The Origin and Evolution of Sex Peptide and Sex Peptide Receptor Interactions

Junhui Peng[1], Nicolas Svetec[1], Henrik Molina[2], Li Zhao[1,*]

[1.] Laboratory of Evolutionary Genetics and Genomics, The Rockefeller University, New York, NY, USA
[2.] Proteomics Resource Center, The Rockefeller University, New York, NY 10065, USA

*Correspondence to: lzhao@rockefeller.edu

Abstract

Post-mating responses play a vital role in successful reproduction across diverse species. In fruit flies, sex peptide (SP) binds to the sex peptide receptor (SPR), triggering a series of post-mating responses. However, the origin of SPR predates the emergence of SP. The evolutionary origins of the interactions between SP and SPR and the mechanisms by which they interact remain enigmatic. In this study, we used ancestral sequence reconstruction, AlphaFold2 predictions, and molecular dynamics simulations to study SP-SPR interactions and their origination. Using AlphaFold2 and long-time molecular dynamics (MD) simulations, we predicted the structure and dynamics of SP-SPR interactions. We show that SP potentially binds to the ancestral states of Diptera SPR. Notably, we found that only a few amino acid changes in SPR are sufficient for the formation of SP-SPR interactions. Ancestral sequence reconstruction and MD simulations further reveal that SPR interacts with SP through residues that are mostly involved in the interaction interface of an ancestral ligand, myoinhibitory peptides (MIPs). We propose a potential mechanism whereby SP-SPR interactions arise from the pre-existing MIP-SPR interface as well as early chance events both inside and outside the pre-existing interface that created novel SP-specific SP-SPR interactions. Our findings provide new insights into the origin and evolution of SP-SPR interactions and their relationship with MIP-SPR interactions.

Introduction

Reproductive success plays a crucial role for individuals, populations, and species. Predominantly, males of many species undergo intense sexual selection because of their necessity to secure mating partners successfully, and higher intrasexual variance in mating success among males. Inter-male competition emerges as a potent force propelling sexual selection. This competition is manifested at the phenotypic level through distinct male attributes such as large horns in beetles or vivid colors in birds (Eberhard, 1985; Hopkins & Kopp, 2021; Parker, 1970). Moreover, at the molecular level, this competitive landscape stimulates the accelerated evolution of genes, which may subsequently contribute to the development of phenotypes that either precipitate or resolve sexual conflict (Ferris et al., 1997; Haerty et al., 2007; Kokko & Jennions, 2014), where genotypical traits beneficial by one sex might be harmful to the other (Chapman et al., 2003).

In many insects, the mating behavior of females, including their potential to mate again, affects the evolution of male reproductive and genotypical traits. Males aim to increase their chance of mating and sperm numbers or competitivity, and if possible, prevent their female partners from mating again. For example, in fruit flies, males have evolved molecular mechanisms that cause changes in females that increase male reproductive success under competitive mating conditions. This is achieved by transferring seminal fluid proteins (SFPs) and other components to the female during mating. Within mated females, SFPs can induce a range of crucial post-mating responses and functions, such as increasing egg production, stimulating the immune system, and reducing the female's sexual receptivity (Chapman, 2001; Clark et al., 1995; Findlay et al., 2014; Haerty et al., 2007; Ram & Wolfner, 2007, 2009; Sirot et al., 2015; Wigby et al., 2020; Wolfner, 1997, 2002). These seminal fluid proteins are often rapidly evolving and undergo substantial adaptive changes within a species (Begun et al., 2000; Begun & Lindfors, 2005; Findlay et al., 2008, 2009; Kern et al., 2004; Wagstaff & Begun, 2005).

In the fruit fly, *D. melanogaster*, the sex peptide (SP) is a seminal fluid protein that plays a pivotal role in initiating post-mating responses in mated females (Chen et al., 1988; Hopkins & Perry, 2022). After cleavage of the 19 amino-acid N-terminal signal peptide in the precursor, mature SP is a small protein with 36 amino acids. Despite its small size, it possesses multiple functional domains. In the mature 36 amino-acid peptide, the N-terminal part is responsible for juvenile hormone induction and for binding to sperm (Peng, Chen, et al., 2005). The proline-rich motif in the middle part is necessary for stimulating the immune system (Domanitskaya et al., 2007; Peng, Zipperlen, et al., 2005). The C-terminal end enables changes in oviposition and receptivity by binding to the sex peptide receptor (SPR) and further interacting with neuronal circuits through sensing neurons (SPSNs) (Feng et al., 2014; Rezával et al., 2012). In addition to enhancing oviposition and reducing the receptivity of females (Aigaki et al., 1991; Avila et al., 2010; Chapman, Bangham, et al., 2003; Chen et al., 1988; Liu & Kubli, 2003), SP can induce numerous other behavioral and physiological changes (Chen et al., 1988; Hopkins & Perry, 2022; Wigby & Chapman, 2005), such as enhancing long time memory (Scheunemann et al., 2019), stimulating immune responses (Peng, Zipperlen, et al., 2005), affecting midgut morphology and physiology (White et al., 2021), and regulating circadian rhythms (Delbare et al., 2023) in females.

Despite SP's fundamental role in these post-mating scenarios, its evolutionary origins and the intricate mechanisms governing its interaction with SPR remain enigmatic. Adding to this complexity is that the existence of SPR predates the emergence of SP (Kim et al., 2010; Poels et al., 2010; Yapici et al., 2008), which raises further questions about the origin of SP and its interactions with SPR. On the other hand, SPR, a gene conserved among various *Insecta* species, such as in mosquitoes, beetles, and moths, is known to bind with conserved myoinhibitory peptides (MIPs) (Kim et al., 2010; Poels et al., 2010). MIPs are small peptides acting as neuromodulators that inhibit muscle activity while simultaneously serving roles in several other physiological processes (Oh et al., 2014).

The transition from a singular MIPs-SPR interaction to a dual MIPs-SPR and SP-SPR interaction in *Drosophila* poses a significant unanswered question. To address these queries, we utilized ancestral sequence reconstruction, AlphaFold2 multimer predictions (Evans et al., 2022; Jumper et al., 2021), and molecular dynamics (MD)

simulations to investigate the biophysical interactions between SP and SPR, and to trace their evolutionary origins. Our objective was to elucidate the complex dynamics and evolutionary trajectory of SP and SPR interactions, thereby enhancing our understanding of post-mating responses and the underlying molecular and biochemical alterations.

Result

C terminus of SP binds stably in the SPR extracellular pocket from AlphaFold2 multimer predictions and MD simulations

SP has been shown to bind to SPR through its C-terminus (Kim et al., 2010; Schmidt et al., 1993; Yapici et al., 2008). However, the detailed molecular mechanism was unclear. Here, we used AlphaFold2 to predict the binding mode of SP to SPR. From our predictions, SP binds to SPR with relatively high confidence, with interface pTM (ipTM) of 55, predicted aligned error (PAE) of 2.3 Å (Figure S1), and pDockQ of 0.47. A pDockQ value exceeding 0.23 and close to 0.5 indicates that our SP-SPR complex model is highly acceptable (Bryant et al., 2022; Burke et al., 2023). In the predicted SP/SPR model, the C terminus of SP (designated as SP_17-36 in Figure 1, corresponding to SP residues 17-36) inserted into the extracellular ligand-binding pocket of SPR (Figure 1A, 1B, Fig. S1A, S1B), which is consistent with earlier biochemical and molecular biology studies (Kim et al., 2010; Schmidt et al., 1993; Yapici et al., 2008). To further evaluate the stability and dynamics of SP/SPR interactions, we conducted three independent MD simulations totaling 14 μs starting from the predicted SP_17-36/SPR complex (Fig S1C, *Methods*). Throughout the MD simulations, SP remained tightly bound to the SPR extracellular pocket (Fig. 1C, Fig. S1D), demonstrating that SP_17-36/SPR is overall stable. We computed the root mean square deviation (RMSD) of SP after aligning the trajectory using SPR transmembrane domains. The RMSD values of SP_17-36 ranged from 2 to 6 Å, indicating some degree of flexibility of SP in the bound state (Figure 1C, Fig. S1D). Intriguingly, when we further computed the RMSD of SP_22-36, we observed that this core region was extremely constrained and rigid with RMSD around 2 Å (Figure 1C, Fig S1D). This suggests that the stable SP/SPR interactions were mostly regulated by the most C-terminus 15 residues, which we term the C-terminal core regions. This finding is in line with earlier biochemical experiments (Ding et al., 2003; Kim et al., 2010; Schmidt et al., 1993). Upon further detailed structural analysis, we found about 24 residues in SPR frequently interacted with the SP core regions at high frequencies (>50%, Figure 1D, *Methods*) in all the three independent MD simulations. These interactions include salt-bridge, pi-cation, aromatic ring stacking, and hydrophobic interactions. The salt-bridge interactions include D82, E172, and D266 of SPR with R35, R35, and K22 of SP, respectively, while the pi-cation interactions include R276 and H352 of SPR with W23 and W32 of SP, respectively. Notably, several residues interact with more than one SP residues (Figure 1D and 1E). For example, Y273 of SPR interacts with both W23 and W32 in SP (Figure 1D-E) through aromatic ring stacking interactions, and the two W residues in SP were previously suggested to be important for SP-SPR interactions (Kim et al., 2010). Besides the two tryptophan residues, Y273 also interacts with G35, R35, and C36 at high frequencies (Figure 1D-E) through hydrophobic interactions. Another example is H352, which frequently

interacts with residue W23 of SP through pi-cation interactions and residues K22 and L26 of SP through hydrophobic interactions. Collectively, these extensive interactions enable SP to bind tightly to SPR.

Sex peptide genes are widespread in *Drosophila* genus and older than previously estimated

The *SP* gene, which codes for sex peptide (SP), has been reported for decades (Chen et al., 1988). It was later reported to be specific to a subset of *Drosophila* species (Kim et al., 2010; Tsuda & Aigaki, 2016; Yapici et al., 2008). To gain a deeper understanding of the origin of the *SP* gene, we took advantage of the conservation of the C-terminal sequences of SP. We applied *tblastn* (Altschul et al., 1990) and gene-structure aware alignments (Iwata & Gotoh, 2012) to search against NCBI eukaryotes representative genomes (Sayers et al., 2022) (Material and Methods). We identified 61 *SP* genes and their duplications in 33 representative genomes (Figure 2A, supplementary file S1). Our results show that the *SP* gene can be found mostly in *Drosophila* species and date its origin to at least the common ancestor with *Scaptodrosophila lebanonensis*. This suggests that the *SP* gene age is much older than previously reported (Tsuda & Aigaki, 2016). Our findings are consistent with a recent study by Hopkins et al. (Hopkins et al., 2024), who also identified one copy of *SP* gene in *S. lebanonensis.* Interestingly, by analyzing RNA-seq data (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA554780), we found that the *SP* gene in *S. lebanonensis* is expressed specifically in male gonads (Figure S2). Although it is unclear about the function of SP in *S. lebanonensis* (Tsuda et al., 2015; Tsuda & Aigaki, 2016), the apparent gonad-specific expression suggests that it may function partially similarly to *D. melanogaster* SP.

Across the identified species, the *SP* genes preserve similar gene structures, each with two exons and one short intron. The length of the short intron ranges from 51 bp to 75 bp. Interestingly, while the first exon is diverged, all *SP* genes are predicted to harbor a 5' region coding for an N-terminal signal peptide (Figure 2B, Figure S3). The region coding for the sperm binding motif (Peng, Chen, et al., 2005) *in D. melanogaster* follows the N-terminal signal peptide and is highly diverged in *Drosophila* subgenus species (Figure 2B, Figure S3). This might suggest that SP in these species: (1) may not have evolved the ability to bind to sperm; or (2) may bind to sperm through different mechanisms as reported in *D. melanogaster*, where a different set of protein interactions could be involved (Singh et al., 2018). The second exon is short and highly conserved, and the peptide products were reported to be responsible for SPR binding in *D. melanogaster* (Kim et al., 2010). The highly conserved sequences suggest that SP from other species may have the capability to bind to SPR, provided that SPR is expressed in the female reproductive tract or both SP and SPR are expressed in the same tissue (Figure 2B, Figure S3). A similar conservation pattern has been observed in a set of genome assemblies from ~145 *Drosophila* species in a recent study (Hopkins et al., 2024).

Most species within the *Drosophila* subgenus have only one copy of *SP* gene, whereas most species in *Sophophora* subgenus contain two or more copies, suggesting a recent expansion of SP genes in these species. Some species in *Sophophora* subgenus even have three or more copies. For example, *D. bipectinata* in the *ananassae* subgroup have 6 copies, and most of the *suzukii* subgroup species contain three copies. The results suggest frequent gene copy number variation events in *Sophophora* subgenus species.

To gain a deeper insight into the origins of *SP* gene copy number variations, we constructed the phylogenetic tree of all the identified 61 *SP* genes (Figure S4). The phylogenetic tree revealed that *Dup99B gene* copy in *D. melanogaster* could date back to anannassae subgroup species, as all *SP* genes in anannassae subgroup (2 in *D. anannassae* and 6 in *D. bipectinata*) were closer to *Dup99B* than *SP* copy in *D. melanogaster*. Our results also suggested a complex evolutionary history of *SP* genes, where Dup99B copy might be lost in *D. elegans* and SP copy might experience multiple duplication events in suzukii subgroup, gunungcola, and some other *Drosophila* species (Figure S4). Furthermore, we found that two of the three copies in suzukii subgroup species are recent suzukii subgroup specific tandem duplications, whereas the third copy in suzukii subgroup species shares a similar origin as *Dup99B*. Interestingly, we found that the high numbers of copies in *D. bipectinata* were anannassae subgroup specific. These results suggest that *SP* genes were likely to experience a substantial expansion in the anannassae subgroup species. Many of these high copy numbers might be lost due to weaker selective constraints. Two copies were maintained in melanogaster group species, i.e., the *SP* gene copy and the *Dup99B* gene copy. Among the two copies, the *Dup99B* copies were highly diverged while the *SP* copies were highly conserved and were further duplicated in the suzukii subgroup.

Within the suzukii species complex, three copies of the sex peptides are present. To investigate whether they are translated, we conducted mass spectrometry and targeted mass spectrometry on accessory gland samples from *D. suzukii* and *D. subpublchrella*. For *D. suzukii*, we identified one unique peptide corresponding to XP_016924330.1, two unique peptides for XP_016924329.1, and two peptides shared between XP_016924329.1 and XP_016924330.1 (Figure S5). In the case of *D. subpubchrella*, four peptides matched both XP_037720360.1 and XP_037721024.1 (Figure S5). These findings indicate that duplicated copies of sex peptides are translated, either at similar levels or with one copy being more highly transcribed or translated than the other. Future research could intriguingly explore the potential functional divergence of these genes.

MIP-SPR share a similar binding interface with SP-SPR

MIPs have been identified as ancestral ligands for SPR (Kim et al., 2010). The MIP precursor can be cleaved into five MIP peptides, MIP1, MIP2, MIP3, MIP4, and MIP5. Here, we used AlphaFold2 to predict the structures of MIP-SPR complexes and ran MD simulations to study the dynamics of the interactions. To compare the SPR interface for SP and MIP peptides, we extracted residues in SPR that interact with SP and MIP peptides at a frequency greater than 5%, respectively. Interestingly, we found that all five MIP-SPR complexes share similar binding interfaces to the SP-SPR complex (Figure 3). In general, SP shares many interacting sites in SPR with MIP peptides (Figure 3A and 3B). For example, D82, E172, R230, R234, L253, T255, V265, D266, Y269, T270, Y273, H352, S356, Y364, N368, and I371 interact with SP and all the five MIPs (Figure 3C). Many of the residues are shared by SP and at least one MIP peptide, such as W77, L79, S83, Y92, T137, P141, L145, N171, P175, A176, H179, I345, V348, T349, I353, L357, E360, L362, A367, M372, and N375 (Figure 3C). By examining SP-interacting residues in SPR during the MD simulations, we pinpointed some SPR residues that interact with SP (frequency > 5%), with no or very low-frequency interactions with MIP peptides. These SPR

residues include M69, E70, N78, T173, Q227, D233, Y272, R276, H281 (Figure 3C). Although MIP-SPR and SP-SPR interactions are mostly similar, the fact that SP and MIP potentially have their unique interface was supported by a recent study examining ligand selectivity against SPR in *D. melanogaster* and *Aedes aegypti*, where the authors found a few amino acids can be important for SP and MIP sensitivity, respectively (J.-H. Lee et al., 2020). We extracted SPR residues that frequently interacts with those two residues to investigate whether the two conserved tryptophan residues of the $W(X)_nW$ motifs in SP and MIP peptides share similar interaction patterns. Interestingly, we found that the first tryptophan residue in SP and all five MIP peptides showed similar interaction patterns (Figure S6). For the second tryptophan residue, all the five MIP peptides share similar interaction patterns, while SP displayed a different pattern but shared some interacting residues in SPR (Figure S6). Overall, our findings further support that the origin of SP-SPR interactions might arise from a combination of exploiting pre-existing MIP-SPR interfaces and the evolution of new SP-SPR interactions.

A few pre-*Drosophila* SPR mutations might create favorable conditions for the origination of SP/SPR interactions

We collected and aligned SPR sequences across *Metazoa* species from orthoDB (Kuznetsov et al., 2023) (Material and Methods). The SPR sequences showed a highly diverged N-terminus and relatively conserved core regions (Figure S7). We then extracted the residues in SPR that frequently interact with SP and MIP during MD simulations (see section *MIP-SPR share a similar binding interface with SP-SPR* in Results). We found that most of these residues were highly conserved in the 38 *Drosophila* species examined in our study, suggesting strong selective constraints in SPR against SP binding so that the predicted SP-interacting residues in SPR were constrained and highly conserved. The number of conserved amino acids decreased sequentially in *Diptera*, *Insecta*, *Arthropoda*, and *Metazoa* species (Figure 4A), suggesting that these sites might undergo decreased selective constraints.

To find out which changes in SPR from old lineages to *Drosophila* lineage might result in SP-SPR interactions, we first reconstructed the ancestral states of SPR of *Drosophila*, *Diptera*, *Insecta*, *Arthropoda*, and *Metazoa* (Material and Methods). For each of the five ancestors, we extracted the most probable ancestral sequences and randomly sampled another one hundred alternative ancestral sequences, respectively (Material and Methods). We used AlphaFold2 to predict whether the ancestral SPRs could bind to SP or MIPs. Interestingly, we found that most of the ancestral *Insecta* SPRs could potentially bind to SP (pDockQ greater than 0.23), and the SP binding potential decreased drastically for ancestral *Arthropoda* and *Metazoa* SPRs (Figure 4B). In contrast, a similar trend was not observed in the case of MIP peptides binding potentials (Figure 4C).

We then computed whether these predicted SP-interacting sites in SPR showed high co-varying scores with their SP binding states (illustrated in Figure 4D) by an in-silico co-varying approach (Material and Methods). We found that several residues showed high co-varying scores (Figure 4E), suggesting that these sites co-varied with SP binding states at high probabilities. Among these residues, A367 and A176 showed the highest scores. In the SP binding states, the two sites were both Alanine. However, in the none-SP binding states, the two sites are more

likely to be other amino acids, for example, Leu, Met, and Val at site 367, and Ile and Thr at site 176 (Figure 4F). We then used AlphaFold2 to predict whether the top-scored residues could affect SP-SPR binding. Interestingly, with only the top scored six amino acids changed to their alternative ancestral states, the mutated sequence did not show SP binding potential, with pDockQ lower than 0.23 (Figure S8). A pDockQ value of 0.23 is the acceptable pDockQ cutoff to define interaction (Bryant et al., 2022). These results indicate the co-varied residues alone could potentially switch on/off SP-SPR binding.

Since A367 and A176 were in the SP binding interface (Figure 5A), we analyzed the interactions between A367 and A176 in SPR with each of the SP residues during the 10-μs MD simulations to further check whether the two sites could affect SP-SPR interactions (Material and Methods). We found that A367 showed frequent interactions with L26 and L28 with the minimum distances frequently ranging between 6 Å and 3.5 Å (Figure 1D, Figure 5B). A distance greater than 6 Å suggests a lack of interaction, while a distance below 3.5 Å is often indicative of polar interactions. Notably, distances of nonpolar residues that lie between 6 Å and 3.5 Å are characteristic of hydrophobic interactions (Material and Methods). Meanwhile, A176 showed frequent interactions with P30, A31, W32, and G33 (Figure 1D, Figure 5C). The results suggest that A367 and A176 form a binding pocket for SP and closely interact with SP through hydrophobic interactions (Material and Methods). Mutations of these two residues from Alanine to other residues, e.g., Leucine, Methionine as shown in Figure 4F, could potentially introduce steric hindrance between SP and SPR due to their large side chains compared to Alanine. The steric hindrance could further reduce the SP binding potentials for SPR harboring those mutations. Interestingly, in the ancestral states of *Arthropoda* and *Metazoa* SPR, where most of them did not show SP binding potential, the two sites were mostly not Alanine (Figure 5D, Figure S9). In *Drosophila* the two sites were both Alanine, and in *Diptera* and *Insecta*, the two sites were highly conserved and were mostly Alanine (Figure 5D). The results further indicate that the pre-Drosophila changes of these residues to Alanine could potentially create changes for the origination of SP-SPR interactions.

Discussion

The evolutionary history of SP-SPR interaction

The sex peptide is one of the most well-studied seminal fluid proteins in nature. However, the sequence and functional conservation of the gene, along with its interaction with the SPR, have not been fully understood to date. In this work, we used a comprehensive set of computational tools to examine the interaction between the SP and SPR, as well as their origins. Our study revealed that SP has the potential to bind to the ancestral states of *Diptera* or even *Insecta* SPR (Figure 4). Indeed, an early study showed that SP could be a potent activator of SPR in moth (Yapici et al., 2008). This finding suggests that there were changes in SPR which could have allowed it to bind SP prior to SP's speculated origin. This leads us to propose that these changes in SPR may have facilitated the origination and functional innovation of SP. This result is in line with previous findings that SP of *D. melanogaster* can have the ability to stimulate juvenile hormone synthesis and depress sex pheromone production in the moth *Helicoverpa armiger*a, which does not have a SP identified (Fan et al., 1999). Interestingly, besides

SPR, several studies suggested it might be quite common for membrane receptors to appear early in evolution compared to their ligands across different species and taxa (Bridgham et al., 2006; Grandchamp & Monget, 2020; Ramilowski et al., 2015; Thornton, 2001).

Changes in SPR predate the origin of SP

Additionally, we discovered that just a few amino acid modifications in SPR could potentially forge SP-SPR interactions. This discovery implies that the shift from a single MIPs-SPR interaction to a dual MIPs-SPR and SP-SPR interaction could have been a relatively uncomplicated evolutionary occurrence. Furthermore, we found that SP-SPR interactions predominantly occur through residues located at the MIP-SPR interface, which is a more ancestral function of SPR. This observation suggests that the co-evolution of the existing MIP-SPR interface may have presented opportunities for the emergence of novel interactions with new, or additional ligands, such as SP, when it appeared. Although AlphFold2 multimer prediction is generally highly accurate (Burke et al., 2023), our method currently has room for enrichment through the incorporation of biochemical and biophysical experiments – an avenue that presents an exciting prospect for future work. Recent studies have emphasized that AlphaFold2 has the potential to predict single mutation effects for protein monomers (Buel & Walters, 2022; McBride et al., 2023; Pak et al., 2023). In the context of this study, we posit that AlphaFold2 multimer prediction has the capability to predict how mutations in SPR might influence SP-SPR binding. To prove this, we carried out a *in silico* mutagenesis assay by generating random mutations in *D. melanogaster* SPR or SP sequence and performing AlphaFold2 multimer predictions using *D. melanogaster* SP and the mutated SPR sequences as well as using *D. melanogaster* SPR and mutated SP sequences (Material and Methods). We found that if random mutations occurred at the predicted SP-binding residues of SPR or C-terminal core regions of SP, the predicted SP-binding potential of the mutated sequences decreased drastically compared to the wild type *D. melanogaster* SPR (Figure S10). This highlights the potential and robustness of AlphaFold2 in our study. Overall, our findings shed light on a potential mechanism through which minor changes in the MIP-SPR interface could have created opportunities for a novel SP to interact and function. Note that the binding sites and interaction residues derived from our study are based on AlphaFold2 multimer predictions and MD simulations. A recent study showed that AlphaFold2 multimer predictions may not fully capture the influence of other variables on protein conformation and binding, but rather an exceptionally powerful hypothesis generator (Terwilliger et al., 2024), which is in line with the purpose of the current study. We anticipate future biochemical and biophysical studies to fully uncover the fascinating molecular details.

SP and MIP share some interface with SPR – synergistic or antagonistic?

Our results show that SP and MIP share a common interface with SPR, while having some unique SP-SPR interactions. These interactions suggest a possible synergistic or antagonistic relationship between SP and MIP in the SPR interface, which is yet to be studied. However, it is important to note that SP and MIPS are expressed in different tissues and biological processes (Tsuda & Aigaki, 2016), and that MIP-SPR interactions originated

earlier than SP-SPR interactions. This historical context suggests that the SPR initially evolved to accommodate MIP-SPR associations, which were later built upon by the emergence of SP-SPR interactions (Hopkins et al., 2024; Kim et al., 2010; Poels et al., 2010). The earlier establishment of MIP-SPR interactions might have set the foundation for the more complex network of interactions that we observe today.

When SP-SPR interactions occur, the immediate question that arises pertains to the fate of the MIP-SPR relationship. Does the SP-SPR interaction disrupt the MIP-SPR one, causing a conflict? Or is the issue effortlessly resolved because MIP generally exhibits tissue-specific expression different from SP? Given the tissue-specific expression of MIP, it's plausible to suggest that the arrival of SP-SPR interactions might not drastically disrupt the existing MIP-SPR dynamic. Instead, it might lead to a more intricate network of interactions, each specialized for distinct tissue-specific functions.

Finally, the divergence of MIP sequences is particularly intriguing when examining the transition from *Arthropoda* to *Insecta* (Figure S11). By analyzing the differences in MIP sequence motifs (Nettling et al., 2015), we found that MIP sequences diverged much more significantly during the transition from *Arthropoda* to *Insecta* than from *Insecta* to *Drosophila* (Figure S11), indicating that the protein's functions and interactions may have evolved in parallel. This divergence hints at adaptive changes to optimize the MIP-SPR interaction, reflecting the organism's shifting needs and constraints over evolutionary timescales.

The origin of SP may date back to the common ancestry of *Ephydroidea* or earlier

Previous results suggest that the sex peptide is a lineage-specific gene found in a subset of *Drosophila* species. However, our findings indicate that the sex peptide is more ancient than previously believed. The discovery of this gene in *S. lebanonensis* suggests that it may be conserved across all *Drosophila* species, despite occasional gene loss in some. This is in line with the recent findings from Hopkins et al. (Hopkins et al., 2024). The presence of the gene in other flies, such as the melon flies *Zeugodacus cucurbitae*, remains uncertain as the potential candidate gene in *Z. cucurbitae* lacks a signal peptide (Figure S12). In vinegar fly *Phortica variegata*, we identified a sequence that, while lacking a signal peptide, contains two $W(X)_{11}W$ motifs in the middle (Figure S13). Since the $W(X)_{11}W$ structure is not at the C-terminal, and the sequence length between the two Ws is 11 amino acids, rather than the typical 8, we argue that this might not be a sex peptide protein. It could represent the ancestral form of the sex peptide, or an unrelated gene with a similar $W(X)_nW$ motif. In either case, it appears that the sex peptide may have originated around the time of divergence between these species. So far, there is no solid evidence to suggest the sex peptide exists in mosquitoes or moths (Tsuda & Aigaki, 2016). However, we were able to discover a bunch of possible $W(X)_nW$ motifs in their genomes. In our study, we only assigned a SP gene if it followed a strict SP gene structure pattern, i.e., the two exons and one intron pattern observed in *Drosophila* species. A limitation would be that if the gene in older lineages evolved to have new gene structures, we could fail to identify orthologs in older lineages.

The origin of SP - *de novo* or divergence?

The SP is produced in male accessory glands of *D. melanogaster* and plays a crucial role in the post-mating behavioral and physiological responses in female fruit flies. Despite its importance in some fruit fly species, the gene is not found in other lineages (Tsuda & Aigaki, 2016). This observation sparks a fascinating discussion about its origin and evolutionary path. In principle, the pattern could be explained by two main theories: *de novo* origin and divergence.

De novo gene birth refers to the process where new genes arise from non-genic DNA sequences (Begun et al., 2006, 2007; Levine et al., 2006). These genes do not have any ancestral gene copies, which can explain their absence in other species. Instead, they evolve from previously non-genic sequences that gain function, usually due to changes in the regulatory regions that start to drive the expression of these sequences in a context where they can impact fitness. This phenomenon has been recognized in various species and is considered an important source of evolutionary novelty. Although hard to prove, SP could indeed be a product of *de novo* gene birth. If true, it suggests that at some point in the evolution of the fruit fly lineage, a non-genic sequence gained function, leading to the formation of SP.

Two pieces of evidence support this idea. First, there is evidence that accessory gland proteins may be easily originated *de novo* (Peng & Zhao, 2024). In fact, the first several works identifying *de novo* genes by the Begun group (Begun et al., 2006, 2007; Levine et al., 2006) were primarily accessory gland proteins. Additionally, several other *de novo* genes have also been shown to be involved in *Drosophila* fertility (Gubala et al., 2017; Lange et al., 2021; Reinhardt et al., 2013; Rivard et al., 2021). Second, although SP has a critical function in *D. melanogaster*, various studies suggest that it is less important in some other *Drosophila* species (Tsuda et al., 2015). This is in line with the simple assumption for novel gene origination, that a new gene is born with a function that can be selected by natural selection. Over time, the new gene may be recruited to complex functional networks and generate pleiotropy.

How likely is it that a gene similar to sex peptide could be born *de novo*? The sex peptide contains a signal peptide, a middle part that functions in sperm binding and juvenile hormone response, and a conserved C-terminal region that binds to SPR. Signal peptides, similar to transmembrane sequences, can be easily formed by hydrophobic sequences (Aspden et al., 2014; Tassios et al., 2023). The conserved C-terminal function $W(X)_8W$ partly relies on the two tryptophan residues (Kim et al., 2010). We asked, how likely intergenic sequences would produce a sequence with two tryptophan residues and several other amino acids in between. The answer is that there are thousands of such sequences in intergenic sequences (Figure S14). Thus, when opportunities arise, if these sequences are transcribed or translated in the right cell types, they may produce a function that interacts with the preexisting SPR.

Alternatively, the lineage-specific pattern could be a result of divergence. Divergence here refers to the process in which an existing gene evolves beyond recognition, i.e., it changes so much over time that it is impossible to identify it as the same gene in different species. Given that the SP is relatively short (36 amino acids), and assuming little to no selective pressure maintaining its sequence, it could have diverged significantly from its ancestral state, rendering it unrecognizable in other species. In addition to genetic drift caused by near neutral processes, natural selection and positive selection can also cause such patterns. If this is the case, it suggests a scenario where the SP or its predecessors existed in other lineages (e.g., mosquitoes, moths, butterflies), but over time, due to lack of constraint, it has changed to the point of being unique to fruit flies.

How likely is a protein to become unrecognizable after a certain period of time? For a long protein, the probability is low. However, for a short protein, it is entirely possible. Overall, signal peptides have a lower sequence constraint, and their substitution rate is three to four times higher than that of mature proteins (Canale et al., 2018). Given a timeframe longer than 60 million years, it is feasible for signal peptides to change without becoming recognizable (Venkataraman et al., 2023). As for C-terminal sequences, speculation becomes more challenging. Even for a gene with entirely different sequences, they still need to bind to SPR. SPR has a pocket that can only bind to certain sequence motifs, making it hard to envision a scenario where sequences change completely but can still bind to SPR. Future structural modeling might offer new insights into the binding potential of SPR. Currently, tools like AlphaFold2 can be used to predict interaction affinity between SPR and a random amino acid. However, executing the probability would require a scale of 20^10, which is not feasible in a short time frame but might be straightforward in the future. Alternatively, there might soon be a way to predict all the possible binding sequences of a protein pocket - such a method does not exist yet, but recent computational methods such as large language models (Lin et al., 2023) or building ultra large candidate libraries (Lyu et al., 2019), could make such an idea possible. This advancement could help us answer whether genes with completely different sequences can still serve the same role in functional networks.

In conclusion, both the *de novo* origin and divergence theories provide plausible explanations for the apparent lineage-specificity of the SP gene. It is also possible that both mechanisms have played a role to some extent. Determining which one is the most probable would necessitate a more profound examination of the evolutionary history of the SP gene. This could be achieved through advanced computational analyses and phylogenetic studies, aided by the generation of high-quality genomes and annotations in a variety of non-*Drosophila* insect species.

If SP turns out to be a lineage-specific new gene, then studying its origin and interaction with SPR also provides an intriguing example of how a novel ligand can quickly incorporate into a functional network and become functionally important. Given that there are many novel and lineage-specific ligands in various species

(Ramilowski et al., 2015), including humans, understanding the origin of SP and the SP-SPR interaction provides fundamental insight into ligand-receptor evolution and the evolution of complex intercellular communications.

Material and Methods

AlphaFold2 mulitmer predictions

We used AlphaFold2 multimer (Evans et al., 2022), version 2.2.2, with default settings for all the multimer predictions in this study. For each prediction, we extracted interface pTM (ipTM) and predicted aligned error (PAE) from the first ranked model. We adapted the scripts from Burke et al (Burke et al., 2023) to compute pDockQ value using a cutoff of 7 Å. To gain a better understanding of the predictions, we plotted in Figure S1D-S1E the statistics of the multiple sequence alignments and PLDDT curves from the AlphaFold2 multimer prediction of *D. melanogaster* SP-SPR.

Identification of potential SP genes

We used the partial sequence of *D. melanogaster* SP as the query and performed a *tblastn* (Altschul et al., 1990) search against NCBI eukaryotes representative genomes (Sayers et al., 2022), which contains high quality genome assemblies of 1642 eukaryotic species. Similar to a recent study by Hopkins et al (Hopkins et al., 2024), we used the most C-terminal sequence of SP, KWCRLNLGPAWGGRC, since the C-terminus is highly conserved, while N-terminus is highly diverged (Figure 1). We collected possible similar sequences with an arbitrarily large *tblastn* E-value cutoff of 9999 in order to include as many *tblastn* hits as possible. At this step, there are 191 *tblastn* hits with query coverage ranging from 60% to 100% and sequence identity ranging from 47% to 100%. We then used a gene structure aware spliced aligner, SPALN (Iwata & Gotoh, 2012), to verify if the hits could potentially be SP genes. We assigned a *tblastn* hit to a potential SP gene only when SPALN predicted a regular SP gene structure with canonical start/stop codons and no frameshifts. A regular SP gene in *Drosophila* species means that it contains two exons and one relatively short intron. In the *Drosophila* species, the length of the short intron ranges from ~30 to ~70 bp. While for the two exons, the length was ~80 to ~130 bp for the first exon, and ~50 to ~60 bp for the second exon. Following this step, we were able to identify 61 potential SP genes across 33 species. We used MAFFT (Katoh et al., 2002) to align the DNA sequences and protein sequences. We further visualized the alignments by MVIEW (Brown et al., 1998) and sequence motifs by WebLogo (Crooks et al., 2004). We then used iqtree2 (Minh et al., 2020) to infer the phylogenetic relations of the 61 SP genes. We further performed SignalP6.0 (Teufel et al., 2022) analysis to predict whether the SP proteins contain N-terminal cleavage signals.

Accessory gland mass-spectrometry

Mixed-age accessory glands from both *D. suzukii* and *D. subpulchrella* were dissected and placed on ice-cold 1X PBS. Each set of accessory glands was then transferred to a volume ranging from 50-110 µl of lysis buffer, homogenized, and the resulting mixture was collected. Proteins were precipitated using ice-cold acetone. Acetone was removed and the pellets were dissolved in 8M Urea/50mM/20mM DTT ammonium bicarbonate and

incubated at room temperature, alkylated (iodoacetamide), and digested with Endoproteinase LysC at <4M urea for 4 hours. Hereafter, the samples were diluted to <2M urea and trypsinized overnight. Peptides were micro-solid phase extracted (Rappsilber et al., 2003) and analyzed by LC-MS/MS using either a 90-minute gradient separated using a 12cm 100um/15cm packed emitter or a 120-minute gradient separated using an EasySprayer (75um/25cm) column. Data were recorded in high resolution/high mass accuracy mode using either a QExactive-HF or a Fusion Lumos. For each sample set, a 2nd run was designed to target several copies of sex peptides using Parallel Reaction Monitoring (PRM) (Peterson et al., 2012). The data were processed using ProteomeDiscoverer v1.4 or v2.5 and searched (Mascot) against a custom database concatenated with LysC and trypsin. Tandem MS spectra of selected matched peptides were compared to in-silico calculated fragmentation (Prosit (Original model, 2019) (Gessulat et al., 2019).

Molecular dynamics simulations

We used AlphaFold2 to predict the structural model of the *Drosophila melanogaster* SP-SPR complex and MIP-SPR complexes as described in section *AlphaFold2 mulitmer predictions* in *Methods*. We used CHARMM-GUI membrane builder (Jo et al., 2008; J. Lee et al., 2016; Wu et al., 2014) to build a heterogeneous bilayer membrane environment for the complexes with cholesterol (30%), POPC (35%), and POPE (35%) lipids (Figure S4). The system was then set up by the additive CHARMM36 protein force field. We added TIP3P water with a length of 12 Å to the top and bottom of the system. We then added additional Na^+ and Cl^- ions to neutralize the system and to reach an ion concentration of 0.15 mM. We further minimized the potential energy of the system by the steepest descent method followed by the conjugate gradient method until the maximum force was smaller than 200 kJ mol^{-1} nm^{-1}. We then used GROMACS (Abraham et al., 2015) to perform subsequent MD simulations. For the SP-SPR complex, we carried out a 10-µs and two additional 2-µs MD simulations. For each of the five MIP-SPR complexes, we carried out a 4-µs and two additional 2-µs MD simulations. The simulation trajectories were recorded every 100 ps. For each frame of each trajectory, we used getcontacts (getcontacts.github.io) to extract interactions between SP or MIP peptides and SPR with the option "--itypes all --vdw_epsilon 1.5". We then computed the frequency of each interaction during the MD simulations. For each of the SP-SPR and MIP-SPR complexes, we considered two residues to be interacting only when the frequency of their interaction is greater than 5% and frequently interacting only when the frequency is greater than 50% in all the three replicate MD simulations.

Ancestral sequence reconstruction

We searched through orthoDB (Kuznetsov et al., 2023) to obtain SPR genes and orthologs. We first collected 328 SPR protein sequences in 228 *Metazoa* species. We used MAFFT (Katoh et al., 2002) to align the sequences. We further excluded the sequences or fragments that have less than 50% coverage compared to SPR gene in *D. melanogaster*. Finally, we obtained 254 SPR genes across different *Metazoa* species. We then used iqtree2 (Minh et al., 2020) to construct the SPR gene tree. We used the maximum likelihood method in PAML (Yang, 2007) to

infer the ancestral sequences of SPRs in different clades, including *Drosophila*, *Diptera*, *Insecta*, *Arthropoda*, and *Metazoa* species. We extracted the most probable ancestral sequences, and alternative states with probability greater than 5%.

In-silico SP-SPR interacting pattern co-varying analysis

For each of the *Drosophila*, *Diptera*, *Insecta*, *Arthropoda*, and *Metazoa* clades, we generated 100 alternative ancestral sequences of SPR. For each of these alternative ancestral SPR sequences, we randomly selected half of the sites with alternative states and mutated the selected sites to one of their alternative states. By doing this, we aimed to avoid almost identical alternative ancestral SPR sequences, which could be challenging for AlphaFold2 (Buel & Walters, 2022). We then used AlphaFold2 (Evans et al., 2022) to predict whether the alternative ancestral SPR sequences could bind *D. melanogaster* SP. We used the pDockQ metric to determine whether SP could bind or not bind to ancestral SPRs. The pDockQ metric has been applied to distinguish interacting from non-interacting proteins (Bryant et al., 2022) and to structurally resolve human PPI network (Burke et al., 2023). Here, we defined SP binding state of an ancestral SPR sequence to be true if the corresponding AlphaFold2 prediction has a pDockQ value greater than 0.3, which is more stringent than the proposed cutoff value of 0.23 (Bryant et al., 2022). Whereas if the pDockQ value was smaller than 0.23, we defined its SP binding state to be false. We extracted the SPR sites that had contacts with either SP or MIP peptides in the MD simulations with frequencies greater than 5%. We took the advantage of Gremlin, a program that could accurately capture co-evolution patterns in protein families (Ovchinnikov et al., 2014), to infer the residues that co-varied with the SP binding state of ancestral SPRs. A simplified scheme of the analysis can be found in Figure 4D. Specifically, we constructed a pseudo alignment, [X_1, X_2, …, X_n; X_{n+1}], from the extracted SPR sites, where X_1 to X_n (n=64) represent the predicted SPR interface as described above, while X_{n+1} represents the predicted binding potential of the corresponding alternative ancestral sequence. We arbitrarily used "H" to stand for potentially binding events and "P" to stand for potentially non-binding events. We then used Gremlin (Ovchinnikov et al., 2014) to compute the co-varying scores between X_{n+1} and other sites from the predicted SPR interface using a maximum likelihood method.

AlphaFold2 predictions of SP-SPR interactions with in silico random mutations in SP or SPR

To investigate whether AlphaFold2 could predict how mutations in SPR affect SP-SPR interactions, we generated 100 mutated SPR sequences, each with 30 random mutations. We chose 30 random mutations to mimic the minimum number of sites that have alternative ancestral states in each of the *Drosophila*, *Diptera*, *Insecta*, *Arthropoda*, and *Metazoa* clades. We used AlphaFold2 to predict the interactions between SP and the mutated SPR sequences. We also checked whether the mutations were in the SP binding interface. Meanwhile, to investigate whether AlphaFold2 could predict how mutations in SP affect SP-SPR interactions, we generated 200 mutated SP sequences. Specifically, the 200 mutated SP sequences include 50 with 3 random mutations in the N-terminal regions, 50 with 6 random mutations in the N-terminal regions, 50 with 3 random mutations in the C-

terminal core regions, and 50 with 6 random mutations in the C-terminal core regions. We used AlphaFold2 to predict the interactions between SPR and the mutated SP sequences.

Identification of the W(X)nW motif in the *D. melanogaster* genome

We used ORFfinder (Sayers et al., 2022) to search for all the intergenic ORFs in the *D. melanogaster* genome by a six-frame translation of all the intergenic regions. We then translated the ORFs and searched for $W(X)_nW$ motif in the translated sequences. We counted the length of the motif and the number of motifs in each of the ORFs. We were able to identify thousands of W(X)nW motifs with the motif length n of 8 to 9. Some of the sequences have high sequence identity compared to the W(X)nW motif in MIP peptides and SP (Figure S14).

Data availability

Primary data have been uploaded to PRIDE database (www.ebi.ac.uk/pride/) with with project accession: [PXD044633], username: reviewer_pxd044633@ebi.ac.uk, and password: LKWAT1Qk. The top ranked predicted structural models of SP-SPR complex and five MIP-SPR complexes along with the MD simulation trajectories are available at Figshare (https://doi.org/10.6084/m9.figshare.25245223.v1). Code and scripts for this study are available on GitHub (https://github.com/LiZhaoLab/DrosophilaSexPeptide).

Author contribution

J.P. and L.Z. conceived the study, designed the experiments, and formulated the analyses. J.P. conducted all analyses except those noted below. N.S. identified gene duplications in the suzukii species complex and conceived mass-spectrometry analysis. H.M. analyzed the mass-spectrometry results. J.P. and L.Z. authored the paper.

Acknowledgement

We thank Christopher B. Langer and Vivian Yan for dissecting the accessory glands. We thank Jason Banfelder, Bala Jayaraman, and The Rockefeller University High Performance Computing (HPC) Center for their support in computation. We are grateful to Dr. Mariana Wolfner for critically reading the manuscript and providing valuable feedback. The authors also thank Drs. Ben Hopkins and Artyom Kopp for their comments and suggestions on a previous version of the manuscript.

Funding

This work was supported by National Institutes of Health (NIH) MIRA R35GM133780, the Robertson Foundation, an Allen Distinguished Investigator Award from Paul G. Allen Family Foundation, a Rita Allen Foundation Scholar Program, and a Vallee Scholar Program (VS-2020-35) to L.Z. J.P. was supported by a C. H. Li Memorial Scholar Fund Award at The Rockefeller University. The content of this study is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

Reference

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Aigaki, T., Fleischmann, I., Chen, P.-S., & Kubli, E. (1991). Ectopic expression of sex peptide alters reproductive behavior of female D. melanogaster. *Neuron*, *7*(4), 557–563. https://doi.org/10.1016/0896-6273(91)90368-A

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Aspden, J. L., Eyre-Walker, Y. C., Phillips, R. J., Amin, U., Mumtaz, M. A. S., Brocard, M., & Couso, J.-P. (2014). Extensive translation of small Open Reading Frames revealed by Poly-Ribo-Seq. *ELife*, *3*. https://doi.org/10.7554/eLife.03528

Avila, F. W., Ravi Ram, K., Bloch Qazi, M. C., & Wolfner, M. F. (2010). Sex Peptide Is Required for the Efficient Release of Stored Sperm in Mated Drosophila Females. *Genetics*, *186*(2), 595–600. https://doi.org/10.1534/genetics.110.119735

Begun, D. J., & Lindfors, H. A. (2005). Rapid Evolution of Genomic Acp Complement in the melanogaster Subgroup of Drosophila. *Molecular Biology and Evolution*, *22*(10), 2010–2021. https://doi.org/10.1093/molbev/msi201

Begun, D. J., Lindfors, H. A., Kern, A. D., & Jones, C. D. (2007). Evidence for de novo evolution of testis-expressed genes in the Drosophila yakuba/Drosophila erecta clade. *Genetics*, *176*(2), 1131–1137. https://doi.org/10.1534/genetics.106.069245

Begun, D. J., Lindfors, H. A., Thompson, M. E., & Holloway, A. K. (2006). Recently Evolved Genes Identified From *Drosophila yakuba* and *D. erecta* Accessory Gland Expressed Sequence Tags. *Genetics*, *172*(3), 1675–1681. https://doi.org/10.1534/genetics.105.050336

Begun, D. J., Whitley, P., Todd, B. L., Waldrip-Dail, H. M., & Clark, A. G. (2000). Molecular Population Genetics of Male Accessory Gland Proteins in Drosophila. *Genetics*, *156*(4), 1879–1888. https://doi.org/10.1093/genetics/156.4.1879

Bridgham, J. T., Carroll, S. M., & Thornton, J. W. (2006). Evolution of Hormone-Receptor Complexity by Molecular Exploitation. *Science*, *312*(5770), 97–101. https://doi.org/10.1126/science.1123348

Brown, N. P., Leroy, C., & Sander, C. (1998). MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics*, *14*(4), 380–381. https://doi.org/10.1093/bioinformatics/14.4.380

Bryant, P., Pozzati, G., & Elofsson, A. (2022). Improved prediction of protein-protein interactions using AlphaFold2. *Nature Communications*, *13*(1), 1265. https://doi.org/10.1038/s41467-022-28865-w

Buel, G. R., & Walters, K. J. (2022). Can AlphaFold2 predict the impact of missense mutations on structure? *Nature Structural & Molecular Biology*, *29*(1), 1–2. https://doi.org/10.1038/s41594-021-00714-2

Burke, D. F., Bryant, P., Barrio-Hernandez, I., Memon, D., Pozzati, G., Shenoy, A., Zhu, W., Dunham, A. S., Albanese, P., Keller, A., Scheltema, R. A., Bruce, J. E., Leitner, A., Kundrotas, P., Beltrao, P., & Elofsson, A. (2023). Towards a structurally resolved human protein interaction network. *Nature Structural & Molecular Biology*, *30*(2), 216–225. https://doi.org/10.1038/s41594-022-00910-8

Canale, A. S., Venev, S. V., Whitfield, T. W., Caffrey, D. R., Marasco, W. A., Schiffer, C. A., Kowalik, T. F., Jensen, J. D., Finberg, R. W., Zeldovich, K. B., Wang, J. P., & Bolon, D. N. A. (2018). Synonymous Mutations at the Beginning of the Influenza A Virus Hemagglutinin Gene Impact Experimental Fitness. *Journal of Molecular Biology*, *430*(8), 1098–1115. https://doi.org/10.1016/j.jmb.2018.02.009

Chapman, T. (2001). Seminal fluid-mediated fitness traits in Drosophila. *Heredity*, *87*(5), 511–521. https://doi.org/10.1046/j.1365-2540.2001.00961.x

Chapman, T., Arnqvist, G., Bangham, J., & Rowe, L. (2003). Sexual conflict. *Trends in Ecology & Evolution*, *18*(1), 41–47. https://doi.org/10.1016/S0169-5347(02)00004-6

Chapman, T., Bangham, J., Vinti, G., Seifried, B., Lung, O., Wolfner, M. F., Smith, H. K., & Partridge, L. (2003). The sex peptide of *Drosophila melanogaster:* Female post-mating responses analyzed by using RNA interference. *Proceedings of the National Academy of Sciences*, *100*(17), 9923–9928. https://doi.org/10.1073/pnas.1631635100

Chen, P. S., Stumm-Zollinger, E., Aigaki, T., Balmer, J., Bienz, M., & Böhlen, P. (1988). A male accessory gland peptide that regulates reproductive behavior of female D. melanogaster. *Cell*, *54*(3), 291–298. https://doi.org/10.1016/0092-8674(88)90192-4

Clark, A. G., Aguadé, M., Prout, T., Harshman, L. G., & Langley, C. H. (1995). Variation in sperm displacement and its association with accessory gland protein loci in Drosophila melanogaster. *Genetics*, *139*(1), 189–201. https://doi.org/10.1093/genetics/139.1.189

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator: Figure 1. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Delbare, S. Y. N., Venkatraman, S., Scuderi, K., Wells, M. T., Wolfner, M. F., Basu, S., & Clark, A. G. (2023). Time series transcriptome analysis implicates the circadian clock in the *Drosophila melanogaster* female's response to sex peptide. *Proceedings of the National Academy of Sciences*, *120*(5). https://doi.org/10.1073/pnas.2214883120

Domanitskaya, E. V., Liu, H., Chen, S., & Kubli, E. (2007). The hydroxyproline motif of male sex peptide elicits the innate immune response in *Drosophila* females. *FEBS Journal*, *274*(21), 5659–5668. https://doi.org/10.1111/j.1742-4658.2007.06088.x

Eberhard, W. G. (1985). *Sexual Selection and Animal Genitalia*. Harvard University Press. https://doi.org/10.4159/harvard.9780674330702

Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., … Hassabis, D. (2022). Protein complex prediction with AlphaFold-Multimer. *BioRxiv*, 2021.10.04.463034. https://doi.org/10.1101/2021.10.04.463034

Fan, Y., Rafaeli, A., Gileadi, C., Kubli, E., & Applebaum, S. W. (1999). Drosophila melanogaster sex peptide stimulates juvenile hormone synthesis and depresses sex pheromone production in Helicoverpa armigera. *Journal of Insect Physiology*, *45*(2), 127–133. https://doi.org/10.1016/S0022-1910(98)00106-1

Ferris, P. J., Pavlovic, C., Fabry, S., & Goodenough, U. W. (1997). Rapid evolution of sex-related genes in *Chlamydomonas*. *Proceedings of the National Academy of Sciences*, *94*(16), 8634–8639. https://doi.org/10.1073/pnas.94.16.8634

Findlay, G. D., MacCoss, M. J., & Swanson, W. J. (2009). Proteomic discovery of previously unannotated, rapidly evolving seminal fluid genes in *Drosophila*. *Genome Research*, *19*(5), 886–896. https://doi.org/10.1101/gr.089391.108

Findlay, G. D., Sitnik, J. L., Wang, W., Aquadro, C. F., Clark, N. L., & Wolfner, M. F. (2014). Evolutionary Rate Covariation Identifies New Members of a Protein Network Required for Drosophila melanogaster Female Post-Mating Responses. *PLoS Genetics*, *10*(1), e1004108. https://doi.org/10.1371/journal.pgen.1004108

Findlay, G. D., Yi, X., MacCoss, M. J., & Swanson, W. J. (2008). Proteomics Reveals Novel Drosophila Seminal Fluid Proteins Transferred at Mating. *PLoS Biology*, *6*(7), e178. https://doi.org/10.1371/journal.pbio.0060178

Gessulat, S., Schmidt, T., Zolg, D. P., Samaras, P., Schnatbaum, K., Zerweck, J., Knaute, T., Rechenberger, J., Delanghe, B., Huhmer, A., Reimer, U., Ehrlich, H.-C., Aiche, S., Kuster, B., & Wilhelm, M. (2019). Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nature Methods*, *16*(6), 509–518. https://doi.org/10.1038/s41592-019-0426-7

Grandchamp, A., & Monget, P. (2020). The membrane receptors that appeared before their ligand: The different proposed scenarios. *PLOS ONE*, *15*(5), e0231813. https://doi.org/10.1371/journal.pone.0231813

Gubala, A. M., Schmitz, J. F., Kearns, M. J., Vinh, T. T., Bornberg-Bauer, E., Wolfner, M. F., & Findlay, G. D. (2017). The *goddard* and *saturn* genes are essential for *Drosophila* male fertility and may have arisen *de novo*. *Molecular Biology and Evolution*, msx057. https://doi.org/10.1093/molbev/msx057

Haerty, W., Jagadeeshan, S., Kulathinal, R. J., Wong, A., Ravi Ram, K., Sirot, L. K., Levesque, L., Artieri, C. G., Wolfner, M. F., Civetta, A., & Singh, R. S. (2007). Evolution in the Fast Lane: Rapidly Evolving Sex-Related Genes in Drosophila. *Genetics*, *177*(3), 1321–1335. https://doi.org/10.1534/genetics.107.078865

Hopkins, B. R., Angus-Henry, A., Kim, B. Y., Carlisle, J. A., Thompson, A., & Kopp, A. (2024). Decoupled evolution of the *Sex Peptide* gene family and *Sex Peptide Receptor* in Drosophilidae. *Proceedings of the National Academy of Sciences*, *121*(3). https://doi.org/10.1073/pnas.2312380120

Hopkins, B. R., & Kopp, A. (2021). Evolution of sexual development and sexual dimorphism in insects. *Current Opinion in Genetics & Development*, *69*, 129–139. https://doi.org/10.1016/j.gde.2021.02.011

Hopkins, B. R., & Perry, J. C. (2022). The evolution of sex peptide: sexual conflict, cooperation, and coevolution. *Biological Reviews*, *97*(4), 1426–1448. https://doi.org/10.1111/brv.12849

Iwata, H., & Gotoh, O. (2012). Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Research*, *40*(20), e161. https://doi.org/10.1093/nar/gks708

Jo, S., Kim, T., Iyer, V. G., & Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry*, *29*(11), 1859–1865. https://doi.org/10.1002/jcc.20945

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2

Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Kern, A. D., Jones, C. D., & Begun, D. J. (2004). Molecular Population Genetics of Male Accessory Gland Proteins in the Drosophila simulans Complex. *Genetics*, *167*(2), 725–735. https://doi.org/10.1534/genetics.103.020883

Kim, Y.-J., Bartalska, K., Audsley, N., Yamanaka, N., Yapici, N., Lee, J.-Y., Kim, Y.-C., Markovic, M., Isaac, E., Tanaka, Y., & Dickson, B. J. (2010). MIPs are ancestral ligands for the sex peptide receptor. *Proceedings of the National Academy of Sciences*, *107*(14), 6520–6525. https://doi.org/10.1073/pnas.0914764107

Kokko, H., & Jennions, M. D. (2014). The Relationship between Sexual Selection and Sexual Conflict. *Cold Spring Harbor Perspectives in Biology*, *6*(9), a017517–a017517. https://doi.org/10.1101/cshperspect.a017517

Kumar, S., Suleski, M., Craig, J. M., Kasprowicz, A. E., Sanderford, M., Li, M., Stecher, G., & Hedges, S. B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. *Molecular Biology and Evolution*, *39*(8). https://doi.org/10.1093/molbev/msac174

Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E. V., & Zdobnov, E. M. (2023). OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research*, *51*(D1), D445–D451. https://doi.org/10.1093/nar/gkac998

Lange, A., Patel, P. H., Heames, B., Damry, A. M., Saenger, T., Jackson, C. J., Findlay, G. D., & Bornberg-Bauer, E. (2021). Structural and functional characterization of a putative de novo gene in Drosophila. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-21667-6

Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., MacKerell, A. D., Klauda, J. B., & Im, W. (2016).

CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation*, *12*(1), 405–413. https://doi.org/10.1021/acs.jctc.5b00935

Lee, J.-H., Lee, N.-R., Kim, D.-H., & Kim, Y.-J. (2020). Molecular characterization of ligand selectivity of the sex peptide receptors of Drosophila melanogaster and Aedes aegypti. *Insect Biochemistry and Molecular Biology*, *127*, 103472. https://doi.org/10.1016/j.ibmb.2020.103472

Levine, M. T., Jones, C. D., Kern, A. D., Lindfors, H. A., & Begun, D. J. (2006). Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proceedings of the National Academy of Sciences*, *103*(26), 9935–9939. https://doi.org/10.1073/pnas.0509809103

Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., dos Santos Costa, A., Fazel-Zarandi, M., Sercu, T., Candido, S., & Rives, A. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, *379*(6637), 1123–1130. https://doi.org/10.1126/science.ade2574

Liu, H., & Kubli, E. (2003). Sex-peptide is the molecular basis of the sperm effect in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, *100*(17), 9929–9933. https://doi.org/10.1073/pnas.1631700100

Lyu, J., Wang, S., Balius, T. E., Singh, I., Levit, A., Moroz, Y. S., O'Meara, M. J., Che, T., Algaa, E., Tolmachova, K., Tolmachev, A. A., Shoichet, B. K., Roth, B. L., & Irwin, J. J. (2019). Ultra-large library docking for discovering new chemotypes. *Nature*, *566*(7743), 224–229. https://doi.org/10.1038/s41586-019-0917-9

McBride, J. M., Polev, K., Abdirasulov, A., Reinharz, V., Grzybowski, B. A., & Tlusty, T. (2023). AlphaFold2 Can Predict Single-Mutation Effects. *Physical Review Letters*, *131*(21), 218401. https://doi.org/10.1103/PhysRevLett.131.218401

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Nettling, M., Treutler, H., Grau, J., Keilwagen, J., Posch, S., & Grosse, I. (2015). DiffLogo: a comparative visualization of sequence motifs. *BMC Bioinformatics*, *16*(1), 387. https://doi.org/10.1186/s12859-015-0767-x

Oh, Y., Yoon, S.-E., Zhang, Q., Chae, H.-S., Daubnerová, I., Shafer, O. T., Choe, J., & Kim, Y.-J. (2014). A Homeostatic Sleep-Stabilizing Pathway in Drosophila Composed of the Sex Peptide Receptor and Its Ligand, the Myoinhibitory Peptide. *PLoS Biology*, *12*(10), e1001974. https://doi.org/10.1371/journal.pbio.1001974

Ovchinnikov, S., Kamisetty, H., & Baker, D. (2014). Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. *ELife*, *3*. https://doi.org/10.7554/eLife.02030

Pak, M. A., Markhieva, K. A., Novikova, M. S., Petrov, D. S., Vorobyev, I. S., Maksimova, E. S., Kondrashov, F. A., & Ivankov, D. N. (2023). Using AlphaFold to predict the impact of single mutations on protein stability and function. *PLOS ONE*, *18*(3), e0282689. https://doi.org/10.1371/journal.pone.0282689

Parker, G. A. (1970). Sperm competition and its evolutionary consequences in the insects. *Biological Reviews*, *45*(4), 525–567. https://doi.org/10.1111/j.1469-185X.1970.tb01176.x

Peng, J., Chen, S., Büsser, S., Liu, H., Honegger, T., & Kubli, E. (2005). Gradual Release of Sperm Bound Sex-Peptide Controls Female Postmating Behavior in Drosophila. *Current Biology*, *15*(3), 207–213. https://doi.org/10.1016/j.cub.2005.01.034

Peng, J., & Zhao, L. (2024). The origin and structural evolution of de novo genes in Drosophila. *Nature Communications*, *15*(1), 810. https://doi.org/10.1038/s41467-024-45028-1

Peng, J., Zipperlen, P., & Kubli, E. (2005). Drosophila Sex-Peptide Stimulates Female Innate Immune System after Mating via the Toll and Imd Pathways. *Current Biology*, *15*(18), 1690–1694. https://doi.org/10.1016/j.cub.2005.08.048

Peterson, A. C., Russell, J. D., Bailey, D. J., Westphall, M. S., & Coon, J. J. (2012). Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Molecular & Cellular Proteomics*, *11*(11), 1475–1488. https://doi.org/10.1074/mcp.O112.020131

Poels, J., Van Loy, T., Vandersmissen, H. P., Van Hiel, B., Van Soest, S., Nachman, R. J., & Vanden Broeck, J. (2010). Myoinhibiting peptides are the ancestral ligands of the promiscuous Drosophila sex peptide receptor. *Cellular and Molecular Life Sciences*, *67*(20), 3511–3522. https://doi.org/10.1007/s00018-010-0393-8

Ram, K. R., & Wolfner, M. F. (2007). Sustained Post-Mating Response in Drosophila melanogaster Requires Multiple Seminal Fluid Proteins. *PLoS Genetics*, *3*(12), e238. https://doi.org/10.1371/journal.pgen.0030238

Ram, K. R., & Wolfner, M. F. (2009). A network of interactions among seminal proteins underlies the long-term postmating response in Drosophila. *Proceedings of the National Academy of Sciences*, *106*(36), 15384–15389. https://doi.org/10.1073/pnas.0902923106

Ramilowski, J. A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V. P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., & Forrest, A. R. R. (2015). A draft network of ligand–receptor-mediated multicellular signalling in human. *Nature Communications*, *6*(1), 7866. https://doi.org/10.1038/ncomms8866

Rappsilber, J., Ishihama, Y., & Mann, M. (2003). Stop and Go Extraction Tips for Matrix-Assisted Laser Desorption/Ionization, Nanoelectrospray, and LC/MS Sample Pretreatment in Proteomics. *Analytical Chemistry*, *75*(3), 663–670. https://doi.org/10.1021/ac026117i

Reinhardt, J. A., Wanjiru, B. M., Brant, A. T., Saelao, P., Begun, D. J., & Jones, C. D. (2013). De Novo ORFs in Drosophila Are Important to Organismal Fitness and Evolved Rapidly from Previously Non-coding Sequences. *PLoS Genetics*, *9*(10), e1003860. https://doi.org/10.1371/journal.pgen.1003860

Rezával, C., Pavlou, H. J., Dornan, A. J., Chan, Y.-B., Kravitz, E. A., & Goodwin, S. F. (2012). Neural Circuitry Underlying Drosophila Female Postmating Behavioral Responses. *Current Biology*, *22*(13), 1155–1165. https://doi.org/10.1016/j.cub.2012.04.062

Rivard, E. L., Ludwig, A. G., Patel, P. H., Grandchamp, A., Arnold, S. E., Berger, A., Scott, E. M., Kelly, B. J., Mascha, G. C., Bornberg-Bauer, E., & Findlay, G. D. (2021). A putative de novo evolved gene required for spermatid chromatin condensation in Drosophila melanogaster. *PLOS Genetics*, *17*(9), e1009787. https://doi.org/10.1371/journal.pgen.1009787

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/nar/gkab1112

Scheunemann, L., Lampin-Saint-Amaux, A., Schor, J., & Preat, T. (2019). A sperm peptide enhances long-term memory in female *Drosophila*. *Science Advances*, *5*(11). https://doi.org/10.1126/sciadv.aax3432

Schmidt, T., Choffat, Y., Klauser, S., & Kubli, E. (1993). The Drosophila melanogaster sex-peptide: A molecular analysis of structure-function relationships. *Journal of Insect Physiology*, *39*(5), 361–368. https://doi.org/10.1016/0022-1910(93)90023-K

Singh, A., Buehner, N. A., Lin, H., Baranowski, K. J., Findlay, G. D., & Wolfner, M. F. (2018). Long-term interaction between Drosophila sperm and sex peptide is mediated by other seminal proteins that bind only transiently to sperm. *Insect Biochemistry and Molecular Biology*, *102*, 43–51. https://doi.org/10.1016/j.ibmb.2018.09.004

Sirot, L. K., Wong, A., Chapman, T., & Wolfner, M. F. (2015). Sexual Conflict and Seminal Fluid Proteins: A Dynamic Landscape of Sexual Interactions. *Cold Spring Harbor Perspectives in Biology*, *7*(2), a017533. https://doi.org/10.1101/cshperspect.a017533

Tassios, E., Nikolaou, C., & Vakirlis, N. (2023). Intergenic Regions of Saccharomycotina Yeasts are Enriched in Potential to Encode Transmembrane Domains. *Molecular Biology and Evolution*, *40*(3). https://doi.org/10.1093/molbev/msad059

Terwilliger, T. C., Liebschner, D., Croll, T. I., Williams, C. J., McCoy, A. J., Poon, B. K., Afonine, P. V., Oeffner, R. D., Richardson, J. S., Read, R. J., & Adams, P. D. (2024). AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nature Methods*, *21*(1), 110–116. https://doi.org/10.1038/s41592-023-02087-4

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology*, *40*(7), 1023–1025. https://doi.org/10.1038/s41587-021-01156-3

Thornton, J. W. (2001). Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proceedings of the National Academy of Sciences*, *98*(10), 5671–5676. https://doi.org/10.1073/pnas.091553298

Tsuda, M., & Aigaki, T. (2016). Evolution of sex-peptide in Drosophila. *Fly*, *10*(4), 172–177. https://doi.org/10.1080/19336934.2016.1193655

Tsuda, M., Peyre, J.-B., Asano, T., & Aigaki, T. (2015). Visualizing Molecular Functions and Cross-Species Activity of Sex-Peptide in *Drosophila*. *Genetics*, *200*(4), 1161–1169. https://doi.org/10.1534/genetics.115.177550

Venkataraman, K., Shai, N., Lakhiani, P., Zylka, S., Zhao, J., Herre, M., Zeng, J., Neal, L. A., Molina, H., Zhao, L., & Vosshall, L. B. (2023). Two novel, tightly linked, and rapidly evolving genes underlie Aedes aegypti mosquito reproductive resilience during drought. *ELife*, *12*. https://doi.org/10.7554/eLife.80489

Wagstaff, B. J., & Begun, D. J. (2005). Molecular Population Genetics of Accessory Gland Protein Genes and Testis-Expressed Genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics*, *171*(3), 1083–1101. https://doi.org/10.1534/genetics.105.043372

White, M. A., Bonfini, A., Wolfner, M. F., & Buchon, N. (2021). *Drosophila melanogaster* sex peptide regulates mated female midgut morphology and physiology. *Proceedings of the National Academy of Sciences*, *118*(1). https://doi.org/10.1073/pnas.2018112118

Wigby, S., Brown, N. C., Allen, S. E., Misra, S., Sitnik, J. L., Sepil, I., Clark, A. G., & Wolfner, M. F. (2020). The *Drosophila* seminal proteome and its role in postcopulatory sexual selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1813), 20200072. https://doi.org/10.1098/rstb.2020.0072

Wigby, S., & Chapman, T. (2005). Sex Peptide Causes Mating Costs in Female Drosophila melanogaster. *Current Biology*, *15*(4), 316–321. https://doi.org/10.1016/j.cub.2005.01.051

Wolfner, M. F. (1997). Tokens of love: Functions and regulation of drosophila male accessory gland products. *Insect Biochemistry and Molecular Biology*, *27*(3), 179–192. https://doi.org/10.1016/S0965-1748(96)00084-7

Wolfner, M. F. (2002). The gifts that keep on giving: physiological functions and evolutionary dynamics of male seminal proteins in Drosophila. *Heredity*, *88*(2), 85–93. https://doi.org/10.1038/sj.hdy.6800017

Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., Dávila-Contreras, E. M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R. M., Klauda, J. B., & Im, W. (2014). CHARMM-GUI *Membrane Builder* toward realistic biological membrane simulations. *Journal of Computational Chemistry*, *35*(27), 1997–2004. https://doi.org/10.1002/jcc.23702

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yapici, N., Kim, Y.-J., Ribeiro, C., & Dickson, B. J. (2008). A receptor that mediates the post-mating switch in Drosophila reproductive behaviour. *Nature*, *451*(7174), 33–37. https://doi.org/10.1038/nature06483

Figure 1. Structure and dynamics of SP/SPR complex. (A) Structural model of SP/SPR complex predicted by AlphaFold2, where SP is colored in blue and SPR is colored in different colors according to the transmembrane segments, starting from TM1 to TM7. The extracellular loop 2 (ECL2) of SPR is also highlighted. (B) Top view of (A). (C) Time resolved RMSD of the C-terminus (blue) and core regions (magenta) during MD simulations. The small RMSD values of the core regions suggest a stable SP/SPR interaction. (D) Frequently interacting residues between SP (residues labeled on Y axis) and SPR (residues labeled on X axis) during MD simulations (frequency > 50% in all three independent MD simulations). (E) A representative structure at t = 4.31 μs from the 10-μs MD simulation, showing the interactions between SPR residues (magenta) with W23 and W32 in SP (green).

Figure 2. Sequence conservation of *SP* genes. (A) Phylogenetic tree of the species that have *SP* genes identified. The tree was extracted from timetree.org (Kumar et al., 2022). The left panel shows the phylogenetic tree of the 33 species in NCBI representative genomes with *SP* genes identified. The copy number of *SP* genes in each species is listed after the species name. The right panel shows the alignment of the two exons and intron of the representative *SP* genes. For species that have more than one copy of *SP* genes, the representative *SP* gene in each species has the highest sequence similarity to the *SP* gene in *D. melanogaster*. Alignment gaps were shown in gray, while diverged and conserved nucleotides were shown in brown and blue, respectively. The bottom panel shows the conservation of the alignments, where the functional regions that code for signal peptide, sperm binding sites, and SPR binding sites in *D. melanogaster* are highlighted. (B) Sequence of the SP proteins across the 33 species. The N-terminal signal peptides were enriched in hydrophobic residues (left panel). The putative sperm binding motif in *D. melanogaster* is WEWPWNR, while it can be quite different in other species (middle panel). The SPR binding motif is highly conserved with two tryptophan residues (right panel).

Figure 3. Mip/SPR shares a similar binding interface with SP/SPR. (A) Top view of SP-SPR complex. SP is colored in blue and the SPR residues that interact with SP in MD simulations are colored in orange. (B) Top view of the five Mip-SPR complexes overlaid together. Mip peptides are colored in blue and the SPR residues that interact with SP in MD simulations are colored in orange as in (A). (C) The frequency of SPR residues interacting with SP or MIPs during MD simulations. SPR residues that interact with either SP or MIPs at 5% frequency or higher during MD simulations were shown.

Figure 4. In silico coevolution between SPR residues and SP binding states. (A) Conservation of SP and Mip binding sites in SPR. (B) AlphaFold predictions of SP and ancestral SPRs binding. (C) AlphaFold predictions of Mips and ancestral SPRs binding. (D) Simplified scheme of the *in-silico* coevolution analysis. We computed whether there are sites that co-varied with SP binding states. Binding state of 1 suggests acceptable binding potential between the ligand and receptor, 0 for not binding. In this simplified example, site 2 has a high chance to co-vary with binding state, followed by site 3, but not sites 1, 4, and 5. (E) Co-varying scores of SPR residues and SP binding states. A367 and A176 showed the highest score. (F) The frequency of each amino acid at residue sites 367 and 176 in different SP binding states.

Figure 5. Pre-*Drosophila* mutations to A176 and A367 in D. melanogaster SPR (DmSPR) might create favorable conditions for the origination of SP-SPR interactions. (A) Structural overview of A176 and A367 in DmSPR and the frequently contacting residues within SP during MD simulations, including L26, L28, G29, P30, A31, W32, and G33. The structure is the representative conformation as shown in Figure 1E. (B) A367 interacts with L26 and L28 through hydrophobic interactions during MD simulations (Material and Methods). (C) A176 interacts with P30, A31, W32, and G33 through hydrophobic interactions during MD simulations (Material and Methods). (D) SPR sites 367, 176, 371, 352, and 276 in different ancestral sequences. Sites 367 and 176 showed the highest co-varying signals with SP binding states (Figure 4E). The two sites were highly conserved in *Drosophila*, *Diptera*, and *Insecta*. Sites 371, 272, 352, and 276 co-varying signals with SP binding states. The four sites were conserved across different clades.

Figure 1

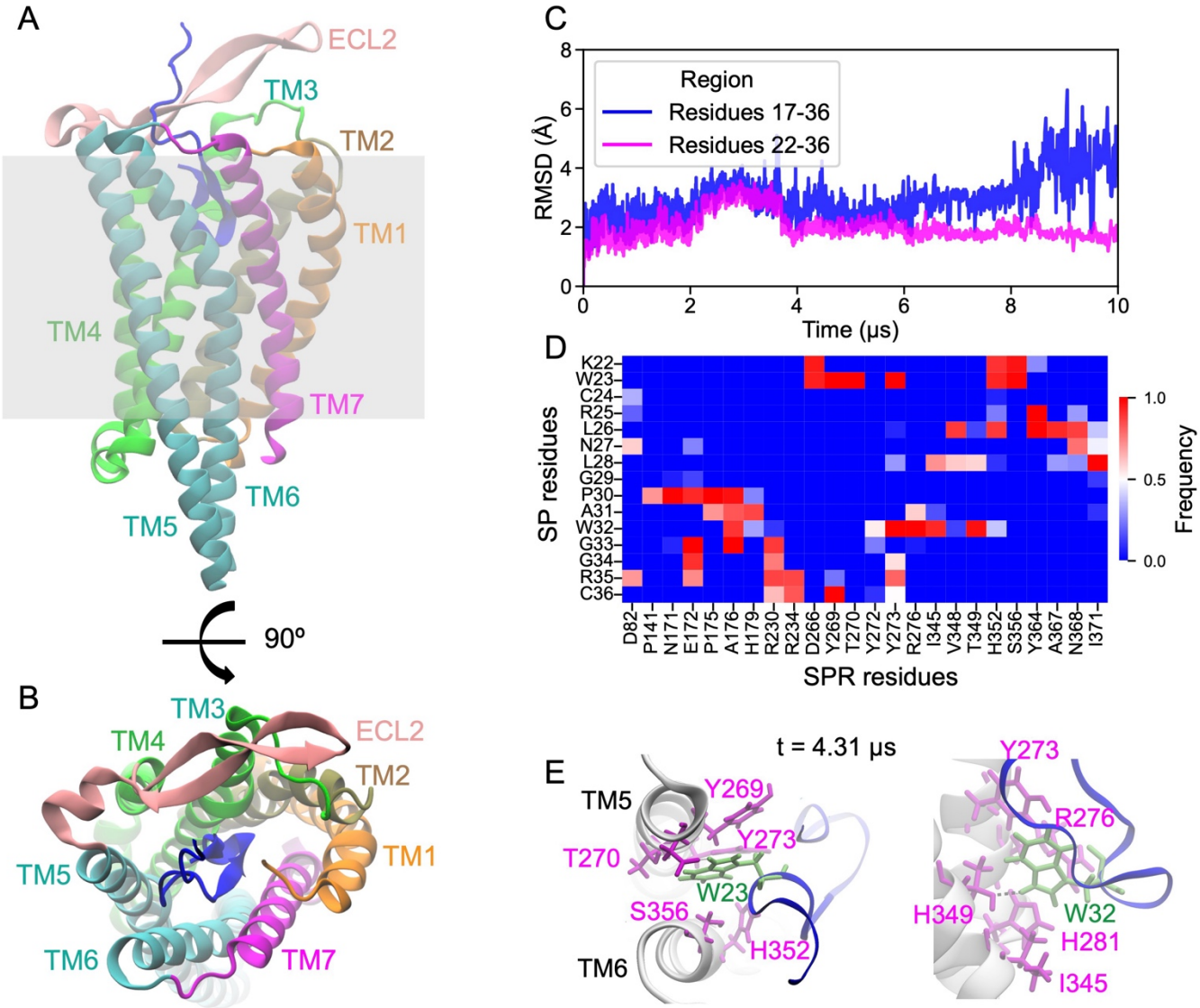

A
ECL2
TM3
TM2
TM1
TM4
TM7
TM6
TM5
90º
B
TM3
TM4
ECL2
TM2
TM5
TM1
TM6
TM7
C
Region
Residues 17-36
Residues 22-36
RMSD (Å)
Time (μs)
D
SP residues
SPR residues
Frequency
E
t = 4.31 μs
TM5
Y269
Y273
T270
W23
S356
H352
TM6
Y273
R276
H349
W32
H281
I345

Figure 2

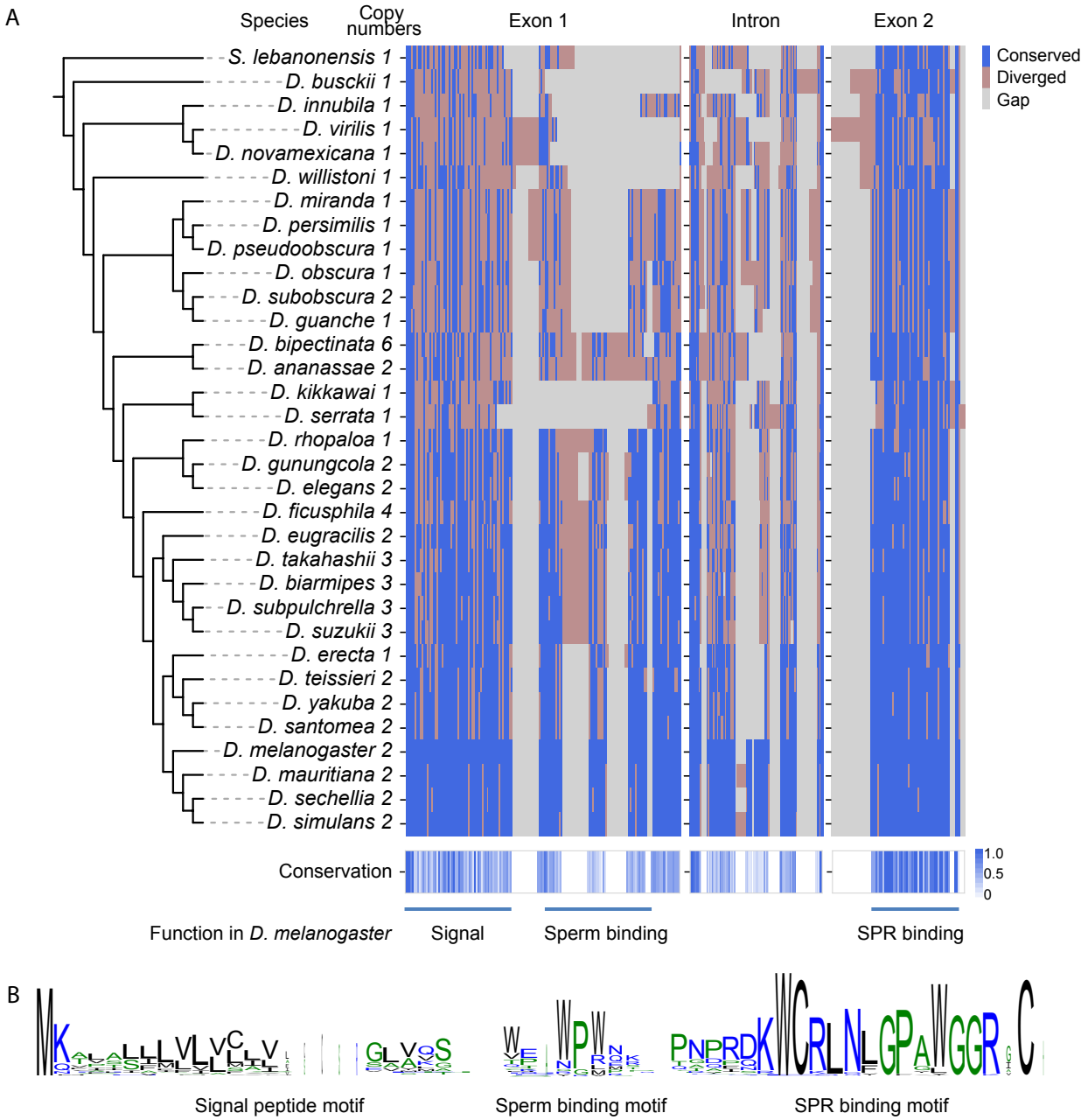

A
Species
Copy numbers
Exon 1
Intron
Exon 2
Conserved
Diverged
Gap
S. lebanonensis 1
D. busckii 1
D. innubila 1
D. virilis 1
D. novamexicana 1
D. willistoni 1
D. miranda 1
D. persimilis 1
D. pseudoobscura 1
D. obscura 1
D. subobscura 2
D. guanche 1
D. bipectinata 6
D. ananassae 2
D. kikkawai 1
D. serrata 1
D. rhopaloa 1
D. gunungcola 2
D. elegans 2
D. ficusphila 4
D. eugracilis 2
D. takahashii 3
D. biarmipes 3
D. subpulchrella 3
D. suzukii 3
D. erecta 1
D. teissieri 2
D. yakuba 2
D. santomea 2
D. melanogaster 2
D. mauritiana 2
D. sechellia 2
D. simulans 2
Conservation
1.0
0.5
0
Function in D. melanogaster
Signal
Sperm binding
SPR binding
B
Signal peptide motif
Sperm binding motif
SPR binding motif

Figure 3

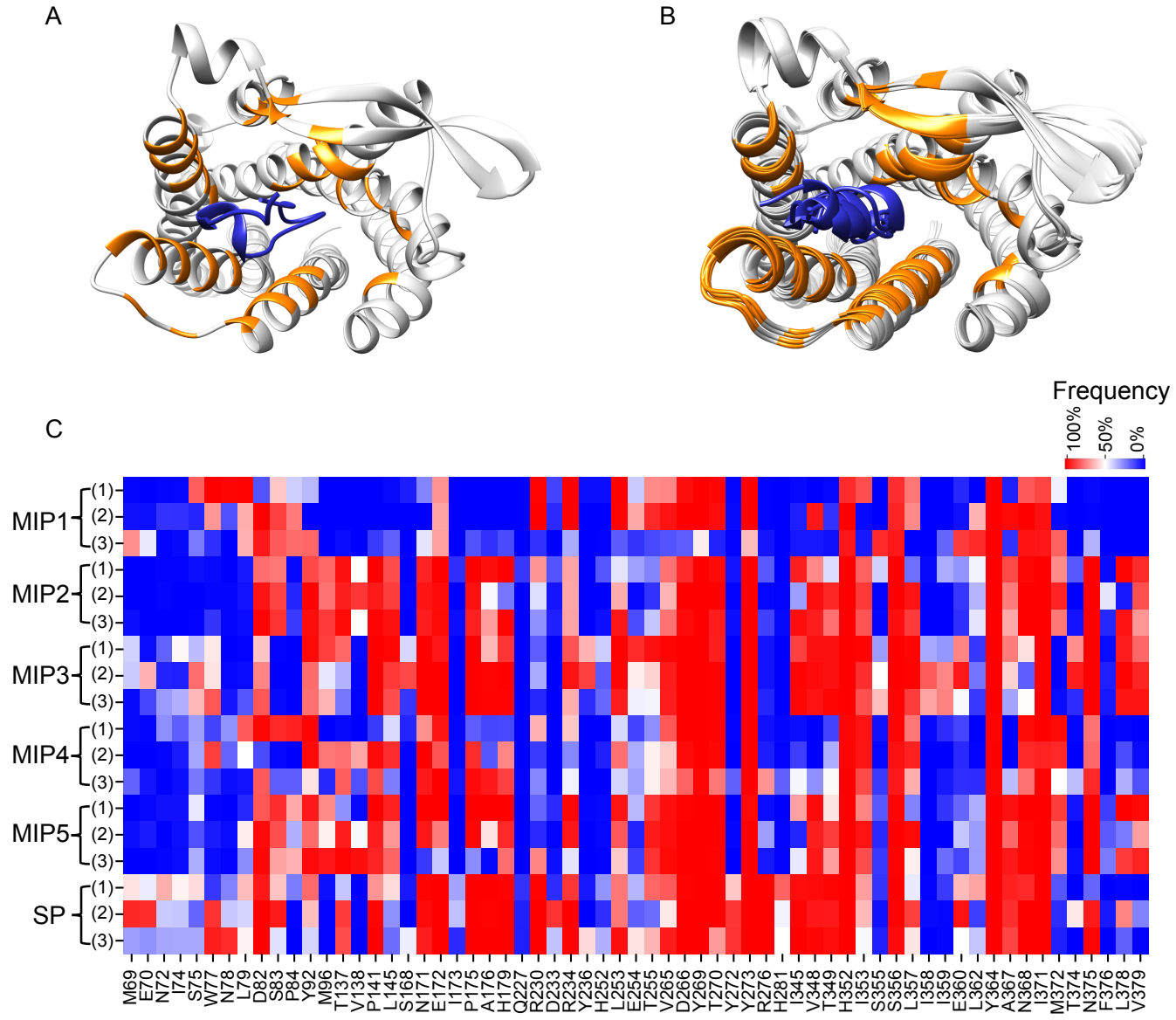

A
B
C
Frequency
100%
50%
0%
MIP1
MIP2
MIP3
MIP4
MIP5
SP
(1)
(2)
(3)
M69
E70
N72
I74
S75
W77
N78
L79
D82
S83
P84
Y92
M96
T137
V138
P141
L145
S168
N171
E172
I173
P175
A176
H179
Q227
R230
D233
R234
Y236
H252
L253
E254
T255
V265
D266
Y269
T270
Y272
Y273
R276
H281
I345
V348
T349
H352
I353
S355
S356
L357
I358
I359
E360
L362
Y364
A367
N368
I371
M372
T374
N375
F376
L378
V379

Figure 4

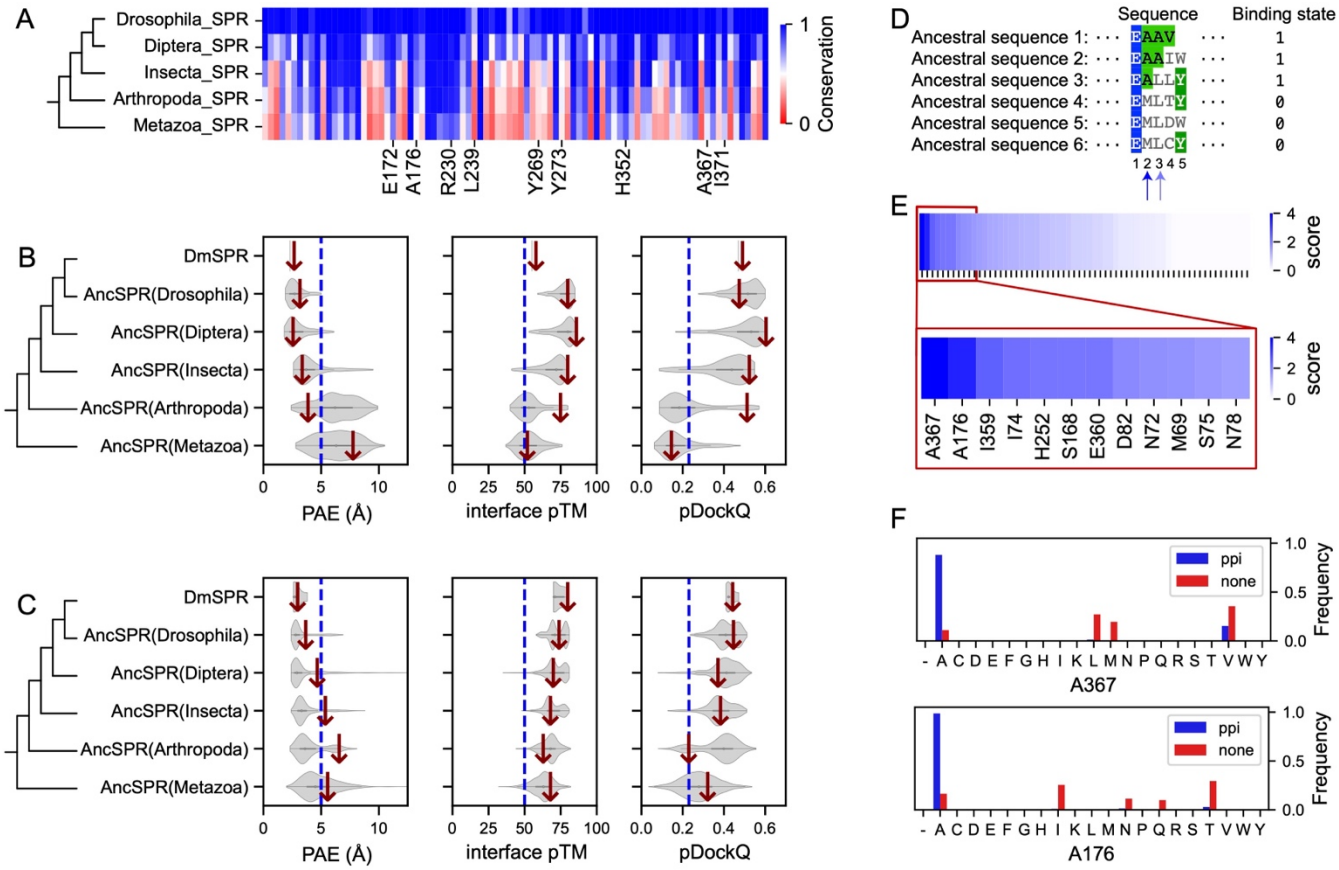
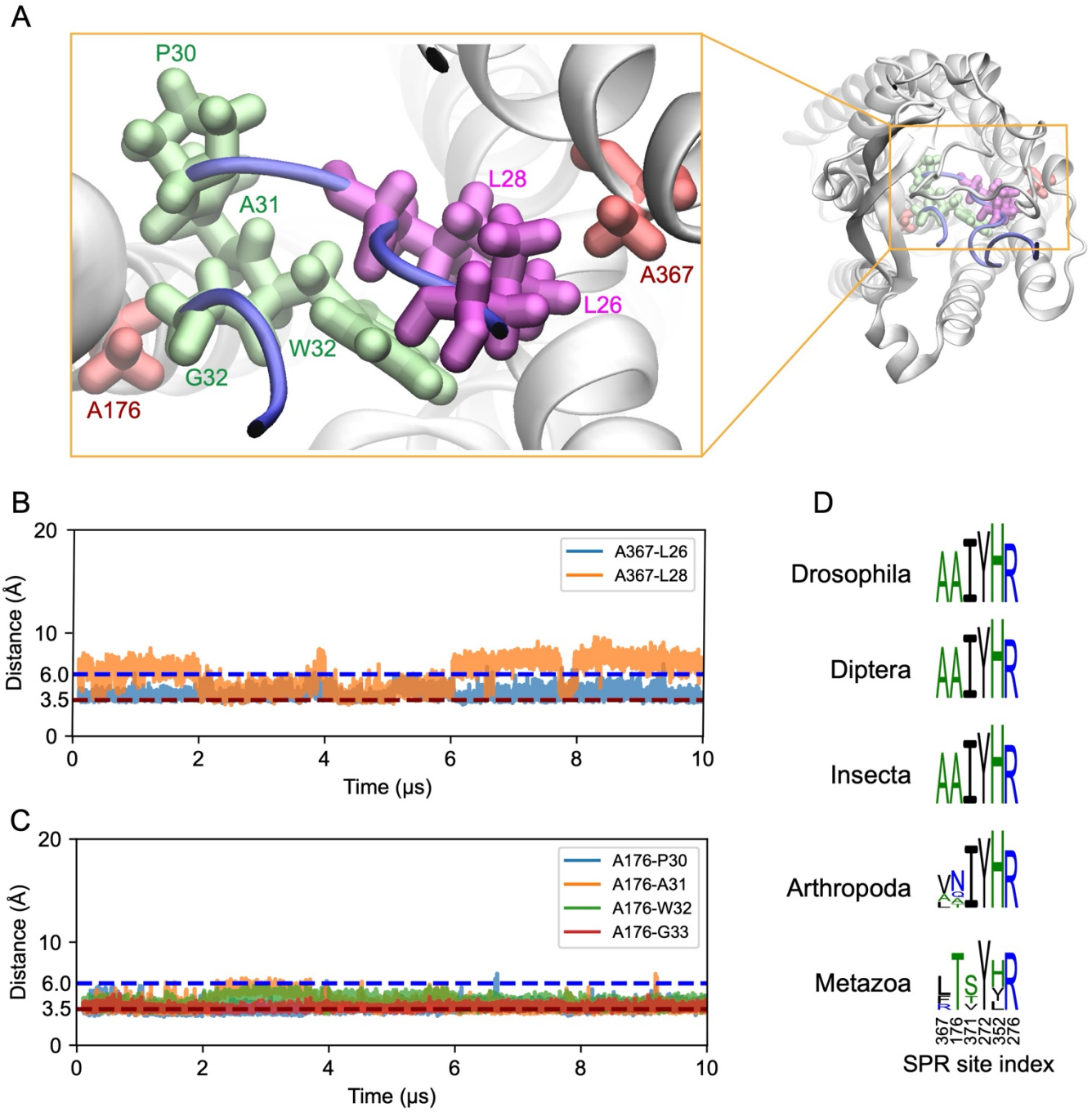

A
Drosophila_SPR
Diptera_SPR
Insecta_SPR
Arthropoda_SPR
Metazoa_SPR
Conservation
E172
A176
R230
L239
Y269
Y273
H352
A367
I371
B
DmSPR
AncSPR(Drosophila)
AncSPR(Diptera)
AncSPR(Insecta)
AncSPR(Arthropoda)
AncSPR(Metazoa)
PAE (Å)
interface pTM
pDockQ
C
D
Sequence
Binding state
Ancestral sequence 1:
Ancestral sequence 2:
Ancestral sequence 3:
Ancestral sequence 4:
Ancestral sequence 5:
Ancestral sequence 6:
E
score
A367
A176
I359
I74
H252
S168
E360
D82
N72
M69
S75
N78
F
ppi
none
Frequency
A367
A176

Figure 5


A
P30
A31
L28
A367
L26
W32
G32
A176
B
Distance (Å)
Time (μs)
A367-L26
A367-L28
C
Distance (Å)
Time (μs)
A176-P30
A176-A31
A176-W32
A176-G33
D
Drosophila
Diptera
Insecta
Arthropoda
Metazoa
SPR site index