DIP/Dpr interactions and the evolutionary design of specificity in protein families

Alina P. Sergeeva, Phinikoula S. Katsamba, Filip Cosmanescu, Joshua J. Brewer, Goran Ahlsen, Seetha Mannepalli, Lawrence Shapiro, Barry Honig

Affiliations

Department of Systems Biology, Columbia University, New York, NY, USA

Alina P. Sergeeva, Barry Honig

Zuckerman Mind, Brain and Behavior Institute, Columbia University, New York, NY, USA

Phinikoula S. Katsamba, Filip Cosmanescu, Goran Ahlsen, Seetha Mannepalli, Lawrence Shapiro, Barry Honig

Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA

Joshua J. Brewer, Lawrence Shapiro, Barry Honig

Department of Medicine, Columbia University, New York, NY, USA

Barry Honig

Abstract

Differential binding affinities among closely related protein family members underlie many biological phenomena, including cell-cell recognition. *Drosophila* DIP and Dpr proteins mediate neuronal targeting in the fly through highly specific protein-protein interactions. We show here that DIPs/Dprs segregate into seven specificity subgroups defined by binding preferences between their DIP and Dpr members. We then describe a novel sequence-, structure- and energy-based computational approach, combined with experimental binding affinity measurements, to reveal how specificity is coded on the canonical DIP/Dpr interface. We show that binding specificity of DIP/Dpr subgroups is controlled by “negative constraints”, which interfere with binding. To achieve specificity, each

27 subgroup utilizes a different combination of negative constraints, which are broadly distributed and
28 cover the majority of the protein-protein interface. We discuss the structural origins of negative
29 constraints, and potential general implications for the evolutionary origins of binding specificity in
30 multi-protein families.
31
32 **Introduction**
33 Over the course of evolution, gene duplications followed by sequence divergence have
34 generated numerous protein families whose homologous members have distinct binding
35 specificities. In many cases there are only subtle differences in the sequence and structure of closely
36 related family members, yet these can have profound functional consequences. Families of cell-cell
37 adhesion proteins offer many such examples where seemingly small changes in sequence generate
38 precise protein-protein interaction specificities. For example, invertebrate Dscam[1,2] and vertebrate
39 clustered protocadherin[3,4] neuronal "barcoding" proteins each display strict intra-family homophilic
40 specificity. While one Dscam (or Pcdh) isoform may have >90% sequence identity to another family
41 member, the few differences between them ensure that only homophilic recognition occurs. This is
42 critical to their function in the diversification of neuronal identities and self/non-self-discrimination[1-4].
43 In other cases, for example the type I and type II classical cadherins which pattern epithelia and
44 other tissue structures, specificity is less strict[5-7]. While classical cadherins also exhibit homophilic
45 binding, in addition they show strong heterophilic binding to other select family members[5-7]. In
46 another example, the nectin adhesion protein family encodes interactions between members that are
47 mainly heterophilic, contributing to the formation of a checkerboard pattern between two cell types,
48 each expressing a cognate nectin.[8] The specificity patterns of these protein families underlie their
49 biological functions, and are conserved in evolution. Understanding how specificity is coded on multi-
50 protein adhesion protein families is thus critical.
51 Here we carry out a comprehensive computational and experimental study of specificity
52 determinants in two interacting families of *Drosophila melanogaster* neuronal recognition proteins,

53 the 21-member Dpr (Defective proboscis extension response) and the 11-member DIP (Dpr
54 Interacting Proteins). These proteins have been extensively characterized structurally[9-11], and their
55 interactions were characterized quantitatively with biophysical measurements[11]. They thus offer an
56 ideal system to study the evolutionary design of specificity on protein-protein interfaces. DIPs and
57 Dprs are expressed in cell-specific patterns throughout the developing nervous system[12]. DIPs
58 preferentially bind Dprs, and a network of specific heterophilic interactions is formed between
59 members of the two families. This molecular binding network is correlated with synaptic specificity in
60 the fly retina, suggesting that DIP/Dpr interactions play an important role in neuronal patterning[9,13].
61 The extracellular regions of DIP and Dpr family members consist of three and two tandem Ig-
62 like domains, respectively[13]. Homodimerization is observed for some DIPs and Dprs, and homo- and
63 hetero-dimerization is mediated by an interface formed between the membrane-distal Ig1 domains
64 (Fig. 1A). Surface plasmon resonance (SPR) showed that members of both families have distinct
65 binding profiles, with DIPs and Dprs initially classified as forming four specificity subgroups.[11] In the
66 current work we extended the number of subgroups to seven based primarily on the strongest
67 heterophilic binding preferences but also on DIP/Dpr sequence similarity (see color-coded subgroup
68 assignment in Fig. 1B). Our DIP/Dpr grouping is somewhat different than that published by Cheng et
69 al.[10] due in part to the fact that these authors did not include DIP-κ and DIP-λ, whose binding
70 preferences had been previously mapped[11]. Additional differences could be due to the biophysical
71 approaches used to measure DIP/Dpr binding affinities in Cosmanescu et al.[11] and Cheng et al.[10]
72 (see Methods).
73 The Ig1 domains of the *Drosophila melanogaster* DIPs and Dprs have intra-family pairwise
74 sequence identities greater than about 50% and 40%, respectively, while the average identity
75 between individual DIPs and Dprs is about 30%. Binding interfaces for crystallographically
76 determined hetero-dimer structures are essentially identical – superimposing to within 1Å (Fig. 1C)[11].
77 The central question we address here is how DIPs and Dprs that are so closely related in sequence
78 and structure can exhibit such highly specific pairwise interactions. Previous studies have identified

79 specificity residues for select DIP/Dpr interactions[9-11]. Here, we analyze specificity for the family as a
80 whole. Our results reveal the central role of “negative constraints”, used here to denote an amino
81 acid in a cognate interface that interferes with binding to a non-cognate partner. The term negative
82 constraint has been used in the field of protein design[14-17] to denote a domain that must be designed
83 against, in effect an “anti-target”. In contrast, our use of the term here focuses on individual amino
84 acids rather than entire domains.
85

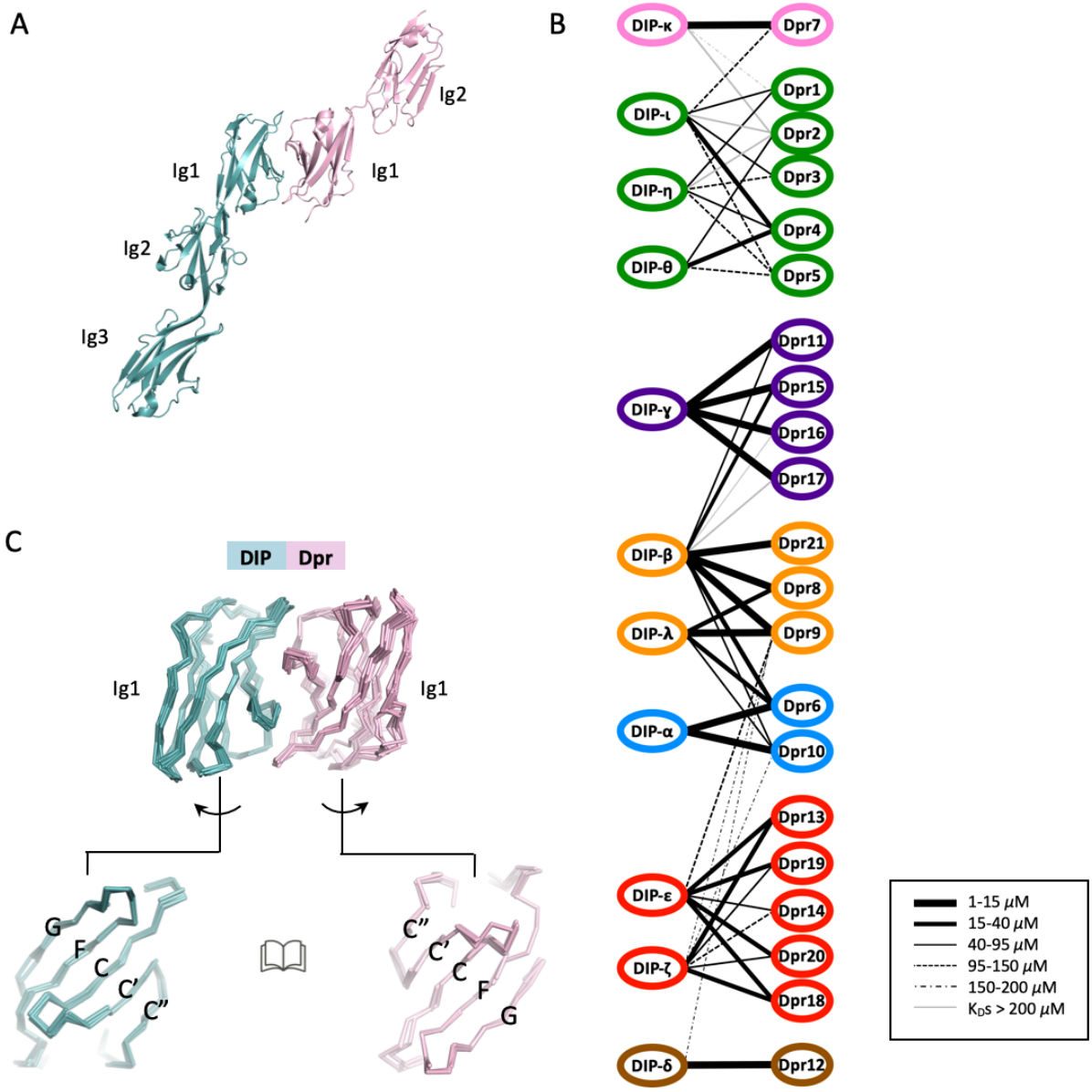


Fig. 1: Structure and interaction properties of DIPs and Dprs. (A) Ribbon representation of the DIP/Dpr heterodimer (PDBID: 6EG0)[11] – DIP shown in cyan, Dpr in pink. **(B)** Affinity-based binding interactome of DIPs and Dprs of *Drosophila melanogaster*. Line thickness indicates the K_D for a given pairwise interaction, as specified in the boxed inset. Groupings are color-coded based on their heterophilic binding preference. The interactome is based on previously published SPR data[11]. **(C)** Structural alignment of Ig1 domains in DIP/Dpr complexes (see Methods for PDBIDs). DIPs and Dprs interact via their CC'C"FG surfaces.

Since there are a total of forty-nine possible DIP/Dpr subfamily pairs and only seven bind strongly, there must be forty-two sets of negative constraints that preclude incorrect pairing. These are coded on a pseudo-symmetric Ig1-Ig1 interface of about 1900 $Å^2$ buried surface area and comprising 33 interfacial residues on the DIP side and 33 interfacial residues on the Dpr side. We

97 recognize that non-interfacial residues may also contribute to specificity but the primary
98 determinants will, in most cases, be part of the interface and these are the focus of the current work.
99 We begin by asking what can be learned from sequence alone and find that this information
100 is useful but incomplete. Our structure-based approach involves building homology models of
101 hypothetical complexes formed between all DIPs and Dprs and, using solved complexes where
102 available, to calculate the energetic consequences of mutating interfacial residues in one family
103 member to those of every other. We tested various programs that calculate binding free energy
104 changes ($\Delta\Delta$Gs) resulting from mutations and found that FoldX[18] offered the best combination of
105 accuracy and computational efficiency. Using FoldX, we identified negative constraints for every
106 DIP/Dpr pair and then confirmed a subset of our predictions using SPR.
107 Our results provide a detailed account of the design of complex specificity on a canonical
108 protein-protein interface. Most of the interface is used to create negative constraints with virtually all
109 expected energetic terms (steric hindrance, Coulombic repulsion, unsatisfied buried charges and/or
110 polar groups) used to weaken undesired complex formation between different subgroup pairs. Some
111 locations on the interface are used by multiple pairs while others are used in only a small number of
112 cases. In general, more than one negative constraint is required to significantly weaken binding. We
113 also discuss the trade-off between specificity and affinity, a phenomenon that appears essential for
114 the generation of multiple specificity groups[14,15,19]. Overall, our analysis reveals general principles
115 about the design of protein families whose members are very similar in sequence and structure and
116 yet exhibit exquisitely controlled binding affinities and specificities.
117
118 **Results**
119 Sequence analysis of DIPs and Dprs
120 Fig. 2 plots an alignment of sequence logos corresponding to interfacial residues of DIP and
121 Dpr *Insecta* orthologs (see Methods for details). Interfacial positions of DIPs and Dprs were assigned
122 based on non-zero change in accessible surface area per residue upon complex formation in

123 experimentally solved crystal structures. Interfacial positions of DIP/Dpr complexes of unknown
124 structure were inferred from multiple sequence alignment (residue correspondence with interfacial
125 positions of solved complexes). The family-wide numbering of interfacial positions (*i* = 1-33,
126 designated above the logos in Fig. 2) is used throughout the paper (see correspondence between
127 the family-wide numbering and Uniprot numbering (given throughout this paper in curly brackets) in
128 Supplementary Table 1).

129 Four positions shown in yellow are occupied almost exclusively by hydrophobic residues
130 which form the hydrophobic core of the DIP/Dpr interface[10] (Fig. 2A). Additionally, shown in purple is
131 a conserved Gln in both subfamilies and a conserved Asn in DIPs, which interlock the Ig1-Ig1
132 interface by forming buried sidechain-to-backbone hydrogen bonds (Fig. 2B). Residues that are
133 conserved across subfamilies cannot be responsible for subgroup specificity but changes in the
134 identity of individual conserved hydrophobic residues can play a role (see below). Rather, potential
135 specificity determinants will generally correspond to positions that are conserved in at least one
136 subfamily but differ in between subfamilies.

137 We tested the ability of a number of sequence-based methods to identify specificity
138 determinants in the seven specificity subgroups. These include GroupSim[20], SDPpred[21], SPEER[22]
139 and Multi-Harmony[23].

140

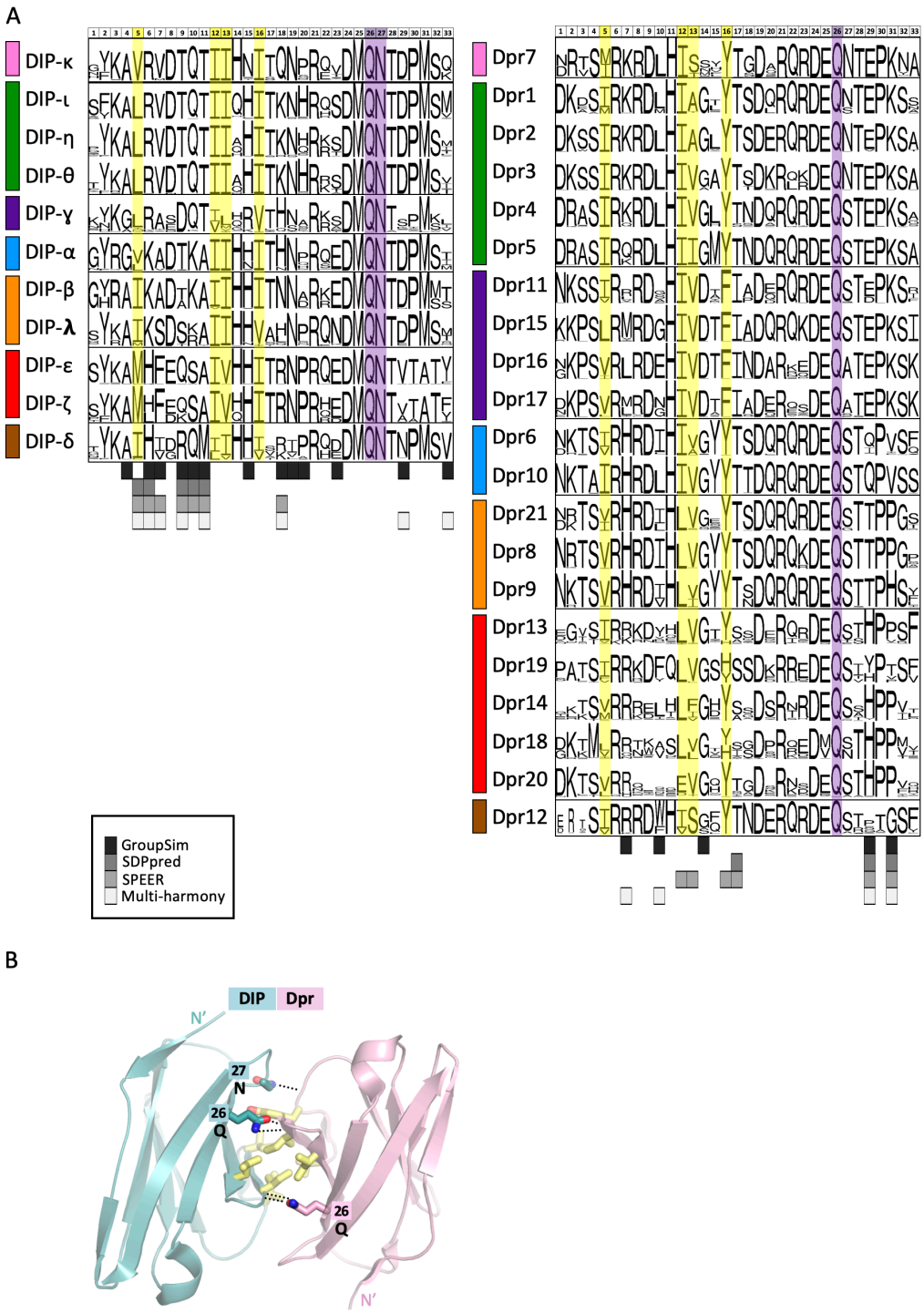


Fig. 2: Sequence-based predictions of DIP/Dpr specificity residues and common binding determinants. (A) Alignment of sequence logos corresponding to interfacial residues of DIP and Dpr Insecta orthologs. Predictions of specificity determining positions using sequence-based methods (GroupSim, SDPpred, SPEER, and Multi-Harmony) are given below logos in different shades of grey as specified in boxed inset. Bar on the left of logos indicates color of the DIP/Dpr subgroup (Fig. 1B). Family-wide numbering of interfacial positions is given above the alignment. The correspondence between protein specific (UniProt) and family-wide numbering is given in Supplementary Table 1.See Methods for details on sequence logo preparation. **(B)** Common binding determinants of DIP/Dpr dimers depicted on ribbon Ig1-Ig1 interfaces in sticks. Hydrophobic interfacial residues are in yellow. Conserved buried hydrogen bonds are depicted by dotted lines. Interfacial positions enclosed in boxes follow family-wide numbering and are annotated with identities of conserved amino-acids.

152 As input, we used a multiple sequence alignment of just interfacial residues with protein sequences
153 of 1732 DIP or 2570 Dpr orthologs segregated into seven specificity subgroups (see Fig. 2 and
154 Methods for details). Nine specificity determining positions were identified based on a consensus of
155 at least 3 out of 4 of these methods, and 14 additional positions, were predicted by at least one
156 method (see predictions below logos in grey, Fig. 2). These expand the number of predicted
157 specificity determining sites compared to our previous study[11] where we relied on visual inspection
158 of a multiple sequence alignment of DIP/Dpr sequences in *Drosophila melanogaster* alone. Below
159 we compare the predictions of these sequence-based methods to an energy-based analysis of DIP
160 and Dpr structures.
161
162 Identifying negative constraints
163 We hypothesized that negative constraints are the essential factor that defines DIP/Dpr
164 subgroups and developed a computational strategy to identify and characterize these putative
165 constraints. Since each subgroup (e.g. *i* and *j*) has both DIP and Dpr members, for purposes of
166 discussion we can define four sets of proteins, DIP-*i*, DIP-*j*, Dpr*i* and Dpr*j*. We need to explain why
167 members of set DIP-*i* don't bind to members of set Dpr*j* and, in turn, why members of set DIP-*j* don't
168 bind to members of set Dpr*i*. Since Dpr*i* binds to DIP-*i* while Dpr*j* does not, we first use FoldX
169 calculations to identify residues of Dpr*i* that weaken binding to DIP-*i* when mutated into a residue of
170 Dpr*j*. Thus, we identify negative constraints of Dpr*j* by studying their predicted effects on the binding
171 of mutated Dpr*i* to DIP-*i*. The experimental tests follow the same logic: the K_Ds of mutant Dpr*i*
172 binding to DIP-*i* are used to validate predicted negative constraints on Dpr*j*.
173 The overall procedure was to start with one or more structures of DIP/Dpr complexes for
174 each subgroup (the template complex) and; 1) to mutate every interfacial residue in the DIP
175 protomer to all residues appearing at the same position in DIPs of other specificity subgroups and
176 then 2) calculate ΔΔGs for the binding of each DIP to the Dpr in the complex. 3) Repeat the
177 procedure for the Dpr in the complex and then calculate ΔΔGs for its binding to each DIP.

178 Templates that were used include: *blue subgroup* - crystal structure of Dpr6/DIP-α and two
179 conformations of Dpr10/DIP-α[9,10]; *purple subgroup* – crystal structure of Dpr11/DIP-γ[10]; *green*
180 *subgroup* – crystal structures of Dpr4/DIP-η and for two conformations each for Dpr1/DIP-η and
181 Dpr2/DIP-θ)[10,11]; *pink subgroup* - homology model of Dpr7/DIP-κ; *orange subgroup* - homology
182 model of Dpr8/DIP-β; *brown subgroup* - homology model of Dpr12/DIP-δ; *red subgroup* -
183 homology model of Dpr13/DIP-ε. Overall, we predicted the effects 3044 mutations of 33 interfacial
184 residues of 13 template complexes on the DIP and Dpr side into residues occurring in DIPs and
185 Dprs of other subfamilies (see Supplementary excel file). Notably, the results depend heavily on the
186 use of homology models (see Methods for details of how they were constructed).
187 Negative constraints were defined as positions predicted to cause destabilizing effects for
188 *every* DIP or Dpr fly subgroup member in the context of binding to a member of another Dpr or DIP
189 subgroup, respectively. Our energy-based filter also requires that the effect is destabilizing for every
190 subfamily template complex used in the FoldX calculations (see Energy filter section in Methods for
191 details). In addition, we introduced a second filtering step, whereby we only defined a negative
192 constraint as a position that would also be predicted for other *Insecta* species so as to focus on
193 negative constraints that are evolutionarily conserved. The evolutionary filter assumes that
194 orthologs in the same subfamily conserve their interaction specificity (see Methods for data
195 supporting the assumption). To clarify, the energy-based filter is based on FoldX calculations of
196 structures of fly proteins whereas the evolutionary filter is based on comparison of ortholog
197 sequences of different species of insects. The evolutionary filter involved an analysis of interfacial
198 residues at corresponding positions of DIP-*i* and DIP-*j (*or Dpr*i* and Dpr*j)* in *Insecta* multiple
199 sequence alignments to see if the identities, size or biophysical properties of amino acids in all
200 members of subfamily "*i*" were different from those of subfamily "*j*" at the predicted position (see
201 Evolutionary filter section in Methods for details).

202 A subset of our predictions (see below) was tested experimentally. Overall, we identified
203 negative constraints for 42 combinations of non-interacting DIP and Dpr subgroups (see
204 Supplementary excel file).
205
206 Evaluation of methods to calculate ΔΔG
207 We tested six algorithms that calculate binding affinity changes upon mutation of individual
208 residues (ΔΔGs) based on reports of their performance in the literature[24-30]. These include FoldX,[18,31]
209 mCSM,[30] BeAtMusic,[27] MutaBind,[29] Rosetta flex ddG,[24] and BindProfX[32]. FoldX, mCSM, and
210 BeAtMusic require from seconds to few minutes of CPU time per single point mutation, while
211 MutaBind, Rosetta flex ddG, and BindProfX take more than an hour. Our evaluation tests were
212 based on 25 experimental ΔΔG values obtained via SPR or AUC measurements of DIP/Dpr and
213 DIP/DIP dimers. Of these, two were published[11] and twenty-three are new (see Supplementary Fig.
214 1 for supporting AUC and SPR data). Data on the performance of these methods is summarized in
215 Table S2.
216 As shown in Supplementary Table 2, FoldX, MutaBind, and Rosetta flex ΔΔG perform best
217 on our dataset as they feature the highest Pearson correlation coefficients (PCC). MutaBind is the
218 best performer but it fails to identify stabilizing mutations (shown in green, see Supplementary Table
219 2) - a failure of all machine learning methods we tested, which is probably reflective of an
220 overrepresentation of destabilizing mutations in training datasets. In addition, we preferred to use
221 methods based on physics-based force fields since these were easiest to interpret in structural
222 terms. Given its performance in terms of speed and accuracy (PCC=0.57, Supplementary Table 2),
223 we settled on FoldX as a computational guide for the identification of negative constraints.
224
225 Negative constraints between blue and purple DIP/Dpr subgroups
226 We discuss Dpr10/DIP-α(blue subgroup) and Dpr11/DIP-γ (purple subgroup) in detail since
227 these pairwise DIP/Dpr interactions have revealed clear associated phenotypes in the central and

228 peripheral nervous systems of *Drosophila melanogaster*[10,33,34]. We consider four sets of negative
229 constraints that weaken blue/purple inter-subgroup DIP/Dpr binding: a) On purple subgroup Dprs; b)
230 On blue subgroup DIP-α; c) On blue subgroup Dprs; and d) On purple subgroup DIP-γ.
231
232 a) *Negative constraints on purple subgroup Dprs that weaken binding to blue subgroup DIP-*
233 *α.* Using the protocol summarized above, we identify negative constraints on purple Dprs by
234 computationally mutating every interfacial amino acid on blue Dprs to those of purple Dprs and
235 calculating $\Delta\Delta$Gs for binding to DIP-α. The protocol is applied for every template complex of the blue
236 subfamily listed above (Supplementary Fig. 2). Fig. 3A summarizes results using one of the DIP-
237 α/Dpr10 crystal conformations (chains C and D of the asymmetric unit). The figure shows an
238 alignment of the four purple Dprs with blue Dpr10. Each column in the alignment is color-coded
239 based on the $\Delta\Delta$G for mutating a residue in Dpr10 to the aligned residue in a purple Dpr.

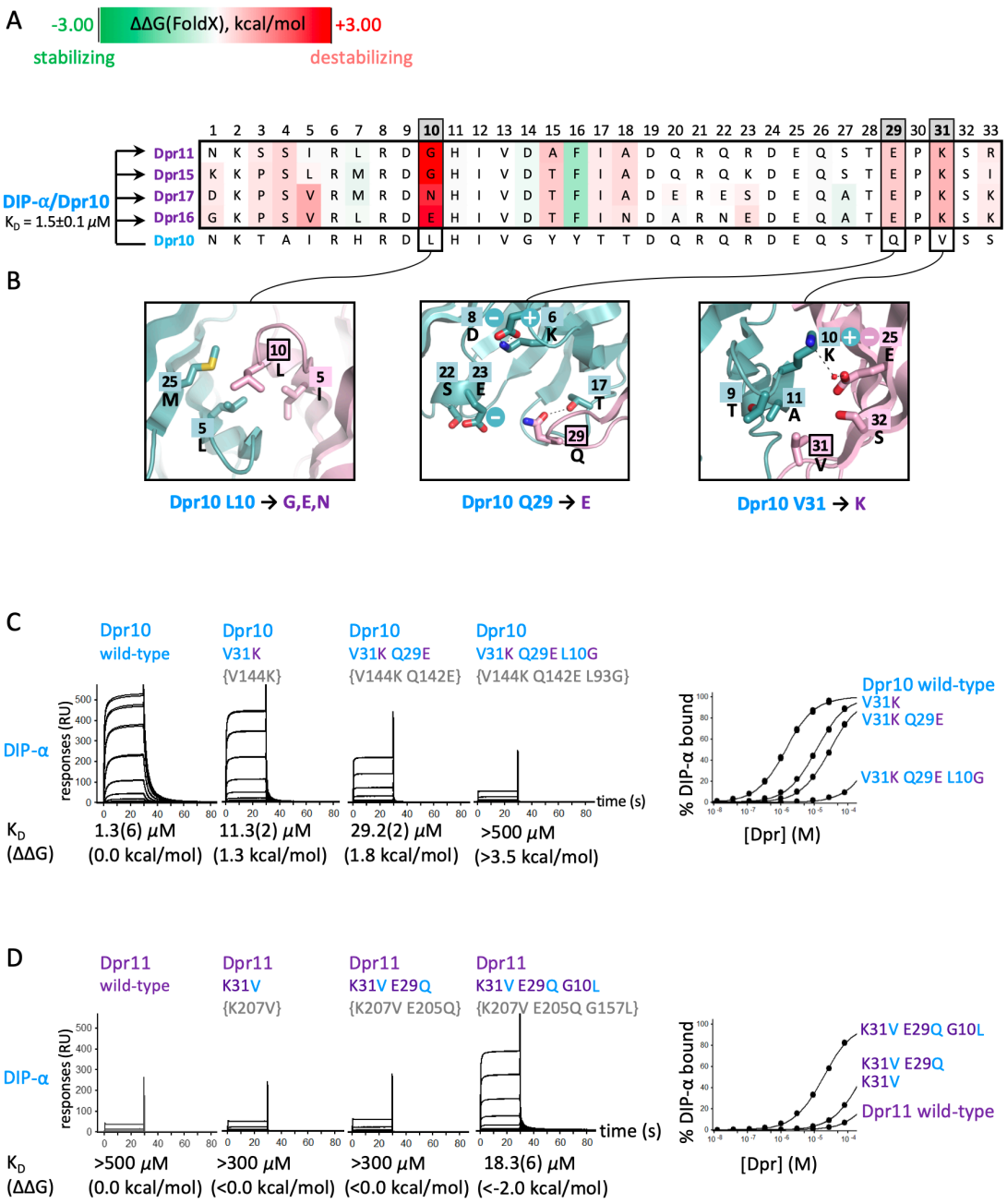


Fig. 3: Negative constraints on purple Dprs 11/15/16/17 which prevent binding to the blue DIP-α. (A) Alignment of interfacial residues of the four purple group Dprs and of Dpr10. Residues are color-coded based on the calculated ΔΔG, as specified in the color bar at the top of the figure. Predicted negative constraint positions are in grey. The average K_D and the experimental error for wild-type DIP-α/Dpr10 interaction based on six independent experiments is given on the left of the alignment. All calculations are carried out on the DIP-α/Dpr10 complex indicated to the left of the figure. The arrows indicate that residues in Dpr10 were computationally mutated to those of the four purple Dprs, as indicated. **(B)** Structural details of wild-type environment for each position where a mutation is predicted to cause a negative constraint (enclosed in a bold box). **(C) and (D).** SPR binding analysis of Dpr10 and Dpr11 mutants, representing negative constraints, to DIP-α. Each row shows SPR sensorgrams of Dpr analytes binding over a DIP-α immobilized surface. An overlay of the binding isotherms for each surface is shown to the right. K_D (ΔΔG) values for each DIP/Dpr interaction are listed below the SPR sensorgrams. For each K_D, the number in parenthesis represents the fitting error in the last significant figure, in μM for a single experiment with an expected experimental error up to 15%. Family-wide numbering of interfacial positions is used for all the mutants. Uniprot numbering is given in curly brackets.

255
256 We identified three interfacial positions, 10, 29, and 31 (Fig. 3A), as containing residues that
257 function as negative constraints. Notably, stabilizing mutations (shown in green) suggest that
258 DIP/Dpr complexes are not fully optimized for binding (see below). The destabilizing effect of the
259 different negative constraints can be understood in biophysical terms (Fig. 3B). Dpr10 has a
260 conserved hydrophobic residue at position 10 whose interaction with buried hydrophobic residues of
261 DIP-α contributes favorably to binding (first panel in Fig. 3B). When this interaction is lost via
262 mutation to glycine (Dpr11 and Dpr15), a negatively charged glutamate (Dpr16), or a polar
263 asparagine (Dpr17), the DIP/Dpr complexes are destabilized either through the creation of a cavity in
264 the hydrophobic interface or from the desolvation penalty associated with placing charged or polar
265 atoms in the hydrophobic environment. Two other negative constraints correspond to conserved
266 charged Glu and Lys at interfacial positions 29 and 31. These residues form salt bridges in the
267 purple DIP/Dpr subfamily, but would create unsatisfied charges in the context of binding to DIP-α.
268 We tested these predictions with SPR measurements of wild type and mutant proteins. At
269 least five independent measurements for each of three wild type interactions, Dpr10/DIP-α,
270 Dpr6/DIP-α, and Dpr11/DIP-γ, resulted in average K_Ds of 1.5±0.1 μM for (15% experimental error),
271 2.0±0.2 μM for (10% of experimental error) and 7.9±0.9μM for (11.4% experimental error)
272 respectively. A similar level of experimental error (10-15%) is expected for the K_Ds of all other Dpr or
273 DIP mutants determined in this study, where each concentration was tested in duplicate but in a
274 single experiment. Fitting errors and experimental errors are displayed in each figure containing SPR
275 data.
276 Fig. 3C displays SPR sensorgrams of Dpr10 wild-type and its mutants passed over a chip
277 immobilized with DIP-α. As can be seen in the figure, the V31K single mutant increases the K_D
278 relative to wild-type by about an order of magnitude; adding a second mutation, Q29E, further
279 increases the K_D by about a factor of 2 while adding a third mutation L10G results in a triple mutant
280 with a $K_D > 500$ μM. Thus, introducing Dpr11-like negative constraints to Dpr10 at these three

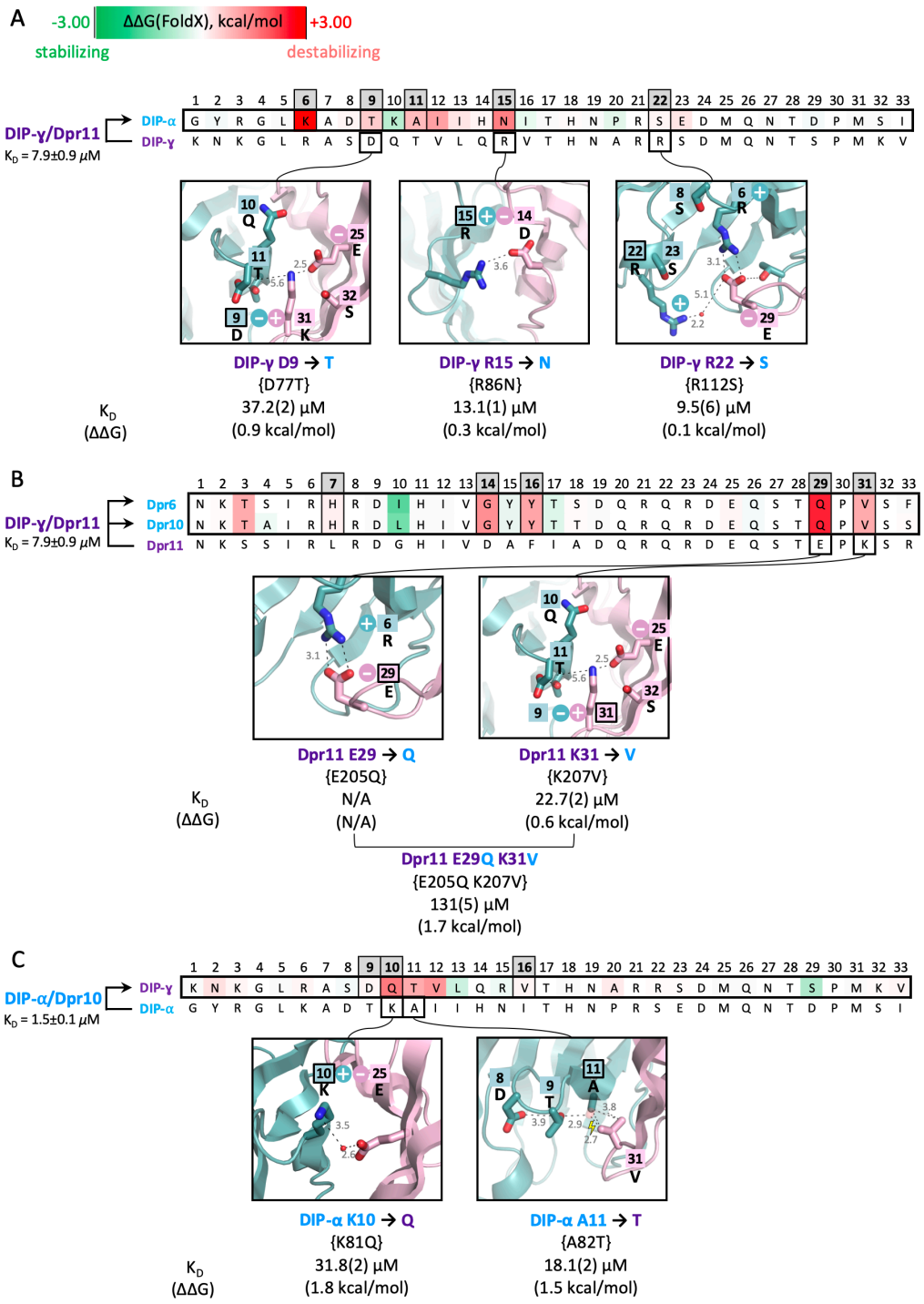
281 positions is sufficient to kill binding between Dpr10 and DIP-α. We then carried out a reverse set of
282 experiments aimed at removing negative constraints on Dpr11 to enhance its binding to DIP-α. As
283 can be seen in Fig. 3D, the K31V single mutant and the K31V,E29Q double mutant have very little
284 effect, while the triple Dpr11 mutant binds to DIP-α with a K_D of 18.3 μM – weaker than wild-type
285 Dpr10 but still quite strong. Overall, our results indicate that we have successfully identified the
286 negative constraints on the Dpr side of the interface that preclude binding of purple group Dprs to
287 blue group DIPs.
288
289 *b) Negative constraints on the blue subgroup DIP-α that weaken binding to purple subgroup*
290 *Dprs.* We identify positions 6,9,11,15,22 as containing negative constraints (Fig. 4A). SPR
291 measurements showed that the residues at positions 9, 15 and 22, all weakened the binding of DIP-γ
292 to Dpr11, although only position 9 had a substantial effect, increasing the K_D of DIP-γ/Dpr11 binding
293 from 7.9 to 37.2 μM (Supplementary Fig. 3A). As can be seen in Fig. 4A, there is a charged residue
294 at each of these positions in DIP-γ that forms an ion pair with a complementary charge on purple
295 group Dprs. Each of these positions contains a neutral polar residue in DIP-α which would then
296 disrupt the ion pair and leave an unsatisfied partially buried charge in a hypothetical complex with
297 purple group Dprs.
298
299 c) *Negative constraints on blue subgroup Dprs that weaken binding to purple subgroup DIP-γ.*
300 Positions 7,14,16,29,31 on Dpr6 and Dpr10 are predicted to contain residues that serve as negative
301 constraints (Fig. 4B). Positions 29 and 31 were tested and, as can be seen in the figure, the single
302 K31V mutant increased the K_D of wild-type Dpr11 binding to DIP-γ from 7.9 μM to 22.7 μM while the
303 K31V,E29Q double mutant with two negative constraints had a K_D of 131 μM, almost a factor of 20
304 greater than wild-type (Supplementary Fig. 3B). The source of the destabilization is evident from the
305 structure figures shows in Fig. 4B.The K to V mutation at position 31 would create a steric clash with

306 Thr at position 11 of DIP-γ and leave two unsatisfied negative charges in the DIP-γ/Dpr11 interface

307 (Asp9 of DIP-γ and Glu25 of Dprs), one of which will also be in a hypothetical complex with blue

308 Dprs (Glu25). The E29Q mutation breaks a salt bridge with Arg6 of DIP-γ leaving an unsatisfied

309 positive charge.



310

Fig. 4: Negative constraints on A) DIP-α, (B), blue Dprs 6/10, (C), DIP-γ that inhibit purple/blue inter-subgroup binding. A) Alignment of interfacial residues of DIP-α and DIP-γ followed by structural details used to explain negative constraint. The average K_D and the experimental error for wild-type DIP-γ/Dpr11 interaction based on five independent experimental measurements is given on the left of the alignment. Experimental K_Ds and ΔΔGs relative to wild type DIP-α/Dpr11 are shown below each insert. All symbols as in Fig. 3. B) Alignment of interfacial residues of Dpr6 and Dpr10 with Dpr11. All other details as in 4A. C) Alignment of interfacial residues of DIP-α and DIP-γ. All other details as in 4A. The supporting SPR data for the mutants presented in A, B, and C can be found in Supplementary Figs. 3A-C. For each K_D, the number in parenthesis represents the fitting error in the last significant figure, in μM for a single experiment with an expected experimental error up to 15%.

d) Negative constraints on the purple subgroup DIP-γ that weaken binding to blue subgroup Dprs. Residues at positions 9, 10 and 16 in DIP-γ are predicted to serve as negative constraints based on energy and evolutionary filters (Fig. 4C). Position 11 would also be predicted but only two of the three crystal structures used as templates yielded positive values for ΔΔG. The third corresponded to one of the two recently determined crystal conformations of DIP-α/Dpr10[10] reported after experimental validation of this position (Supplementary Fig. 3D). Mutations at Dpr10 positions 10 and 11, K10Q and A11T, were tested for binding to DIP-γ and each increase the K_D of binding to by about an order of magnitude, thus confirming the prediction. Again, the physical basis of the negative constraint is evident from the structure: K10 forms a water-mediated salt-bridge with sequence invariant Glu at position 25 of light blue Dprs that would be disrupted with a neutral Gln; The A to T mutation at position 11 of DIP-γ would introduce a steric clash with Val31 of light blue Dprs.

<u>Negative constraints between blue and green DIP/Dpr subgroups</u>

In a previous study, we tested mutants designed to switch the binding preference of the blue subgroup Dpr6 from its cognate DIP-α to the green subgroup DIP-η and, in turn, switch the binding preference of the green subgroup Dpr4 from DIP-η to DIP-α.[11] The relevant SPR data from that study are compared to our current FoldX predictions in Supplementary Table 3. As shown in the table, Dpr6 mutations at positions 7 and 31 weaken binding to DIP-α (Supplementary Table 3A) while reverse mutations at these positions in Dpr4 increase binding to DIP-α (Supplementary Table 3B). These results are in good qualitative agreement with the FoldX calculations and clearly indicate that positions 7 and 31 in blue subgroup Dprs serve as negative constraints that weaken binding to

342 green subgroup DIPs. On the other hand, the SPR data show that position 31 in green Dpr4, but not
343 position 7, contains a negative constraint that weakens binding to its cognate partner DIP-η
344 (Supplementary Table 3C) and, consistently, a mutation at position 31, but not 7, increases binding
345 to non-cognate blue DIP-α (Supplementary Table 3D). As can be seen in the table, these
346 experimental results are also in good agreement with FoldX calculations.
347
348 DIP/Dpr family overview
349 Fig. 5 displays negative constraints identified on the surface of hypothetical interfaces that
350 would be formed by a number of non-interacting subgroups. Most of the interface is used to create
351 negative constraints but different regions are exploited by different subgroups (Fig. 5). Overall, our
352 predictions implicate 21 positions on the Dpr side of an interface and 17 positions on the DIP side as
353 a source of negative constraints (see positions highlighted in grey, Fig. 6). These positions are buried
354 to different degrees in the interface, are located both in loops and secondary structure elements and
355 have different physical-chemical properties (hydrophobic, polar or charged). The majority of the
356 negative constraints reside in loop regions and code for polar or charged residues (21 positions) with
357 10 positions burying more than 50Å^2 of accessible surface area upon complex formation. Each non-
358 cognate DIP/Dpr subfamily pair was found to be destabilized by a set of about 7 interfacial residues
359 involving both sides of the interface. The physical source of the negative constraint can be shape non-
360 complementarity in the hydrophobic core (Fig. 2) (through steric clashes or cavity creation),
361 electrostatic effects (Coulombic repulsion or a desolvation penalty associated with burying unsatisfied
362 charges or polar residues), or both (see examples in Fig. 5). Notably, our SPR results suggest that
363 one negative constraint is not sufficient to ablate binding but, at least two are required to assure that
364 members of different subgroups do not form a stable complex.

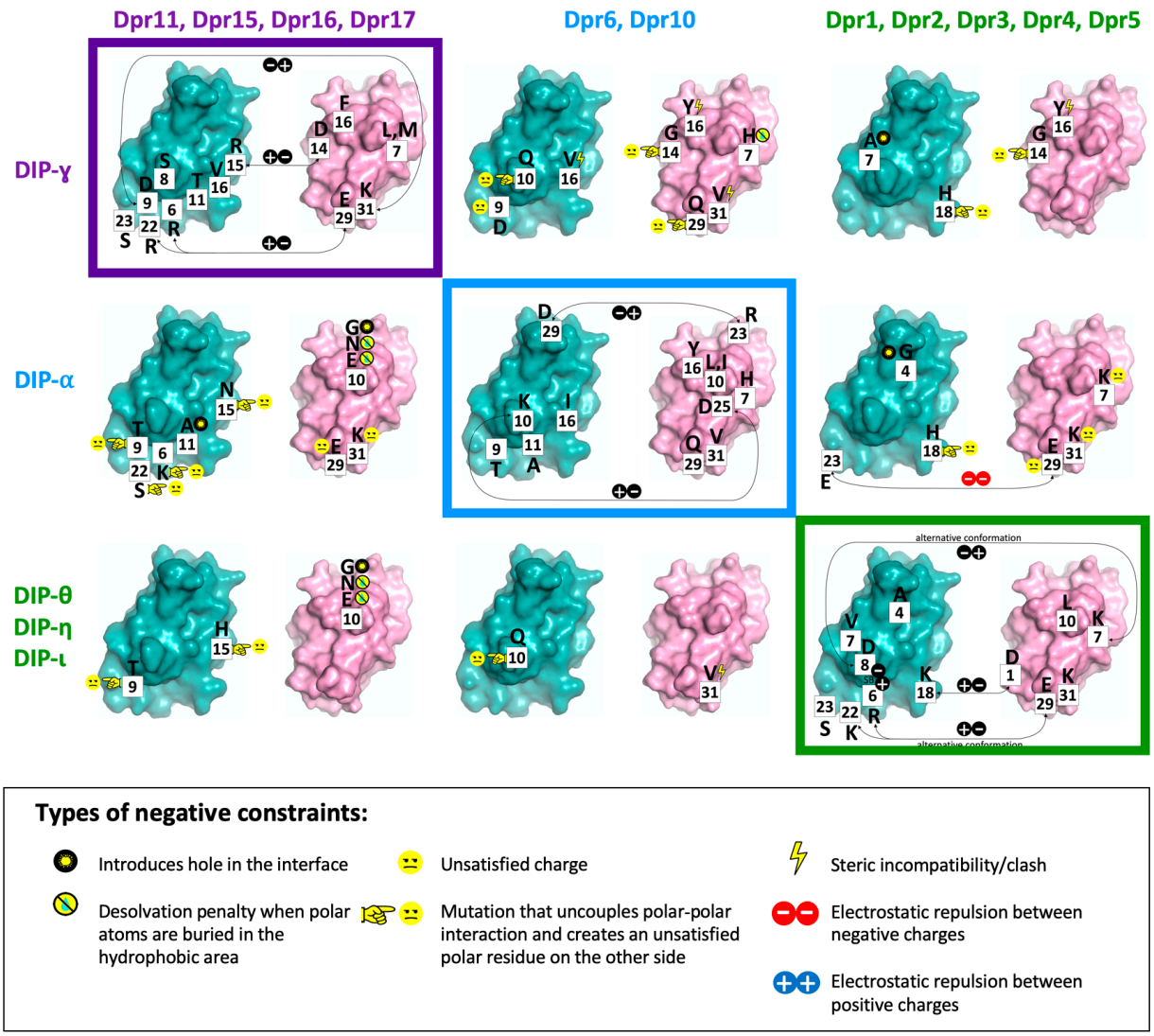


Fig. 5: Negative constraints in DIP/Dpr non-cognate pairs. Negative constraints in three DIP/Dpr subfamilies predicted using the protocol are shown on the DIP (in cyan) and Dpr (in pink) interacting surfaces of purple, blue, and green subfamilies using an open book representation, with cognate interacting pairs shown enclosed in boxes coded with subfamily color. Family-wide interfacial position number are in filled white boxes. Amino acid residue identity at a negative constraint position is given in a one letter code. The physical origin of each negative constraint is shown with a pictogram as detailed in the enclosed rectangle at bottom.

Our structure-based predictions of negative constraints include all positions predicted by sequence-based methods, but identify about twice as many positions. Sequence-based methods often

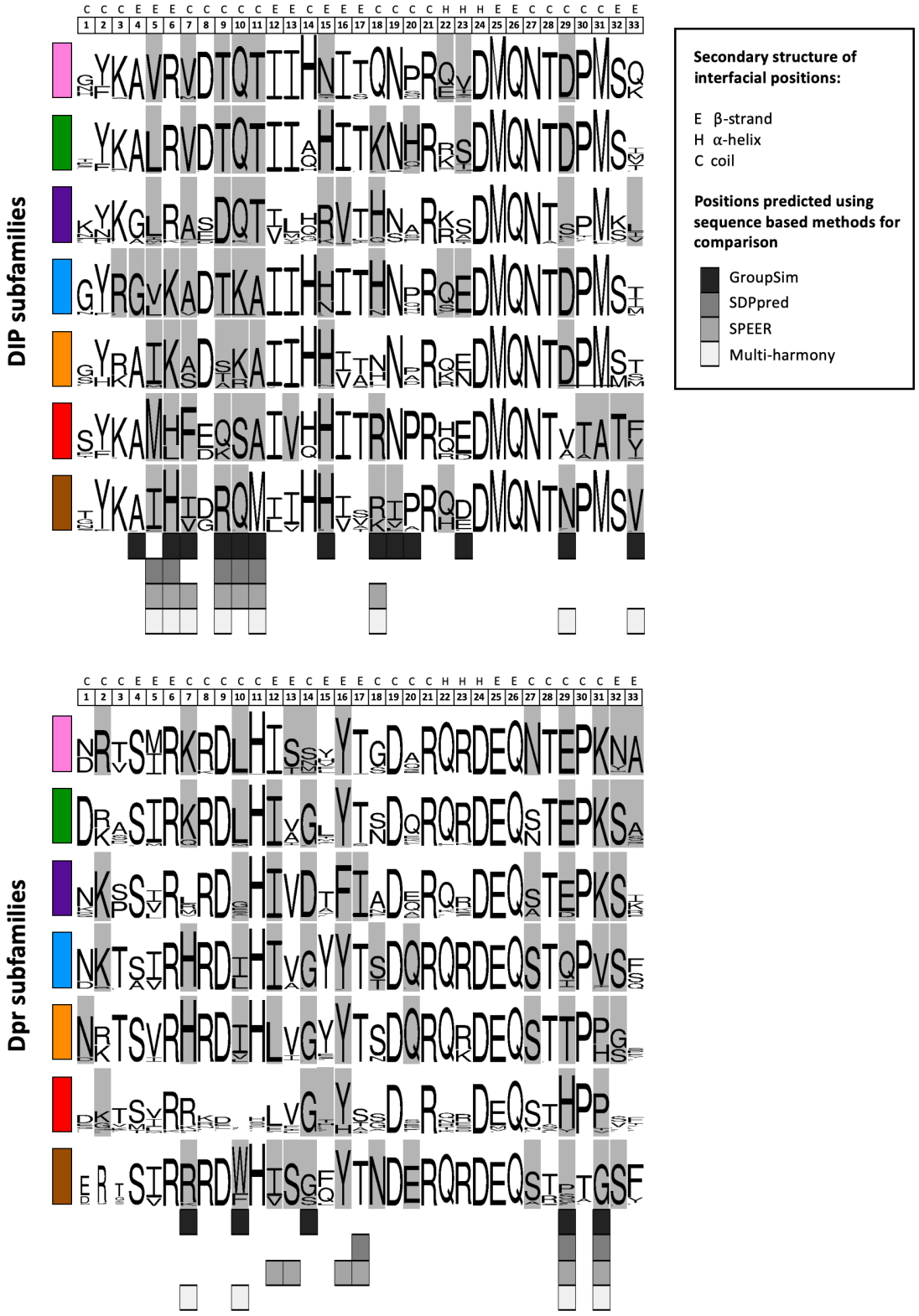


Fig. 6: Predicted positions of negative constraints DIP/Dpr subgroups. Logos created for Insecta with residues that function as negative constraints highlighted in grey. Subfamily color-code as in Fig. 1B. Each interfacial position (1-33) is annotated above logos with associated secondary structure, as detailed in the boxed legend. Sequence-based predictions of specificity determining positions using GroupSim, SDPpred, SPEER, and Multi-Harmony methods (same as in Fig. 2A) are given below logos as specified in boxed inset.

381
382 fail to identity positions not fully conserved within a subfamily[22,35]. We also note that the majority of
383 sequence-based methods encounter difficulties in identifying sites that are specific to only one
384 subfamily. For example, positions 14 and 16 correspond to conserved Asp and Phe residues in the
385 purple Dpr subfamily, while other Dprs have Gly and Tyr in these positions, respectively; see Fig. 2).
386 Similarly, we predict several residues in the hydrophobic core as potential negative constraints (at
387 positions 12, 13, 16 in Dprs and 5, 13, 16 in DIPs), whereas sequence-based methods predict only
388 one of these positions (Fig. 2). Of note, these positions were not among those we experimentally
389 confirmed in this study.

390 Balancing affinity and specificity

391 In some interfacial positions, FoldX predicts that mutations to residues found in other
392 subgroups would actually increase affinity (Figs. 3 and 4 and Supplementary excel file). As an
393 example, position 10 in purple Dprs, which is either a Gly or a polar residue, is predicted by FoldX to
394 disrupt nonpolar interactions with DIP s in other subgroups. SPR experiments confirmed this
395 prediction (Supplementary Fig. 4); L10G in Dpr10 decreased binding to DIP-α and DIP-β ($\Delta\Delta G$ >1.6
396 kcal mol^{-1}), Notably, G10L increased cognate binding between Dpr11 and DIP-γ by about the same
397 amount, indicating that some cognate binding affinity has been "sacrificed" so as to weaken non-
398 cognate binding.

399 In a second example, position 7 is predicted to be used by the green Dpr subfamily to
400 weaken non-cognate binding to DIP-α by placing an unsatisfied charge, K7, in non-cognate
401 interfaces (Fig. 5). Confirming this prediction, the H7K {H110K} mutation in Dpr6 weakens cognate
402 Dpr6/DIP-α binding by >1.9 kcal mol^{-1} [11]. In contrast, the K7H {K82H} mutation in Dpr4 strengthens
403 cognate binding to DIP-η by 0.3 kcal mol^{-1} [11], a small effect but outside the error bars of our
404 measurements. This is another case where cognate binding appears designed to be suboptimal.
405

406 **Discussion**

407 We have described the structural and energetic origins of the partition of DIPs and Dprs into
408 orthogonal specificity groups defined by SPR-derived binding affinity measurements. We previously
409 analyzed specificity determinants in type II cadherins[5], nectins[8] and DIPs and Dprs[11], primarily
410 through visual inspection of sequences guided by structural data. Here we have adopted a far more
411 extensive and quantitative approach as required by the complexity of the problem we set out to
412 address. Specifically, how specificity is coded on seven subgroups of DIPs and Dprs so that out of
413 the 49 possible subgroup pairings of closely related subfamilies of DIPs and Dprs, only seven form
414 strong pairwise interactions. The problem is complicated by the fact that each subgroup contains
415 between 1 and 5 members on the Dpr side and 1 and 3 members on the DIP side. Asking how 42
416 combinations of DIP and Dpr subfamilies are designed not to interact, or in some cases to interact
417 weakly, poses a unique set of challenges. Notably, sequence-based methods were able to provide
418 only partial answers to this question.
419 The approach we have adopted is essentially to build a large set of homology models of
420 complexes that form and do not form and to ask, using FoldX calculations, SPR measurements and
421 visual inspection what is wrong with those that do not form. Overall, we identified 38 positions which
422 are used for negative constraints and carried out 23 validation tests with SPR at 8 of those positions
423 (more than one mutation was tested at each position). In addition, our predictions are in agreement
424 with previously published SPR experiments on the blue and green subfamilies,[11] which validate
425 three residues that serve as negative constraints and confirm one true negative prediction that a
426 residue does not serve as a negative constraint based on twelve SPR measurements listed in
427 Supplementary Table 3.
428 The negative constraints we identified for the most part exploit strong localized unfavorable
429 energetic contributions rather, for example, than more subtle effects that are distributed over many
430 residues (as is the case for type I cadherins where entropic effects involving the movement of entire
431 domains play a role in determining dimerization affinities)[7]. The physical origins of negative
432 constraints include: replacing a charge in an ion pair with one of opposite sign leading to Coulombic

433 repulsion; replacing one member of an ion pair with a neutral group, which has the effect of burying
434 an unsatisfied charge; mutations in the hydrophobic core to larger amino acids that create steric
435 clashes or to smaller ones that create cavities which lead to packing defects and to weaker
436 hydrophobic contributions to binding. Of course, the same "trick" cannot be reused for 42 different
437 subgroup pairs. In fact, every set of negative constraints on the DIP and Dpr side of the 42 non-
438 cognate pairs is unique in the sense that the same set is not reused (see Supplementary Table 4).
439 This is why as discussed above, most of the interface needs to be exploited. Notably, in cases we
440 tested,[11] at least three negative constraints had to be added or removed to essentially switch the
441 specificity of a particular protein from one subgroup to another. As can be seen in Fig. 1B, intra-
442 subgroup K_Ds are generally in the 10 μM range (although some are as weak as 40 μM) while inter-
443 subgroup K_Ds (with a few exceptions) are undetectable, implying a K_D > 500 μM. This suggests that
444 the generation of a new DIP/Dpr specificity requires a change in binding free energy of ~2.5 kcal
445 mol^{-1}, a value that is difficult (but not impossible) to reach with one or two mutations.
446 How is the DIP/Dpr interface designed to achieve affinity and specificity? First, we note that,
447 in common with most cell-cell recognition proteins, DIP/Dpr affinities are in the μM range, a
448 thousand-fold weaker, for example, than many antibody-antigen complexes with substantially
449 smaller binding interfaces. That lower affinities are a general feature of adhesion receptors, even
450 those which are not members of large families, suggest that there are other factors involved that
451 have little to do with negative constraints, for example cellular motility. Moreover, cell-cell avidity
452 involves the interactions of multiple receptors so that there may not be a need to evolve proteins with
453 nanomolar affinities. This may account for the observation above that even cognate interfaces are
454 not optimized for affinity. Moreover, a trade-off between affinity and specificity is to be expected
455 in large protein families with only limited sequence divergence. If an interaction between two
456 family members, say '*a*' and '*b*', is very strong, then it is more of a challenge, for example, to
457 generate an isoform '*c*' that is similar in sequence to '*b*', but interacts very weakly with '*a*'. Thus,
458 the higher the specificity requirements, the greater the constraints on absolute affinities. That

459 some residues in the interface are not optimized for cognate binding (see above) may then
460 reflect the need to ensure that binding is not too tight.

461 We find that more than a half of the DIP/Dpr interface (38 of 66 interfacial residues) is utilized
462 for negative constraints across different insect species. However, these only play a role in the
463 context of non-canonical subgroup-subgroup interactions, whereas in any given cognate DIP/Dpr
464 complex, a majority of the interfacial residues likely play a stabilizing role. One can imagine that an
465 ancestral DIP/Dpr complex relied entirely on stabilizing interactions in the hydrophobic core along
466 with additional stabilizing or energy neutral interactions in the periphery. As multiple family members
467 evolved the need for specificity was satisfied through the introduction of negative constraints
468 throughout the interface. This evolutionary strategy is likely used by other large protein families. Of
469 note, it has been found that specificity can be achieved through the insertion and/or deletion of
470 loops[36] but this strategy does not appear to have been used by DIPs and Dprs.

471 Detailed binding affinity measurements have now been reported for a number of families of
472 adhesion receptors. In some cases, for example type II cadherins,[5] the family can be divided into
473 subgroups whose members exhibit considerable promiscuity in their intra-subgroup binding
474 properties while not binding to members of other subgroups. This behavior is reminiscent of the
475 subgroup characteristics of DIPs and Dprs although, since there are fewer cadherin specificity
476 groups, their evolutionary design is less of a challenge. In both cases binding promiscuity may well
477 be a source of functional redundancy but this has not been established. In contrast, Dscams[1,2] and
478 clustered protocadherins[3,4] are strictly homophilic which, as mentioned above, is an absolute
479 requirement for their role of establishing a unique identity for every neuron. In contrast to DIPs and
480 Dprs, the exquisite specificity of these protein families is established through multi-domain interfaces
481 that allow for a greater degree of fine-tuning[2,37-41]. This could result from simply a larger surface area
482 on which to place negative constraints and/or from constraints[39] imposed by a multi-domain
483 structure. Negative constraints are obviously a central component of evolutionary design in Pcdh
484 and DSCAM protein families, but these have not yet been characterized in detail.

485 The range of DIP/Dpr binding affinities indicated in Fig. 1B raises an additional set of
486 interesting questions regarding the biological role of weak interactions. In a number of cases there is
487 a clearly established connection between molecular affinities and cellular phenotype[8,42] but in general
488 very little is known about this topic. The situation becomes more complex for cells that express more
489 than one family member and/or more than one subgroup member. The ability to design mutants of
490 altered affinities and specificities combined with modern gene editing techniques opens the door to a
491 new type of experiment which probes the molecular basis of cell-cell recognition. We have reported
492 preliminary studies on this problem involving DIP and Dpr mutants which abolished homophilic or both
493 homophilic and heterophilic interactions in the *Drosophila* retina[11,43]. The next logical step will be to
494 test how subtle changes in affinity and specificity would affect biological phenotype.
495

496 ## Methods

497 **Protein structure alignment**

498 The DIP/Dpr structures used in alignment presented in Fig. 1C were, *blue subgroup* - crystal
499 structures of Dpr6/DIP-α (PDBID: 5EO9) and two conformations of Dpr10/DIP-α (PDBID: 6NRQ)[9,10];
500 *purple subgroup* – crystal structure of Dpr11/DIP-ɣ (PDBID: 6NRR)[10]; *green subgroup* – crystal
501 structures of Dpr4/DIP-η (PDBID: 6EG0) and for two conformations each for Dpr1/DIP-η (PDBID:
502 6NRW) and Dpr2/DIP-θ (PDBID: 6EG1)[10,11].
503

504 **Construction of homology models for all superfamily members**

505 We built homology models for each DIP and Dpr Ig1 domain (using the crystal structure of the DIP-
506 η/Dpr4 complex as a template, PDBID: 6EG0). The models are expected in principle to be high
507 quality since there is considerable sequence and structural similarity between intra-family DIP and
508 Dpr Ig1 domains. A few modeled proteins have longer loops than those of the template but these
509 regions do not appear critical for Ig1-Ig1 interface. Homology models of DIP and Dpr monomers built

510 using MODELLER[44] were superimposed onto Dpr4/DIP-η crystal structures yielding 231 DIP/Dpr
511 heterodimers. The side chains of the dimer models were further minimized using the Scwrl4
512 algorithm[45].
513
514 **FoldX protocol to evaluate ΔΔG(binding)**
515 Although FoldX can be used as a "black box" we summarize here the procedures used to
516 calculate ΔΔG(binding). We first refine the structure of a protein complex using the FoldX utility
517 called "RepairPDB" (http://foldxsuite.crg.eu/command/RepairPDB). This procedure "repairs"
518 improbable dihedral angles and van der Waals clashes in a protein. We then run five rounds of
519 "RepairPDB" and confirm that total energy of a protein complex reaches a plateau. The complex
520 whose structure has now been optimized with respect to the FoldX energy function is then used as a
521 starting point for a computational mutagenesis procedure with "BuildModel"
522 (http://foldxsuite.crg.eu/command/BuildModel). A desired mutation is made to the repaired structure
523 and then energy-minimized by sampling rotamers for the mutated residue and its neighboring amino
524 acids. The wild type protein is then energy-minimized with respect to rotamers of the same set of
525 residues. The difference in interaction energy between the mutant and wild type is then calculated
526 between the two protomers in the complex using the 'AnalyseComplex' FoldX utility
527 (http://foldxsuite.crg.eu/command/AnalyseComplex).
528 To test if FoldX achieves convergence for a given mutation we run "BuildModel" ten times by
529 setting the "numberOfRuns" parameter to 10 (http://foldxsuite.crg.eu/parameter/numberOfRuns),
530 which produces ten pairs of wild type and mutant structures, as different neighbors of the mutated
531 residue could be moved during minimization. The average value of the interaction energy obtained
532 via energy decomposition of the total energy using 'AnalyseComplex' defines ΔΔG(binding).
533 Since we carried out 10 separate runs for each FoldX prediction, this yields 30,440 FoldX
534 single point mutation calculations (1 week of computing time on 1 CPU).

535 It is important to note that our homology models give comparable FoldX results to those
536 obtained from crystal structures. This is perhaps not surprising since we use high-quality homology
537 models based on high sequence identities and the fact that all DIP/Dpr crystal structures
538 superimpose to within 1Å. To illustrate this point, we carried out FoldX calculations on homology
539 models and compared the results to those obtained from crystal structures of Dpr11/DIP-γ,
540 Dpr6/DIP-α, and Dpr10/DIP-α. The Pearson correlation coefficient (and root mean square error) with
541 experimental values was 0.56 (1.65) for the crystal structures and 0.53 (1.10) for the homology
542 models based on 25 experimental measurements of binding energy differences.
543
544 **Calculating the effects of mutations on protein-protein binding**
545 The following is a summary of the methods evaluated for the calculation of the effects of mutations
546 on binding free energies. ***FoldX*** evaluates the effects of mutations using an empirical force-field that
547 allows side chains to move but keeps the backbone rigid[18,31]; ***mCSM*** is a machine learning method
548 based on the assumption that the impact of a mutation is correlated with atomic-distance patterns in
549 surrounding amino acids[30]; ***BeAtMusic*** relies on a set of coarse-grained statistical potentials derived
550 from known protein structures[27]. BeAtMusic is the fastest of all the methods we tested – it takes
551 seconds to assess all possible mutations in a protein chain or at the interface; ***MutaBind*** utilizes a
552 combination of molecular mechanics force fields, statistical potentials and a fast side-chain
553 optimization algorithm; ***Rosetta flex ddG***[24] samples conformational diversity using "backrub" to
554 generate an ensemble of models and then applies torsional minimization, side-chain repacking, and
555 averaging across this ensemble to calculate $\Delta\Delta G$; ***BindProfX***[32] combines a sequence profile
556 conservation score of structural homologues with the FoldX potential. Data on the performance of
557 these methods is summarized in Supplementary Table 2.
558
559
560 **Sequence-logo preparation**

561 To get orthologs of DIPs and Dprs in Insecta species we performed the search using reciprocal best
562 hit BLAST[46] with an e-value cutoff of 10^{-35} using in-house developed scripts on non-redundant
563 BLAST database. The resulting 2950 Dpr sequences and 2005 DIP sequences (121 species) were
564 screened using CD-HIT[47] with 100% threshold to get rid of redundancy arising from multiple GI
565 (GenInfo Identifier) numbers pointing to the same protein product. The obtained non-redundant set
566 of full-length sequences corresponding to each DIP or Dpr was combined with a reference sequence
567 (DIP-α or Dpr6 respectively), and aligned using Clustal-Omega[48]. The multiple sequence alignments
568 (MSA) of each DIP or Dpr were further processed to remove sequences that had non-complete Ig1
569 domains and to extract interfacial positions based on the reference sequences. Sequence logos[49]
570 based on the MSAs presented in Fig. 2 were created separately for each *Drosophila* DIP and Dpr
571 based on 2570 Dpr and 1732 DIP sequences. Sequence logos were also prepared on a subfamily
572 level by combining sequences of all its members, see Fig. 6.
573
574 **Sequence-based methods used to predict specificity residues**
575 GroupSim[20] analyzes evolutionary conservation patterns using physico-chemical properties of amino
576 acids by means of similarity matrices; SDPpred[21] uses information theory schemes based on
577 Shannon entropy and frequency scores; SPEER[22] uses amino acid properties, entropy and
578 evolutionary rates by analyzing quantitative measures of the conservation patterns of protein sites
579 based on their physico-chemical properties and the heterogeneity of evolutionary changes between
580 and within the protein subfamilies. Multi-Harmony[23] uses a combination of Shannon entropy and
581 machine learning.
582 All the sequence-based methods we use for determination of specificity positions require
583 predefined subgrouping of sequences based on their binding preferences. We grouped the above
584 2570 Dpr and 1732 DIP sequences (truncated to 33 interfacial residues only) into seven sets (DIP-
585 θ/η/ι/, DIP-κ, DIP-α, DIP-β/λ, DIP-γ, DIP-δ, DIP-ε,ζ for DIPs, and Dpr1/2/3/4/5, Dpr7, Dpr6/10,
586 Dpr8/9/21, Dpr11/15/16/17, Dpr12, Dpr13,14,18,19,20 for Dprs).

587 We rely on the assumption that orthologs in the same subfamily conserve their interaction
588 specificity. This assumption would not work for comparison of proteins in species that are too far
589 removed from one another on the tree of life, but since we consider orthologs within one class,
590 *Insecta*, the assumption is likely to be correct. To justify this, we calculated the average pairwise
591 sequence identities within *Insecta* orthologs for interfacial residues of each DIP and Dpr. The
592 average pairwise sequence identity for interfacial residues of DIP (86%) and Dpr (83%) orthologs in
593 *Insecta* is 84%. This value is similar to the average intra-subgroup pairwise sequence identity for
594 each DIP and Dpr *Insecta* subgroup (78%) and the intra-subgroup pairwise sequence identities of
595 interfacial residues in *Drosophila melanogaster* (~80%). These values are also consistently higher
596 than inter-subgroup pairwise sequence identities in *Drosophila melanogaster* or *Insecta* (~50%).
597 Only one subfamily (red Dpr) is an exception to the rule - it has relatively low conservation both in
598 *Insecta* (48%), and in *Drosophila melanogaster* (40%).
599
600 **Protocol for identifying negative constraints.**
601 *Energy filter*
602 FoldX calculations on the effects of mutations were carried out on X-ray structures of DIP/Dpr
603 complexes of *Drosophila melanogaster* where available and on homology models where crystal
604 structures were not available. If mutations were predicted to cause destabilizing effects ($\Delta\Delta G >= 0.05$
605 kcal mol^{-1}) for every DIP or Dpr fly subgroup member in the context of binding to a member of
606 another Dpr or DIP subgroup, the position were accepted provided $\Delta\Delta G >= 0.05$ kcal mol^{-1} for every
607 subfamily template complex used in the FoldX calculations. The cutoff value for $\Delta\Delta G$ was chosen
608 based on the estimated standard deviation we found for mutating an interfacial amino acid into self
609 ($\Delta\Delta G = 0.00 \pm 0.05$ kcal mol^{-1}). Positions that passed the energy filter were further subject to the
610 evolutionary filter (see below).
611
612 *Evolutionary filter*

613 Positions were filtered using multiple sequence alignments (MSAs) for DIP and Dpr subfamilies
614 based on sequences of insect orthologs (see above). For every non-cognate DIP-*i*/Dpr-*j* family, to
615 filter positions on the DIP side we compared corresponding positions from DIP-*i* and DIP-*j*
616 subfamilies in MSAs. To filter positions on the Dpr side we compared positions from Dpr*i* and Dpr*j*
617 subfamilies in MSAs. The position was accepted if at least one of the three conditions outlined below
618 was true.
619 1) If positions in "*i*" and "*j*" subgroup were both conserved (Shannon entropy, S<0.23) and different in
620 identity – the position was accepted. S is given by - $\boldsymbol{S} = -\sum_{i=1}^{N} \boldsymbol{f_i log_{21} f_i}$
621 where f_i is the fraction (frequency of occurrence) of residues of amino acid type present at a position
622 in MSA and N is the number of amino acid types (base of 21 was picked to account for 20 amino
623 acids types and a gap).
624
625 2) If residues in the "*i*" and "*j*" subgroups had difference in their biophysical properties. We binned
626 the 20 amino acids into four property groups: hydrophobic (I, V, L, M, F, W, A, C, G, P), polar
627 (S,T,Q,N,H,Y), positively charged (K,R), and negatively charged (E,D). Then, for every amino acid at
628 a position of subfamily "*i*" (aa_i) and for every amino acid at the same position of subfamily "*j*" (aa_j), we
629 checked if aa_i and aa_j belonged to the same or different biophysical property groups defined above.
630 When aa_i and aa_j were different in property, we included the contribution of the amino acid pair
631 weighted by amino acid occurrences in MSAs ($f_{aa_i} \times f_{aa_j} \times 100\%$) to the overall percent difference.
632 When aa_i and aa_j had the same property, the contribution was not included. The sum of all included
633 weighted contributions ($\sum_{k,l}^{N} f_{aa_{ik}} \times f_{aa_{jl}} \times 100\%$, where N is the number of amino acid pairs where
634 (property of aa_i) ≠ (property of aa_j), while *k* and *l* run over all amino acids present at a position in
635 subfamilies "*i*" and "*j*", respectively) defined the total biophysical property difference in %. The
636 positions were accepted at >90% cutoff.
637

638 3) If residues in the *i*" and "*j*" subgroups had differences in size (amino acids of one subgroup
639 always larger or always smaller in size than those appearing in the other subgroup. We binned the
640 20 amino acids into five size groups, which were based on the volume of amino acids: tiny (G, A, S),
641 small (C, D, P, N, T), medium (E, V, Q, H), large (M, I, L, K, R), bulky (F, Y, W). We computed two
642 percent differences - the percent difference when aa_i belonged to a larger size group than aa_j and
643 the percent difference when aa_i belonged to a smaller size group than aa_j (relative size differences
644 between groups were encoded in the group names as tiny < small < medium < large < bulky). Each
645 percent difference was computed as a sum of the weighted amino acid contributions
646 ($\sum_{k,l}^{N} f_{aa\,ik} \times f_{aa\,jl} \times \ 100\%$, where N is the number of amino acid pairs where (size of aa_i group) ≠
647 (size of aa_j group), while *k* and *l* run over all amino acids present at a position in subfamilies "*i*" and
648 "*j*", respectively) when the condition above was satisfied (size(aa_i) > size(aa_j) in the first case and
649 size(aa_i) < size(aa_j) in the second case). The maximum of the two percent differences defined the
650 total size difference in %. The positions were accepted at >90% cutoff.
651
652 **Plasmid construction and protein expression**
653 All proteins were produced by expression in human embryonic kidney cells (HEK293F). DNA
654 sequences encoding DIP and Dpr extracellular regions were amplified by PCR and sub-cloned into
655 the mammalian expression vector VRC-8400[50] between the NotI and BamHI sites. Sequence
656 boundaries for each DIP and Dpr were the same as those defined in Cosmanescu et al.[11] All
657 sequences were preceded by the signal sequence of human binding immunoglobulin protein BiP
658 (MKLSLVAAMLLLLSAARA), and a kozak sequence (GCCACC). Constructs were followed by a C-
659 terminal hexa-histidine tag. Point mutations were introduced using the QuickChange method
660 (Agilent). HEK293F cells were transiently transfected with each expression construct using the
661 polyethylenimine method[51], and cells were expanded and grown in shake flasks in a CO2 incubator
662 for 3-6 days.
663

664 **Protein purification**

665 Conditioned media was equilibrated to 10mM Tris-HCl pH 8.0, 500mM NaCl, 3mM $CaCl_2$ and 5mM
666 Imidazole pH 8.0 and incubated with Ni^{2+} charged IMAC Sepharose 6 Fast Flow resin (GE
667 Healthcare) for 1 hr at 25°C. Resin was washed with at least 20 column volumes of buffer containing
668 10mM Imidazole pH 8.0. SDS gel electrophoresis was used to detect which elution fractions
669 contained the desired protein.

670 Proteins were further purified by size-exclusion chromatography (Superdex 200 HiLoad 26/60 or
671 Superdex S200 Increase 10/300 GL; GE Healthcare) on an AKTA pure fast protein liquid
672 chromatography system (GE Healthcare). Most proteins were stored in a buffer of 10mM Bis-Tris pH
673 6.6 and 150mM NaCl. UV absorbance at 280nm was used to determine protein concentration and
674 verification of purity was determined by gel electrophoresis. Accurate molecular weights were
675 determined through MALDI-TOF mass spectrometry at the Proteomics Shared Resource facility at
676 Columbia University.

677

678 **Sedimentation equilibrium by analytical ultracentrifugation**

679 Experiments were performed in a Beckman XL-A/I analytical ultracentrifuge (Beckman-Coulter, Palo
680 Alto CA, USA), utilizing six-cell centerpieces with straight walls, 12 mm path length and sapphire
681 windows. Protein samples were dialyzed to 10mM Bis-Tris pH 6.6, 150mM NaCl. The samples were
682 diluted to an absorbance of 0.65, 0.43 and 0.23 at a 10 mm path length and 280 nm wavelength in
683 channels A, B and C, respectively. Dilution buffer were used as blank. The samples were run at four
684 speeds. Most proteins were run at 15000, 19000, 23000, and 27000 RPM. For all runs the lowest
685 speed was held for 20hr and then four scans were taken with a 1hr interval, the second lowest held
686 for 10hr then four scans with a 1hr interval, and the third lowest and highest speed measured as the
687 second lowest speed. Measurements were done at 25°C, and detection was by UV at 280 nm.
688 Solvent density and protein v-bar were determined using the program SednTerp. (Alliance Protein
689 Laboratories) To calculate the K_D and apparent molecular weight, data were fit to a global fit model,

690 using HeteroAnalysis software package, obtained from University of Connecticut[52]
691 (http://www.biotech.uconn.edu/auf).
692
693 **Surface Plasmon Resonance (SPR) binding experiments**
694 SPR binding assays were performed using a Biacore T100 biosensor equipped with a Series S CM4
695 sensor chip. To minimize artificial binding resulting from enhanced- avidity effects of oligomers
696 interacting with an immobilized ligand, DIPs and their respective mutants, were consistently used as
697 ligands rather than analytes, and immobilized over independent flow cells using amine-coupling
698 chemistry in HBS pH 7.4 (10mM HEPES, 150mM NaCl) buffer at 25°C using a flow rate of 20 µL
699 min^{-1}. Dextran surfaces were activated for 7 minutes using equal volumes of 0.1M NHS(N-
700 Hydroxysuccinimide) and 0.4M EDC(1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide). Each protein
701 of interest was immobilized at ~30 µg mL^{-1} in 10 mM sodium acetate, pH 5.5 until the desired
702 immobilization level was achieved. The immobilized surface was blocked using a 4-minute injection
703 of 1.0 M ethanolamine, pH 8.5. Typical immobilization levels ranged between 760-980 RU. To
704 minimize nonspecific binding the reference flow cell was blocked by immobilizing BSA in 10 mM
705 sodium acetate, pH 4.25 for 3 minutes using a similar amine-coupling protocol as described above.
706 For each experiment where DIP mutants were immobilized to the chip surface, the wild type
707 molecule was immobilized on an adjacent flow cell as a positive control.
708 Binding analysis was performed at 25°C in a running buffer of 10 mM Tris-HCl, pH 7.2, 150mM NaCl,
709 1mM EDTA, 1 mg mL^{-1} BSA and 0.01% (v/v) Tween-20. Analytes were prepared in running buffer and
710 tested at nine concentrations using a three-fold dilution series ranging from 81 to 0.012 µM for data
711 shown in Figs. 3C and 3D, Supplementary Figs. 3A-C, Supplementary Fig. 4, with the of exception
712 Dpr10 binding to DIP-α A11T in Supplementary Fig. 3C, which was tested at nine concentration
713 ranging from 27 to 0.004 µM, prepared in a three-fold dilution series. Dpr6 was tested at eight
714 concentrations ranging from 9 to 0.004 µM, prepared in a running buffer using a three- fold dilution
715 series for data presented in Supplementary Fig. 1A, with the exception of experiments over the DIP-α

716 I16A and DIP-α K10Q G4S surfaces where Dpr6 was used at ten concentrations ranging from 81-
717 0.004 μM, using a three-fold dilution series. Dpr10 binding experiments in Supplementary Fig. 1B, was
718 tested at nine concentrations ranging from 27 to 0.004 μM, prepared in a running buffer using a three-
719 fold dilution series, with the exception of experiments over surfaces immobilized with DIP-α I16A,
720 K10Q G4S and K10Q D29S where the analyte was used at ten concentrations ranging from 81-0.004
721 μM, using a three-fold dilution series. In each experiment, every concentration was tested in duplicate.
722 Within each experiment, there two technical replicates starting with a single concentration series,
723 where samples are tested in order of increasing concentration, followed by a repeat of the same
724 concentration series, performed again from low to high concentration. During a binding cycle, the
725 association phase between each analyte and the immobilized molecule was monitored for either 30
726 or 40 seconds as indicated by the plotted sensorgrams, followed by 120-second dissociation phase,
727 each at 50 μL min^{-1}. At the end of the dissociation phase the signal returned to baseline thus
728 eliminating the need for a regeneration step. The last step was buffer wash injection at 100 μL min^{-1}
729 for 60 seconds. The analyte was replaced by buffer every two or three binding cycles to double-
730 reference the binding signals by removing systematic noise and instrument drift. The duplicate binding
731 responses were fit globally, using an 1:1 interaction model and a single K_D was calculated as the
732 analyte concentration that would yield 0.5 R_{max}[53] and a fitting error. K_Ds < 100 μM were calculated
733 using an independent R_{max}. For K_Ds >100 μM, the R_{max} was fixed to a global value determined by the
734 R_{max} of a different Dpr analyte tested over the same DIP surface during the same experiment that
735 showed binding above 50% and therefore produced a more accurate R_{max}.[53] For K_Ds >300 μM, a
736 lower limit is listed since at the analyte concentrations used we could not measure accurate K_Ds even
737 when the R_{max} is fixed. The data was processed using Scrubber 2.0 (BioLogic Software).
738 For several wild type interactions discussed in this manuscript, Dpr10/DIP-α, Dpr6/DIP-α, and
739 Dpr11/DIP-γ, we have also determined K_Ds from independent experiments (see the source data file).
740 For the Dpr10/DIP-α binding pair, the K_D from six independent experiments is 1.5±0.1 μM, and
741 similarly for the Dpr6/DIP-α, the K_D from six independent experiments is 2.0±0.2 μM. The K_D from

742 five independent experiments for Dpr11/DIP-γ is 7.9±0.9μM. The standard deviation in ΔΔG values
743 computed based on K_D measurements of the above three cases do not exceed 0.1 kcal mol^{-1}.
744 Therefore, mutations resulting in ΔΔG (SPR) > | 0.2 kcal mol^{-1} | are likely to be significant.
745 In a note added in proof, Cheng et al.[10] recently criticized K_Ds we previously determined via SPR
746 experiments[11], specifically for the interacting pairs Dpr6/DIP-α, Dpr11/DIP-γ and Dpr1/DIP-η, for
747 being consistently weaker than K_Ds they reported. The following discussion explains how differences
748 in experimental methodology likely explains these differences and raises issues about the approach
749 used by Cheng et al.[10]. The underlying challenge in both sets of measurements is that some DIPs
750 homodimerize, a problem that is discussed by Rich and Myszka in their 2006 review of SPR data[54].
751 We addressed this problem by immobilizing DIP proteins to the chip surface, using Dprs as analytes,
752 to minimize artificially lower K_Ds resulting from homodimers used as analytes. It has been suggested
753 that an alternate "fix" is to use the free monomer analyte concentration in K_D calculations[10].
754 However, the monomer/dimer equilibrium in the analyte can shift the moment the analyte is injected
755 over a surface immobilized with a heterophilic binding partner, which changes the monomeric
756 analyte concentration available for a heterophilic binding reaction over the course of the experiment.
757 Therefore, this correction does not adequately solve the problem.
758 Moreover, to calculate the free monomer analyte concentration of DIP-η, Cheng et al. used size
759 exclusion chromatography and SPR to calculate a homophilic K_D of 23 μM and 14 μM,
760 respectively,[10] compared to our analytical ultracentrifugation-determined K_D of 56.2 μM[11]. Given the
761 reliability of AUC measurements, this highlights the problems associated with Chang et al.'s
762 approach. Most importantly these numbers suggest that the differences in K_D reported in both
763 studies are due in part to the use of different biophysical techniques used to study these molecules.
764 In addition, for many of the results that differ between our studies and Cheng et al.[10], these authors
765 relied on ECIA experiments with artificially multimerized molecules. As previously discussed in detail
766 in Cosmanescu et al.[11] although ECIA can be used to detect interactions in a high-throughput
767 fashion and without the need for purified proteins, it involves the use of multimerized forms of both

768 the prey and the bait molecules, and thus introduces artificial avidity to amplify binding signals,
769 enhancing the likelihood of positive detection, but masking the real binding affinities between
770 interactants. In addition, the use of unpurified protein supernatants[10] could introduce experimental
771 bias toward identifying interactions between molecules that are more easily expressed at higher
772 levels compared to proteins that can be difficult to express.
773

774 **Acknowledgments**

775 This work was supported by NIH grants R01 GM30518 (B.H.), R01 MH114817 (L.S.) and NSF grant
776 MCB-1914542 (B.H.). We thank Engin Ozkan for providing DIP-α/Dpr10 and DIP-γ/Dpr11 structures
777 and density maps prior to their appearance in the PDB database.
778

779 **Author Contributions**

780 A.P.S., L.S. and B.H. conceived the project, designed experiments, and analyzed data. A.P.S.
781 carried out all the theoretical and computational work. P.S.K. performed and analyzed SPR
782 experiments. G.A. performed AUC experiments. J.J.B, F.C., S.M. produced expression vectors and
783 proteins. A.P.S. and F.C. designed mutant constructs. A.P.S., B.H., and L.S. wrote the manuscript.
784

785 **Data and code availability**

786 The raw data that support the findings of this study, and the code used to generate it, are available
787 from the corresponding authors upon request.
788

789 **Competing interests**

790 The authors declare no competing interests.
791

792 **References**

793

794 1. Zipursky, S.L., Wojtowicz, W.M. & Hattori, D. Got diversity? Wiring the fly brain with
795 Dscam. *Trends Biochem Sci* **31**, 581-8 (2006).
796 2. Wojtowicz, W.M. et al. A vast repertoire of Dscam binding specificities arises from
797 modular interactions of variable Ig domains. *Cell* **130**, 1134-45 (2007).
798 3. Thu, C.A. et al. Single-cell identity generated by combinatorial homophilic interactions
799 between alpha, beta, and gamma protocadherins. *Cell* **158**, 1045-1059 (2014).
800 4. Schreiner, D. & Weiner, J.A. Combinatorial homophilic interaction between gamma-
801 protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proc*
802 *Natl Acad Sci U S A* **107**, 14893-8 (2010).
803 5. Brasch, J. et al. Homophilic and Heterophilic Interactions of Type II Cadherins Identify
804 Specificity Groups Underlying Cell-Adhesive Behavior. *Cell Rep* **23**, 1840-1852 (2018).
805 6. Katsamba, P. et al. Linking molecular affinity and cellular specificity in cadherin-
806 mediated adhesion. *Proc Natl Acad Sci U S A* **106**, 11594-9 (2009).
807 7. Vendome, J. et al. Structural and energetic determinants of adhesive binding specificity
808 in type I cadherins. *Proc Natl Acad Sci U S A* **111**, E4175-84 (2014).
809 8. Harrison, O.J. et al. Nectin ectodomain structures reveal a canonical adhesive interface.
810 *Nat Struct Mol Biol* **19**, 906-15 (2012).
811 9. Carrillo, R.A. et al. Control of Synaptic Connectivity by a Network of Drosophila IgSF Cell
812 Surface Proteins. *Cell* **163**, 1770-1782 (2015).
813 10. Cheng, S. et al. Molecular basis of synaptic specificity by immunoglobulin superfamily
814 receptors in Drosophila. *Elife* **8**(2019).
815 11. Cosmanescu, F. et al. Neuron-Subtype-Specific Expression, Interaction Affinities, and
816 Specificity Determinants of DIP/Dpr Cell Recognition Proteins. *Neuron* (2018).
817 12. Tan, L. et al. Ig Superfamily Ligand and Receptor Pairs Expressed in Synaptic Partners in
818 Drosophila. *Cell* **163**, 1756-69 (2015).
819 13. Ozkan, E. et al. An extracellular interactome of immunoglobulin and LRR proteins reveals
820 receptor-ligand networks. *Cell* **154**, 228-39 (2013).
821 14. Havranek, J.J. & Harbury, P.B. Automated design of specificity in molecular recognition.
822 *Nat Struct Biol* **10**, 45-52 (2003).
823 15. Humphris, E.L. & Kortemme, T. Design of multi-specificity in protein interfaces. *PLoS*
824 *Comput Biol* **3**, e164 (2007).
825 16. Leaver-Fay, A. et al. Computationally Designed Bispecific Antibodies using Negative
826 State Repertoires. *Structure* **24**, 641-651 (2016).
827 17. Mandell, D.J. & Kortemme, T. Computer-aided design of functional protein interactions.
828 *Nat Chem Biol* **5**, 797-807 (2009).
829 18. Schymkowitz, J. et al. The FoldX web server: an online force field. *Nucleic Acids Res* **33**,
830 W382-8 (2005).
831 19. Grigoryan, G., Reinke, A.W. & Keating, A.E. Design of protein-interaction specificity gives
832 selective bZIP-binding peptides. *Nature* **458**, 859-64 (2009).
833 20. Capra, J.A. & Singh, M. Characterization and prediction of residues determining protein
834 functional specificity. *Bioinformatics* **24**, 1473-80 (2008).
835 21. Kalinina, O.V., Mironov, A.A., Gelfand, M.S. & Rakhmaninova, A.B. Automated selection
836 of positions determining functional specificity of proteins by comparative analysis of
837 orthologous groups in protein families. *Protein Sci* **13**, 443-56 (2004).

838 22. Chakrabarti, S., Bryant, S.H. & Panchenko, A.R. Functional specificity lies within the
839 properties and evolutionary changes of amino acids. *J Mol Biol* **373**, 801-10 (2007).
840 23. Brandt, B.W., Feenstra, K.A. & Heringa, J. Multi-Harmony: detecting functional
841 specificity from sequence alignment. *Nucleic Acids Res* **38**, W35-40 (2010).
842 24. Barlow, K.A. et al. Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-
843 Protein Binding Affinity upon Mutation. *J Phys Chem B* **122**, 5389-5399 (2018).
844 25. Buss, O., Rudat, J. & Ochsenreither, K. FoldX as Protein Engineering Tool: Better Than
845 Random Based Approaches? *Comput Struct Biotechnol J* **16**, 25-33 (2018).
846 26. Choi, Y., Furlon, J.M., Amos, R.B., Griswold, K.E. & Bailey-Kellogg, C. DisruPPI: structure-
847 based computational redesign algorithm for protein binding disruption. *Bioinformatics*
848 **34**, i245-i253 (2018).
849 27. Dehouck, Y., Kwasigroch, J.M., Rooman, M. & Gilis, D. BeAtMuSiC: Prediction of changes
850 in protein-protein binding affinity on mutations. *Nucleic Acids Res* **41**, W333-9 (2013).
851 28. Geng, C., Vangone, A., Folkers, G.E., Xue, L.C. & Bonvin, A. iSEE: Interface structure,
852 evolution, and energy-based machine learning predictor of binding affinity changes
853 upon mutations. *Proteins* (2018).
854 29. Li, M., Simonetti, F.L., Goncearenco, A. & Panchenko, A.R. MutaBind estimates and
855 interprets the effects of sequence variants on protein-protein interactions. *Nucleic Acids*
856 *Res* **44**, W494-501 (2016).
857 30. Pires, D.E., Ascher, D.B. & Blundell, T.L. mCSM: predicting the effects of mutations in
858 proteins using graph-based signatures. *Bioinformatics* **30**, 335-42 (2014).
859 31. Guerois, R., Nielsen, J.E. & Serrano, L. Predicting changes in the stability of proteins and
860 protein complexes: a study of more than 1000 mutations. *J Mol Biol* **320**, 369-87 (2002).
861 32. Xiong, P., Zhang, C., Zheng, W. & Zhang, Y. BindProfX: Assessing Mutation-Induced
862 Binding Affinity Change by Protein Interface Profiles with Pseudo-Counts. *J Mol Biol* **429**,
863 426-434 (2017).
864 33. Ashley, J. et al. Transsynaptic interactions between IgSF proteins DIP-alpha and Dpr10
865 are required for motor neuron targeting specificity. *Elife* **8**(2019).
866 34. Venkatasubramanian, L. et al. Stereotyped terminal axon branching of leg motor
867 neurons mediated by IgSF proteins DIP-alpha and Dpr10. *Elife* **8**(2019).
868 35. Chakraborty, A. & Chakrabarti, S. A survey on prediction of specificity-determining sites
869 in proteins. *Brief Bioinform* **16**, 71-88 (2015).
870 36. Akiva, E., Itzhaki, Z. & Margalit, H. Built-in loops allow versatility in domain-domain
871 interactions: lessons from self-interacting domains. *Proc Natl Acad Sci U S A* **105**, 13292-
872 7 (2008).
873 37. Goodman, K.M. et al. Structural Basis of Diverse Homophilic Recognition by Clustered
874 alpha- and beta-Protocadherins. *Neuron* **90**, 709-23 (2016).
875 38. Goodman, K.M. et al. Molecular basis of sidekick-mediated cell-cell adhesion and
876 specificity. *Elife* **5**(2016).
877 39. Li, S.A., Cheng, L., Yu, Y., Wang, J.H. & Chen, Q. Structural basis of Dscam1
878 homodimerization: Insights into context constraint for protein recognition. *Sci Adv* **2**,
879 e1501118 (2016).
880 40. Rubinstein, R. et al. Molecular logic of neuronal self-recognition through protocadherin
881 domain interactions. *Cell* **163**, 629-42 (2015).

882 41. Sawaya, M.R. et al. A double S shape provides the structural basis for the extraordinary
883 binding specificity of Dscam isoforms. *Cell* **134**, 1007-18 (2008).
884 42. Togashi, H. et al. Nectins establish a checkerboard-like cellular pattern in the auditory
885 epithelium. *Science* **333**, 1144-7 (2011).
886 43. Xu, S. et al. Interactions between the Ig-Superfamily Proteins DIP-alpha and Dpr6/10
887 Regulate Assembly of Neural Circuits. *Neuron* (2018).
888 44. Eswar, N. et al. Comparative protein structure modeling using Modeller. *Curr Protoc*
889 *Bioinformatics* **Chapter 5**, Unit-5 6 (2006).
890 45. Krivov, G.G., Shapovalov, M.V. & Dunbrack, R.L., Jr. Improved prediction of protein side-
891 chain conformations with SCWRL4. *Proteins* **77**, 778-95 (2009).
892 46. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment
893 search tool. *J Mol Biol* **215**, 403-10 (1990).
894 47. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-
895 generation sequencing data. *Bioinformatics* **28**, 3150-2 (2012).
896 48. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence
897 alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
898 49. Crooks, G.E., Hon, G., Chandonia, J.M. & Brenner, S.E. WebLogo: a sequence logo
899 generator. *Genome Res* **14**, 1188-90 (2004).
900 50. Barouch, D.H. et al. A human T-cell leukemia virus type 1 regulatory element enhances
901 the immunogenicity of human immunodeficiency virus type 1 DNA vaccines in mice and
902 nonhuman primates. *J Virol* **79**, 8828-34 (2005).
903 51. Baldi, L., Hacker, D.L., Meerschman, C. & Wurm, F.M. Large-scale transfection of
904 mammalian cells. *Methods Mol Biol* **801**, 13-26 (2012).
905 52. Cole, J.L., Lary, J.W., T, P.M. & Laue, T.M. Analytical ultracentrifugation: sedimentation
906 velocity and sedimentation equilibrium. *Methods Cell Biol* **84**, 143-79 (2008).
907 53. Rich, R.L., Myszka, D.G. Extracting affinity constants from biosensor binding responses.
908 in *Label-free biosensors : techniques and applications* (ed. Cooper, M.A.) pp. 48-84
909 (Cambridge; New York: Cambridge University Press, 2009).
910 54. Rich, R.L. & Myszka, D.G. Survey of the year 2006 commercial optical biosensor
911 literature. *J Mol Recognit* **20**, 300-66 (2007).
912