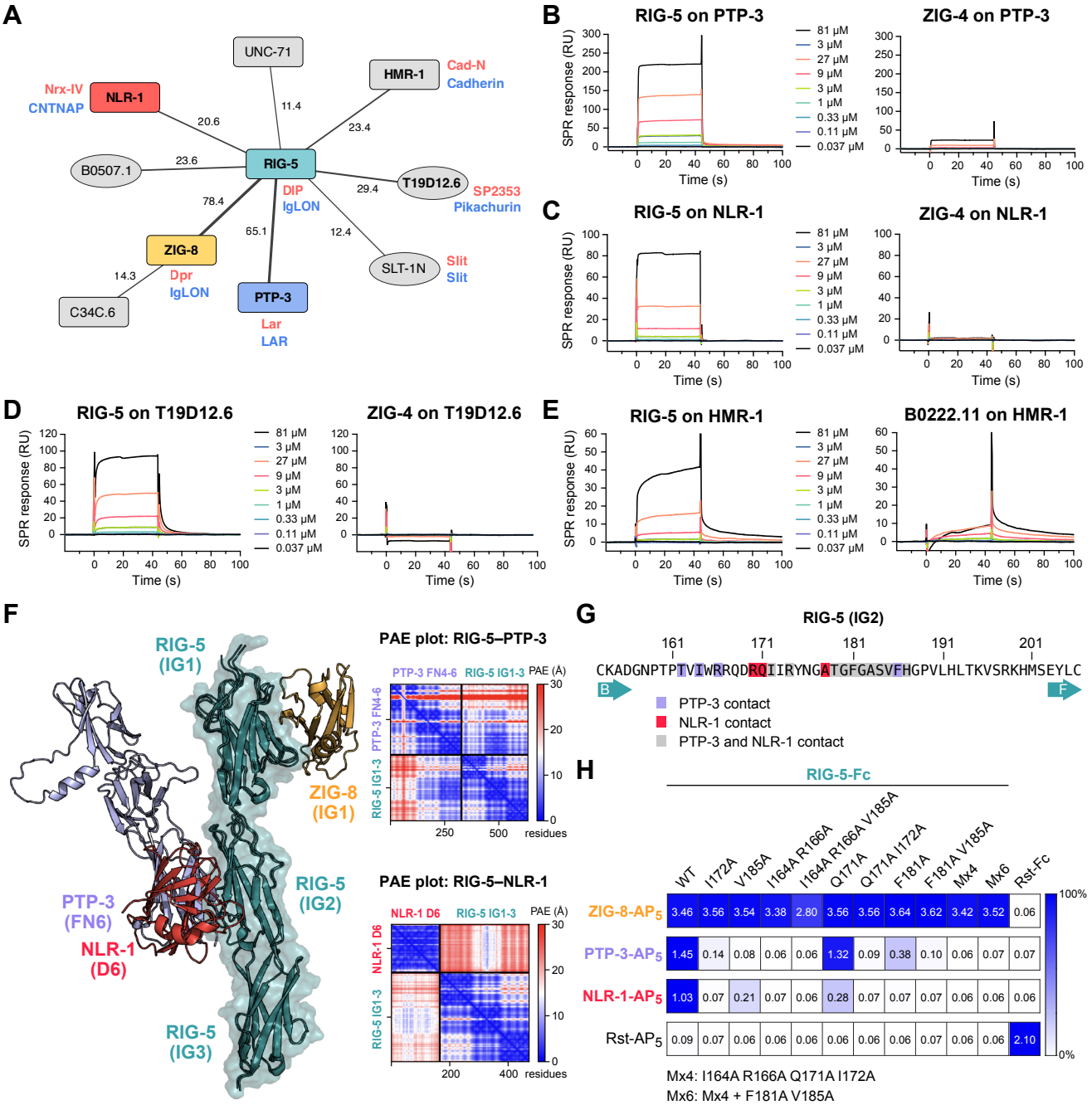



**Figure 6.** RIG-5 interacts with molecules with neuronal and synaptic functions.
**A.** Interactions of ZIG-8/Dpr and RIG-5/DIP according to our screen. Red and blue names represent orthologs in *D. melanogaster* and mammals, respectively. Line thickness scales with the symmetrized *z*-scores for each interaction. The numbers next to the lines indicate the *z*-score values.
**B-E.** SPR sensorgrams for interactions of RIG-5 with immobilized PTP-3 (B), NLR-1 (C), T19D12.6 (D), and HMR-1 (E).
**F.** Overlay of the ZIG-8–RIG-5 structure (PDB ID: 6ON9) (Cheng *et al.*, 2019) and AlphaFold-multimer models for RIG-5 IG1-3+NLR-1 D6 and RIG-5 IG1-3+PTP-3 FN4-6. ZIG-8/Dpr binding to RIG-5/DIP does not overlap with PTP-3 and NLR-1, which share a binding site on the RIG-5 IG2 domain. See Figure S6 for details of the AlphaFold-predicted interfaces. Predicted aligned error (PAE) plots for AlphaFold predictions are shown for both RIG-5 (ECD)-PTP-3 (FN4-6) and RIG-5 (ECD)-NLR-1 (D6) complexes.
**G.** RIG-5 IG2 residues identified to interact with PTP-3 and NLR-1 in AlphaFold models.
**H.** ECIA for RIG-5 ECD mutants against ZIG-8, PTP-3 and NLR-1 ECDs. Mutations in RIG-5 IG2 that break PTP-3 and NLR-1 binding do not affect ZIG-8 binding, which is known to happen through the IG1 domain. The raw readout for the assay (absorbance at λ = 650 nm) is noted in each square. The homodimeric Rst interaction is a positive control.

372 formation of high-density ZIG-8/Dpr–RIG-5/DIP clusters.

373 To gain further insights into some of these interactions, we used the Colabfold implementation of
374 AlphaFold2-Multimer.[61,62] We predicted models for both the RIG-5–PTP-3 and RIG-5–NLR-1
375 complexes with reasonable ipTM (interface predicted template modeling) values (0.72 and 0.49,
376 respectively) and predicted aligned error (PAE) values (**Figure 6F**), where the second IG domain
377 of RIG-5/DIP interacts with these signaling receptors using highly overlapping interfaces (**Figures**
378 **6F, 6G and S6**). We designed point mutations of RIG-5 that are likely to break these interactions,
379 and tested them for binding via ECIA (**Figure 6H**): We observed that the mutations broke the
380 PTP-3 and NLR-1 complexes of RIG-5 as predicted, but none affected ZIG-8/Dpr binding, which
381 depends on an epitope on the first IG domain.[45,55] These results support our RIG-5 interaction
382 discovery, as well as the AlphaFold prediction of the binding interfaces. We suggest that the study
383 of Dprs and DIPs in the fly model should include the orthologs Lar, Nrx-4 and CadN, to reveal
384 Dpr-DIP signaling at the synapse.

385 **Limitations of the ECIA screen and further insights**

386 While we observed many interactions that were previously known, there were some expected
387 complexes that we did not observe. As mentioned above, the FGF-FGFR complexes were missed
388 likely as a result of lack of FGF expression in our expression system. 23% of our constructs did
389 not yield detectable expression, as judged by western blotting (**Table S1**), which is likely the
390 largest source of false negatives. However, there are other unexpected negatives that cannot be
391 explained by lack of expression based on known binding data with vertebrate homologs. One
392 such case is the lack of binding between LAT-1 and TEN-1, the nematode homologs of
393 Latrophilins and Teneurins, which form a synapse-instructive complex in mammals. A previous
394 study suggested that this interaction may not exist in invertebrates on the basis of genetic data,
395 and therefore the nematode and mammalian Latrophilins and Teneurins may act through different
396 ligands.[63] We report the Toll-like receptor TOL-1 and the LRR protein LRON-11 as interaction
397 partners for LAT-1 and TEN-1 in *C. elegans*, respectively, and show in an accompanying paper
398 that the LAT-1–TOL-1 complex is needed during early embryo development.[64] Another
399 unexpected observation from the high-throughput assay was the lack of an interaction hit between
400 Neurexin (NRX-1) and Neuroligin (NLG-1). Neurexins and Neuroligins are major regulators of

401 synapse formation and function, and interact strongly with each other.[65] However, biochemical
402 proof of a direct interaction between *C. elegans* NRX-1 and NLG-1 is limited.[66] To scrutinize our
403 unexpected negative result, we performed surface plasmon resonance experiments with the
404 ectodomain of NLG-1 against the LNS6 domain of NRX-1, previously established as the domain
405 responsible for Neuroligin interactions.[67] Unlike studies demonstrating nM affinity with comparable
406 constructs of mammalian homologs,[68,69] we observed binding between nematode NRX-1 and
407 NLG-1 with a ~48 μM dissociation constant, about three orders of magnitude weaker than the
408 mammalian orthologs (**Figures S7A and S7B**). Similarly, size-exclusion chromatography
409 experiments showed no stable complex formation, unlike previous observations for mammalian
410 Neurexin-Neuroligin complexes, but in agreement with a very weak complex (**Figure S7C**).[70]
411 Therefore, it is possible that while nematode Neurexins and Neuroligins may have preserved their
412 neuronal functions, they may mediate it through interactions via other proteins, such as with
413 MADD-4,[71] via novel interaction partners we identified in our assay, such as TMEM132 for NRX-
414 1, or heavily depend on other factors, such as heparan sulfate modifications of Neurexin as
415 recently discovered[72] for strong complex formation.

416 **AlphaFold-Multimer predictions for interactions observed by ECIA**

417 Since we observed that AlphaFold models have proven useful in revealing novel interactions from
418 our dataset, we took a systematic approach to study if AlphaFold results may indicate complex
419 formation. We calculated AlphaFold-multimer complexes for 72 high-confidence interactions we
420 observed for complexes with <2500 amino acids, as well as 46 complexes created by randomly
421 pairing proteins from the same set of proteins. We used the ipTM values reported by AlphaFold
422 as a measure of its confidence for the complex model. Our observed complexes had an average
423 ipTM value of 0.495 ± 0.241, while random protein pairs gave an average ipTM value of 0.273 ±
424 0.139 (**Figure 7A and Table S3**). The remarkable differences in the distribution of ipTM values
425 for observed complexes vs. randomly paired proteins attest to the usefulness of AlphaFold as a
426 predictor of protein complex formation, while also supporting our discovery results.

427 To analyze AlphaFold predictions further, we investigated whether observing higher ipTM values
428 correlate with any property of our complexes. We observed no correlation between sequence
429 length and ipTM (**Figure 7B**). However, ipTM values were significantly higher for complexes

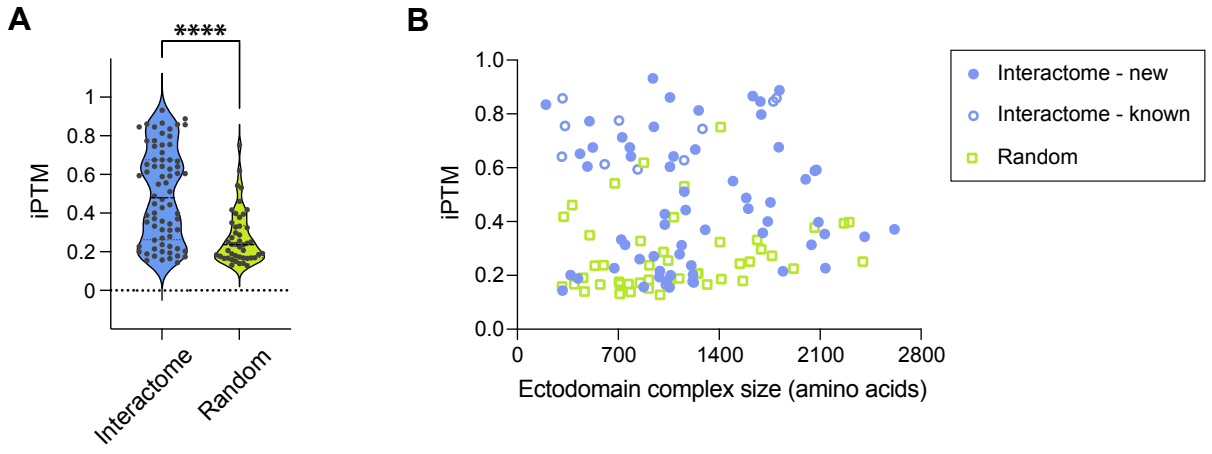



**Figure 7.** AlphaFold predictions of extracellular protein complexes and chemical synapses corresponding to PPIs.
**A.** ipTM (interface predicted Template Modelling) score values of AlphaFold-predicted complexes from our interactome (blue) and random pairings of proteins (green).
**B.** Size of the protein complex does not correlate with ipTM values. ECIA hits with previously determined homologous

which had homologous (mostly mammalian) complex structures in the Protein Data Bank (0.676 ± 0.190). This is not surprising, as these structures were likely in the training set for AlphaFold-multimer. However, even after removing protein pairs with homologous complex structures in the PDB, the complexes we detected had significantly higher ipTM values on average (0.459 ± 0.236) than randomly paired complexes, demonstrating that AlphaFold can be a useful tool in identifying novel protein complexes, and for studying the structures of these complexes. Computational pipelines using this idea have been established recently.[73]

**Chemical synapses corresponding to the observed interactions.**

Protein-protein interactions (PPIs) play a crucial role in forming chemical synapses between neurons.[74] Therefore, we investigated the chemical synapse connectome[75] associated with PPIs. Specifically, a PPI is considered as evidence that supports a chemical synapse if the corresponding genes are expressed in the neurons forming a chemical synapse. We used the gene expression data from CeNGEN.[76] Among the 134 PPIs that has available gene expression data related to chemical synapses, there are only 4 PPIs that do not serve as evidence of any known chemical synapses, namely ZC374.2–T02C5.1A, INS-6–ZIG-2, SGCA-1–IRLD-51, and LRON-7–T02C5.1A. The remaining 130 PPIs can potentially explain the entire known chemical synapse connectome. While 99.9% of chemical synapses have more than 5 evidences, there are a few examples of more specific chemical synapses. For example, HMR-1–HMR-1, IGCM-3–IGCM-3, and MAB-20–PLX-2 support the chemical synapses between the AVKL and AVKR neurons. Additionally, we have included the 185 experimental PPIs with the number of supporting synapses in **Table S7**. To establish a robust statistical framework for evaluating the significance of these interactions associated with the formation of chemical synapses, we implemented a randomized neuron connectome as a control. Our analysis revealed two distinct categories of noteworthy PPIs: (1) PPIs with the highest number of supporting synapses: TMEM-132–NRX-1, CASY-1–T19D12.6, PTP-3–NLR-1, CASY-1–PLX-1, SAX-3–PLX-1. These interactions are of particular interest due to their prevalence within the synaptic connectome, potentially indicating their functional importance in neuronal communication; (2) PPIs demonstrating the highest statistical significance when compared to the randomized control: SLT-1N–SLT-1C, LECT-2–MIG-13, UNC-5–PXN-1, ZIG-4–DAF-28, F15B9.8–ZK856.6. The statistical significance of these

459 PPIs suggests that their occurrence is non-random and may represent biologically relevant
460 interactions worthy of further investigation. These results highlight significant putative
461 contributions of experimental PPIs in unraveling the intricate complexities of chemical synapses.

462

463 **DISCUSSION**

464 Protein-protein interactions in the extracellular space control the development and functioning of
465 all aspects of multicellular physiology. With this study, we report the results of an extracellular
466 interaction screen of 379 *C. elegans* cell surface receptors and secreted proteins, the largest for
467 a non-human species to-date, that covers a diverse range of the protein fold space. The
468 interactions we report are mostly novel, even when counting those identified for orthologs in other
469 taxa as previously known (86% previously unknown).

470 **Novel interactions among guidance cues and receptors for neuronal development and**
471 **animal morphogenesis**

472 In our ectodomain collection, we have included known axon guidance cues and receptors, and to
473 our surprise, observed many novel interactions that link separate cue-receptor axes. This brings
474 up the possibility that these interactions might build "supra-signaling complexes" able to
475 incorporate cues and signaling receptors from multiple pathways. Based on our map of domain-
476 domain interactions (**Figure 3H**), the Robo-Ephrin-Semaphorin (SAX-3–EFN-4–MAB-20), Robo-
477 Ephrin-Eph (SAX-3–EFN-4–VAB-1) or Robo-Plexin-Semaphorin (SAX-3–PLX-1–SMP-1/2)
478 supercomplexes may be possible, as the domains needed to form these complexes do not
479 overlap. How these supra-molecular complexes would signal needs to be studied *in vivo*,
480 especially in the light of expression patterns of these various proteins. It is also possible that some
481 of the new interactions we identified act to silence the canonical pathways by blocking the
482 formation of the known complexes: this may be the case for SAX-3/Robo, as all its binding
483 partners interact with its IG domains. These interactions can be studied readily, thanks to the
484 extensive collection of mutant strains, published genetic interactions, and expression data already
485 available in *C. elegans*.

The interactions we identified between the separate axon guidance axes also highlight the multiple roles these receptors and ligands play. For example, morphogenesis of the intestine in *C. elegans* is known to be mediated by movements of cells governed by the actions of MAB-20/Semaphorin, SAX-3/Robo and EFN-4/Ephrin, which we have now identified to interact with each other.[77] Similarly, *efn-4* and *mab-20* mutant animals have highly similar phenotypes in axonal growth and guidance in the same cells,[78,79] as well as in epidermal enclosure of embryos,[80] and male tail morphogenesis.[81,82] Genetic interactions and shared phenotypes identified for these genes are likely due to the direct physical interactions we report here.

**Growth factors and their receptors in *C. elegans***

We have identified cystine-knot binding partners for the *C. elegans* Trk-like receptor, TRK-1. Previous efforts to identify nematode neurotrophins and their receptors by sequence have not been successful.[83] Despite lack of apparent sequence similarity for both the vertebrate neurotrophins and the Trk receptor ectodomain, AlphaFold prediction of our newly identified complexes at 2:2 stoichiometry, show strong resemblances to the known structures of mammalian Trk receptors, neurotrophins and their complexes (**Figure S4**).[84–86]

We report additional interactions for growth factor-like molecules, one of which is the worm ortholog of the Neuron-derived neurotrophic factor, NDNF-1, interacting with EGL-15/FGFR. We suggest that this interaction may be the underlying reason for why NDNF and FGFR mutations are both found in human patients with congenital hypogonadotropic hypogonadism and Kallmann syndrome. The fruit fly NDNF, Nord, also binds with and regulates degradation of Dally, a glypican-type heparan sulfate proteoglycan.[87,88] Since FGF-FGFR interactions are strengthened by heparan sulfate (**Figure 5D**),[56,57] and that heparin does not break the NDNF-1-EGL-15/FGFR interaction (**Figure 5E**), it is reasonable to speculate that NDNF may interact with Dally/Glypican as well as FGF receptors, and prevent FGF from turning on FGFR signaling.

**Interactomes for model organisms**

One important takeaway message we learned is that proper study of biological processes in a model organism requires dedicated effort in revealing the relevant biochemistry for that model organism. As one example, insulin signaling in *C. elegans* has been heavily studied and proven

514 insightful for understanding aging, neuronal and embryonic development. However, as we show
515 here, more may need to be done to reveal the biochemical components of the system (such as
516 ZIG-2 to -5) and their effects on signaling, which may or may not be shared with other model
517 organisms.

518 Conversely, our interactome dataset will likely prove useful across many model organisms. For
519 example, we revealed several new binding partners for the two *C. elegans* homologs of
520 *Drosophila* Dprs and DIPs, which are a 32-member receptor family known for their roles in
521 neuronal wiring in the fruit fly. Our findings will facilitate study of neuronal wiring in both model
522 systems, given the novel signaling co-receptors we propose for this receptor family. In addition,
523 the interactome dataset can be used in conjunction with expression data to identify candidate
524 proteins and complexes with roles in any cell surface process in specific cell types, such as
525 synapse formation. For example, expression of ZIG-8/Dpr and RIG-5/DIP are enriched in pairs of
526 cells known to form synapses (**Table S7**).

527 **Towards a complete extracellular interactome**

528 Here, we report significant improvements to our extracellular interactome methodology, including
529 a statistically rigorous method for data analysis, applicable to related interactome strategies and
530 high-throughput ELISA-like assays. We also implemented enhancements to improve throughput;
531 Additional gains in throughput can also be achieved by pooling of bait or prey, as recently done
532 in Wojtowicz *et al.*[8] The largest obstacle for us has been the time and cost of creating the plasmid
533 collection, while the entire assay could then be completed by two full-time researchers in only two
534 months. As gene synthesis costs decrease significantly, complete interactomes of all receptors
535 and secreted proteins will become cheaper to execute for several model organisms of interest, to
536 support a more complete mechanistic understanding of development, health and disease.

537 **ACKNOWLEDGMENTS**

538 We would like to thank Kang Shen for guidance and sharing reagents, and Jingxian Li, Demet
539 Araç, and Joseph S. Pak for technical and critical discussions, and Mateusz Krzyscik and Dengke
540 Ma for sharing reagents. We acknowledge Hyun Lee and the University of Illinois at Chicago
541 Biophysics Core Facility, and Elena Solomaha and the UChicago Biophysics Core Facility for

SPR access and help. E.Ö. and I.A.K. acknowledge both support from the National Institute for Theory and Mathematics in Biology through the National Science Foundation (grant number DMS-2235451) and the Simons Foundation (grant number MP-TMPS-00005320). This study used resources of the Advanced Photon Source (APS), a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Crystallographic data was collected at Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences from the NIH (P30 GM124165). The Eiger 16M detector on the 24-ID-E beam line is funded by a NIH-ORIP HEI grant (S10OD021527).

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## REFERENCES


1. Nichols, S.A., Dirks, W., Pearse, J.S., and King, N. (2006). Early evolution of animal cell signaling and adhesion genes. Proc Natl Acad Sci U S A *103*, 12451–12456. https://doi.org/10.1073/pnas.0604065103.

2. Santos, R., Ursu, O., Gaulton, A., Bento, A.P., Donadi, R.S., Bologa, C.G., Karlsson, A., Al-Lazikani, B., Hersey, A., Oprea, T.I., et al. (2017). A comprehensive map of molecular drug targets. Nat Rev Drug Discov *16*, 19–34. https://doi.org/10.1038/nrd.2016.230.

3. Bushell, K.M., Söllner, C., Schuster-Boeckler, B., Bateman, A., and Wright, G.J. (2008). Large-scale screening for novel low-affinity extracellular protein interactions. Genome Res *18*, 622–630. https://doi.org/10.1101/gr.7187808.

4. Özkan, E., Carrillo, R.A., Eastman, C.L., Weiszmann, R., Waghray, D., Johnson, K.G., Zinn, K., Celniker, S.E., and Garcia, K.C. (2013). An Extracellular Interactome of Immunoglobulin and LRR Proteins Reveals Receptor-Ligand Networks. Cell *154*, 228–239. https://doi.org/10.1016/j.cell.2013.06.006.

5. Wojtowicz, W.M., Wu, W., Andre, I., Qian, B., Baker, D., and Zipursky, S.L. (2007). A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. Cell *130*, 1134–1145. https://doi.org/10.1016/j.cell.2007.08.026.

6. Visser, J.J., Cheng, Y., Perry, S.C., Chastain, A.B., Parsa, B., Masri, S.S., Ray, T.A., Kay, J.N., and Wojtowicz, W.M. (2015). An extracellular biochemical screen reveals that FLRTs and Unc5s mediate neuronal subtype recognition in the retina. eLife *4*, e08149. https://doi.org/10.7554/eLife.08149.

7. Smakowska-Luzan, E., Mott, G.A., Parys, K., Stegmann, M., Howton, T.C., Layeghifard, M., Neuhold, J., Lehner, A., Kong, J., Grünwald, K., et al. (2018). An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. Nature *553*, 342–346. https://doi.org/10.1038/nature25184.

8. Wojtowicz, W.M., Vielmetter, J., Fernandes, R.A., Siepe, D.H., Eastman, C.L., Chisholm, G.B., Cox, S., Klock, H., Anderson, P.W., Rue, S.M., et al. (2020). A Human IgSF Cell-Surface Interactome Reveals a Complex Network of Protein-Protein Interactions. Cell *182*, 1027-1043.e17. https://doi.org/10.1016/j.cell.2020.07.025.

9. Verschueren, E., Husain, B., Yuen, K., Sun, Y., Paduchuri, S., Senbabaoglu, Y., Lehoux, I., Arena, T.A., Wilson, B., Lianoglou, S., et al. (2020). The Immunoglobulin Superfamily Receptome Defines Cancer-Relevant Networks Associated with Clinical Outcome. Cell *182*, 329-344.e19. https://doi.org/10.1016/j.cell.2020.06.007.

10. Li, H., Watson, A., Olechwier, A., Anaya, M., Sorooshyari, S.K., Harnett, D.P., Lee, H.-K.P., Vielmetter, J., Fares, M.A., Garcia, K.C., et al. (2017). Deconstruction of the Beaten Path-Sidestep interaction network provides insights into neuromuscular system development. eLife *6*, e28111. https://doi.org/10.7554/eLife.28111.

11. Söllner, C., and Wright, G.J. (2009). A cell surface interaction network of neural leucine-rich repeat receptors. Genome Biol *10*, R99. https://doi.org/10.1186/gb-2009-10-9-r99.

12. Corsi, A.K., Wightman, B., and Chalfie, M. (2015). A Transparent Window into Biology: A Primer on Caenorhabditis elegans. Genetics *200*, 387–407. https://doi.org/10.1534/genetics.115.176099.

13. Vogel, C., and Chothia, C. (2006). Protein family expansions and biological complexity. PLoS Comput Biol *2*, e48. https://doi.org/10.1371/journal.pcbi.0020048.

14. Wright, G.J. (2009). Signal initiation in biological systems: the properties and detection of transient extracellular protein interactions. Mol Biosyst *5*, 1405–1412. https://doi.org/10.1039/B903580J.

15. Vogel, C., Teichmann, S.A., and Chothia, C. (2003). The immunoglobulin superfamily in Drosophila melanogaster and Caenorhabditis elegans and the evolution of complexity. Development *130*, 6317–6328. https://doi.org/10.1242/dev.00848.

16. Hobert, O., Hutter, H., and Hynes, R.O. (2004). The immunoglobulin superfamily in Caenorhabditis elegans and Drosophila melanogaster. Development *131*, 2237–2238; author reply 2238-2240. https://doi.org/10.1242/dev.01183.

17. Huynh, C.Q., and Zieler, H. (1999). Construction of modular and versatile plasmid vectors for the high-level expression of single or multiple genes in insects and insect cell lines. J Mol Biol *288*, 13–20. https://doi.org/10.1006/jmbi.1999.2674.

18. Lee, D.F., Chen, C.C., Hsu, T.A., and Juang, J.L. (2000). A baculovirus superinfection system: efficient vehicle for gene transfer into Drosophila S2 cells. J Virol *74*, 11873–11880. https://doi.org/10.1128/jvi.74.24.11873-11880.2000.

19. Kovács, I.A., Barabási, D.L., and Barabási, A.-L. (2020). Uncovering the genetic blueprint of the C. elegans nervous system. Proc Natl Acad Sci U S A *117*, 33570–33577. https://doi.org/10.1073/pnas.2009093117.

20. Spirin, V., and Mirny, L.A. (2003). Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci U S A *100*, 12123–12128. https://doi.org/10.1073/pnas.2032324100.

21. Lambiotte, R., Delvenne, J.-C., and Barahona, M. (2014). Random Walks, Markov Processes and the Multiscale Modular Organization of Complex Networks. IEEE Trans Netw Sci Eng *1*, 76–90. https://doi.org/10.1109/TNSE.2015.2391998.

22. Alther, T.A., Domanitskaya, E., and Stoeckli, E.T. (2016). Calsyntenin 1-mediated trafficking of axon guidance receptors regulates the switch in axonal responsiveness at a choice point. Development *143*, 994–1004. https://doi.org/10.1242/dev.127449.

23. Ruiz de Almodovar, C., Fabre, P.J., Knevels, E., Coulon, C., Segura, I., Haddick, P.C.G., Aerts, L., Delattin, N., Strasser, G., Oh, W.-J., et al. (2011). VEGF mediates commissural axon chemoattraction through its receptor Flk1. Neuron *70*, 966–978. https://doi.org/10.1016/j.neuron.2011.04.014.

24. Bashaw, G.J., and Klein, R. (2010). Signaling from axon guidance receptors. Cold Spring Harb Perspect Biol *2*, a001941. https://doi.org/10.1101/cshperspect.a001941.

25. Yu, T.W., Hao, J.C., Lim, W., Tessier-Lavigne, M., and Bargmann, C.I. (2002). Shared receptors in axon guidance: SAX-3/Robo signals via UNC-34/Enabled and a Netrin-independent UNC-40/DCC function. Nat Neurosci *5*, 1147–1154. https://doi.org/10.1038/nn956.

26. Lee, J., Bandyopadhyay, J., Lee, J.I., Cho, I., Park, D., and Cho, J.H. (2015). A role for peroxidasin PXN-1 in aspects of C. elegans development. Mol Cells *38*, 51–57.

https://doi.org/10.14348/molcells.2015.2202.

27. Kimura, K.D., Tissenbaum, H.A., Liu, Y., and Ruvkun, G. (1997). daf-2, an insulin receptor-like gene that regulates longevity and diapause in *Caenorhabditis elegans*. Science *277*, 942–946. https://doi.org/10.1126/science.277.5328.942.

28. Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., and Hafen, E. (2001). An evolutionarily conserved function of the Drosophila insulin receptor and insulin-like peptides in growth control. Curr Biol *11*, 213–221. https://doi.org/10.1016/s0960-9822(01)00068-9.

29. Hua, Q.-X., Nakagawa, S.H., Wilken, J., Ramos, R.R., Jia, W., Bass, J., and Weiss, M.A. (2003). A divergent INS protein in Caenorhabditis elegans structurally resembles human insulin and activates the human insulin receptor. Genes Dev *17*, 826–831. https://doi.org/10.1101/gad.1058003.

30. Duret, L., Guex, N., Peitsch, M.C., and Bairoch, A. (1998). New insulin-like proteins with atypical disulfide bond pattern characterized in *Caenorhabditis elegans* by comparative sequence analysis and homology modeling. Genome Res *8*, 348–353. https://doi.org/10.1101/gr.8.4.348.

31. Murphy, C.T., and Hu, P.J. (2013). Insulin/insulin-like growth factor signaling in *C. elegans*. WormBook, 1–43. https://doi.org/10.1895/wormbook.1.164.1.

32. Pierce, S.B., Costa, M., Wisotzkey, R., Devadhar, S., Homburger, S.A., Buchman, A.R., Ferguson, K.C., Heller, J., Platt, D.M., Pasquinelli, A.A., et al. (2001). Regulation of DAF-2 receptor signaling by human insulin and ins-1, a member of the unusually large and diverse C. elegans insulin gene family. Genes Dev *15*, 672–686. https://doi.org/10.1101/gad.867301.

33. Zheng, S., Chiu, H., Boudreau, J., Papanicolaou, T., Bendena, W., and Chin-Sang, I. (2018). A functional study of all 40 Caenorhabditis elegans insulin-like peptides. J Biol Chem *293*, 16912–16922. https://doi.org/10.1074/jbc.RA118.004542.

34. Aurelio, O., Hall, D.H., and Hobert, O. (2002). Immunoglobulin-domain proteins required for maintenance of ventral nerve cord organization. Science *295*, 686–690. https://doi.org/10.1126/science.1066642.

35. Hutter, H. (2019). Formation of longitudinal axon pathways in Caenorhabditis elegans. Semin Cell Dev Biol *85*, 60–70. https://doi.org/10.1016/j.semcdb.2017.11.015.

36. Roed, N.K., Viola, C.M., Kristensen, O., Schluckebier, G., Norrman, M., Sajid, W., Wade, J.D., Andersen, A.S., Kristensen, C., Ganderton, T.R., et al. (2018). Structures of insect Imp-L2 suggest an alternative strategy for regulating the bioavailability of insulin-like hormones. Nat Commun *9*, 3860. https://doi.org/10.1038/s41467-018-06192-3.

37. Menting, J.G., Whittaker, J., Margetts, M.B., Whittaker, L.J., Kong, G.K.-W., Smith, B.J., Watson, C.J., Záková, L., Kletvíková, E., Jiráček, J., et al. (2013). How insulin engages its primary binding site on the insulin receptor. Nature *493*, 241–245. https://doi.org/10.1038/nature11781.

38. Uchikawa, E., Choi, E., Shang, G., Yu, H., and Bai, X.-C. (2019). Activation mechanism of the insulin receptor revealed by cryo-EM structure of the fully liganded receptor-ligand complex. eLife *8*, e48630. https://doi.org/10.7554/eLife.48630.

39. Patapoutian, A., and Reichardt, L.F. (2001). Trk receptors: mediators of neurotrophin action. Curr Opin Neurobiol *11*, 272–280. https://doi.org/10.1016/s0959-4388(00)00208-7.

40. Sun, P.D., and Davies, D.R. (1995). The cystine-knot growth-factor superfamily. Annu Rev Biophys Biomol Struct *24*, 269–291. https://doi.org/10.1146/annurev.bb.24.060195.001413.

41. Vozdek, R., Long, Y., and Ma, D.K. (2018). The receptor tyrosine kinase HIR-1 coordinates HIF-independent responses to hypoxia and extracellular matrix injury. Sci Signal *11*, eaat0138. https://doi.org/10.1126/scisignal.aat0138.

42. Schlessinger, J., Plotnikov, A.N., Ibrahimi, O.A., Eliseenkova, A.V., Yeh, B.K., Yayon, A., Linhardt, R.J., and Mohammadi, M. (2000). Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization. Mol Cell *6*, 743–750. https://doi.org/10.1016/s1097-2765(00)00073-3.

43. Messina, A., Pulli, K., Santini, S., Acierno, J., Känsäkoski, J., Cassatella, D., Xu, C., Casoni, F., Malone, S.A., Ternier, G., et al. (2020). Neuron-Derived Neurotrophic Factor Is Mutated in Congenital Hypogonadotropic Hypogonadism. Am J Hum Genet *106*, 58–70. https://doi.org/10.1016/j.ajhg.2019.12.003.

44. Pitteloud, N., Acierno, J.S., Meysing, A., Eliseenkova, A.V., Ma, J., Ibrahimi, O.A., Metzger, D.L., Hayes, F.J., Dwyer, A.A., Hughes, V.A., et al. (2006). Mutations in fibroblast growth factor receptor 1 cause both Kallmann syndrome and normosmic idiopathic hypogonadotropic hypogonadism. Proc Natl Acad Sci U S A *103*, 6281–6286. https://doi.org/10.1073/pnas.0600962103.

45. Cheng, S., Park, Y., Kurleto, J.D., Jeon, M., Zinn, K., Thornton, J.W., and Özkan, E. (2019). Family of neural wiring receptors in bilaterians defined by phylogenetic, biochemical, and structural evidence. Proc Natl Acad Sci U S A *116*, 9837–9842. https://doi.org/10.1073/pnas.1818631116.

46. Lobb-Rabe, M., Nawrocka, W.I., Carrillo, R.A., and Özkan, E. (2023). Neuronal wiring receptors Dprs and DIPs are GPI anchored and this modification contributes to their cell surface organization. Preprint at bioRxiv, https://doi.org/10.1101/2023.03.02.530872 https://doi.org/10.1101/2023.03.02.530872.

47. Baumgartner, S., Littleton, J.T., Broadie, K., Bhat, M.A., Harbecke, R., Lengyel, J.A., Chiquet-Ehrismann, R., Prokop, A., and Bellen, H.J. (1996). A Drosophila neurexin is required for septate junction and blood-nerve barrier formation and function. Cell *87*, 1059–1068. https://doi.org/10.1016/s0092-8674(00)81800-0.

48. Meng, L., and Yan, D. (2020). NLR-1/CASPR Anchors F-Actin to Promote Gap Junction Formation. Dev Cell *55*, 574-587.e3. https://doi.org/10.1016/j.devcel.2020.10.020.

49. Einheber, S., Zanazzi, G., Ching, W., Scherer, S., Milner, T.A., Peles, E., and Salzer, J.L. (1997). The axonal membrane protein Caspr, a homologue of neurexin IV, is a component of the septate-like paranodal junctions that assemble during myelination. J Cell Biol *139*, 1495–1506. https://doi.org/10.1083/jcb.139.6.1495.

50. Poliak, S., Gollan, L., Martinez, R., Custer, A., Einheber, S., Salzer, J.L., Trimmer, J.S., Shrager, P., and Peles, E. (1999). Caspr2, a new member of the neurexin superfamily, is localized at the juxtaparanodes of myelinated axons and associates with K+ channels. Neuron *24*, 1037–1047. https://doi.org/10.1016/s0896-6273(00)81049-1.

720 51. Anderson, G.R., Galfin, T., Xu, W., Aoto, J., Malenka, R.C., and Südhof, T.C. (2012).
721 Candidate autism gene screen identifies critical role for cell-adhesion molecule CASPR2 in
722 dendritic arborization and spine development. Proc Natl Acad Sci U S A *109*, 18120–18125.
723 https://doi.org/10.1073/pnas.1216398109.

724 52. Varea, O., Martin-de-Saavedra, M.D., Kopeikina, K.J., Schürmann, B., Fleming, H.J.,
725 Fawcett-Patel, J.M., Bach, A., Jang, S., Peles, E., Kim, E., et al. (2015). Synaptic
726 abnormalities and cytoplasmic glutamate receptor aggregates in contactin associated
727 protein-like 2/Caspr2 knockout neurons. Proc Natl Acad Sci U S A *112*, 6176–6181.
728 https://doi.org/10.1073/pnas.1423205112.

729 53. Kaufmann, N., DeProto, J., Ranjan, R., Wan, H., and Van Vactor, D. (2002). Drosophila
730 liprin-alpha and the receptor phosphatase Dlar control synapse morphogenesis. Neuron *34*,
731 27–38. https://doi.org/10.1016/s0896-6273(02)00643-8.

732 54. Johnson, K.G., Tenney, A.P., Ghose, A., Duckworth, A.M., Higashi, M.E., Parfitt, K., Marcu,
733 O., Heslip, T.R., Marsh, J.L., Schwarz, T.L., et al. (2006). The HSPGs Syndecan and
734 Dallylike bind the receptor phosphatase LAR and exert distinct effects on synaptic
735 development. Neuron *49*, 517–531. https://doi.org/10.1016/j.neuron.2006.01.026.

736 55. Carrillo, R.A., Özkan, E., Menon, K.P., Nagarkar-Jaiswal, S., Lee, P.-T., Jeon, M.,
737 Birnbaum, M.E., Bellen, H.J., Garcia, K.C., and Zinn, K. (2015). Control of Synaptic
738 Connectivity by a Network of Drosophila IgSF Cell Surface Proteins. Cell *163*, 1770–1782.
739 https://doi.org/10.1016/j.cell.2015.11.022.

740 56. Broadbent, I.D., and Pettitt, J. (2002). The C. elegans hmr-1 gene can encode a neuronal
741 classic cadherin involved in the regulation of axon fasciculation. Curr Biol *12*, 59–63.
742 https://doi.org/10.1016/s0960-9822(01)00624-8.

743 57. Barnes, K.M., Fan, L., Moyle, M.W., Brittin, C.A., Xu, Y., Colón-Ramos, D.A., Santella, A.,
744 and Bao, Z. (2020). Cadherin preserves cohesion across involuting tissues during C.
745 elegans neurulation. eLife *9*, e58626. https://doi.org/10.7554/eLife.58626.

746 58. Cebul, E.R., Marivin, A., Wexler, L.R., Perrat, P.N., Bénard, C.Y., Garcia-Marcos, M., and
747 Heiman, M.G. (2024). SAX-7/L1CAM acts with the adherens junction proteins MAGI-1,
748 HMR-1/Cadherin, and AFD-1/Afadin to promote glial-mediated dendrite extension. Preprint
749 at bioRxiv, https://doi.org/10.1101/2024.01.11.575259
750 https://doi.org/10.1101/2024.01.11.575259.

751 59. Yonekura, S., Xu, L., Ting, C.-Y., and Lee, C.-H. (2007). Adhesive but not signaling activity
752 of Drosophila N-cadherin is essential for target selection of photoreceptor afferents. Dev Biol
753 *304*, 759–770. https://doi.org/10.1016/j.ydbio.2007.01.030.

754 60. Pecot, M.Y., Tadros, W., Nern, A., Bader, M., Chen, Y., and Zipursky, S.L. (2013). Multiple
755 interactions control synaptic layer specificity in the Drosophila visual system. Neuron *77*,
756 299–310. https://doi.org/10.1016/j.neuron.2012.11.007.

757 61. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022).
758 ColabFold: making protein folding accessible to all. Nat Methods *19*, 679–682.
759 https://doi.org/10.1038/s41592-022-01488-1.

760 62. Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R.,
761 Blackwell, S., Yim, J., et al. (2022). Protein complex prediction with AlphaFold-Multimer.
762 Preprint at bioRxiv, https://doi.org/10.1101/2021.10.04.463034

https://doi.org/10.1101/2021.10.04.463034.

63. Schöneberg, T., and Prömel, S. (2019). Latrophilins and Teneurins in Invertebrates: No Love for Each Other? Front Neurosci *13*, 154. https://doi.org/10.3389/fnins.2019.00154.

64. Carmona-Rosas, G., Li, J., Smith, J.J., Cheng, S., Baltrusaitis, E., Nawrocka, W.I., Zhao, M., Kratsios, P., Araç, D., and Özkan, E. (2023). Structural basis and functional roles for Toll-like receptor binding to Latrophilin adhesion-GPCR in embryo development. Preprint at bioRxiv, https://doi.org/10.1101/2023.05.04.539414 https://doi.org/10.1101/2023.05.04.539414.

65. Südhof, T.C. (2017). Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits. Cell *171*, 745–769. https://doi.org/10.1016/j.cell.2017.10.024.

66. Hu, Z., Hom, S., Kudze, T., Tong, X.-J., Choi, S., Aramuni, G., Zhang, W., and Kaplan, J.M. (2012). Neurexin and neuroligin mediate retrograde synaptic inhibition in C. elegans. Science *337*, 980–984. https://doi.org/10.1126/science.1224896.

67. Ichtchenko, K., Hata, Y., Nguyen, T., Ullrich, B., Missler, M., Moomaw, C., and Südhof, T.C. (1995). Neuroligin 1: a splice site-specific ligand for beta-neurexins. Cell *81*, 435–443. https://doi.org/10.1016/0092-8674(95)90396-8.

68. Araç, D., Boucard, A.A., Özkan, E., Strop, P., Newell, E., Südhof, T.C., and Brunger, A.T. (2007). Structures of neuroligin-1 and the neuroligin-1/neurexin-1 beta complex reveal specific protein-protein and protein-Ca2+ interactions. Neuron *56*, 992–1003. https://doi.org/10.1016/j.neuron.2007.12.002.

69. Comoletti, D., Flynn, R., Jennings, L.L., Chubykin, A., Matsumura, T., Hasegawa, H., Südhof, T.C., and Taylor, P. (2003). Characterization of the interaction of a recombinant soluble neuroligin-1 with neurexin-1beta. J Biol Chem *278*, 50497–50505. https://doi.org/10.1074/jbc.M306803200.

70. Ko, J., Zhang, C., Arac, D., Boucard, A.A., Brunger, A.T., and Südhof, T.C. (2009). Neuroligin-1 performs neurexin-dependent and neurexin-independent functions in synapse validation. EMBO J *28*, 3244–3255. https://doi.org/10.1038/emboj.2009.249.

71. Maro, G.S., Gao, S., Olechwier, A.M., Hung, W.L., Liu, M., Özkan, E., Zhen, M., and Shen, K. (2015). MADD-4/Punctin and Neurexin Organize C. elegans GABAergic Postsynapses through Neuroligin. Neuron *86*, 1420–1432. https://doi.org/10.1016/j.neuron.2015.05.015.

72. Zhang, P., Lu, H., Peixoto, R.T., Pines, M.K., Ge, Y., Oku, S., Siddiqui, T.J., Xie, Y., Wu, W., Archer-Hartmann, S., et al. (2018). Heparan Sulfate Organizes Neuronal Synapses through Neurexin Partnerships. Cell *174*, 1450-1464.e23. https://doi.org/10.1016/j.cell.2018.07.002.

73. Schweke, H., Pacesa, M., Levin, T., Goverde, C.A., Kumar, P., Duhoo, Y., Dornfeld, L.J., Dubreuil, B., Georgeon, S., Ovchinnikov, S., et al. (2024). An atlas of protein homo-oligomerization across domains of life. Cell *187*, 999-1010.e15. https://doi.org/10.1016/j.cell.2024.01.022.

74. Sanes, J.R., and Zipursky, S.L. (2020). Synaptic Specificity, Recognition Molecules, and Assembly of Neural Circuits. Cell *181*, 536–556. https://doi.org/10.1016/j.cell.2020.04.008.

75. Varshney, L.R., Chen, B.L., Paniagua, E., Hall, D.H., and Chklovskii, D.B. (2011). Structural properties of the Caenorhabditis elegans neuronal network. PLoS Comput Biol *7*,

e1001066. https://doi.org/10.1371/journal.pcbi.1001066.

76. Taylor, S.R., Santpere, G., Weinreb, A., Barrett, A., Reilly, M.B., Xu, C., Varol, E., Oikonomou, P., Glenwinkel, L., McWhirter, R., et al. (2021). Molecular topography of an entire nervous system. Cell *184*, 4329-4347.e23. https://doi.org/10.1016/j.cell.2021.06.023.

77. Asan, A., Raiders, S.A., and Priess, J.R. (2016). Morphogenesis of the C. elegans Intestine Involves Axon Guidance Genes. PLoS Genet *12*, e1005950. https://doi.org/10.1371/journal.pgen.1005950.

78. Dong, B., Moseley-Alldredge, M., Schwieterman, A.A., Donelson, C.J., McMurry, J.L., Hudson, M.L., and Chen, L. (2016). EFN-4 functions in LAD-2-mediated axon guidance in Caenorhabditis elegans. Development *143*, 1182–1191. https://doi.org/10.1242/dev.128934.

79. Wang, X., Zhang, W., Cheever, T., Schwarz, V., Opperman, K., Hutter, H., Koepp, D., and Chen, L. (2008). The C. elegans L1CAM homologue LAD-2 functions as a coreceptor in MAB-20/Sema2 mediated axon guidance. J Cell Biol *180*, 233–246. https://doi.org/10.1083/jcb.200704178.

80. Nakao, F., Hudson, M.L., Suzuki, M., Peckler, Z., Kurokawa, R., Liu, Z., Gengyo-Ando, K., Nukazuka, A., Fujii, T., Suto, F., et al. (2007). The PLEXIN PLX-2 and the ephrin EFN-4 have distinct roles in MAB-20/Semaphorin 2A signaling in Caenorhabditis elegans morphogenesis. Genetics *176*, 1591–1607. https://doi.org/10.1534/genetics.106.067116.

81. Hahn, A.C., and Emmons, S.W. (2003). The roles of an ephrin and a semaphorin in patterning cell-cell contacts in C. elegans sensory organ development. Dev Biol *256*, 379–388. https://doi.org/10.1016/s0012-1606(02)00129-x.

82. Ikegami, R., Zheng, H., Ong, S.-H., and Culotti, J. (2004). Integration of semaphorin-2A/MAB-20, ephrin-4, and UNC-129 TGF-beta signaling pathways regulates sorting of distinct sensory rays in C. elegans. Dev Cell *6*, 383–395. https://doi.org/10.1016/s1534-5807(04)00057-7.

83. Jaaro, H., Beck, G., Conticello, S.G., and Fainzilber, M. (2001). Evolving better brains: a need for neurotrophins? Trends Neurosci *24*, 79–85. https://doi.org/10.1016/s0166-2236(00)01690-8.

84. Wiesmann, C., Ultsch, M.H., Bass, S.H., and de Vos, A.M. (1999). Crystal structure of nerve growth factor in complex with the ligand-binding domain of the TrkA receptor. Nature *401*, 184–188. https://doi.org/10.1038/43705.

85. Banfield, M.J., Naylor, R.L., Robertson, A.G., Allen, S.J., Dawbarn, D., and Brady, R.L. (2001). Specificity in Trk receptor:neurotrophin interactions: the crystal structure of TrkB-d5 in complex with neurotrophin-4/5. Structure *9*, 1191–1199. https://doi.org/10.1016/s0969-2126(01)00681-5.

86. Wehrman, T., He, X., Raab, B., Dukipatti, A., Blau, H., and Garcia, K.C. (2007). Structural and mechanistic insights into nerve growth factor interactions with the TrkA and p75 receptors. Neuron *53*, 25–38. https://doi.org/10.1016/j.neuron.2006.09.034.

87. Yang, S., Wu, X., Daoutidou, E.I., Zhang, Y., Shimell, M., Chuang, K.-H., Peterson, A.J., O'Connor, M.B., and Zheng, X. (2022). The NDNF-like factor Nord is a Hedgehog-induced extracellular BMP modulator that regulates Drosophila wing patterning and growth. eLife *11*, e73357. https://doi.org/10.7554/eLife.73357.

88. Akiyama, T., Seidel, C.W., and Gibson, M.C. (2022). The feedback regulator Nord controls Dpp/BMP signaling via extracellular interaction with Dally in the Drosophila wing. Dev Biol *488*, 91–103. https://doi.org/10.1016/j.ydbio.2022.05.008.

89. Letunic, I., and Bork, P. (2018). 20 years of the SMART protein domain annotation resource. Nucleic Acids Res *46*, D493–D496. https://doi.org/10.1093/nar/gkx922.

90. Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., Chothia, C., and Gough, J. (2009). SUPERFAMILY--sophisticated comparative genomics, data mining, visualization and phylogeny. Nucleic Acids Res *37*, D380-386. https://doi.org/10.1093/nar/gkn762.

91. Käll, L., Krogh, A., and Sonnhammer, E.L.L. (2004). A combined transmembrane topology and signal peptide prediction method. J Mol Biol *338*, 1027–1036. https://doi.org/10.1016/j.jmb.2004.03.016.

92. Pierleoni, A., Martelli, P.L., and Casadio, R. (2008). PredGPI: a GPI-anchor predictor. BMC Bioinformatics *9*, 392. https://doi.org/10.1186/1471-2105-9-392.

93. Gíslason, M.H., Nielsen, H., Almagro Armenteros, J.J., and Johansen, A.R. (2021). Prediction of GPI-anchored proteins with pointer neural networks. Curr Res Biotechnol *3*, 6–13. https://doi.org/10.1016/j.crbiot.2021.01.001.

94. Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., et al. (2022). AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res *50*, D439–D444. https://doi.org/10.1093/nar/gkab1061.

95. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R., and Finn, R.D. (2018). HMMER web server: 2018 update. Nucleic Acids Res *46*, W200–W204. https://doi.org/10.1093/nar/gky448.

96. Cheng, S., Seven, A.B., Wang, J., Skiniotis, G., and Özkan, E. (2016). Conformational Plasticity in the Transsynaptic Neurexin-Cerebellin-Glutamate Receptor Adhesion Complex. Structure *24*, 2163–2173. https://doi.org/10.1016/j.str.2016.11.004.

97. Kabsch, W. (2010). *XDS*. Acta Crystallogr D Biol Crystallogr *66*, 125–132. https://doi.org/10.1107/S0907444909047337.

98. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). *Phaser* crystallographic software. J Appl Crystallogr *40*, 658–674. https://doi.org/10.1107/S0021889807021206.

99. Liebschner, D., Afonine, P.V., Baker, M.L., Bunkóczi, G., Chen, V.B., Croll, T.I., Hintze, B., Hung, L.W., Jain, S., McCoy, A.J., et al. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. Acta Crystallogr D Struct Biol *75*, 861–877. https://doi.org/10.1107/S2059798319011471.

100. Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with *phenix.refine*. Acta Crystallogr D Biol Crystallogr *68*, 352–367. https://doi.org/10.1107/S0907444912001308.

101. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and

887 development of *Coot*. Acta Crystallogr D Biol Crystallogr *66*, 486–501.
888 https://doi.org/10.1107/S0907444910007493.

889 102. Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J.,
890 Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). *MolProbity*: all-atom structure
891 validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr *66*, 12–21.
892 https://doi.org/10.1107/S0907444909042073.

893

894

895 **METHODS**

896 **Selection of Ectodomains**

897 We collected names of genes annotated to have selected domains from *SMART*[89] and
898 *SUPERFAMILY*[90] databases in the *C. elegans* genome. Transcripts for these genes were
899 identified in the WS252 release of the *C. elegans* genome, and analyzed using *phobius* for signal
900 peptide and transmembrane domains.[91] The identified ectodomains were written into Genbank-
901 formatted files for bait and prey plasmids for cloning and/or gene synthesis. The bioinformatic
902 pipeline described above was performed using tools available in Bioperl modules. Proteins that
903 have no discernable transmembrane domains were also run on the PredGPI[92] server
904 (http://gpcr.biocomp.unibo.it/predgpi/) to assess whether they may be GPI-anchored; we removed
905 the C-terminal GPI anchoring sequences for those proteins with predicted GPI-anchors.
906 Predictions were repeated with the recently released NetGPI[93] to update **Table S1**.

907 Following curation of ectodomain sequences, included in **Table S1**, we re-analyzed domain
908 compositions using updated tools and AlphaFold predictions,[94] which became available as we
909 prepared the manuscript for publication. Based on the results from these tools, we decided not to
910 distinguish EGF domains from WR1 domains (SMART SM00289; Interpro IPR006150), EB
911 modules (Pfam PF01683) and Lustrin-type Cys-rich domains (Pfam PF14625; Interpro
912 IPR028150), which were previously defined using sequence alignments. We observed that
913 domain prediction tools such as *SMART* and *hmmscan*[95] repeatedly gave overlapping predictions
914 for these families, and structure predictions showed strong similarities and a continuum of
915 elaborations above basic features, making a rigid distinction between these folds difficult. Future
916 work using experimentally determined structures, predictions and sequence alignments will be
917 necessary to carefully classify similar Cys-rich small domains into better-defined categories.

918 **Ectodomain Expression Library Construction and Protein Expression**

919 To facilitate protein expression, we built new ECIA vectors carrying the strong constitutively active
920 Actin 5C promoter rather than the inducible metallothionein promoter on pECIA2 and pECIA14
921 described in Özkan *et al.*[4] Briefly, the multiple cloning site (MCS) region of pACTIN-SV[17] was
922 replaced by a BiP signal sequence followed by the cassette attB1-MCS-attB2-3C-Fc-V5 tag-$His_6$

923 (3C: HRV 3C Protease site) to generate the bait vector pECIA75, while the replacement of the
924 MCS of pACTIN-SV by BiP-attB1-MCS-attB2-3C-COMP-AP-FLAG-His6 (COMP: pentameric
925 coiled coil from rat COMP; AP: human placental alkaline phosphatase) resulted in the prey vector
926 pECIA76.

927 To construct the expression library for the proteins of interest listed in Table S1, the ectodomains
928 of 16 proteins from our lab collection were subcloned into the MCS of the new ECIA vectors. We
929 also purchased a *C. elegans* ORF library from GE Healthcare and found 107 proteins in Table S1
930 from the 12611 clones of the ORF library. However, only 63 (including 5 with mutation(s) or
931 intron(s) that were corrected in cloning) were successfully subcloned while the remaining 44
932 clones from the clone collection proved to be wrong or empty vectors. Through RT-PCR, 9 were
933 amplified from a *C.elegans* N2 mRNA pool (gift from Kang Shen at Stanford). The left 288 were
934 synthesized by GenScript.

935 Every bait or prey construct was individually expressed using transient transfection of *Drosophila*
936 Schneider 2 (S2) cells in Schneider's medium with 1.8 g/L L-Glutamine (Gibco, 21720024) with
937 10% Fetal Bovine Serum, 50 units/ml Penicillin, and 50 µg/ml Streptomycin. Briefly, 10 mL of S2
938 cells at 1.8 million per mL were seeded in T75 flasks and incubated at 28°C overnight. The cells
939 were transfected transiently next day with 5 µg of each plasmid using TransIT®-Insect
940 Transfection Reagent (Mirus, MIR 6104) following manufacturer's manuals. Conditioned media
941 were collected 4 days after transfection. Protease inhibitors (Sigma, P8849) and 0.02% $NaN_3$
942 were added to harvested media before storage at 4°C in 15 mL conical tubes.

943 All bait and prey samples collected were run on SDS-PAGE gels, blotted and probed with mouse
944 THE™ His Tag Antibody [iFluor 488] (GenScript, A01800) for assessing protein expression. Bio-
945 Rad ChemiDoc Imaging System was used to quantitate the bait and prey samples based on the
946 reference band of 0.1 µM rat $His_6$-tagged GluD2 ectodomain (ATD+LBD)[96] on each blot. Overall,
947 we observed expression as high as ~0.31 µM (LRON-3-Fc) in conditioned media. The lowest
948 protein concentration we could detect and measure was ~0.45 nM, for R09H10.5-$AP_5$. Expression
949 for 23% of the samples could not be detected. Relative expression levels are reported in **Table**
950 **S1**.

951 **Extracellular Interactome Assay (ECIA)**

952 Clear Nunc 384-well MaxiSorp plates (Thermo Scientific, 464718) were coated with 20 µl 5 µg/mL
953 Protein A (Genscript, Z02201) in 100 mM sodium bicarbonate pH 9.6 at 4°C overnight. Excess
954 protein A was discarded, and the coated plates were blocked with 90 µl 10% SuperBlock™ T20
955 (PBS) Blocking Buffer (ThermoFisher, 37516) in PBS at 500x rpm for 3 hours at room temperature.
956 The blocking buffer was then removed, and the plates were washed three times with 90 µl PBST
957 (PBS + 0.1% Tween 20) using a microplate washer (BioTek). 20 µl of medium containing a
958 secreted Fc-fusion protein (bait) was added into wells of a single plate at 4°C with a shaking at
959 500 rpm on a plate shaker overnight for bait capture. Next day, plates were blocked with 1%
960 Bovine Serum Albumin (BSA) in PBS for three hours at room temperature while shaking at 500
961 rpm, followed by three washes with 90 µl of wash buffer (PBS with 1 mM $CaCl_2$, 1 mM $MgCl_2$, and
962 0.1% BSA). This was followed by addition of 20 µl of different medium containing secreted $AP_5$-
963 fusion proteins (prey) into each well of a plate using a Rainin BenchSmart 96-well multipipetter
964 and incubation at room temperature while shaking at 500 rpm for 3 hours. The plates were then
965 washed three times with 90 µl of wash buffer. Finally, 50 µl of BluePhos Phosphatase Substrate
966 (KPL, 50-88-02) was added to each well using a Rainin BenchSmart 96-well multipipetter at room
967 temperature. Absorbance at 650 nm was measured at 1 and 2 hours using a microplate reader
968 and images of these 384-well plates were scanned.

969 **The contents of the large high-throughput screen data**

970 Each 384-well plate included four control measurements and 379 prey paired against one bait (+
971 one prey, NAS-23, repeated as a control but not used in the analysis). The 379 unique
972 ectodomains correspond to 374 genes, where some genes were represented by more than one
973 constructs, such as the N- and C-terminal fragments of SLT-1 (see Table S1).

974 The large ECIA screen data has each pairwise heterotypic interaction tested twice, with each
975 protein alternatively in the bait vs. prey configuration; this yielded two separate reports of
976 interaction strength. Homotypic interactions were measured once.

977 **Analysis of ECIA Data: MaxEnt statistical model**

978 To assess the significance of interactions, we employ a maximum entropy network ensemble-

based technique[19] to construct a statistical model that captures the absorbance background. Initially, we represent the experimental network by an $n \times m$ matrix $A$ and normalize it as $An = A/A_{max}$ (**Figure S2A**). Here, the rows and columns denote prey and bait, respectively, while entries denote the normalized absorbance of each candidate protein-protein interaction (PPI). To capture the random background given the observed row and column sums of $An$, we maximize the entropy of a random network ensemble, denoted as $S = -\Sigma_G P(G) \ln P(G)$, subject to soft constraints that preserve the average row and column sums of the normalized experimental data $An$. Here, $G$ denotes random networks and $P$ denotes the probability of $G$. With maximized entropy, the model generates an ensemble with the broadest distribution of binary networks (a matrix with only 0 and 1 entries) that on average replicates the row and column sums of $An$. The average matrix of the $P$ serves as the baseline expectation with the absence of any biological information (Figure 1B). Each matrix entry in *P* representing the connection probability of a prey-bait pair as $p_{ij} = 1/(\alpha_i \beta_j + 1)$. In practice, the parameters $\alpha_i, \beta_j$ are iteratively determined to preserve the mean total normalized absorbance of each prey $k_i = \Sigma_j a_{ij}$ and bait $k_j = \Sigma_i a_{ij}$ as

$$\alpha_i = \frac{1}{k_i} \Sigma_j \frac{1}{\beta_j + 1/\alpha_i}$$

$$\beta_j = \frac{1}{k_j} \Sigma_i \frac{1}{\alpha_i + 1/\beta_j}$$

where $a_{ij}$ represents the elements in the normalized absorbance matrix $An$. Here we apply 100 iterations to ensure the convergence of $\alpha_i$ and $\beta_j$. With these optimized $\alpha_i$ ($\beta_j$) for each row (column), we test the self-consistency of the model by verifying the implication that all $a_{ij}$ should satisfy $a_{ij} = 1/(1 + \alpha_i \beta_j)$. However, upon using the optimized $\alpha_i$ and $\beta_j$, we obtain sets of $m$ ($n$) different values of $\alpha_i$ ($\beta_j$) for each $\alpha_i$ ($\beta_j$), calculated as $\{\alpha_i\} = (a_{ij}^{-1} - 1)/\beta_j \ \forall j$ and $\{\beta_j\} = (a_{ij}^{-1} - 1)/\alpha_i \ \forall i$. We observed small fluctuations of these $\{\alpha_i\}$ and $\{\beta_j\}$ around the optimized values, quantified as $\sigma_{\alpha_i} = STD(\{\alpha_i\})$ and $\sigma_{\beta_j} = STD(\{\beta_j\})$. The standard deviation of $p_{ij}$ is then calculated using error propagation as $\sigma_p = \sqrt{\left(\frac{\partial p}{\partial \alpha_i}\sigma_{\alpha_i}\right)^2 + \left(\frac{\partial p}{\partial \beta_j}\sigma_{\beta_j}\right)^2}$. $\sigma_p$ quantifies the systematic error that emerges from the deviation of the experimental data from the statistical model. In

principle, there is also statistical error in $p_{ij}$ that can be calculated by $\sigma_{p_{ij}} \propto p_{ij}(1 - p_{ij})$. However, we find that the systematic error dominates in this case, thus, the statistical error is ignored. *z*-scores are used to identify potential signals that are significantly different from the baseline expectation as $z = \frac{An-P}{\sigma_P}$.

Although the *z*-score matrix can be asymmetric, we anticipate that a large fraction of the PPIs are symmetric, meaning reciprocated signal in the two experimental directions ($s \rightarrow t$ or $t \rightarrow s$). Thus, the reciprocal ratio of interactions works as a metric to reflect the precision of the experiment, akin to a cross-validation scenario. In **Figures S2D-E**, we illustrate how the reciprocal ratio of interactions varies with different z-score thresholds and unique PPIs. We assign z-score to each PPI by symmetrizing z scores in both directions as $z = (z_{A\rightarrow B} + z_{B\rightarrow A})/\sqrt{2}$ (**Table S2**). The scaling factor $\sqrt{2}$ ensures that the combined z-score remains a standard z-score, allowing for direct comparison and interpretation. We use the reciprocal ratio to determine the ideal z-score threshold to distinguish signal from background noise. Our analysis demonstrates that the maximum reciprocal ratio of 0.67, corresponding to 145 unique PPIs, is achieved at a stringent threshold of $z = 12.2$. Additionally, we observe that the reciprocal ratio fluctuates slightly between 0.61 and 0.67 for z-scores between 8.4 and 33.6 (**Figure S2D**). As a result, we propose an intermediate threshold of $z = 8.4$, which includes 185 unique PPIs (**Figure S2C, Table S3**). Selecting an even lower threshold is expected to include more false positives as indicated by a diminishing reciprocal ratio.

**Network Analysis of ECIA Data and Communities**

Network theory offers a particularly powerful framework for mining insight from measurements of physical protein-protein interactions. Here, we conducted a network analysis of our ECIA dataset.

*Community detection:* We identified communities in this PPI network using a multi-scale partition algorithm,[21] which identifies node groupings of different sizes by random walks of various lengths. The scale of the communities obtained by this approach is controlled by a hyperparameter tau (resolution), analogous to the length of the walk. Smaller values of tau correspond to shorter walks, which will on average end relatively close to where they begin. This results in the identification of

1031 relatively small communities. As the length of the walk (tau) increases, increasingly large
1032 neighborhoods are reached, with the likelihood of remaining in any particular neighborhood of the
1033 network as the walk proceeds determined by how insular the connectivity is within that
1034 neighborhood. We used this method across 50-fold variation in tau (0.1 to 5) to compute
1035 community partitions at a variety of scales. For each protein, we recorded the union of all other
1036 nodes that appear in the same community, across all choices of tau. The subset of proteins
1037 grouped in the same community across at least 25% of the tau range and $z_{min}$ range are further
1038 labeled as "canonical" community neighbors, and are made available in Table S4.

1039 Network analyses were performed using custom code written in Python 3.9, using NetworkX and
1040 python-louvain (https://github.com/taynaud/python-louvain). Network visualization was performed
1041 using NetworkX and Gephi. Code, data and output neighbor lists are available at
1042 https://github.com/mattrosen/ECIA-network-analysis.

1043 **Protein Expression and Purification for Interaction Validation**

1044 For crystallography and SPR experiments, proteins were produced using baculoviral infection of
1045 *Trichoplusia ni* cells cultured in Insect-XPRESS (Lonza, BP12-730Q) or ESF 921 (Expression
1046 Systems, 96-001) media. Ectodomains or smaller fragments of proteins were cloned into
1047 baculoviral transfer vector pAcGP67 and baculoviruses viruses were generated with BestBac
1048 linearized DNA (Expression systems, 91-002) using Sf9 cells cultured in Sf-900 III (Gibco,
1049 12658019) with 10% FBS. The constructs were cloned with the addition of C-terminal biotin
1050 acceptor peptide (for biotin capture on strepatividin-based SPR chips) followed by a hexahistidine
1051 tag. The constructs used in crystallography were cloned with C-terminal hexahistidine tags only.
1052 Proteins were purified from insect cell media using affinity chromatography with Ni-NTA agarose
1053 resin (Pierce HisPur, 88223) followed by size-exclusion chromatography using Superdex 200
1054 Increase or Superose 6 Increase 10/300 GL columns (Cytiva) in 10 mM HEPES, pH 7.2 or 7.4,
1055 100 mM NaCl (HEPES-buffered saline; HBS). Proteins used in SPR experiments were usually
1056 biotinylated using BirA ligase (Avidity, BirA500) and purified on size-exclusion columns to remove
1057 free biotin.

1058 **Surface Plasmon Resonance Experiments**

1059 SPR experiments were carried out using SA or CM5 sensor chips with Biacore T200 or 8K models
1060 (Cytiva). Proteins were immobilized in HBSp+ (10 mM HEPES, 150 mM NaCl, 0.05% Tween 20),
1061 pH 7.2 or 7.4 using biotin capture on SA chips, or random amine coupling on CM5 chips when
1062 proteins were not tagged with a biotinylation sequence. Experiments were performed at 25ºC, at
1063 a flow rate of 30 µl/min, with varying ligand immobilization levels (between ~ 500 RUs to 1200
1064 RUs), association and dissociation times, running buffer, and regeneration conditions. Details of
1065 the SPR experiments are listed in the **Table S6**. SPR sensorgrams and isotherms were plotted
1066 using Prism version 10.

1067 For NDNF-1–EGL-15 SPR experiments (Figure 5E), 60 mL of S2 culture (in Schneider's medium
1068 supplemented with Insect media supplement, Sigma I7267, and no serum) was transiently
1069 transfected with the Fc-tagged NDNF-1 (bait) expression plasmid, and conditioned medium was
1070 collected three days post-transfection. Medium containing NDNF-1-Fc was dialyzed overnight
1071 against HBS, pH 7.4, and was captured on Protein A SPR chips (Cytiva, 29127555) to validate
1072 the NDNF-EGL-15 interaction, using purified EGL-15 ectodomain as analyte. We observed slow
1073 dissociation of the NDNF-1-Fc from the SPR chip over several hours, resulting in a slow but
1074 noticeable baseline drift.

1075 Except in cases where kinetics were too slow, we used the 1:1 Langmuir binding model for fitting
1076 binding isotherms to acquire $K_D$ values in Prism version 10. With slow kinetics, we fit sensorgrams
1077 to kinetic models using Biacore's BIAEvaluation software to acquire rate constants; we indicated
1078 which kinetic model was used in the relevant figure legends.

1079 **Structure Determination for the ZIG-4–INS-6 complex**

1080 The ZIG-4–INS-6 complex was formed by co-expression using baculoviral expression of secreted
1081 ZIG-4 and INS-6. Both proteins were C-terminally tagged with hexahistidine tags, which allowed
1082 us to purify the complex from conditioned media using Ni-NTA Agarose resin. Complex was
1083 further purified over a Superdex 200 10/300 Increase (Cytiva) size-exclusion chromatography
1084 column in HBS. Purified complex was concentrated to 5 mg/ml, and screened for crystallization
1085 using the sitting-drop vapor-diffusion method at 22ºC.

1086 The ZIG-4–INS-6 complex crystallized in several conditions, which resulted in crystal structures

1087 resolved in three space groups. The tetragonal crystal form was grown in 0.1 M NaCl, 0.1 M
1088 sodium cacodylate, pH 6.5, 2 M $(NH_4)_2SO_4$. These crystals were cryo-protected in a solution
1089 containing 0.1 M NaCl, 0.1 M sodium cacodylate, pH 6.4, 1.6 M $(NH_4)_2SO_4$ and 24% glycerol, and
1090 diffracted to ~1.3 Å resolution. We also grew crystals The *C*-centered monoclinic crystals were
1091 grown in 0.2 M sodium/potassium phosphate, 0.1 M bis-tris propane, pH 6.8, 28% PEG 3350.
1092 These crystals were cryoprotected in 0.2 M sodium/potassium phosphate, 0.1 M bis-tris propane,
1093 pH 6.8, 28% PEG 3350, and 20% glycerol, and diffracted to ~2.3 Å resolution. Finally, the primitive
1094 monoclinic crystals were grown in 0.2 M sodium iodide, 0.1 M bis-tris propane, pH 6.5, 20% PEG
1095 3350. These crystals were cryoprotected in 0.2 M sodium citrate, 0.1 M bis-tris propane, pH 6.5,
1096 20% PEG 3350, and 25% glycerol, and diffracted to ~2.4 Å resolution.

1097 All crystallographic data were indexed, merged, and scaled using the *XDS* package.[97] Molecular
1098 replacement was performed with *PHASER*[98] in the *PHENIX* package[99] using PDB ID: 6FEY
1099 (ImpL2 + DILP5) as the molecular replacement model[36] with the tetragonal dataset. The refined
1100 higher-resolution model was used for molecular replacement with the other two datasets.
1101 *phenix.refine* was used to refine all models in reciprocal space and water placement,[100] and
1102 *COOT* was used for model building and corrections in real space.[101] Model building and
1103 refinement was guided by *MOLPROBITY* chemical validation tools within *PHENIX*.[102]

1104 The coordinates and structure factors for the ZIG-4–INS-6 complex are deposited at the Protein
1105 Data Bank with the following accession codes: 8TK9 (tetragonal form), 8TKT (C-centered
1106 monoclinic form), 8TKU (primitive monoclinic form). Crystallographic data collection and model
1107 refinement statistics are in **Table S5**.

1108

## FIGURE LEGENDS

**Figure 1. The *C. elegans* ectodomain collection and high-throughput interaction assay design.**

**A.** The distribution of secreted and membrane-anchored proteins in our *C. elegans* ectodomain collection.

**B, C.** The distribution of protein domains in the ectodomain collection.

**D, E.** The design of the ECIA pipeline and data analysis.

**F.** The expression levels of bait and prey proteins in the S2 cell culture media.

**Figure 2. Interactions and the community structure of the extracellular interaction.**

**A.** Network of moderate-confidence ($z_{min} > 8.4$) protein-protein interactions ($N = 185$) identified via the maximum-entropy method; proteins are colored by community grouping.

**B.** Network of community-community interactions. Only *connected communities* (communities with at least one protein making at least one interaction outside the community; $N = 14$) are shown.

**C.** List of proteins in each connected community.

**Figure 3. Axon guidance receptors and cues interact with each other outside the known cue-receptor axes.**

**A.** Schematic of interactions from the high-throughput assay, where line thickness is scaled to symmetrized MaxEnt *z*-scores (Figure 2). Interactions that were not previously known are shown with red lines. The numbers next to the lines indicate *z*-scores.

**B.** Interactions observed in the high-throughput assay are reproduced (see Figure S3 for more). As expected, interactions observed with a bait-prey (Fc-AP) pair are also observed in the reciprocal orientation, resulting in diagonally symmetrical assay results.

**C-G.** ECIA can be used to identify domains required for novel interactions: The first two IG domains of SAX-3 interact with the IPT domains 3 to 6 (C). SAX-3 IG domains also interact with the second domain of EFN-4, as observed by ECIA (D) and size-exclusion chromatography (SEC) (E). The interaction of MAB-20 with EFN-4 is mediated by the first domain (RBD) of EFN-4, as observed by ECIA (F) and SEC (G). SDS-polyacrylamide gels show the presence of both ectodomains in the complex fractions of SEC runs (E, G).

**H.** Summary of interactions between the domains of the axon guidance receptors and cues. Black arrows refer to interactions we observed but also previously characterized in other taxa.

**Figure 4. ZIG-2, -3, -4 and -5 make up a family of insulin-binding IgSF proteins.**

**A.** Pairwise sequence identities show that ZIG-2 to -5 are more closely related to each other and to *Drosophila* ImpL2 than the rest of the ZIG family.

**B.** ECIA for 22 insulin family members against ZIG-2 to -5. For expression levels of ZIGs and insulins, see Figures S4C, S4D.

**C.** ZIG-insulin interaction network inferred from the ECIA shown in B. The thickness of the connecting lines reflects the absorbance values in B for each ZIG-insulin interaction.

**D.** SPR sensorgrams for INS-1 and INS-6 binding to ZIG-5 and ZIG-4 immobilized on SPR chips. Kinetic fits with estimated on- and off-rates and equilibrium constants are shown in Figure S4E.

**E.** The structure of ZIG-4 (purple) bound to INS-6 (green) as observed in our tetragonal crystals.

**F.** The structure of the ZIG-4–INS-6 complex strongly resembles that of *Drosophila* ImpL2 bound to human IGF-1 (PDB ID: 6FF3).

**Figure 5. Networks of interactions with growth factor-like molecules and receptors.**
**A.** Network of interactions between cytokine- and growth factor-like molecules and receptors as observed in our screen (Figure 2). Line thickness corresponds to the symmetrized *z*-scores for each interaction. The numbers next to the lines indicate the *z*-score values.
**B.** SPR sensorgrams for the interaction of TRK-1 with immobilized ZK856.6 and negative control (hFGFR2 against ZK856.6).
**C.** SPR sensorgrams for the interaction of B0222.11 with immobilized HIR-1 and negative control (LET-756 against HIR-1). Kinetic fits and parameters are shown in Figure S5E.
**D.** SPR sensorgrams and binding isotherms for the interaction of EGL-15 with immobilized LET-756 in the presence or absence of 50 μg/mL (3.1 μM) heparin.
**E.** SPR sensorgrams for the interaction of EGL-15 with immobilized Fc-tagged NDNF-1 and negative control (HIR-1 against NDNF-1-Fc).

**Figure 6. RIG-5 interacts with molecules with neuronal and synaptic functions.**
**A.** Interactions of ZIG-8/Dpr and RIG-5/DIP according to our screen. Red and blue names represent orthologs in *D. melanogaster* and mammals, respectively. Line thickness scales with the symmetrized *z*-scores for each interaction. The numbers next to the lines indicate the *z*-score values.
**B-E.** SPR sensorgrams for interactions of RIG-5 with immobilized PTP-3 (B), NLR-1 (C), T19D12.6 (D), and HMR-1 (E).
**F.** Overlay of the ZIG-8–RIG-5 structure (PDB ID: 6ON9) (Cheng *et al*., 2019) and AlphaFold-multimer models for RIG-5 IG1-3+NLR-1 D6 and RIG-5 IG1-3+PTP-3 FN4-6. ZIG-8/Dpr binding to RIG-5/DIP does not overlap with PTP-3 and NLR-1, which share a binding site on the RIG-5 IG2 domain. See Figure S6 for details of the AlphaFold-predicted interfaces. Predicted aligned error (PAE) plots for AlphaFold predictions are shown for both RIG-5 (ECD)-PTP-3 (FN4-6) and RIG-5 (ECD)-NLR-1 (D6) complexes.
**G.** RIG-5 IG2 residues identified to interact with PTP-3 and NLR-1 in AlphaFold models.
**H.** ECIA for RIG-5 ECD mutants against ZIG-8, PTP-3 and NLR-1 ECDs. Mutations in RIG-5 IG2 that break PTP-3 and NLR-1 binding do not affect ZIG-8 binding, which is known to happen through the IG1 domain. The raw readout for the assay (absorbance at λ = 650 nm) is noted in each square. The homodimeric Rst interaction is a positive control.

**Figure 7. AlphaFold predictions of extracellular protein complexes and chemical synapses corresponding to PPIs.**
**A.** ipTM (interface predicted Template Modelling) score values of AlphaFold-predicted complexes from our interactome (blue) and random pairings of proteins (green).
**B.** Size of the protein complex does not correlate with ipTM values. ECIA hits with previously determined homologous structures (open blue circles) have the highest ipTM values on average.

## SUPPLEMENTAL INFORMATION

**Figure S1. Domain collection, protein expression and network analysis.**
**A.** Signal peptide and transmembrane helix analysis of the *C. elegans* proteome (WS252 release) shows that 23% of proteins have predicted signal peptides, and 44% of proteins are either membrane-anchored or secreted.
**B-E.** Expression testing of *D. melanogaster* (blue) and *C. elegans* (green) ectodomains in S2 cells using the Metallothionein (MT) and Actin 5C (Ac) promoters. Rst D1 refers to the fist immunoglobulin domain of Rst. For MT-driven expression, transiently transfected cells were induced with 0.8 mM $CuSO_4$ at 16 hours post-transfection. All transfections were collected 88 hours post-transfection for western blotting using a mouse primary anti-His antibody (1:2000) and an Alexa Fluor 488-coupled donkey anti-mouse IgG secondary antibody (1:5000). Overall, the Actin 5C promoter results in higher expression, but not in every case.
**F.** Network of 185 interactions detected with a cutoff of $z_{min} > 8.4$ drawn using the organic layout algorithm in Cytoscape, where node size relates to node degree (see the legend), and the edge thickness scales to $z_{min}$.
**G.** The degree distribution of all the interactions depicted in F.

**Figure S2. MaxEnt model to filter the experimental data.**
**A.** The normalized experimental data $An$.
**B.** The mean of the statistical background model $P$.
**C.** The difference between $An$ and $P$. PPIs with z-score above intermediate (orange) and stringent (purple) thresholds are shown in matrix form. Reciprocal PPIs are marked with dots (•) and non-reciprocal PPIs are marked with an '×'.
**D.** The reciprocal ratio of interactions as a function of the chosen threshold of *z*-scores. The maximum reciprocal ratio is achieved with $z = 12.2$.
**E.** The reciprocal ratio as a function of the number of unique edges identified. The shading represents $n \pm SE$, where *n* is the number of reciprocal edges. SE is calculated by the shot noise as $SE = \sqrt{n}$.

**Figure S3. Interactions of axon guidance receptors and cues.**
**A.** Image of the 384-well plate and absorbance at 650 nm for the ECIA experiment for selected axon guidance-related proteins in Figure 3B.
**B.** ECIA experiment for other guidance-related proteins. *D. melanogaster* Rst is a homodimeric protein and serves as a control.

**Figure S4. The ZIG-insulin interactome.**
**A.** Sequence alignment of four ZIGs and the fly ortholog, ImpL2. ZIG-2 to -5 carry a disulfide unique to all worm ZIGs.
**B.** The ECIA construct design where ZIGs are depicted as bait and insulins as prey, as used in the experiment presented in Figure 4B.
**C, D.** Expression of all insulin and ZIG constructs used in the experiment presented in Figure 4B. Expression of bait is shown in C and expression of prey in D.
**E.** Kinetic fitting of SPR sensorgrams from Figure 4D with parameters.
**F.** Superposition of three ZIG-4–INS-6 structures solved using three different crystal forms.

**G.** The INS-6–ZIG-4 complex is compatible with insulins interacting with the L1 domains + αCT helix in insulin receptors. hIR: human insulin receptor; PDB ID: 3W11.
**H.** Structure of the active T-like $IR_2$-$insulin_4$ structure from PDB ID: 6PXV. Four insulin-binding sites are shown in red, yellow, blue and pink.
**I, J.** Insulin-bound ZIG-4 would severely clash with dimeric IR, regardless of insulin binding to site 1 (I), or site 2 (J).

**Figure S5. Comparison of AlphaFold models of complexes discovered by the ECIA screen with the structure of human ligand-bound neurotrophin receptor.**
**A.** Structure of human neurotrophin receptor, TrkB (domain 5) bound to NT4/5 (PDB: 1HCF).
**B.** AlphaFold-predicted TRK-1 ectodomain bound to ZK856.6 at a 2:2 stoichiometry.
**C.** AlphaFold-predicted TRK-1 ectodomain bound to B0416.2 at a 2:2 stoichiometry.
**D.** PAE (Predicted Aligned Error) plots corresponding to models shown in B. and C. High ipTM (interface predicted Template Modelling) scores indicate high-confidence predictions.
**E.** Kinetic fitting of SPR sensorgrams collected for the binding of B0222.11 to HIR-1, shown in Figure 5C.

**Figure S6. Interfaces observed in AlphaFold models of RIG-5-NLR-1 and RIG-5-PTP-3 complexes.**
**A.** The AlphaFold-predicted interface of RIG-5 (ECD) bound to NLR-1 (D6).
**B.** The AlphaFold-predicted interface of RIG-5 (ECD) bound to PTP-3 (FN4-6). The RIG-5 residues mutated in the experiment presented in Figure 6H are shown in light cyan in A and B.

**Figure S7. Binding experiments for NLG-1–NRX-1 complex.**
**A.** SPR sensorgrams for soluble NRX-1 LNS6 domain binding to immobilized NLG-1 ECD.
**B.** Binding isotherm and $K_D$ for binding shown in A.
**C.** Size-exclusion chromatography runs for NRX-1 LNS-6 (orange), NLG-1 ECD (green) and the mixed sample (black).

**Table S1. Excel file containing even more data too large to fit in a PDF.**
Ectodomains used in the interactome study by gene, transcript and protein names, sequence, domain composition, signal peptide and membrane anchoring predictions. TM: transmembrane. Relative expression levels are measured and reported in columns P and Q for bait and prey constructs, respectively.

**Table S2. Excel file containing even more data too large to fit in a PDF.**
**A.** Symmetrized *z*-scores using the MaxEnt method.
**B.** Asymmetric *z*-scores using the MaxEnt method.

**Table S3. Excel file containing even more data too large to fit in a PDF.**
List of interactions observed in the high-throughout ECIA experiment using our new MaxEnt method with 2-hour absorbance measurements. Interactions with only one orientation with $z>3$ are labeled pink in column G. For comparison, scoring according to our old method, geometric mean of trimmed z-scores ($\sqrt{(z_1 \times z_2)_{old}}$) (Özkan, et al. *Cell*, 2013), are given in H, where a score of >20 was considered significant. Column I reports if the interaction or an orthologous one was

reported before, based on a literature search. Alphafold-multimer (Colabfold version 1.5.2) iPTM scores for a subset of interacting pairs are in column J.

**Table S4. Excel file containing even more data too large to fit in a PDF.**
**A.** Canonical neighbors for every ectodomain tested, protein/sequence names in top row in bold.
**B.** All neighbors for every ectodomain tested, protein/sequence names in top row in bold.

**Table S5.** Data and refinement statistics for x-ray crystallography of the ZIG-4–INS-6 complex.

**Table S6. Excel file containing even more data too large to fit in a PDF.**
Experimental details and parameters for all surface plasmon experiments included in the manuscript. Biacore chips are purchased from Cytiva. HBSp+: 10 mM HEPES, 150 mM NaCl, 0.05% Tween-20.

**Table S7. Excel file containing even more data too large to fit in a PDF.**
185 experimental PPIs based on the number of chemical synapses associated with each interaction. Interactions where there was no expression data for one or both of the binding partners are labeled N/A. We randomize the neuron connectome as a random control.

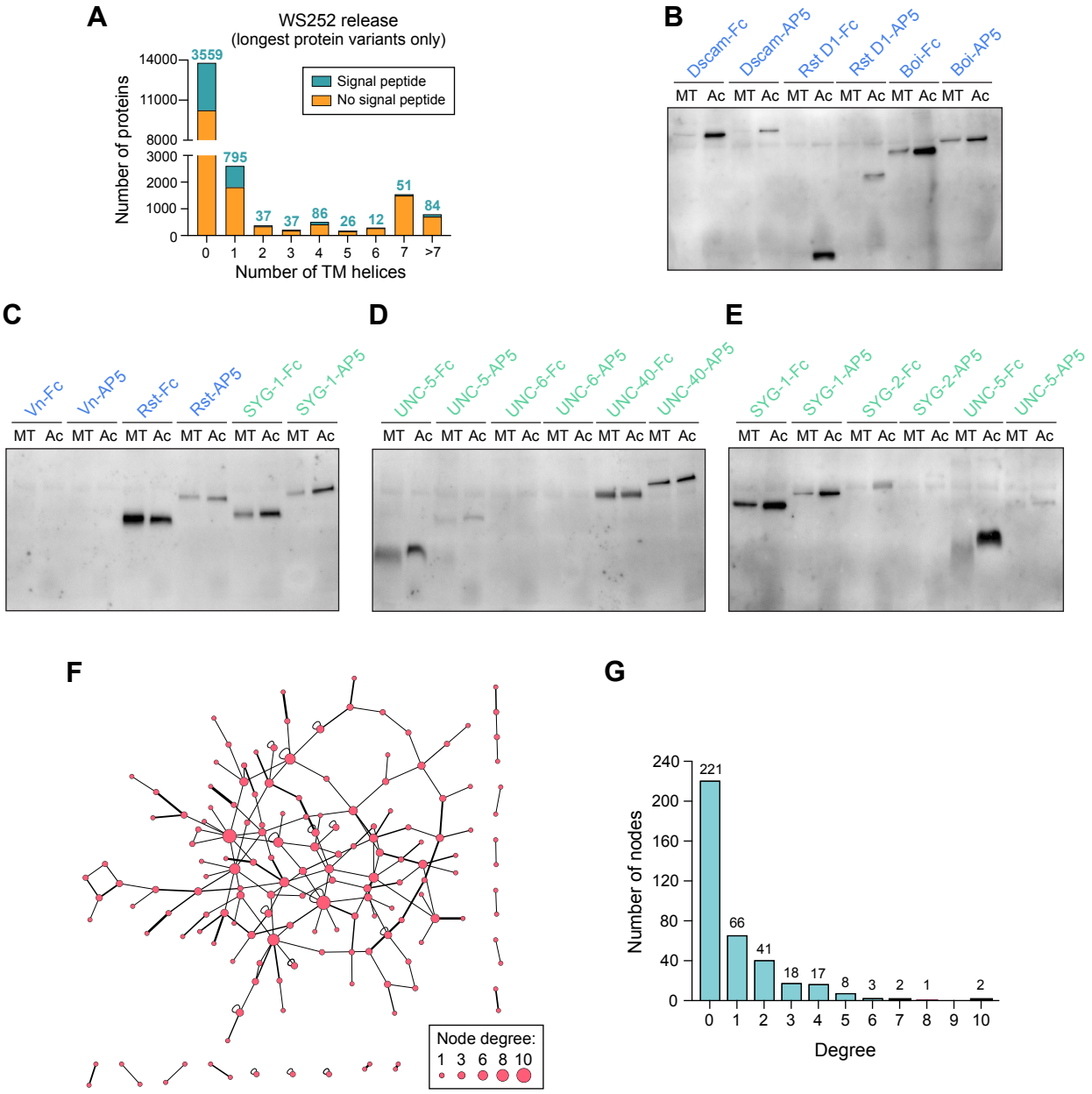



**Figure S1.** Domain collection, protein expression and network analysis.
**A.** Signal peptide and transmembrane helix analysis of the *C. elegans* proteome (WS252 release) shows that 23% of proteins have predicted signal peptides, and 44% of proteins are either membrane-anchored or secreted.
**B-E.** Expression testing of *D. melanogaster* (blue) and *C. elegans* (green) ectodomains in S2 cells using the Metallothionein (MT) and Actin 5C (Ac) promoters. Rst D1 refers to the fist immunoglobulin domain of Rst. For MT-driven expression, transiently transfected cells were induced with 0.8 mM $CuSO_4$ at 16 hours post-transfection. All transfections were collected 88 hours post-transfection for western blotting using a mouse primary anti-His antibody (1:2000) and an Alexa Fluor 488-coupled donkey anti-mouse IgG secondary antibody (1:5000). Overall, the Actin 5C promoter results in higher expression, but not in every case.
**F.** Network of 185 interactions detected with a cutoff of $z_{min} > 8.4$ drawn using the organic layout algorithm in Cytoscape, where node size relates to node degree (see the legend), and the edge thickness scales to $z_{min}$.
**G.** The degree distribution of all the interactions depicted in F.

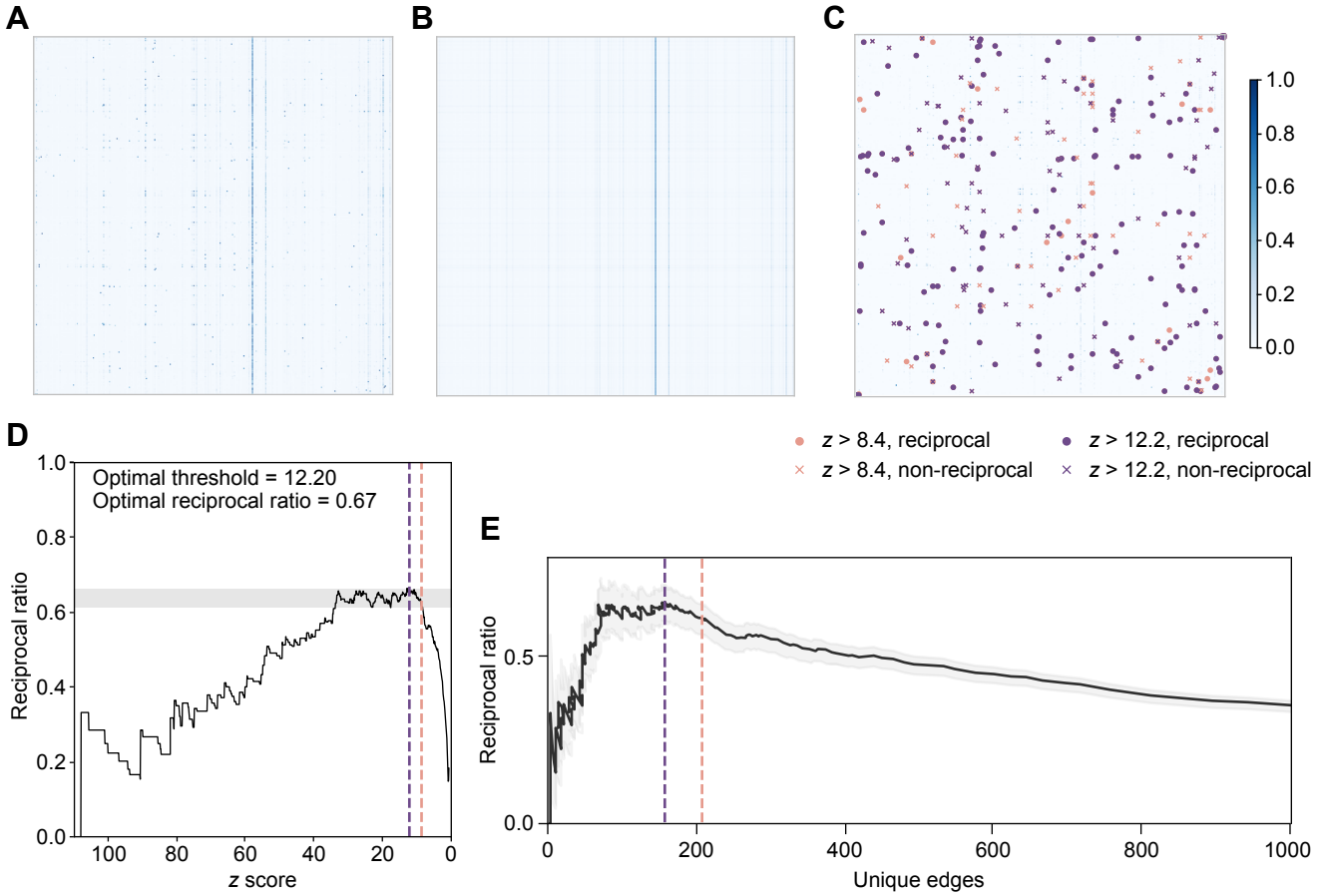
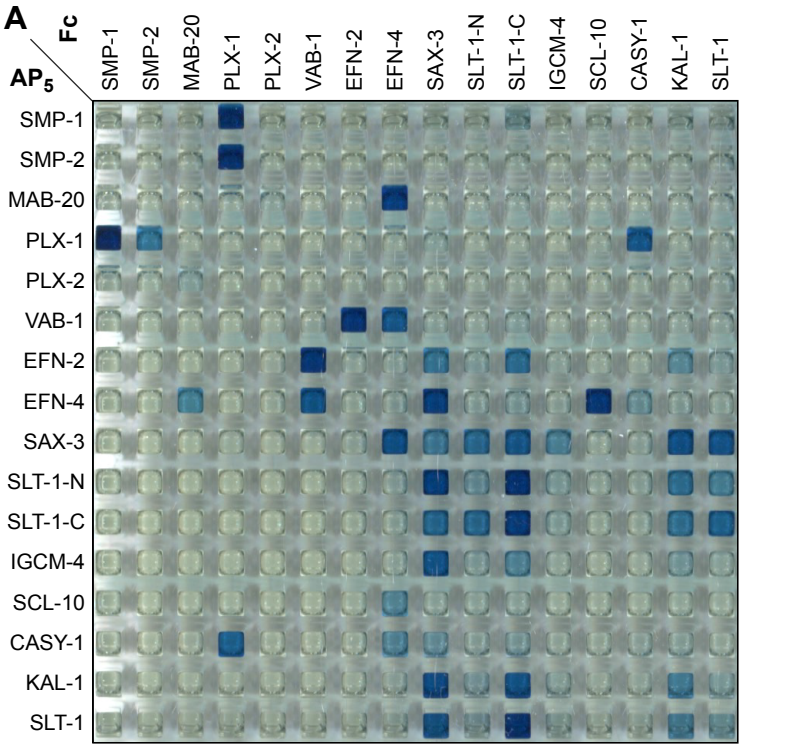
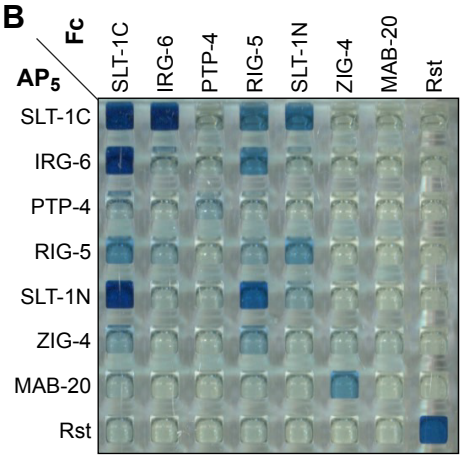



**Figure S2.** MaxEnt model to filter the experimental data.
**A.** The normalized experimental data $An$.
**B.** The mean of the statistical background model $P$.
**C.** The difference between $An$ and P. PPIs with *z*-score above intermediate (orange) and stringent (purple) thresholds are shown in matrix form. Reciprocal PPIs are marked with dots (•) and non-reciprocal PPIs are marked with an '×'.
**D.** The reciprocal ratio of interactions as a function of the chosen threshold of *z*-scores. The maximum reciprocal ratio is achieved with $z = 12.2$.
**E.** The reciprocal ratio as a function of the number of unique edges identified. The shading represents $n \pm SE$, where $n$ is the number of reciprocal edges. $SE$ is calculated by the shot noise as $SE = \sqrt{n}$.




| AP5 \ Fc | SLT-1C | IRG-6 | PTP-4 | RIG-5 | SLT-1N | ZIG-4 | MAB-20 | Rst |
|---|---|---|---|---|---|---|---|---|
| SLT-1C | 3.0353 | 3.4921 | 0.09 | 0.7481 | 1.0363 | 0.076 | 0.0666 | 0.0758 |
| IRG-6 | 3.5482 | 0.099 | 0.0761 | 0.7107 | 0.0992 | 0.0742 | 0.0608 | 0.0738 |
| PTP-4 | 0.1409 | 0.0767 | 0.2187 | 0.0947 | 0.0723 | 0.0606 | 0.0594 | 0.0746 |
| RIG-5 | 0.637 | 0.2263 | 0.0761 | 0.2351 | 0.5565 | 0.0765 | 0.061 | 0.0789 |
| SLT-1N | 3.3737 | 0.103 | 0.0813 | 1.6467 | 0.2212 | 0.0706 | 0.0586 | 0.0749 |
| ZIG-4 | 0.2604 | 0.085 | 0.0729 | 0.296 | 0.0838 | 0.1369 | 0.0937 | 0.0671 |
| MAB-20 | 0.144 | 0.0819 | 0.0647 | 0.113 | 0.0814 | 0.6008 | 0.0609 | 0.0719 |
| Rst | 0.1232 | 0.0762 | 0.0685 | 0.1056 | 0.0769 | 0.0641 | 0.0582 | 1.9455 |

Abs. at 650 nm: 0.5, 1.5, 2.5, 3.5

| AP5 \ Fc | SMP-1 | SMP-2 | MAB-20 | PLX-1 | PLX-2 | VAB-1 | EFN-2 | EFN-4 | SAX-3 | SLT-1-N | SLT-1-C | IGCM-4 | SCL-10 | CASY-1 | KAL-1 | SLT-1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SMP-1 | 0.0783 | 0.0655 | 0.0675 | 3.58 | 0.0629 | 0.065 | 0.0662 | 0.0645 | 0.1063 | 0.0934 | 0.2902 | 0.0788 | 0.1228 | 0.0687 | 0.1425 | 0.1149 |
| SMP-2 | 0.072 | 0.0629 | 0.0618 | 3.547 | 0.0633 | 0.0652 | 0.0668 | 0.0705 | 0.0746 | 0.0908 | 0.1065 | 0.075 | 0.0836 | 0.0751 | 0.0982 | 0.0888 |
| MAB-20 | 0.0784 | 0.0596 | 0.0609 | 0.0593 | 0.1293 | 0.0606 | 0.0635 | 2.894 | 0.0794 | 0.0883 | 0.1441 | 0.0807 | 0.0772 | 0.0675 | 0.1031 | 0.0859 |
| PLX-1 | 3.996 | 0.9262 | 0.0599 | 0.059 | 0.0599 | 0.0679 | 0.0669 | 0.0744 | 0.1684 | 0.0746 | 0.0842 | 0.07 | 0.0838 | 1.812 | 0.0882 | 0.0802 |
| PLX-2 | 0.0614 | 0.0543 | 0.1797 | 0.0541 | 0.0534 | 0.0567 | 0.0566 | 0.0608 | 0.0634 | 0.07 | 0.0845 | 0.0652 | 0.0584 | 0.0595 | 0.0795 | 0.0757 |
| VAB-1 | 0.0633 | 0.0574 | 0.0552 | 0.0564 | 0.0572 | 0.0643 | 3.447 | 1.692 | 0.0976 | 0.0858 | 0.1639 | 0.0692 | 0.0685 | 0.0761 | 0.0999 | 0.0879 |
| EFN-2 | 0.0743 | 0.0616 | 0.0646 | 0.0697 | 0.0615 | 3.281 | 0.068 | 0.0692 | 0.8144 | 0.1364 | 1.032 | 0.1114 | 0.0716 | 0.0674 | 0.4556 | 0.1763 |
| EFN-4 | 0.075 | 0.0605 | 0.7005 | 0.06 | 0.06 | 1.829 | 0.0654 | 0.0739 | 2.654 | 0.0843 | 0.156 | 0.0763 | 3.155 | 0.3077 | 0.1072 | 0.0991 |
| SAX-3 | 0.0657 | 0.0611 | 0.0613 | 0.1024 | 0.0596 | 0.0624 | 0.0832 | 1.863 | 0.6146 | 0.9786 | 1.548 | 0.5666 | 0.0698 | 0.0757 | 1.83 | 1.835 |
| SLT-1-N | 0.0809 | 0.0643 | 0.0637 | 0.0637 | 0.0628 | 0.0693 | 0.0803 | 0.0989 | 2.642 | 0.2194 | 3.435 | 0.1694 | 0.1128 | 0.0891 | 0.9466 | 0.6328 |
| SLT-1-C | 0.0724 | 0.0625 | 0.0629 | 0.0616 | 0.06 | 0.0637 | 0.0739 | 0.0754 | 1.086 | 0.9784 | 2.775 | 0.1545 | 0.0847 | 0.0711 | 0.8995 | 1.353 |
| IGCM-4 | 0.0698 | 0.0617 | 0.0615 | 0.0623 | 0.0635 | 0.0687 | 0.0706 | 0.079 | 1.798 | 0.1159 | 0.5123 | 0.093 | 0.0738 | 0.0725 | 0.2433 | 0.1293 |
| SCL-10 | 0.0871 | 0.0562 | 0.0587 | 0.057 | 0.0555 | 0.0571 | 0.0583 | 0.4851 | 0.066 | 0.0713 | 0.0883 | 0.0626 | 0.0726 | 0.0605 | 0.068 | 0.0689 |
| CASY-1 | 0.0741 | 0.06 | 0.0599 | 1.488 | 0.0619 | 0.0677 | 0.0688 | 0.5234 | 0.3353 | 0.0882 | 0.1917 | 0.0905 | 0.0917 | 0.1227 | 0.1331 | 0.0888 |
| KAL-1 | 0.0761 | 0.0654 | 0.0664 | 0.065 | 0.0627 | 0.065 | 0.0703 | 0.0744 | 2.068 | 0.1545 | 1.641 | 0.14 | 0.0762 | 0.0704 | 0.6798 | 0.2105 |
| SLT-1 | 0.0814 | 0.0723 | 0.0707 | 0.0801 | 0.0682 | 0.0737 | 0.0798 | 0.0834 | 1.645 | 0.1532 | 3.074 | 0.1366 | 0.0788 | 0.0757 | 0.632 | 0.3389 |

Abs. at 650 nm: 0.5, 1, 1.5, 2, 2.5, 3, 3.5

**Figure S3.** Interactions of axon guidance receptors and cues.
**A.** Image of the 384-well plate and absorbance at 650 nm for the ECIA experiment for selected axon guidance-related proteins in Figure 3B.
**B.** ECIA experiment for other guidance-related proteins. *D. melanogaster* Rst is a homodimeric protein and serves as a control.

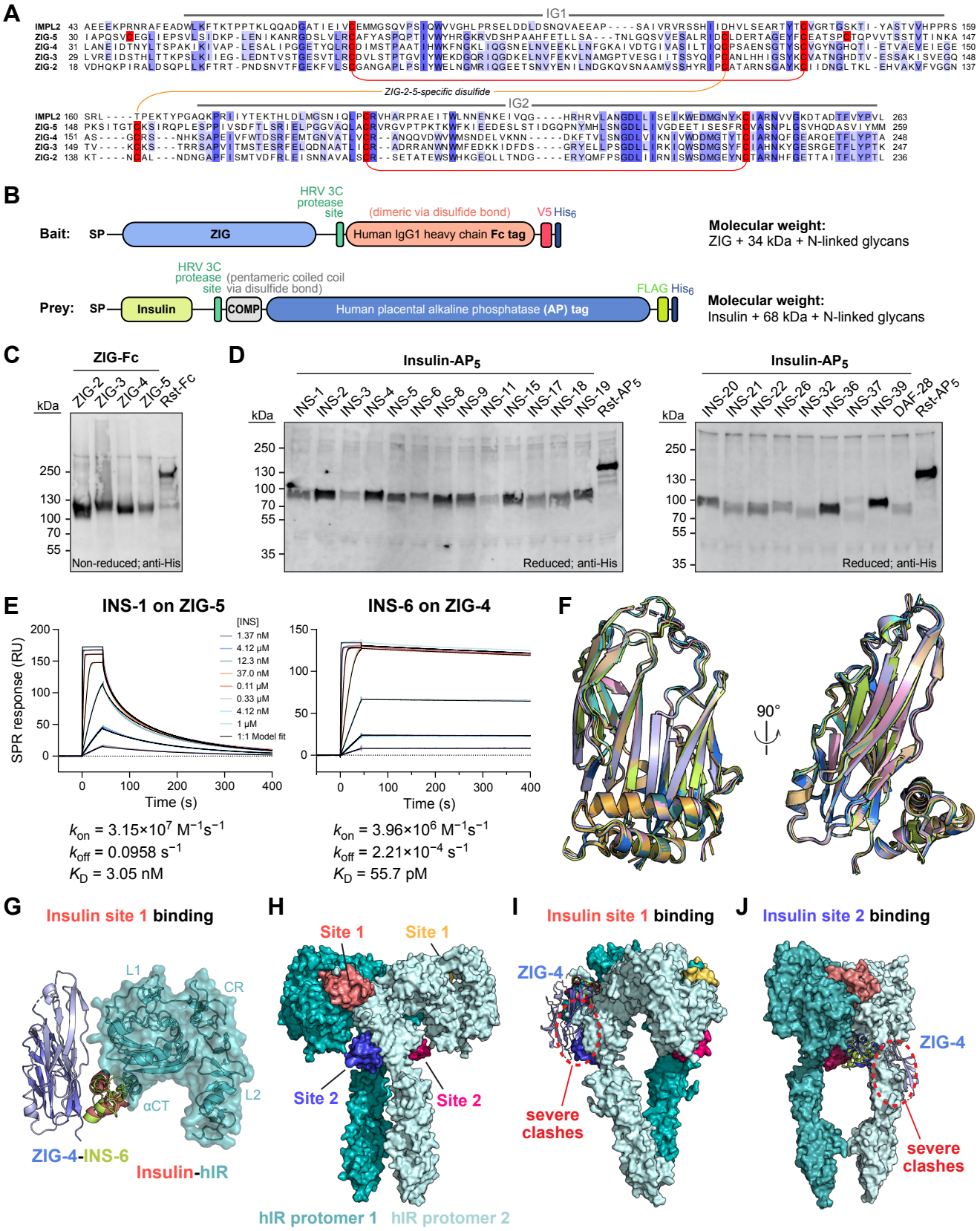



**Figure S4.** The ZIG-insulin interactome.
**A.** Sequence alignment of four ZIGs and the fly ortholog, ImpL2. ZIG-2 to -5 carry a disulfide unique to all worm ZIGs.
*(continued)*

*(continued)*
**B.** The ECIA construct design where ZIGs are depicted as bait and insulins as prey, as used in the experiment presented in Figure 4B.
**C, D.** Expression of all insulin and ZIG constructs used in the experiment presented in Figure 4B. Expression of bait is shown in C and expression of prey in D.
**E.** Kinetic fitting of SPR sensorgrams from Figure 4D with parameters.
**F.** Superposition of three ZIG-4–INS-6 structures solved using three different crystal forms.
**G.** The INS-6–ZIG-4 complex is compatible with insulins interacting with the L1 domains + αCT helix in insulin receptors. hIR: human insulin receptor; PDB ID: 3W11.
**H.** Structure of the active T-like $IR_2$-$insulin_4$ structure from PDB ID: 6PXV. Four insulin-binding sites are shown in red, yellow, blue and pink.
**I, J.** Insulin-bound ZIG-4 would severely clash with dimeric IR, regardless of insulin binding to site 1 (I), or site 2 (J).

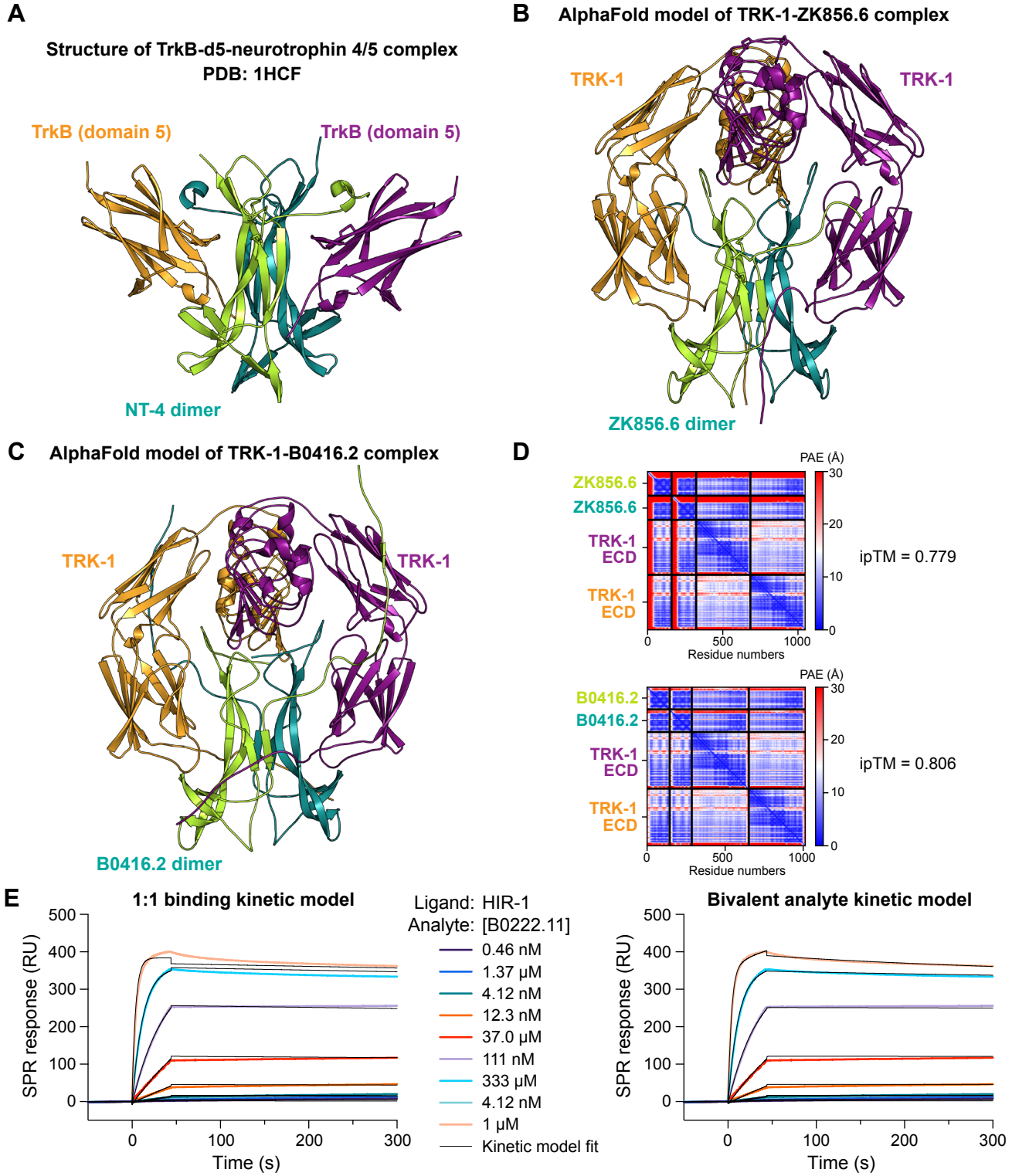



**Figure S5.** Comparison of AlphaFold models of complexes discovered by the ECIA screen with the structure of human ligand-bound neurotrophin receptor.
**A.** Structure of human neurotrophin receptor, TrkB (domain 5) bound to NT4/5 (PDB: 1HCF).
**B.** AlphaFold-predicted TRK-1 ectodomain bound to ZK856.6 at a 2:2 stoichiometry.
**C.** AlphaFold-predicted TRK-1 ectodomain bound to B0416.2 at a 2:2 stoichiometry.
**D.** PAE (Predicted Aligned Error) plots corresponding to models shown in B. and C. High ipTM (interface predicted Template Modelling) scores indicate high-confidence predictions.
**E.** Kinetic fitting of SPR sensorgrams collected for the binding of B0222.11 to HIR-1, shown in Figure 5C.

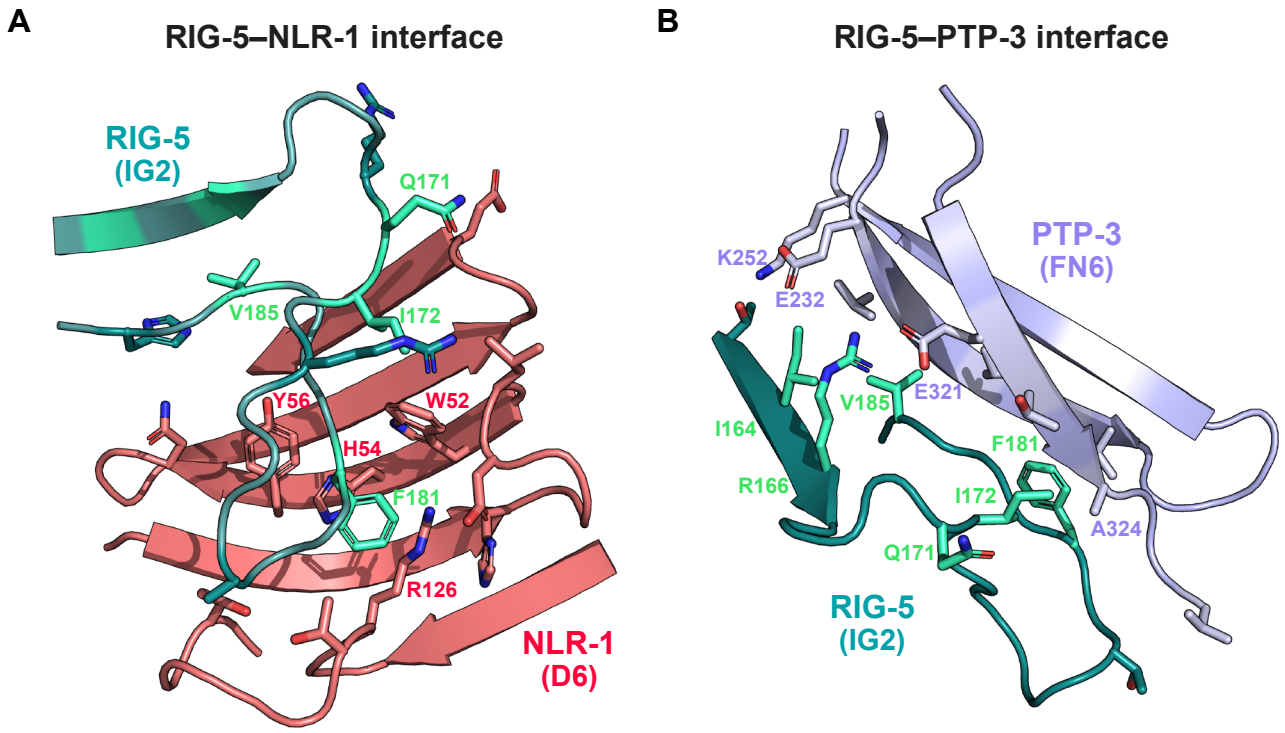



**Figure S6.** Interfaces observed in AlphaFold models of RIG-5-NLR-1 and RIG-5-PTP-3 complexes.
**A.** The AlphaFold-predicted interface of RIG-5 (ECD) bound to NLR-1 (D6).
**B.** The AlphaFold-predicted interface of RIG-5 (ECD) bound to PTP-3 (FN4-6). The RIG-5 residues mutated in the experiment presented in Figure 6H are shown in light cyan in A and B.

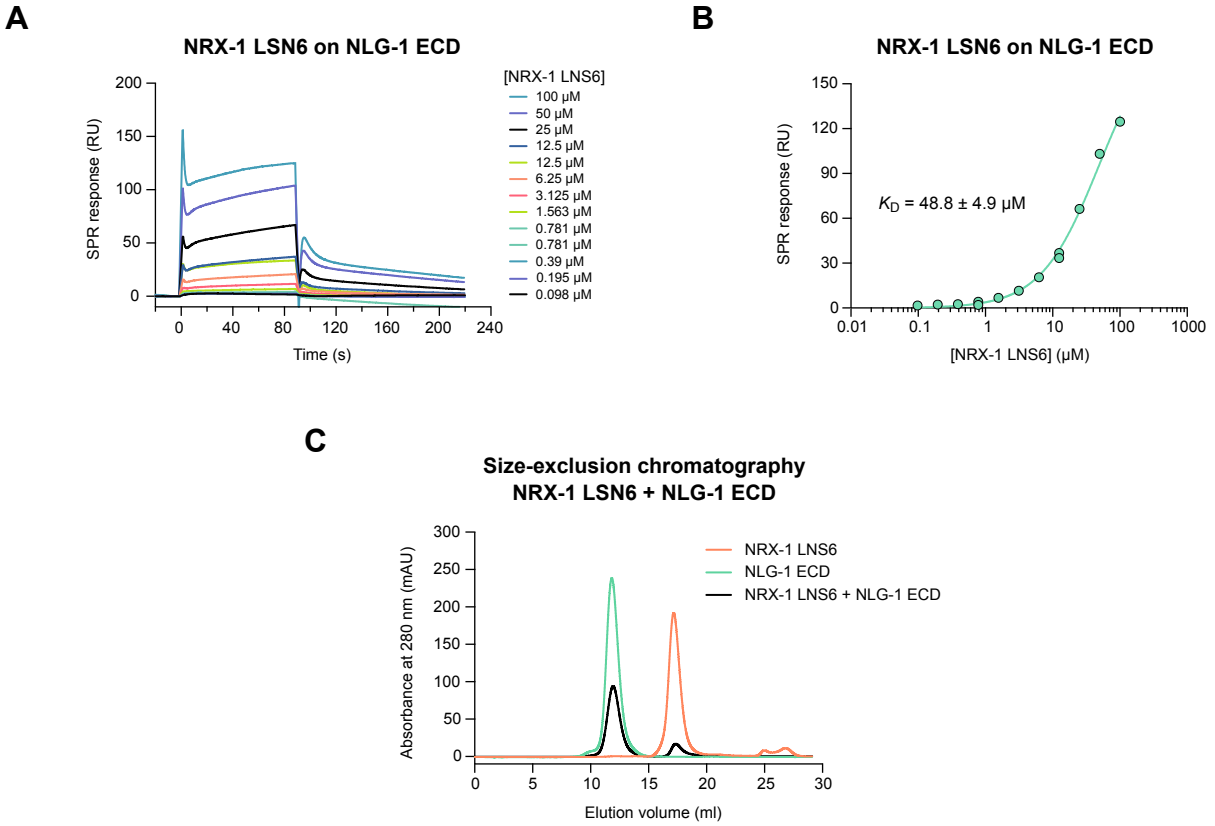



**Figure S7.** Binding experiments for NLG-1–NRX-1 complex.
**A.** SPR sensorgrams for soluble NRX-1 LNS6 domain binding to immobilized NLG-1 ECD.
**B.** Binding isotherm and $K_D$ for binding shown in A.
**C.** Size-exclusion chromatography runs for NRX-1 LNS-6 (orange), NLG-1 ECD (green) and the mixed sample (black).

**Table S5.** Data and refinement statistics for x-ray crystallography of the ZIG-4–INS-6 complex.

| | **Tetragonal form** | ***C*-centered monoclinic** | **Primitive monoclinic** |
|---|---|---|---|
| **Data Collection** | | | |
| Beamline | APS 24-ID-E | APS 24-ID-E | APS 24-ID-E |
| Space Group | $P4_22_12$ | $C2$ | $P2_1$ |
| *Cell Dimensions* | | | |
| $a$, $b$, $c$ (Å) | 74.528, 74.528, 107.058 | 166.430, 56.654, 73.685 | 73.839, 55.739, 149.945 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 90, 90 | 90, 113.247, 90 | 90, 93.775, 90 |
| Resolution (Å)* | 200-1.30 (1.38-1.30) | 200-2.30 (2.44-2.30) | 200-2.35 (2.49-2.35) |
| $R_{sym}$ (%) | 4.6 (120.1) | 7.1 (113.1) | 6.2 (124.8) |
| $<I>/<\sigma(I)>$ | 24.0 (1.78) | 9.7 (1.1) | 10.7 (0.9) |
| $CC_{1/2}$ | 0.999 (0.742) | 0.998 (0.652) | 0.998 (0.463) |
| Completeness (%) | 99.8 (99.4) | 98.4 (96.7) | 97.9 (97.1) |
| Redundancy | 12.8 (12.2) | 3.4 (3.3) | 3.1 (3.1) |
| **Refinement** | | | |
| Resolution (Å)* | 50-1.30 (1.32-1.30) | 53.13-2.30 (2.38-2.30) | 74.81-2.35 (2.60-2.50) |
| Reflections | 73,784 | 27,969 | 50,096 |
| $R_{cryst}$ (%) | 14.47 (27.19) | 21.52 (43.52) | 19.66 (38.72) |
| $R_{free}$ (%)** | 16.89 (31.58) | 25.18 (47.21) | 23.98 (44.33) |
| *Number of atoms* | | | |
| Protein | 2,128 | 3,878 | 7,755 |
| Ligand/Glycans | 11 | 0 | 0 |
| Water | 305 | 7 | 39 |
| *Average B-factors (Å$^2$)* | | | |
| All | 24.8 | 76.2 | 83.1 |
| Protein | 23.0 | 76.3 | 83.2 |
| Ligand | 36.3 | N/A | N/A |
| Solvent | 37.0 | 60.2 | 57.5 |
| *R.m.s. deviations from ideality* | | | |
| Bond Lengths (Å) | 0.012 | 0.003 | 0.008 |
| Bond Angles (°) | 1.291 | 0.631 | 1.002 |
| *Ramachandran plot* | | | |
| Favored (%) | 98.41 | 96.67 | 97.24 |
| Outliers (%) | 0.40 | 0.00 | 0.00 |
| Rotamer Outliers (%) | 0.43 | 0.48 | 0.39 |
| All-atom Clashscore‡ | 3.34 | 3.31 | 3.91 |

* The values in parentheses are for reflections in the highest resolution bin.
** 5% of reflections (3,747) for tetragonal crystals, 5% of reflections (1,373) for *C*-centered monoclinic crystals, and 4% of reflections (1,977) for primitive monoclinic crystals were not used during refinement for cross validation purposes.
‡ Clashscores were calculated by *phenix.refine* (*Phenix* version 1.21).
N/A: Not applicable.