Molecular reconstruction of recurrent evolutionary switching in olfactory receptor specificity

Lucia L. Prieto-Godino[1,2,*], Hayden R. Schmidt[1] and Richard Benton[1,*]

[1]Center for Integrative Genomics
Faculty of Biology and Medicine
Génopode Building
University of Lausanne
CH-1015
Lausanne
Switzerland

[2]The Francis Crick Institute
London
NW1 1BF
United Kingdom

[*]Corresponding authors:
lucia.prietogodino@crick.ac.uk
Richard.Benton@unil.ch

Abstract

Olfactory receptor repertoires exhibit remarkable functional diversity, but how these proteins have evolved is poorly understood. Through analysis of extant and ancestrally-reconstructed drosophilid olfactory receptors from the Ionotropic Receptor (IR) family, we investigated evolution of two organic acid-sensing receptors, IR75a and IR75b. Despite their low amino acid identity, we identify a common "hotspot" in their ligand-binding pocket that has a major effect on changing the specificity of both IRs, as well as at least two distinct functional transitions in IR75a during evolution. Ligand-docking into IR models predicts that the hotspot does not contact odor molecules, suggesting that this residue indirectly influences ligand/receptor interactions. Moreover, we show that odor specificity is refined by changes in additional, receptor-specific sites, including those outside the ligand-binding pocket. Our work reveals how a core, common determinant of ligand-tuning acts within epistatic and allosteric networks of substitutions to lead to functional evolution of olfactory receptors.

Introduction

Amongst the senses, olfaction is particularly flexible over evolutionary time, enabling animals to adapt their recognition of the vast, ever-changing universe of volatile chemicals in the environment (Bargmann, 2006; Ramdya and Benton, 2010). This flexibility is reflected in the evolution of large, divergent families of olfactory receptors with different odor tuning properties. Several functional surveys of receptor repertoires in vertebrates (e.g., humans and the house mouse *Mus musculus*) and insects (e.g., the vinegar fly *Drosophila melanogaster* and the malaria mosquito *Anopheles gambiae*) have identified ligands for many receptors (Carey et al., 2010; Hallem and Carlson, 2006; Saito et al., 2009). Moreover, comparative sequence and functional analyses of orthologous receptors across species have started to identify amino acid differences that can explain species-specific receptor tuning properties (Adipietro et al., 2012; Auer et al., 2020; Butterwick et al., 2018; Del Mármol et al., 2021; Leary et al., 2012; Mainland et al., 2013; Prieto-Godino et al., 2017; Yang et al., 2017). However, the molecular basis of functional changes in receptors over evolutionary timescales – and whether common principles in this process exist between different receptors – remains unclear.

A powerful model to study olfactory receptor evolution is the Ionotropic Receptor (IR) repertoire, a protostomian chemosensory subfamily of ionotropic glutamate receptors (iGluRs) (Benton et al., 2009; Croset et al., 2010; Ni, 2021; Rytz et al., 2013). Although IRs and iGluRs have limited amino acid sequence identity, their overall conserved (predicted) secondary and tertiary structural organization suggests that the chemosensory receptors share many mechanistic similarities with their iGluR ancestors (Abuin et al., 2011). The best-characterized IRs are predicted to be hetero-tetramers formed of two subunits of a conserved co-receptor and two subunits of a "tuning" receptor (Abuin et al., 2019, 2011). The latter are more variable in sequence both within and between species, particularly in the extracellular ligand-binding domain (LBD), consistent with their diverse odor-recognition properties.

Functional characterization of IRs in different drosophilid species has revealed orthologous receptors that have distinct odor recognition properties (Prieto-Godino et al., 2017, 2016), notably IR75a, a receptor that likely originated in the Neodipteran ancestor (>200 million years ago) and its paralog IR75b, which arose through duplication of *Ir75a* in the Drosophilidae ancestor (~60-70 million years ago) (Croset et al., 2010; Prieto-Godino et al., 2017). In *D. melanogaster* and *D. simulans*, two cosmopolitan species that feed on a wide range of fermented fruit, these receptors exhibit different sensitivity towards carboxylic acids: IR75a is tuned predominantly to acetic acid, while IR75b responds maximally to butyric acid (Prieto-Godino et al., 2017, 2016; Silbering et al., 2011). By contrast, in the closely-related island endemic *D. sechellia*, which feeds and breeds exclusively on the ripe noni fruit of the *Morinda citrifolia* shrub (Figure 1A), IR75a and IR75b preferentially respond to butyric acid and hexanoic acid, respectively (Prieto-Godino et al., 2017, 2016). These differences are likely to be ecologically significant: for example, acetic acid is a key product of microbial fermentation of vegetal material and regulates numerous behaviors in *D. melanogaster* including attraction (Becher et al., 2010) sexual receptivity (Gorter et al., 2016) and oviposition (Joseph et al., 2009; Kim et al., 2018). Hexanoic acid is a dominant component of noni fruit and elicits attractive behaviors in *D. sechellia* (Amlou et al., 1998; Dekker et al., 2006; Prieto-Godino et al., 2017). In this work, we combine comparative *in vivo* functional analyses of these receptors across the drosophilid phylogeny, with ancestral sequence reconstruction, site-directed mutagenesis and protein modelling to investigate their evolution.

Results

Evolution of olfactory responses of IR75a across the drosophilid phylogeny

The distinct responses of *D. sechellia* IR75a (*Dsec*IR75a) compared to orthologs in its two generalist cousins (Prieto-Godino et al., 2016) suggested that acetic acid-sensing was the ancestral function of IR75a. We tested this hypothesis by measuring odor-evoked responses of IR75a-expressing olfactory sensory neurons (OSNs) across the drosophilid phylogeny, representing >40 million years divergence time (Figure 1A-C, see Material and methods). As stimuli we used a panel of linear carboxylic acids spanning from one- to six-carbon chains (hereafter abbreviated to C1-C6, where acetic acid is C2 and butyric acid is C4). All tested species within the *melanogaster/obscura* group (except for *D. sechellia*) displayed strongest responses to C2, similar to *D. melanogaster* and *D. simulans* (Figure 1B-C). Unexpectedly, the responses of more divergent species were more similar to those of *D. sechellia*, exhibiting strongest responses to C4 (Figure 1B-C).

To simplify data visualization and discern in an unbiased way which odors contribute maximally to differential tuning of Ir75a neurons across species, we performed principal component analysis (PCA) on their response profiles. The first principal component (PC1) explains 67.5% of the variance in the data, mostly capturing the inverse variation between C2 and C4 (Figure 1D). When plotting the C2 and C4 responses against each other, Ir75a neurons of different species segregated into two clusters with either high responses to C2 and low to C4 or vice versa (Figure 1E). The clustering matched well the phylogeny with the exception of

D. sechellia, which grouped together with *D. willistoni*, *D. mojavensis* and *D. virilis* (Figure 1A and 1E). These observations suggested a new model in which the ancestral drosophilid IR75a was predominantly a C4 sensor that evolved to become a C2 sensor in the last common ancestor of the melanogaster and obscura groups, before reverting to an ancestral-like state in *D. sechellia* (Figure 2A).

Functional analysis of ancestrally-reconstructed IR75a

To test this hypothesis, we "resurrected" the ancestral IR75a receptors at the ancestral drosophilid and *melanogaster/obscura* group phylogenetic nodes by inferring their sequence via maximum likelihood (Randall et al., 2016) from the sequence of orthologs from sixteen extant species (Figure 2A and Figure 2 – figure supplement 1 and Materials and methods). We synthesized genes encoding the inferred ancestral proteins – termed here IR75a^{Dros} and IR75a$^{mel\text{-}obs}$, respectively – and integrated these into a common genomic location to avoid differential positional influence on their expression. These transgenes were expressed individually in the *D. melanogaster* "IR decoder neuron", an OSN that lack the endogenous tuning receptor subunit but expresses the IR8a co-receptor (Abuin et al., 2011; Grosjean et al., 2011; Prieto-Godino et al., 2016). When expressed in this system, the tuning curves of *D. melanogaster* IR75a (*Dmel*IR75a) and *D. sechellia* IR75a (*Dsec*IR75a) recapitulate the response profile of these receptors expressed in their endogenous neurons, with strongest responses to C2 and C4, respectively (Figure 2B-C), consistent with previous observations (Prieto-Godino et al., 2016). As predicted, IR75a$^{mel\text{-}obs}$ responded similarly to *Dmel*IR75a (Figure 2B-C). Importantly, the more ancient, resurrected receptor, IR75a^{Dros}, had a tuning profile that was almost identical to *Dsec*IR75a, with maximal responses to C4 (Figure 2B-C). These results indicate that IR75a has switched tuning profile at least twice during its evolutionary history, from C4- to C2-sensing and, in *D. sechellia*, back again.

Identifying the molecular basis of the functional evolution of IR75a

To determine the molecular basis of the functional evolution of IR75a, we aligned the sequences of IR75a$^{mel\text{-}obs}$ and IR75a^{Dros}. 114 positions in these proteins exhibit different amino acids (82% identity), of which 45 are located within the LBD (Figure 2 – figure supplement 1). As this level of divergence precluded straightforward experimental determination of the relevant sites, we first approached the problem by focusing on the more recent "reverse" transition of C2-sensing to C4-sensing that occurred on the branch leading to *Dsec*IR75a. We had previously narrowed down this tuning switch to three amino acid positions within the internal pocket of the bilobed (S1-S2) LBD (Prieto-Godino et al., 2016) (Figure 3A). Simultaneous substitution of these sites in *Dmel*IR75a with the residues found in *Dsec*IR75a (i.e., T289S, Q536K and F538L) produced a receptor that faithfully recapitulated the response properties of *Dsec*IR75a when assessed in the IR decoder neuron (Figure 3B), as previously described (Prieto-Godino et al., 2016). Furthermore, reverse amino acid substitutions in *Dsec*IR75a (S289T, K536Q and L538L) conferred response properties characteristic of *Dmel*IR75a (Figure 3B). These results indicate that the change in tuning is encoded entirely within these three sites. Thus, there are only $2^3 = 8$ variants to transit between *Dmel*IR75a and

*Dsec*IR75a (including the wild-type sequences) – compared to $2^{114} = 2\times10^{34}$ possible variants between IR75a^{Dros} and IR75a$^{mel\text{-}obs}$ – offering an excellent opportunity to study the functional evolution of an olfactory receptor.

We generated versions of *Dmel*IR75a in which each of these three sites were substituted individually as well the three possible double substitutions. All single amino acid changes had an impact on receptor responses, shifting tuning towards C4 (and C3) to different extents, while still retaining some sensitivity to C2 (Figure 3C). Double substitutions showed further shifts towards the *Dsec*IR75a tuning profiles but to varying degrees: *Dmel*IR75a^{T289S,Q536K} is a broadly-tuned receptor that responds maximally (albeit weakly) to C2, C3 and C4, *Dmel*IR75a^{Q536K,F538L} displays maximum responses to both C3 and C4, and *Dmel*IR75a^{T289S,F538L} responds maximally to C4 alone (Figure 3C). To move beyond simple descriptions of the effects of these mutations, we used visualizations and analyses that give insights into the evolutionary landscape of this receptor, as described below.

Mapping the evolutionary landscape of IR75a

Evolutionary landscapes map genotypes onto a quantitative phenotype to illustrate evolutionary change as a "navigation" process. The roughness of the landscape determines how accessible each of all possible paths are, and therefore the likelihood that evolution proceeds through each path (Aguilar-Rodríguez et al., 2017; Wright, 1932). We reasoned this approach was useful to understand the evolution of IR75a as the lack of informative intraspecific genetic diversity within these receptor genes in drosophilid species (data not shown) precluded the direct identification of evolutionary intermediates.

Responses properties of a receptor are multidimensional, where each odor defines a dimension. However, PCA of the responses of all *Dmel*IR75a receptor mutants to all odors revealed that PC1 explains most of the variance (54.5%) and captures the inverse variation between C2 and C4 – similar to the IR75a neuron response PCA across the drosophilid phylogeny (Figure 1D) – and, to a lesser extent, the co-variation of C5 with C4 (Figure 3D). We therefore plotted the phenotypic data for each single, double and triple *Dmel*IR75a mutant in the C2 versus C4 space alone (Figure 3E) and in the PC1 space (Figure 3F), acknowledging these simplifications represents only part of the functional changes (albeit potentially the most important).

The C2 vs C4 plot enables visualization of all possible evolutionary paths by joining the stepwise mutations from *Dmel*IR75a to *Dmel*IR75a^{T289S,Q536K,F538L}. Most receptor variants are plotted close to the straight line (i.e., the shortest path) that joins the initial and final state, with the *Dmel*IR75a^{F538L} single mutant showing the largest individual "step" along this line (Figure 3E). If we assume that all steps along this line would be favoured by selection, this observation suggests the hypothesis that the F538L change might have been the first one of the three substitutions to occur, as *de novo* mutations with larger effect are typically substituted first followed by those with smaller effect (Holder and Bull, 2001; Orr, 2005). An alternative hypothesis is that T289S variant already existed, perhaps as standing variation in the population: T289S has little phenotypic consequence by itself (Figure 3C,E), but it greatly augments the effect of the F538L substitution (Figure 3C,E). In this

case, T289S and F538L would have reached fixation together in the same genetic background because their combined effect allows for a large adaptive leap from one peak (C2) to the other (C4).The Q536K change alone has an intermediate phenotype, but we suspect is unlikely to have been the first to occur, as combination with either other amino acid change leads to a less specific receptor (*Dmel*IR75a^{Q536K,F538L}, lying to the top-right of the line), or a receptor with overall weak sensitivity (*Dmel*IR75a^{T289S,Q536K}, lying to the bottom-left of the line) (Figure 3C,E).

To formalize this analysis, we calculated epistasis (*i.e.*, non-additive effects of different mutations), which can constrain the available evolutionary paths (Aguilar-Rodríguez et al., 2017; Phillips, 2008). Using the projection of our physiological responses onto PC1 as our phenotypic readout, we calculated for each of the combinations of mutants (*i.e.*, single with single, or single with double) whether the sum of the effects of each of the starting substitutions on the wild-type background was significantly different from the effect of introducing both mutations simultaneously (see Materials and methods). Notably, negative magnitude epistasis was observed in all combinations of Q536K and F538L (whether T289S was also present the background or not) (Figure 3F, red lines). This observation supports our previous hypothesis that whether F538L or T289S was the first mutation to occur, the second one would be the other of these two, because Q536K incurs in negative epistasis when combined with F538L and it does not lead to a monotonic increase in responses when combined with T289S (Figure 3F). Furthermore, this visualization highlights the significant effect on tuning of the F538L substitution, as well as the minimal (statistically non-significant) effect of the T289S substitution, supporting its possible segregation in the population without deleterious effects. It remains unclear what, if any, is the function of Q536K; it is possible it was fixed by drift or affects some other aspect of receptor activity not analyzed here.

A hotspot for evolution of IR odor tuning

We next asked whether the knowledge of the molecular basis of the C2→C4 tuning change on the *D. sechellia* branch offers insight into the ancestral switch from C4 to C2 in the last common ancestor of the *melanogaster/obscura* group. Strikingly, examination of the identity of residues aligned with *Dmel*IR75a F538 in orthologs across the drosophilid phylogeny revealed a perfect correspondence between the identity of this position and the best agonist for the receptor: all IR75a orthologues of species responding most strongly to C2 have an F, while those that respond to C4, have an L (Figure 4A and Figure 2 – figure supplement 1). Such correspondence was not seen for amino acid identities at position 289 and 536: for example, *Dwil*IR75a has a Q at the position equivalent to Q536 of *Dmel*IR75a (Figure 2 – figure supplement 1), but this species' Ir75a neurons respond to C4, like *D. sechellia* (Figure 1C).

These observations suggest that position 538 (or equivalent in orthologous sequences of slightly different lengths) has been a "hotspot" for odor response evolution, changing from L to F in the *melanogaster/obscura* ancestor and then changing back on the *D. sechellia* branch. The reversion of amino acid identity in *Dsec*IR75a is not due to an inverse mutation in the corresponding DNA sequence:

285 a C→T mutation in codon position one led to the L to F substitution in the
286 *melanogaster/obscura* ancestor, while in *D. sechellia* a C→A mutation in codon
287 position three led, convergently, to restoration of the L-encoding codon (Figure 4A).
288 The important contribution of a single amino acid in IR75a in determining the
289 specificity for shorter- or longer-chain acids was reminiscent of our observation of
290 the evolution of the paralogous receptor IR75b: the difference in tuning of
291 *Dmel*IR75b and *Dsec*IR75b to C4 and C6, respectively, is determined in large part
292 by a T523S substitution in the LBD (Prieto-Godino et al., 2017). IR75b and IR75a
293 exhibit only 38% amino acid identity. However, alignment of these receptors
294 revealed that position 523 in *Dmel*IR75b corresponds precisely to the 538 hotspot
295 in IR75a (Figure 4B). Thus, evolution of novel specificities in two different receptors
296 – which diverged from a common ancestor >60 million years ago (Prieto-Godino et
297 al., 2017) – is specified by changes in the same site within their LBDs.
298 To gain insights into the molecular mechanisms underlying the role of the
299 hotspot in determining the functional properties of IR75a, we performed docking
300 analysis of ligands in a protein homology model of the *Dmel*IR75a LBD (Prieto-
301 Godino et al., 2016). The top-ranked pose of C2 (or C4) in this model predicts an
302 ionic bond of the carboxyl group of this acid with R297 (Figure 4C and Figure 4 –
303 figure supplement 1). This arginine residue is conserved in all acid-sensing IRs
304 (Benton et al., 2009), and is essential for odor-evoked activity of IR84a (Abuin et
305 al., 2011); consistently, a similar interaction was predicted in a model of the IR84a
306 LBD with its ligand phenylacetic acid (Figure 4 – figure supplement 1). Moreover,
307 this interaction is equivalent to that of the α-carboxyl group of the glutamate ligand
308 with a conserved arginine in iGluRs (Figure 4C and figure 4 – figure supplement
309 1), suggesting a conserved mode of ligand-recognition in IRs and iGluRs. By
310 contrast, the *Dmel*IR75a hotspot residue F538 is not predicted to interact with
311 ligand molecules, as it is >4 Å away from any of the top-ranked ligand poses (Figure
312 4C and figure 4 – figure supplement 1). These observations suggest that the
313 hotspot is not directly involved ligand binding; how changes in the hotspot might
314 alter odor response properties of IRs is considered in the Discussion.
315
316 **Functional interactions between the IR75b hotspot and surface residues of**
317 **the LBD**
318
319 Although mutation of the hotspot in *Dmel*IR75b is sufficient to confer novel
320 responses to C5 and C6, this single change did not fully recapitulate the response
321 profile of *Dsec*IR75b, as *Dmel*IR75b^{T523S} retained sensitivity to shorter chain acids
322 (Figure 5A). Compared to *Dmel*IR75b and *D. simulans* IR75b (*Dsim*IR75b),
323 *Dsec*IR75b contain three additional changes within the LBD pocket region (P473S,
324 G492S and A520T; like the hotspot, all are within the S2 lobe), but in our previous
325 work none of these appeared to contribute substantially – individually or together –
326 to the refinement of specificity for C6 (Prieto-Godino et al., 2017). The *Dsec*IR75b
327 LBD contains six additional derived residues located in S1 (Prieto-Godino et al.,
328 2017), leading us to test whether these sites contribute to the observed changes in
329 odor responses.
330 Replacement of the entire S1 lobe in *Dmel*IR75b with that of *Dsec*IR75b
331 generated a receptor, *Dmel*IR75b^{DsecS1}, with significantly increased responses to
332 C6 and lower (albeit not statistically significant; Figure 5 source data 1) sensitivity

333 to C3, when compared to *Dmel*IR75b (Figure 5A). Addition of the hotspot
334 substitution (T523S) produced a receptor (*Dmel*IR75b^{DsecS1,T523S}) that is more
335 similar in response profile to *Dsec*IR75b than either *Dmel*IR75b^{DsecS1} or
336 *Dmel*IR75b^{T523S} (Figure 5A). However, this receptor still has robust sensitivity to
337 C4, like *Dmel*IR75b but unlike *Dsec*IR75b (Figure 5A). Further incorporation of the
338 three additional substitutions of residues in the S2 domain (Prieto-Godino et al.,
339 2017), generated a receptor (*Dmel*IR75b^{DsecS1,4mutS2} that, in terms of specificity, is
340 indistinguishable from *Dsec*IR75b (Figure 5A). However, this receptor has overall
341 reduced sensitivity for all acids when compared with *Dsec*IR75b, and its response
342 to C6 is significantly smaller than that of *Dsec*IR75b (Figure 5A). Despite this
343 overall reduced sensitivity, *Dmel*IR75b^{DsecS1,4mutS2} still has a significantly increased
344 sensitivity to C6 when compared with *Dmel*IR75b (Figure 5A).
345 To visualize how the S1 residues might impact receptor function, we
346 mapped the position of the six derived changes within this lobe onto a protein
347 homology model of *Dmel*IR75b (Figure 5B). All of these are located on the surface
348 of the LBD and therefore unlikely to contact ligands directly. IRs and iGluRs are
349 thought to exhibit the same global structure and stoichiometry (Abuin et al., 2019,
350 2011), prompting us to align the modelled *Dsec*IR75b LBD onto a homotetrameric
351 iGluR structure (PDB 3KG2) (Figure 5C). This analysis did not suggest particularly
352 close proximity of most of these residues to the interaction interface between
353 subunits (Figure 5C).
354
355 **Discussion**
356
357 As species adapt to new ecological niches, olfactory receptors evolve to define new
358 relationships between external signals and internal neural representations.
359 Population genetic analyses of *D. melanogaster* strains isolated from diverse global
360 habitats reveal that olfactory receptors (and other chemosensory protein families)
361 display some of the strongest genomic signatures of recent selection (Arguello et
362 al., 2016), suggesting these proteins act as "first responders" in local adaptation to
363 new environments. In this work we have used comparative sequence and
364 functional analyses across the well-defined phylogeny of the *Drosophila* genus to
365 study how members of the organic-acid sensing clade of IRs have changed over
366 evolutionary timescales.
367 Our most important finding is the discovery of a "hotspot" residue in these
368 IRs, whose mutation had a major effect on the odor specificity of one receptor
369 (IR75a) at two different timepoints during species diversification, as well as on the
370 tuning change of a distinct receptor (IR75b). For IR75a, our data support a model
371 in which the ancestral drosophilid IR75a was – contrary to previous assumptions –
372 a C4 sensor that switched, through mutation at the hotspot (and other sites), to C2
373 sensing in the *melanogaster*/*obscura* ancestor before "reverting" to C4-sensing in
374 *D. sechellia*. The ancestral function of IR75b is still unclear: we have not been able
375 to unambiguously identify Ir75b neurons across the drosophilid phylogeny, as they
376 are not easily distinguishable from those expressing a related receptor, IR75c
377 (Prieto-Godino et al., 2017). However, like IR75a, IR75b has adapted through
378 hotspot substitution along the lineage leading to *D. sechellia*. Neither the identity
379 nor even the chemical class of amino acids occupying the hotspot are conserved
380 in these receptors (phenylalanine (F, aromatic) or leucine (L, hydrophobic) in

381 IR75a; threonine (T) or serine (S) (both polar) in IR75b). These observations
382 indicate that the position of the hotspot in these IRs' LBDs, and not its identity *per*
383 *se*, must explain its central role in defining odor-response properties in different
384 receptors. Studies of other families of olfactory receptors in invertebrates and
385 vertebrates have revealed enormous inter- and intra-specific sequence variation,
386 some of which have been linked to differences in odor tuning (Adipietro et al., 2012;
387 Auer et al., 2020; Block, 2018; Butterwick et al., 2018; Del Mármol et al., 2021;
388 Leary et al., 2012; Mainland et al., 2013; Prieto-Godino et al., 2017; Yang et al.,
389 2017). Within this molecular and function diversity, it will be of interest to examine
390 whether analogous hotspots exist, revealing favored (or constrained) mechanisms
391 through which evolution selects for new odor-detection properties.
392 The mechanistic role of the hotspot remains unknown. Our ligand docking
393 into an IR75a LBD model suggest this residue does not contact ligands directly,
394 located at a distance of >4 Å away. While we acknowledge the caveats of
395 interpreting ligand-docking into protein models, our predictions are consistent with
396 experimentally-determined glutamate/iGluRs interactions, as well as our previous
397 structure-function analysis of IR84a (Abuin et al., 2011). IR75a orthologs across
398 the drosophilid phylogeny conserve all three main agonist-binding residues
399 characteristic of iGluRs (R297, T456, E507 in *Dmel*IR75a) (Benton et al., 2009;
400 Mayer, 2006), suggesting that the core contacts between odors and this receptor
401 are unchanged, and that their mutation might be more likely to lead to a non-
402 functional protein. By contrast, mutation of more distant residues that could, for
403 example, (de)stabilize ligand-induced LBD conformational changes to alter the
404 coupling between ligand binding and channel gating, may be a more subtle way of
405 modifying the tuning profile of these receptors, without the risk of drastic loss of
406 function. Future determination of the mechanistic impact of hotspot mutations will
407 be an important priority, which will likely require experimentally-determined
408 structures of odor-free and odor-bound IR LBDs. Such knowledge may also inform
409 our understanding of the mechanism of ligand-induced gating in iGluRs, where the
410 equivalent position (Y732 (Figure 5C)) is thought to have a function in ligand-gating
411 although its precise role is unclear (Armstrong and Gouaux, 2000; Mamonova et
412 al., 2008).
413 While the hotspot is clearly important, its contribution to modification of
414 tuning properties is shaped by additional changes in these receptors. For IR75a,
415 there are two additional substitutions within the ligand-binding pocket, while for
416 IR75b, one or more residues located on the external surface of the LBD are
417 relevant. In both cases, the functional consequences of combining these
418 substitutions with that of the hotspot are not easily predicted from their individual
419 impact, revealing complex epistatic interactions. Moreover, the distance of these
420 additional sites from the predicted odor-binding site, suggest the existence of
421 allosteric effects of certain sites on odor/receptor interactions. These results are of
422 interest in light of molecular evolutionary analyses of olfactory receptor repertoires,
423 which have identified numerous residues under positive selection (or relaxed
424 purifying selection) in different receptors – implying a contribution to functional
425 divergence – in regions far from the predicted ligand-binding pocket (Arguello et
426 al., 2016; Chen et al., 2010; Gardiner et al., 2009; Smadja et al., 2009; Steiger et
427 al., 2010). Together such observations argue that the evolution of olfactory receptor
428 specificity does not simply arise by alterations in direct interactions of receptors

429 with odor ligands, but rather can emerge from a complex network of interactions of
430 amino acid substitutions – with major or minor effects – both near and far from the
431 ligand-binding pocket.
432 One clear limitation of our study is the restriction of our profiling to a set of
433 linear carboxylic acids at a single concentration. While these ligands are found in
434 nature and are the best agonists for these receptors among many screened odors
435 (Silbering et al., 2011), they necessarily only give a partial insight into the functional
436 changes of individual receptors. The C4 to C6 switch of IR75b in *D. sechellia* is
437 likely related to high abundance of C6 in its sole host fruit (noni). For IR75a the
438 ecological framework is less clear: we speculate that the C4 to C2 switch in the
439 *melanogaster*/*obscura* ancestor may be related to the use of host fruits with acetic-
440 acid producing bacteria, which may then have been no longer relevant for *D.*
441 *sechellia* which preferentially feeds upon ripe (non-fermenting) noni fruit (which
442 contains much more C4 than C2 (Auer et al., 2020; Farine et al., 1996)). Future
443 expansion of the odor profiling of these species' receptors will be essential to
444 understand the pressures in the natural world that have selected for olfactory
445 receptor proteins with new chemical recognition properties. Such knowledge,
446 together with mechanistic insight into ligand/receptor interactions, may be useful to
447 re-engineer the ligand-binding specificities of receptors (for example, as
448 chemogenetic tools (Fukabori et al., 2020)) or design pharmacological
449 manipulators to control the olfactory-guided behaviors of pest insects.
450
451 **Acknowledgements**
452
453 We thank Roman Arguello, Jean Hausser, Michael Shahandeh and other members
454 of the Benton laboratory for discussions and comments on the manuscript. We
455 thank Tom Baden for help with Igor analysis and Liliane Abuin and Steeve Cruchet
456 for technical assistance. H.R.S. was supported by an EMBO Long-Term Fellowship
457 (ALTF 940-2019) and a Helen Hay Whitney Foundation Fellowship. L.L.P.-G.'s
458 laboratory was supported by a European Research Council (ERC) Starting
459 Investigator Grant (802531), and the Francis Crick Institute, which receives its core
460 funding from Cancer Research UK (FC001594), the UK Medical Research Council
461 (FC001594), and the Wellcome Trust (FC001594). Work in R.B.'s laboratory was
462 supported by ERC Consolidator and Advanced Grants (615094 and 833548,
463 respectively), an HFSP Young Investigator Award (RGY0073/2011), the Swiss
464 National Science Foundation Nano-Tera Envirobot project (20NA21_143082). For
465 the purpose of Open Access, the authors have applied a CC BY public copyright
466 license to any Author Accepted Manuscript version arising from this submission.
467
468 **Author contributions**
469
470 L.L.P.-G. and R.B. conceived the project. L.L.P.-G. designed and performed all
471 experiments and analyses, except for the ligand docking studies, which were
472 performed by H.R.S. All authors contributed to interpretation of results. L.L.P.-G.
473 and R.B. wrote the paper, with contributions from H.R.S.

Materials and methods

Drosophila strains and culture

Flies were maintained at 25ºC in 12 h light:12 h dark conditions. We used the following published *D. melanogaster* strains: *Ir84a*^{Gal4} (Grosjean et al., 2011), *UAS-DmelIr75a, UAS-DsecIr75a, UAS-DmelIr75a*^{T289S,Q536K,F538L} (Prieto-Godino et al., 2016). Other drosophilid species were obtained from the *Drosophila* Species Stock Center: *D. sechellia* (14021-0248.25), *D. simulans* (14021-0251.195), *D. yakuba* (14021-0261.01), *D. erecta* (14021-0224.01), *D. ananassae* (14024-0371.13), *D. pseudoobscura* (14011-0121.94), *D. willistoni (*14030-0811.24), *D. mojavensis* (15081-1352.22), *D. virilis* (15010-1051.87).

Molecular biology

cDNAs of *D. melanogaster* and *D. sechellia Ir75a* and *Ir75b* were previously described (Prieto-Godino et al., 2017, 2016). Site-directed mutagenesis was performed using standard procedures, and mutant cDNAs were subcloned into *pUAST attB* for transgenesis of *D. melanogaster* using the phiC31 site-specific integration system (landing site attP40) by BestGene and Genetic Services Inc. All transgenes were sequence-verified both before and after transformation.

Ancestral protein reconstruction

The 16 “modern” sequences of IR75a shown in Figure 2 – figure supplement 1 were used to computationally infer the ancestral sequences at all of the nodes of the tree using the known phylogeny. Ancestral sequences were calculated using the bio-informatic tool FastML (Ashkenazy et al., 2012) with rate variation modelled as a gamma distribution. The marginal posterior probability for most of the amino acids of all of the reconstructed nodes was above 0.8. For the hot-spot mutation the marginal posterior probability was 1. To ascertain the robustness of the inferences made by FastML, we also inferred ancestral sequences of the nodes using a different bioinformatic tool, codeml within the PAML package. The inferred sequences had a global identity of 96.5% and were fully identical at the three amino acid substitutions under study. DNA sequences encoding the IR75a^{Dros} and the IR75a$^{obs\text{-}mel}$ sequences predicted by FastML were synthesized by Eurofins-Genomics, subcloned into *pUAST attB* and transformed into flies as described above.

In vivo electrophysiology

Single sensillum extracellular recordings were performed essentially as described in (Benton and Dahanukar, 2011; Prieto-Godino et al., 2017). For all stimuli, 10 μl odor (10% v/v in water) were use, and presented to the animal in a 1 s pulse. ac2 sensilla were identified in different species by targeting sensilla at antennal locations where ac2 are found in *D. melanogaster*, and using pyridine as a diagnostic odor, which activates the ac2 neuron expressing IR41a and appears to be conserved across species (Figure 1B). CAS numbers and sources of odors are

522 as follows: formic acid (64-18-6), acetic acid (64-19-7), propionic acid (79-09-4),
523 butyric acid (107-92-6), pentanoic acid (109-52-4), hexanoic acid (142-62-1),
524 pyridine (110-86-1) (diagnostic for ac2), octanol (111-87-5) (diagnostic for ac3).
525 Odor-evoked responses were calculated by summing the activity of all OSNs in a
526 sensillum to a given stimulus, as reliable spike-sorting is not possible and other
527 neurons housed in the ac2 or ac4 sensilla do not respond to acidic odors (Silbering
528 et al., 2011). We counted the number of spikes in a 500 ms window at stimulus
529 delivery (200 ms after stimulus onset due to a delay introduced by the air path),
530 subtracted the number of spikes in a 500 ms window 2 s before stimulus delivery,
531 and doubled the result to obtain spikes/s. To calculate solvent-corrected responses
532 (as shown in the figures), we subtracted from the response to each diluted odor,
533 the response obtained when stimulating with the corresponding solvent (water for
534 all odors except for pyridine and octanol, which were dissolved in paraffin oil (8012-
535 95-1). A maximum of three sensilla were tested per animal, and individual
536 genotypes were measured, in an interleaved fashion, on multiple independent
537 days. Normalized responses were calculated by dividing solvent-corrected
538 responses of a given sensillum to each odor by the maximal response of that
539 sensillum, such that each sensillum always had one odor whose normalized
540 response was 1. A pre-requisite for PCA analysis is to input z-scored responses;
541 these were calculated with the in-built function of MATLAB, which works according
542 to the definition of z-scoring, i.e., across each sensillum recording, each response
543 was subtracted from the mean response of all sensillar responses and divided by
544 the standard deviation.
545
546 **Statistical analysis**
547
548 Principal component analysis was carried with in-built MATLAB functions.
549 Statistical analyses were carried out with in-built functions in R Studio or Igor. For
550 all statistical tests, a Shapiro test for normally was first performed. If both samples
551 being compared were normally distributed, a t-test was performed; if one of the
552 samples was not normally distributed, a Wilcoxon test was run. When performing
553 multiple comparisons, p-values were corrected using the Bonferroni method.
554 Epistasis was calculated by determining whether the effects of each individual
555 mutation were added linearly by comparing the observed responses of the
556 combination of two genetic manipulations with the expected distribution of
557 individual mutations were added linearly. Briefly, to generate expected distributions
558 for the linear combination of two mutations, we subtracted the mean responses of
559 the initial receptor from the two “intermediate” receptors, this provides the “effect”
560 of each mutation(s) individually, we then we took 1000 random samples from the
561 “effects” of each of the two mutations and summed them to generate a distribution
562 of the expected responses if the two effects of each mutation added linearly. To
563 determine whether two mutations interacted epistatically, we statistically compared
564 the expected distribution with the actual distribution (the double manipulation
565 combination) using a Wilcoxon test. If the result of this test was $p < 0.05$ after
566 correction for multiple comparisons using Bonferroni method, the two mutations
567 were considered to interact epistatically.
568
569

570 **Ligand docking**
571
572 AutoDock Vina version 1.1.2 (Trott and Olson, 2009) was used to dock ligands into
573 homology models of the LBDs of *Dmel*IR75a and *Dmel*IR84a (K205-N354-L451-
574 W583) (Prieto-Godino et al., 2016) . The crystal structure of *Rnor*GluA2 bound to
575 glutamate (PDB accession 1FTJ) was used to optimize the grid conditions and
576 docking procedure for IRs. A12×12×12 Å grid was placed along the axis between
577 the CZ atom in a conserved arginine residue (R96 in *Rnor*GluA2, R100 in
578 *Dmel*IR75a) and the CD atom in a conserved glutamate residue (E193 in
579 *Rnor*GluA2, E197 in the IR75a, this glutamate is not conserved in *Dmel*IR84a). The
580 center of the grid was placed closer to the arginine than to the glutamate, such that
581 if a straight line was drawn from the arginine to the glutamate, the grid center was
582 one-third of the way along this line. For *Dmel*IR84a, the conserved glutamate is not
583 present, and the position of the grid was manually determined using the conserved
584 arginine (R317) as a guide and the grids for RnGluA2 and IR75a as reference.
585 Docking was performed using an exhaustiveness of 8, with the default settings for
586 AutoDock Vina; these were sufficient to accurately predict the position of the co-
587 crystallized glutamate in RnGluA2 (Figure 4 – figure supplement 1). Ligand poses
588 were inspected manually, but in all cases the top-ranked pose was judged to be
589 the most plausible. Structure visualization was performed using PyMol version
590 2.3.3. The *Dsec*IR75b LBD model (Prieto-Godino et al., 2017) was aligned to the
591 GluA2 homotetramer using the “align” command in PyMol.
592
593

Figure Legends

Figure 1. Evolution of olfactory responses of Ir75a neurons across the drosophilid phylogeny.
(**A**) Phylogeny of analyzed drosophilid species and their known ecological niches (Markow, 2015).
(**B**) Representative traces of extracellular recordings of neuronal responses in antennal coeloconic 2 (ac2) sensilla (schematized top left) to the indicated odors in different drosophilid species. ac2 sensilla were identified based on their morphology, their location on the antennal surface – either near the entry to the sacculus or in the distal posterior part intermingled with ac3 sensilla (Silbering et al., 2011) – their pattern of basal firing, and the conserved responses of the Ir41a neuron towards pyridine and the lack of response to octanol (which is detected by the Or35a neuron in ac3 sensilla (Yao et al., 2005)).
(**C**) Heatmap of electrophysiological solvent-corrected responses (see Materials and methods) of ac2 sensilla of all species in (**A**) to a series of linear carboxylic acids; the color-scale is on the right. Each rectangle represents the responses measured in a single sensillum.
(**D**) Loadings of the first principal component of all responses shown in (**C**).
(**E**) Responses of ac2 sensilla for all species plotted as z-scored responses (see Material and methods) of C2 against z-scored responses to C4. This plotting reveals two clear clusters with species responding maximally to C2 (orange) or C4 (blue).

Figure 1 Source data 1. Data for Figure 1 showing the solvent-corrected spikes/second responses of ac2 sensilla for each of the species (Figure 1C). The same responses z-scored (Figure 1E) and the results of the PCA analysis, showing loadings and variance explained by each PC (Figure 1D). Here and in subsequent source data files “NaN” (“not a number”) indicates cases where data was not available, usually due to the electrode coming out of the sensillum before the end of the series of stimulations.

Figure 2. Ancestral sequence reconstruction of IR75a.
(**A**) Phylogeny of the protein sequences of drosophilid IR75a orthologs used for the reconstruction of ancestral nodes ($IR75a^{Dros}$ and $IR75a^{mel\text{-}obs}$) (see Figure 2 – figure supplement 1). The orange and blue dots indicate maximal responses of the corresponding neurons to C2 and C4, respectively (Figure 1C). The branches of the tree have been similarly color-coded according to predicted receptor responses based on parsimony.
(**B**) Representative traces of extracellular recordings of neuronal responses to the indicated odors of receptors expressed in the IR decoder neuron (see text). Genotypes are of the form: *UAS-xxx/UAS-xxx;*$Ir84a^{Gal4}/Ir84a^{Gal4}$, here and in all subsequent figures.
(**C**) Quantification of responses of the indicated receptors expressed in the IR decoder neuron. In this and subsequent similar figure panels, the top row shows barplots of responses normalized by maximal response, and the bottom row shows individual datapoints, mean and SEM of raw solvent-corrected responses to odor stimuli.

641
642 **Figure 2 – figure supplement 1. Alignment of IR75a orthologs.**
643 Multiple sequence alignment of IR75a orthologs used for the ancestral protein
644 reconstruction. The ancestral versions that were functionally tested, IR75a^{Dros} and
645 IR75a$^{obs\text{-}mel}$, are also shown. Dark blue indicates identical residues, light blue
646 indicates similar residues, grey indicates dissimilar residues. The red lines indicate
647 the spans of S1 and S2 lobes of the LBD. The amino acids colored in red are those
648 positions that were estimated by FASTML (see Materials and methods) to have a
649 marginal posterior probability < 0.7.
650
651 **Figure 2 Source data 1.** Data for Figure 1C, responses in spikes/s of ac2 sensilla
652 from each of the genotypes, and normalized with respect the maximal response of
653 each sensilla.
654
655 **Figure 3. Mapping the evolutionary landscape of IR75a.**
656 (**A**) Top: cartoon of the domain organization of IR75a; bottom: protein homology
657 model of the *Dmel*IR75a LBD (adapted from (Prieto-Godino et al., 2016)). The three
658 functionally-important amino acid positions that differ between *Dmel*IR75a and
659 *Dsec*IR75a are highlighted in red.
660 (**B-C**) Quantification of responses of the indicated receptor versions expressed in
661 the IR decoder neuron.
662 (**D**) Loadings of the first principal component of all responses shown in (**C**).
663 (**E**) Responses of each of the indicated receptor versions plotted in the C2 vs C4
664 space.
665 (**F**) Visualization of epistasis and accessible mutational pathways. Each of the
666 individual, double and triple mutations are plotted in the PC1 axis (error bars are
667 standard error of the mean). Possible evolutionary paths join these with lines. Solid
668 lines indicate when the path joins two points that significantly increase PC1 value
669 (i.e. increased responses to C4 and C5 and decreased responses to C2); dashed
670 lines denote paths that, while accessible, do not lead to significantly increased PC1
671 values (see Source Data, for statistical values). Red lines indicate the cases where
672 two mutations interact epistatically when combined, *i.e.*, the combination of the two
673 mutations is not equal to the expected response if their effects added linearly (see
674 Materials and methods for details and Source Data for statistical values).
675
676 **Figure 3 Source data 1.** First tab: Data for Figure 3B and 3C, responses in
677 spikes/second of ac2 sensilla from each of the genotypes indicated, and
678 normalized with respect the maximal response of each sensilla. Second tab:
679 Results from the PCA analysis. Third tab: P-values resulting from the statistical
680 analysis in Figure 3F.
681
682 **Figure 4. A hotspot for tuning IR sensitivity.**
683 (**A**) Phylogeny of species' IR75a receptors used for the reconstruction of ancestral
684 nodes. The inferred identity of hotspot mutation for key nodes is shown as well as
685 for each of the extant receptors at the end of each leaf, together with the measured
686 (solid circle) or predicted (empty circle) sensitivity. On the right, the nucleotides
687 encoding the hotspot amino acid position in IR75a across species. *Drosophila*
688 species abbreviations (where not presented in Figure 1): *Dmir* (*miranda*), *Dper*

(*persimilis*), *Dame* (*americana*), *Dbuz* (*buzzatii*), *Dgri* (*grimshawi*), *Dalb* (*albomicans*).
(**B**) Top: Protein sequence alignment of *Dmel*IR75a and *Dmel*IR75b. Dark blue indicates identical residues, light blue indicates similar residues, grey indicates dissimilar residues. The red lines indicate the spans of S1 and S2 lobes of the LBD. The red box indicates the common amino acid position ("hotspot"). Bottom: separate alignments of the hotspot region in IR75a and IR75b for *D. melanogaster* and *D. sechellia* proteins showing that while the residue position is conserved, the identity of the amino acids is different for these two receptors.
(**C**) Left: homology model of *Dmel*IR75a LBD (blue) with docked ligands; the insets show the top-ranked poses for C2 (cyan) and C4 (purple) in the LBD pocket. Right: crystal structure of the *Rattus norvegicus* (*Rnor*) iGluR GluA2 LBD (gray) in complex with glutamate (yellow) (PDB accession 1FTJ). The hotspot residue is shown in orange in *Dmel*IR75a, and the equivalent residue in *Rnor*GluA2 in magenta. Amino acid numbering corresponds to that of the full-length sequence.

Figure 4 – figure supplement 1. Ligand docking in IRs and iGluRs.
Top-three docked poses for the indicated ligands in the *Dmel*IR75a and *Dmel*IR84a LBD homology models, and the *Rnor*GluA2/glutamate crystal structure (PDB 1FTJ), ranked by AutoDock Vina docking score. In *Dmel*IR75a, all three poses of C2 occupy a similar space in the binding pocket, contacting the conserved arginine R297 and >4 Å from the hotspot residue 538 (orange). Similarly, in *Dmel*IR84a, all three poses of phenylacetic acid occupy a similar position in the binding pocket and make contact with the conserved arginine R317, which is essential for ligand-evoked currents from IR84a *in vivo* (Abuin et al., 2011). In *Rnor*GluA2, the docked poses of docked glutamate (green), particularly the top-ranked pose, are similar to the positioning of co-crystallized glutamate (yellow). Amino acid numbering corresponds to that of the full-length sequences.

Figure 5. Epistasis in the ligand binding domain of IR75b.
(**A**) Quantification of responses of the indicated wildtype and mutated IR75b variants expressed in the IR decoder neuron. Note that the *Dmel*IR75b^{DsecS1} variant includes the six derived changes between *D. sechellia* and *D. simulans/D. melanogaster* as well as a seventh amino acid substitution common to *D. sechellia* and *D. simulans* receptors.
(**B**) Protein homology model of the *Dsec*IR75b LBD (adapted from (Prieto-Godino et al., 2017)). The S1 and S2 lobes are colored pale and dark green, respectively. Surface-located amino acids that differ between *Dmel*IR75b and *Dsec*IR75b in the S1 domain are depicted in magenta; the hotspot (position 523) and other residues in the ligand-binding pocket in the S2 domain are depicted in red.
(**C**) Homology model of the *Dmel*IR75b LBD (colored as in (**B**)) aligned to the full-length homotetrameric crystal structure of *Rnor*GluA2 (PDB accession 3KG2), which is colored by chain in faded gray, red, blue, and yellow. The *Rnor*GluA2 amino-terminal domain (which is not present in most IRs (Rytz et al., 2013)) has been removed for clarity.

Figure 5 Source data 1. First tab: Data for Figure 5A, responses in spikes/second of ac2 sensilla from each of the genotypes indicated, and normalized with respect

to the maximal response of each sensilla. Second tab: P-values resulting from the statistical analysis, as discussed in the text.

Supplementary file 1. Nucleotide sequences encoding the reconstructed ancestral IR75a orthologs.

References

Abuin L, Bargeton B, Ulbrich MH, Isacoff EY, Kellenberger S, Benton R. 2011. Functional Architecture of Olfactory Ionotropic Glutamate Receptors. *Neuron*.

Abuin L, Prieto-Godino LL, Pan H, Gutierrez C, Huang L, Jin R, Benton R. 2019. In vivo assembly and trafficking of olfactory Ionotropic Receptors. *BMC Biol* **17**. doi:10.1186/s12915-019-0651-7

Adipietro KA, Mainland JD, Matsunami H. 2012. Functional evolution of Mammalian odorant receptors. *PLoS Genet* **8**:e1002821. doi:10.1371/journal.pgen.1002821

Aguilar-Rodríguez J, Payne JL, Wagner A. 2017. A thousand empirical adaptive landscapes and their navigability. *Nat Ecol Evol*. doi:10.1038/s41559-016-0045

Amlou M, Moreteau B, David JR. 1998. Genetic analysis of Drosophila sechellia specialization: Oviposition behavior toward the major aliphatic acids of its host plant. *Behav Genet*. doi:10.1023/A:1021689312582

Arguello JR, Cardoso-Moreira M, Grenier JK, Gottipati S, Clark AG, Benton R. 2016. Extensive local adaptation within the chemosensory system following Drosophila melanogaster's global expansion. *Nat Commun* **7**:11855. doi:10.1038/ncomms11855

Armstrong N, Gouaux E. 2000. Mechanisms for Activation and Antagonismof an AMPA-Sensitive Glutamate Receptor: Crystal Structures of the GluR2 Ligand Binding Core. *Neuron*.

Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O, Pupko T. 2012. FastML: A web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res*. doi:10.1093/nar/gks498

Auer TO, Khallaf MA, Silbering AF, Zappia G, Ellis K, Álvarez-Ocaña R, Arguello JR, Hansson BS, Jefferis GSXE, Caron SJC, Knaden M, Benton R. 2020. Olfactory receptor and circuit evolution promote host specialization. *Nature*. doi:10.1038/s41586-020-2073-7

Bargmann CI. 2006. Comparative chemosensation from receptors to ecology. *Nature*. doi:10.1038/nature05402

Becher PG, Bengtsson M, Hansson BS, Witzgall P. 2010. Flying the fly: Long-range flight behavior of drosophila melanogaster to attractive odors. *J Chem Ecol*. doi:10.1007/s10886-010-9794-2

Benton R, Dahanukar A. 2011. Electrophysiological recording from Drosophila olfactory Sensilla. *Cold Spring Harb Protoc*. doi:10.1101/pdb.prot5630

Benton R, Vannice KS, Gomez-Diaz C, Vosshall LB. 2009. Variant ionotropic glutamate receptors as chemosensory receptors in Drosophila. *Cell*.

Block E. 2018. Molecular Basis of Mammalian Odor Discrimination: A Status Report. *J Agric Food Chem*. doi:10.1021/acs.jafc.8b04471

Butterwick JA, del Mármol J, Kim KH, Kahlson MA, Rogow JA, Walz T, Ruta V. 2018. Cryo-EM structure of the insect olfactory receptor Orco. *Nature*. doi:10.1038/s41586-018-0420-8

Carey AF, Wang G, Su C-YY, Zwiebel LJ, Carlson JR. 2010. Odorant reception in the malaria mosquito Anopheles gambiae. *Nature* 1–7. doi:10.1038/nature08834

Chen M, Peng ZG, He SP. 2010. Olfactory receptor gene family evolution in stickleback and medaka fishes. *Sci China Life Sci*. doi:10.1007/s11427-010-0025-4

Croset V, Rytz R, Cummins SF, Budd A, Brawand D, Kaessmann H, Gibson TJ, Benton R. 2010. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet*.

Dekker T, Ibba I, Siju KP, Stensmyr MC, Hansson BS. 2006. Olfactory shifts parallel superspecialism for toxic fruit in Drosophila melanogaster sibling, D. sechellia. *Curr Biol* **16**:101–109. doi:10.1016/j.cub.2005.11.075

Del Mármol J, Yedlin M, Ruta V. 2021. The structural basis of odorant recognition in insect olfactory receptors. *bioRxiv*.

Farine J-PP, Legal L, Moreteau B, Le Quere jean-LL. 1996. Volatile components of ripe fruits of Morinda citrifolia and their effects on Drosophila. *Phytochemistry* 1–6. doi:10.1016/0031-9422(95)00455-6

Fukabori R, Iguchi Y, Kato S, Takahashi K, Eifuku S, Tsuji S, Hazama A, Uchigashima M, Watanabe M, Mizuma H, Cui Y, Onoe H, Hikishima K, Yasoshima Y, Osanai M, Inagaki R, Fukunaga K, Nishijo T, Momiyama T, Benton R, Kobayashi K. 2020. Enhanced retrieval of taste associative memory by chemogenetic activation of locus coeruleus norepinephrine neurons. *J Neurosci*. doi:10.1523/JNEUROSCI.1720-20.2020

Gardiner A, Butlin RK, Jordan WC, Ritchie MG. 2009. Sites of evolutionary divergence differ between olfactory and gustatory receptors of Drosophila. *Biol Lett*. doi:10.1098/rsbl.2008.0723

Gorter JA, Jagadeesh S, Gahr C, Boonekamp JJ, Levine JD, Billeter J-CC. 2016. The nutritional and hedonic value of food modulate sexual receptivity in Drosophila melanogaster females. *Sci Rep* **6**:19441. doi:10.1038/srep19441

Grosjean Y, Rytz R, Farine J-P, Abuin L, Cortot J, Jefferis GSXE, Benton R. 2011. An olfactory receptor for food-derived odours promotes male courtship in Drosophila. *Nature*.

Hallem EA, Carlson JR. 2006. Coding of Odors by a Receptor Repertoire. *Cell* **125**:143–160. doi:10.1016/j.cell.2006.01.050

Holder KK, Bull JJ. 2001. Profiles of adaptation in two similar viruses. *Genetics*.

Joseph RM, Devineni A V., King IFGG, Heberlein U. 2009. Oviposition preference for and positional avoidance of acetic acid provide a model for competing behavioral drives in Drosophila. *Proc Natl Acad Sci U S A* **106**:11352–11357. doi:10.1073/pnas.0901419106

Kim G, Huang JH, McMullen JG, Newell PD, Douglas AE. 2018. Physiological responses of insects to microbial fermentation products: Insights from the interactions between Drosophila and acetic acid. *J Insect Physiol*. doi:10.1016/j.jinsphys.2017.05.005

Leary GP, Allen JE, Bunger PL, Luginbill JB, Linn CE, Macallister IE, Kavanaugh

832 MP, Wanner KW. 2012. Single mutation to a sex pheromone receptor
833 provides adaptive specificity between closely related moth species. *Proc Natl*
834 *Acad Sci U S A* **109**:14081–14086. doi:10.1073/pnas.1204661109
835 Mainland JD, Keller A, Li YR, Zhou T, Trimmer C, Snyder LL, Moberly AH,
836 Adipietro KA, Liu WLL, Zhuang H, Zhan S, Lee SS, Lin A, Matsunami H.
837 2013. The missense of smell: Functional variability in the human odorant
838 receptor repertoire. *Nat Neurosci* **17**:114–120. doi:10.1038/nn.3598
839 Mamonova T, Speranskiy K, Kurnikova M. 2008. Interplay between structural
840 rigidity and electrostatic interactions in the ligand binding domain of GluR2.
841 *Proteins Struct Funct Genet* **73**:656–671. doi:10.1002/prot.22090
842 Markow TA. 2015. The secret lives of Drosophila flies. *Elife* **4**.
843 doi:10.7554/eLife.06793
844 Mayer ML. 2006. Glutamate receptors at atomic resolution. *Nature*.
845 doi:10.1038/nature04709
846 Ni L. 2021. The Structure and Function of Ionotropic Receptors in Drosophila.
847 *Front Mol Neurosci*. doi:10.3389/fnmol.2020.638839
848 Orr HA. 2005. The genetic theory of adaptation: A brief history. *Nat Rev Genet*.
849 doi:10.1038/nrg1523
850 Phillips PC. 2008. Epistasis - The essential role of gene interactions in the
851 structure and evolution of genetic systems. *Nat Rev Genet*.
852 doi:10.1038/nrg2452
853 Prieto-Godino LL, Rytz R, Bargeton B, Abuin L, Arguello JR, Peraro MD, Benton
854 R. 2016. Olfactory receptor pseudo-pseudogenes. *Nature* **539**.
855 doi:10.1038/nature19824
856 Prieto-Godino LL, Rytz R, Cruchet S, Bargeton B, Abuin L, Silbering AF, Ruta V,
857 Dal Peraro M, Benton R. 2017. Evolution of Acid-Sensing Olfactory Circuits
858 in Drosophilids. *Neuron* **93**. doi:10.1016/j.neuron.2016.12.024
859 Ramdya P, Benton R. 2010. Evolving olfactory systems on the fly. *Trends Genet*.
860 Randall RN, Radford CE, Roof KA, Natarajan DK, Gaucher EA. 2016. An
861 experimental phylogeny to benchmark ancestral sequence reconstruction.
862 *Nat Commun*. doi:10.1038/ncomms12847
863 Rytz R, Croset V, Benton R. 2013. Ionotropic Receptors (IRs): Chemosensory
864 ionotropic glutamate receptors in Drosophila and beyond. *Insect Biochem*
865 *Mol Biol*.
866 Saito H, Chi Q, Zhuang H, Matsunami H, Mainland JD. 2009. Odor coding by a
867 mammalian receptor repertoire. *Sci Signal*. doi:10.1126/scisignal.2000016
868 Silbering AF, Rytz R, Grosjean Y, Abuin L, Ramdya P, Jefferis GSXE, Benton R.
869 2011. Complementary function and integrated wiring of the evolutionarily
870 distinct Drosophila olfactory subsystems. *J Neurosci*.
871 Smadja C, Shi P, Butlin RK, Robertson HM. 2009. Large gene family expansions
872 and adaptive evolution for odorant and gustatory receptors in the pea aphid,
873 Acyrthosiphon pisum. *Mol Biol Evol*. doi:10.1093/molbev/msp116
874 Steiger SS, Fidler AE, Mueller JC, Kempenaers B. 2010. Evidence for adaptive
875 evolution of olfactory receptor genes in 9 bird species. *J Hered*.
876 doi:10.1093/jhered/esp105
877 Trott O, Olson AJ. 2009. AutoDock Vina: Improving the speed and accuracy of
878 docking with a new scoring function, efficient optimization, and
879 multithreading. *J Comput Chem*. doi:10.1002/jcc.21334

880 Wright S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in
881 evolution. *Sixth Int Congr Genet*.
882 Yang K, Huang LQ, Ning C, Wang CZ. 2017. Two single-point mutations shift the
883 ligand selectivity of a pheromone receptor between two closely related moth
884 species. *Elife*. doi:10.7554/eLife.29100
885 Yao CA, Ignell R, Carlson JR. 2005. Chemosensory coding by neurons in the
886 coeloconic sensilla of the Drosophila antenna. *J Neurosci* **25**:8359–8367.
887 doi:10.1523/JNEUROSCI.2432-05.2005
888

Figure 1

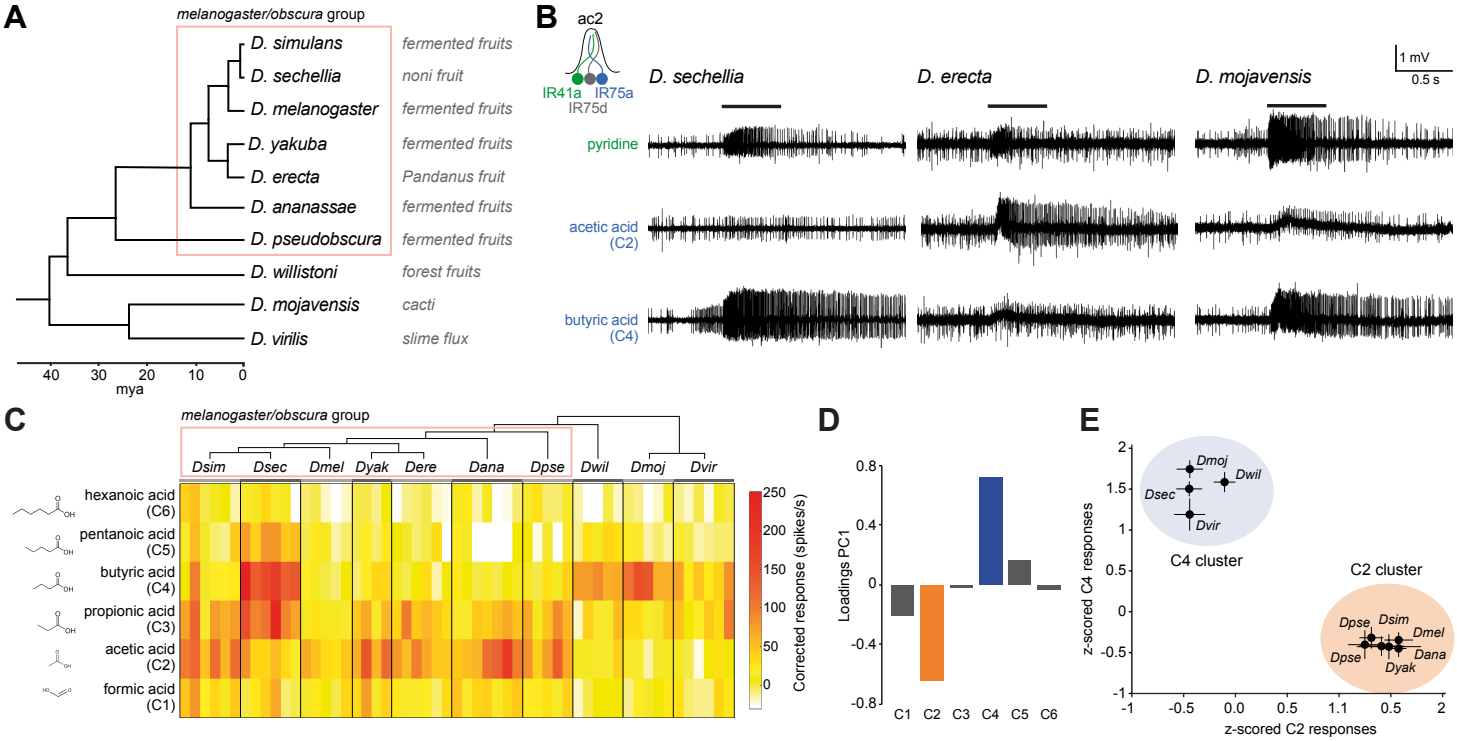

A
melanogaster/obscura group
D. simulans
D. sechellia
D. melanogaster
D. yakuba
D. erecta
D. ananassae
D. pseudobscura
D. willistoni
D. mojavensis
D. virilis
fermented fruits
noni fruit
fermented fruits
fermented fruits
Pandanus fruit
fermented fruits
fermented fruits
forest fruits
cacti
slime flux
40 30 20 10 0
mya
B
ac2
IR41a IR75a
IR75d
D. sechellia
D. erecta
D. mojavensis
1 mV
0.5 s
pyridine
acetic acid (C2)
butyric acid (C4)
C
melanogaster/obscura group
Dsim Dsec Dmel Dyak Dere Dana Dpse Dwil Dmoj Dvir
hexanoic acid (C6)
pentanoic acid (C5)
butyric acid (C4)
propionic acid (C3)
acetic acid (C2)
formic acid (C1)
250 200 150 100 50 0
Corrected response (spikes/s)
D
Loadings PC1
0.8 0.4 0 -0.4 -0.8
C1 C2 C3 C4 C5 C6
E
z-scored C4 responses
z-scored C2 responses
Dmoj
Dwil
Dsec
Dvir
C4 cluster
C2 cluster
Dpse
Dsim
Dmel
Dana
Dyak
Dpse

Figure 2

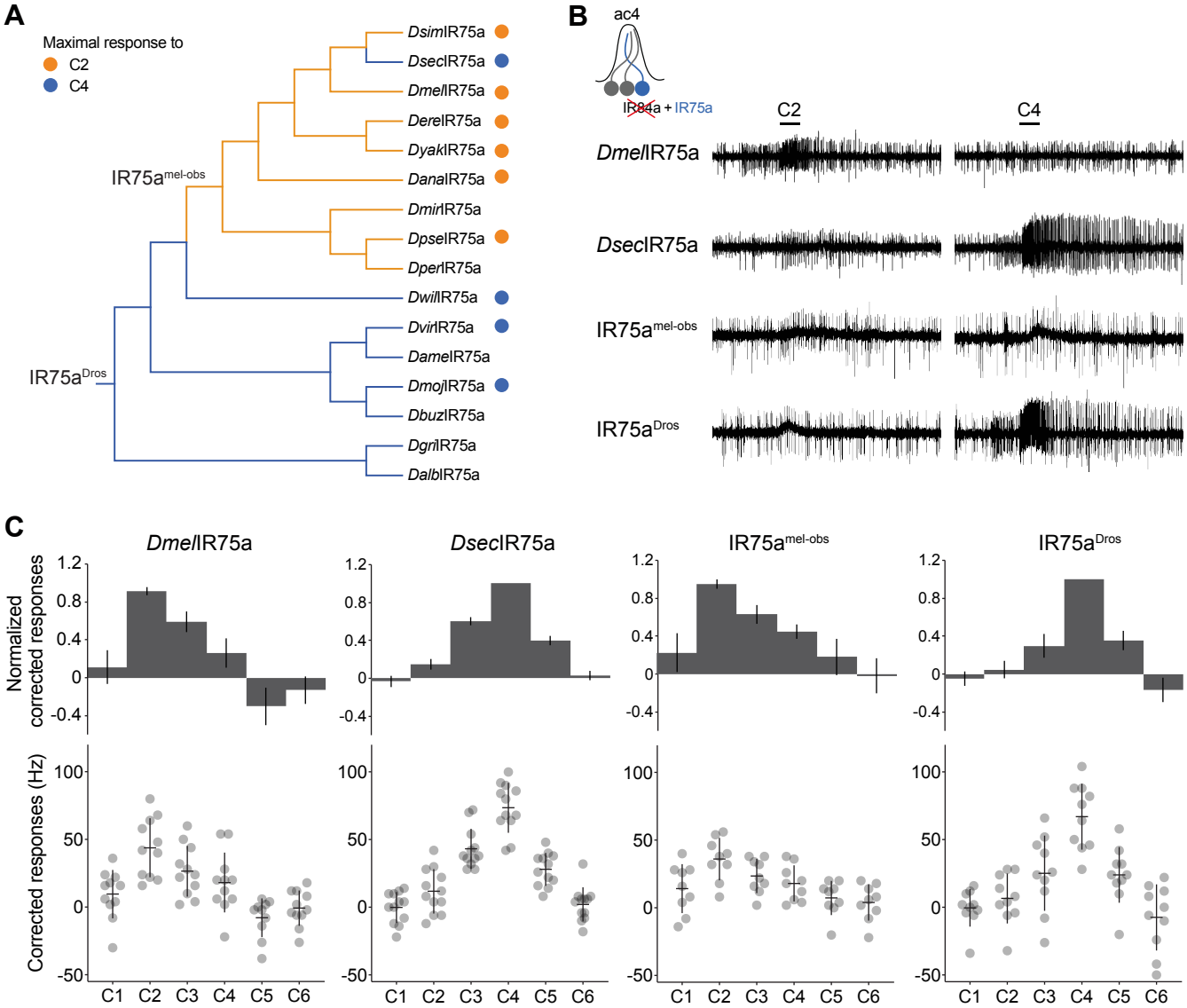

A
Maximal response to
C2
C4
DsimIR75a
DsecIR75a
DmelIR75a
DereIR75a
DyakIR75a
DanaIR75a
DmirIR75a
DpseIR75a
DperIR75a
DwilIR75a
DvirIR75a
DamelIR75a
DmojIR75a
DbuzIR75a
DgriIR75a
DalbIR75a
IR75a mel-obs
IR75a Dros
B
ac4
IR84a + IR75a
C2
C4
DmelIR75a
DsecIR75a
IR75a mel-obs
IR75a Dros
C
DmelIR75a
DsecIR75a
IR75a mel-obs
IR75a Dros
Normalized corrected responses
Corrected responses (Hz)
C1 C2 C3 C4 C5 C6

Figure 2 - figure supplement 1

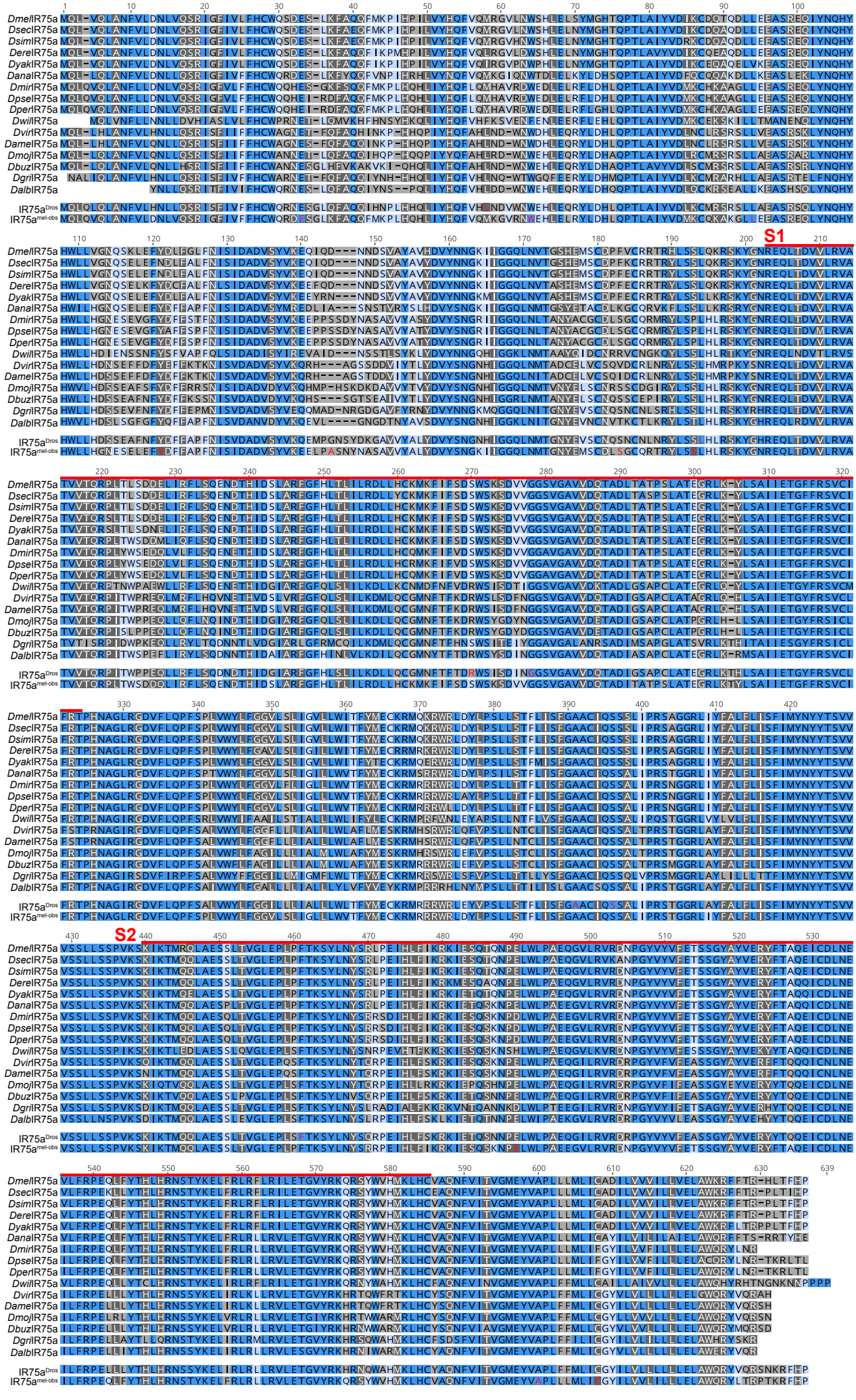


Figure 3

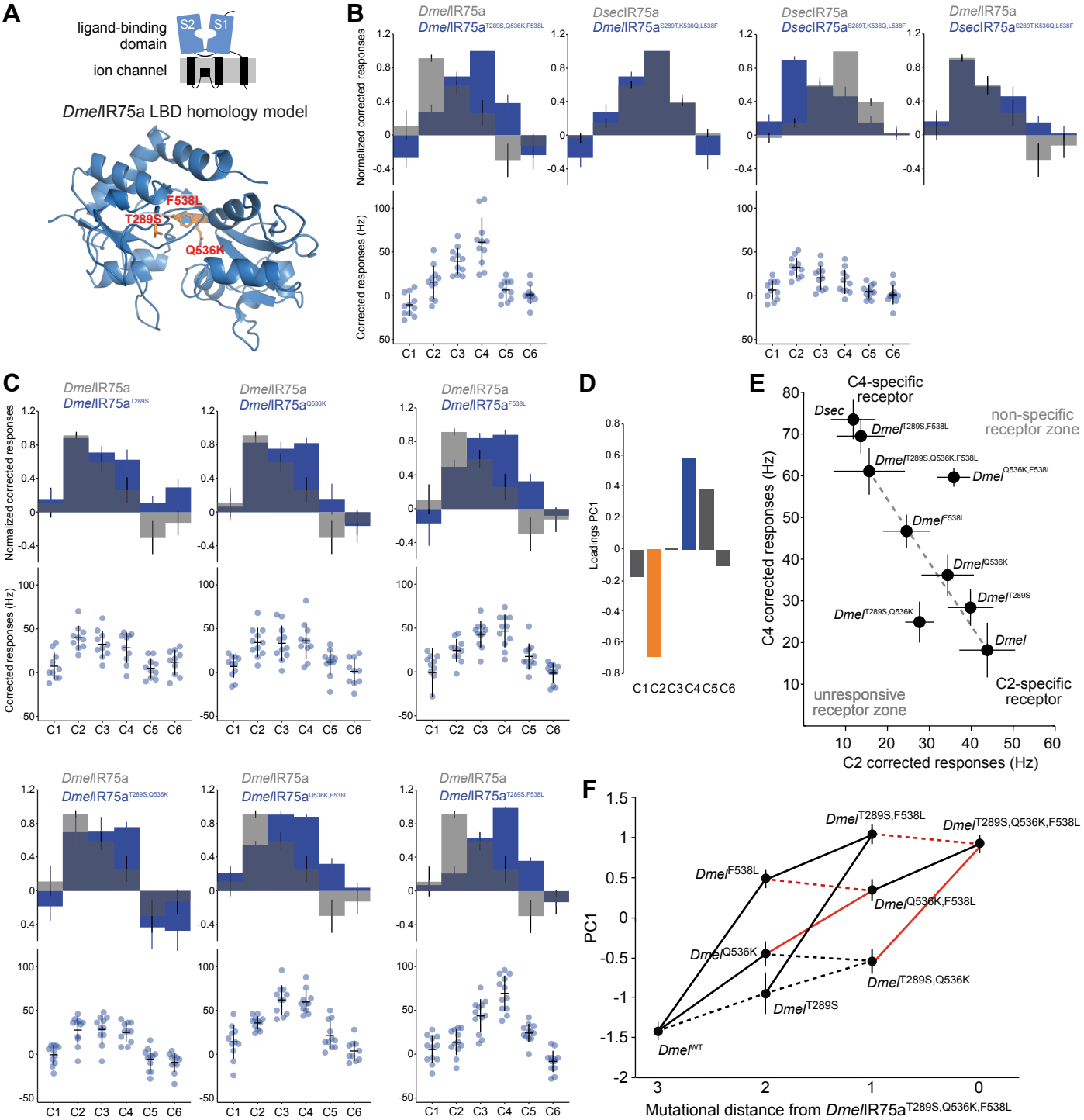

A
ligand-binding domain
S2
S1
ion channel
DmelIR75a LBD homology model
F538L
T289S
Q536K
B
DmelIR75a
DmelIR75a T289S,Q536K,F538L
DsecIR75a
DmelIR75a S289T,K536Q,L538F
DsecIR75a
DsecIR75a S289T,K536Q,L538F
DmelIR75a
DsecIR75a S289T,K536Q,L538F
Normalized corrected responses
Corrected responses (Hz)
C1 C2 C3 C4 C5 C6
C
DmelIR75a
DmelIR75a T289S
DmelIR75a
DmelIR75a Q536K
DmelIR75a
DmelIR75a F538L
DmelIR75a
DmelIR75a T289S,Q536K
DmelIR75a
DmelIR75a Q536K,F538L
DmelIR75a
DmelIR75a T289S,F538L
D
Loadings PC1
C1 C2 C3 C4 C5 C6
E
C4-specific receptor
non-specific receptor zone
unresponsive receptor zone
C2-specific receptor
Dsec
DmelT289S,F538L
DmelT289S,Q536K,F538L
DmelQ536K,F538L
DmelF538L
DmelQ536K
DmelT289S
DmelT289S,Q536K
Dmel
C4 corrected responses (Hz)
C2 corrected responses (Hz)
F
DmelT289S,F538L
DmelT289S,Q536K,F538L
DmelF538L
DmelQ536K,F538L
DmelQ536K
DmelT289S,Q536K
DmelT289S
DmelWT
PC1
Mutational distance from DmelIR75a T289S,Q536K,F538L

Figure 4

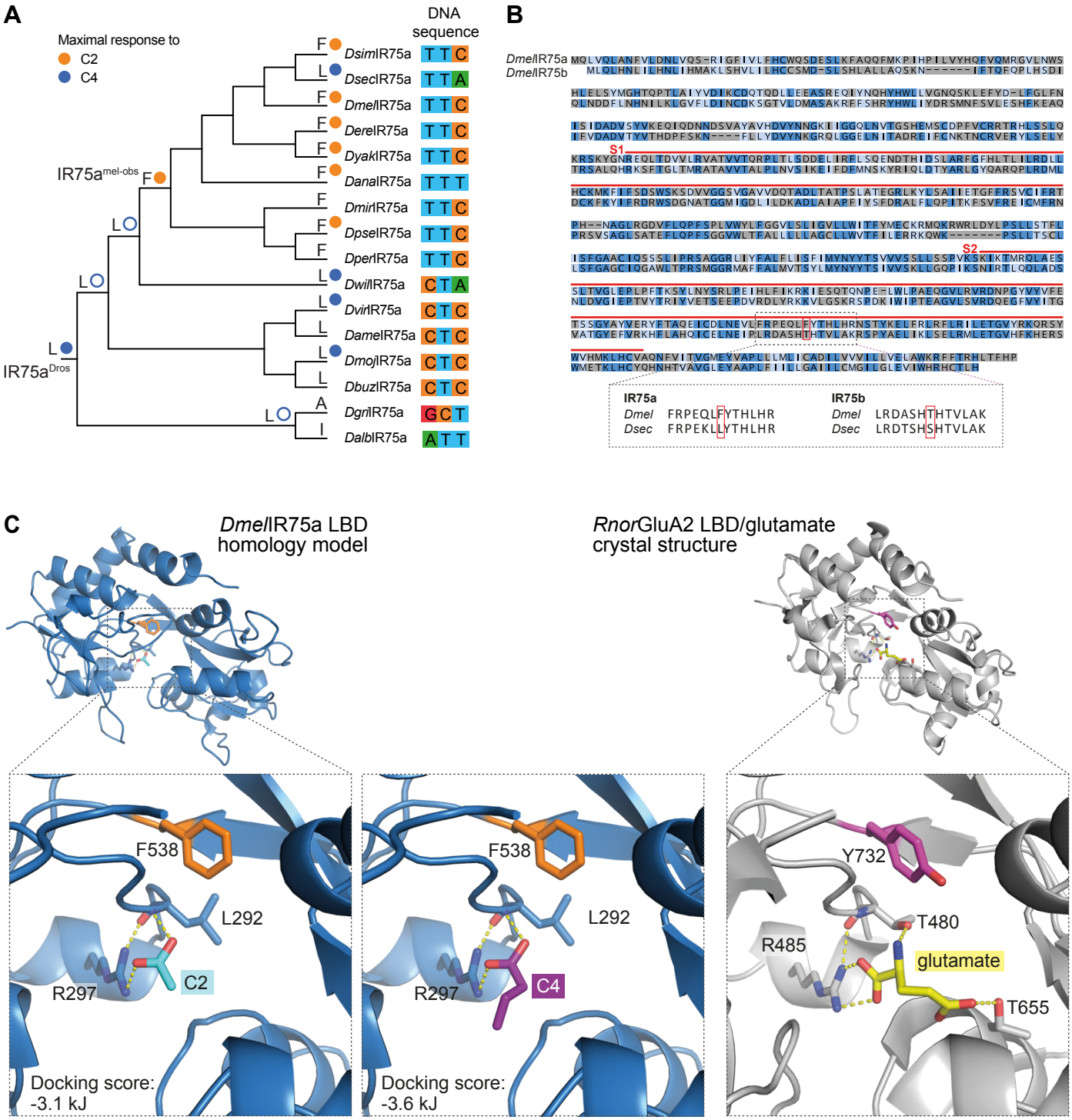

A
Maximal response to
C2
C4
DNA sequence
IR75a^mel-obs
IR75a^Dros
DsimIR75a T T C
DsecIR75a T T A
DmelIR75a T T C
DereIR75a T T C
DyakIR75a T T C
DanaIR75a T T T
DmirIR75a T T C
DpseIR75a T T C
DperIR75a T T C
DwilIR75a C T A
DvirIR75a C T C
DameIR75a C T C
DmojIR75a C T C
DbuzIR75a C T C
DgriIR75a G C T
DalbIR75a A T T
B
DmelIR75a
DmelIR75b
S1
S2
IR75a
Dmel FRPEQLFYTHLHR
Dsec FRPEKLLYTHLHR
IR75b
Dmel LRDASHTHTVLAK
Dsec LRDTSHSHTVLAK
C
DmelIR75a LBD homology model
RnorGluA2 LBD/glutamate crystal structure
F538
L292
R297
C2
C4
Docking score: -3.1 kJ
Docking score: -3.6 kJ
Y732
T480
R485
glutamate
T655

Figure 4 - figure supplement 1

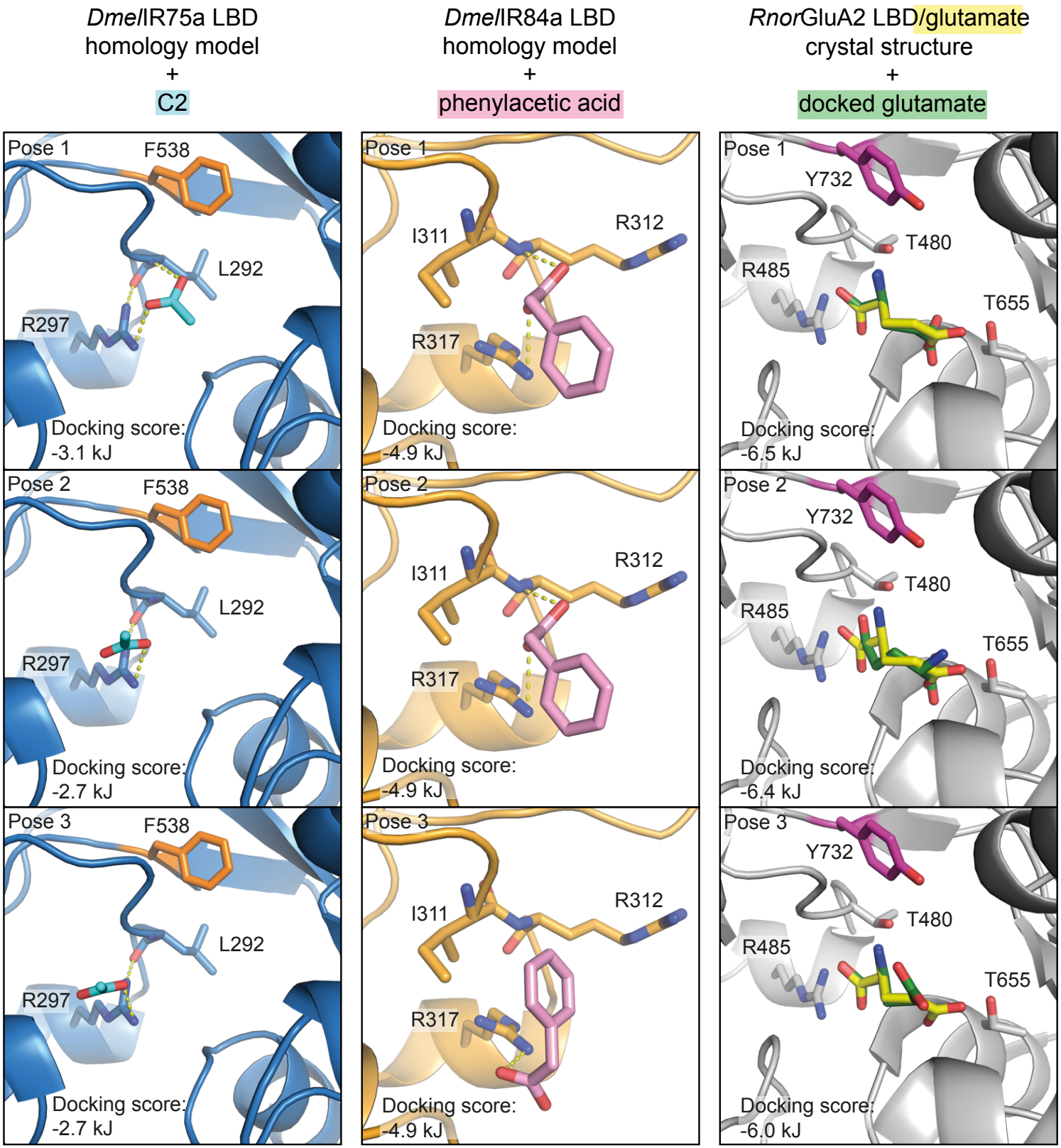
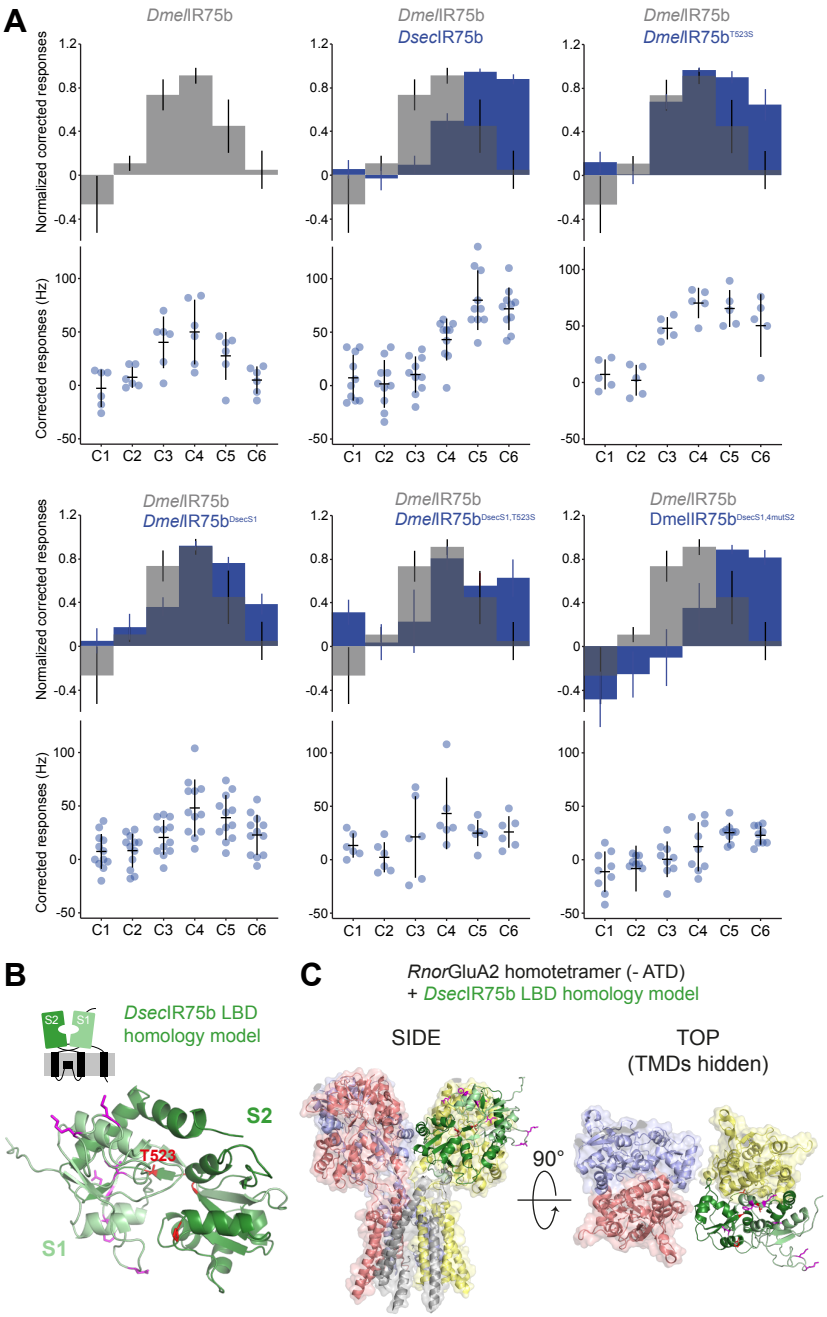


Figure 5


A
DmelIR75b
DmelIR75b
DsecIR75b
DmelIR75b
DmelIR75b T523S
DmelIR75b
DmelIR75b DsecS1
DmelIR75b
DmelIR75b DsecS1,T523S
DmelIR75b
DmelIR75b DsecS1,4mutS2
Normalized corrected responses
Corrected responses (Hz)
C1 C2 C3 C4 C5 C6
B
DsecIR75b LBD homology model
S2
S1
T523
C
RnorGluA2 homotetramer (- ATD)
+ DsecIR75b LBD homology model
SIDE
TOP
(TMDs hidden)
90°