

Drosophila proteins as human surrogates for early drug discovery: Findings from detailed structural analysis

Anjali Shankar¹, Rajeshri Karki¹

¹ Wardlaw+Hartridge School, Edison, New Jersey

SUMMARY

Drug discovery is a complex process involving various methods used to identify biological or chemical entities that can modify the function of specific proteins believed to play a role in disease. *Drosophila melanogaster*, or fruit fly, has been used by researchers for over a century to study human biology, genetics of human disease, and to find new medicines. Proteins from non-human species are utilized in the early stages of drug discovery when they share high sequence similarity with human homologs and are more accessible than human proteins. This study examined the sequence and structural similarity between *Drosophila* and human proteins to address the question of whether *Drosophila* proteins can be used as human surrogates for early drug discovery. We hypothesized that the sequence identity within the drug binding pocket of *D. melanogaster* proteins would exceed 50% when compared to their human homologs. To assess this hypothesis, we compared five protein sequences and crystal structures from *D. melanogaster* complexed with approved drugs or potential drug candidates to their human homologs. The overall sequence identity between the 2 species ranged from 38 to 71%, with good structural overlap. In the drug binding pocket, the sequence identity was 55 to 100%, supporting the hypothesis. However, we also identified cases where amino acids were not conserved in functionally critical regions, limiting the use of certain *Drosophila* proteins for identifying human therapeutic agents. The approach described here sheds light on the diverse aspects that must be considered when selecting proteins to address drug discovery questions.

INTRODUCTION

The common fruit fly, *Drosophila melanogaster*, is a widely used organism for investigating the molecular basis of human diseases (1). Scientists have used *D. melanogaster* to understand how gene alterations lead to disease, examine what happens when genes mutate, study the role of different proteins in human biology, screen for new therapeutic compounds (drug discovery), and analyze the mechanism of action of drugs or potential drugs using structural biology (1). Scientists finished mapping the *D. melanogaster* genome sequence in 2000 (2). It contains approximately 13,379 protein-coding genes (2). Previous research suggests that approximately 65% of the genes that cause diseases in humans have a matching functional homolog in *D. melanogaster*, and many of these homologs are expressed in tissues that perform the function of the equivalent human tissue (3). *D. melanogaster* is cheap and easy to maintain because it grows on simple food, like fruit, and requires

minimal space, which is one reason scientists began using this species in research studies (1). Each female can produce as many as 100 eggs daily and several thousand over the course of her life, offering researchers an abundant supply of organisms for diverse research activities (1).

Drug discovery is a complex process that involves multiple stages. One stage consists of the identification of biological or chemical agents that alter the function of a protein or another biological entity (e.g., DNA, RNA, etc.) thought to be involved in disease (4). When targeting a protein, a key focus is on protein binding pockets, which are specific regions where natural substrates bind. Scientists often look for biological or chemical agents that fit precisely into these pockets, blocking or activating the protein's function in a way that can treat disease (5). Scientists design biological assays to test whether protein function can be modulated by screening chemical libraries of compounds. In this approach, large quantities of proteins are required (5). If feasible, homologous proteins from other species such as *Saccharomyces cerevisiae*, *Escherichia coli*, mouse, rat, and *Drosophila*, are used for the initial screening (1, 6). Structural information about *D. melanogaster* proteins can also help scientists understand how human proteins work and how to design drugs that can block or activate those proteins (4). A high degree of sequence and structural similarity in the binding pocket between species suggests that a drug designed to target a *Drosophila* protein may bind similarly to its human homolog. This makes *Drosophila* an advantageous model for early-stage drug screening and structural studies before moving to validation in more complex human systems.

Given the potential of using *D. melanogaster* as a model organism for early drug discovery, we wanted to understand the structural similarity of proteins between *D. melanogaster* and humans. We hypothesized that the sequence identity within the drug binding pocket of *D. melanogaster* proteins would exceed 50% compared with the corresponding human homologs. The rationale for this prediction is that the drug binding pocket is, in most cases, the natural substrate binding pocket of a protein and is crucial for the protein's function. Homologous proteins from different species with the same function bind to the same or similar substrates, requiring similar amino acids for tight binding of the substrate and for carrying out this function. In addition to the conservation in the drug binding pocket, high overall sequence similarity may result in similar structural features, protein dynamics, and protein-protein interactions. Conservation of structure and sequence may therefore allow the use of proteins from different species, including *Drosophila*, as human surrogates for early drug discovery. To test our hypothesis, we conducted an extensive analysis of five protein sequences and crystal structures from *D. melanogaster* that were complexed with

either approved drugs or potential drug candidates, alongside their human homologs. Each selected protein target possessed a distinct biological function in humans. In this study, we analyzed five proteins: angiotensin-converting enzyme (ACE), phosphatidylinositol 3-kinase catalytic subunit type 3 (VPS34), acetylcholinesterase, neuronal calcium sensor 1 (NCS-1, also known as frequenin), and deoxynucleoside kinase (DNK) in *D. melanogaster* and deoxycytidine kinase (DCK) in humans.

ACE helps control blood pressure and the balance of salts in the body by breaking down certain circulating hormones such as angiotensin I, bradykinin or enkephalins (7, 8). ACE inhibitors are used as medicines to help lower blood pressure, treat heart problems, and protect the kidneys in people with diabetes (7).

VPS34 is a lipid kinase, which is a type of enzyme found across all eukaryotic organisms (9). It plays critical roles in the natural processes of recycling and repairing damaged cells, communicating between cells (referred to as cell signaling), and moving cellular components in different compartments within a cell (9). The ability of VPS34 inhibitors to modulate cellular processes has made them an area of interest for treating human diseases (10).

Acetylcholinesterase is an enzyme that rapidly hydrolyzes the acetylcholine neurotransmitter released into the synaptic cleft, allowing it to terminate signal transduction at the neuromuscular junction (11). It also plays a role in controlling the death of nerve cells (11). Donepezil, a medicine used to treat dementia associated with Alzheimer's disease, reversibly inhibits acetylcholinesterase (12).

NCS-1 is a highly conserved calcium binding protein that is expressed in neurons (13). It regulates the connectivity and communication between nerve cells, playing a vital role in cognitive functions such as learning and memory (13, 14). This protein is a potential target for treating mental illnesses like schizophrenia and bipolar disorder, as well as other conditions like severe behavioral problems in children (15, 16).

The DNK in *D. melanogaster* has a broad substrate specificity and can phosphorylate diverse substrates with high turnover rates (17). Mammals have four different DNKs while *D. melanogaster* has only one DNK. Medicines phosphorylated by DNK play a key role in the treatment of cancer and viral diseases, such as the chemotherapeutic drug gemcitabine (17, 18).

We observed that the drug binding pockets of the *Drosophila* and human proteins analyzed in this study showed high conservation, with sequence identity ranging from 55 to 100%, thereby supporting our hypothesis. Nevertheless, careful evaluation of the amino acids is needed, as even a single amino acid difference in a functionally significant position of the drug binding pocket can influence the suitability of a *Drosophila* protein for early drug discovery efforts. These findings emphasize the importance of analyzing both sequence and structural similarities when selecting model proteins for drug discovery research.

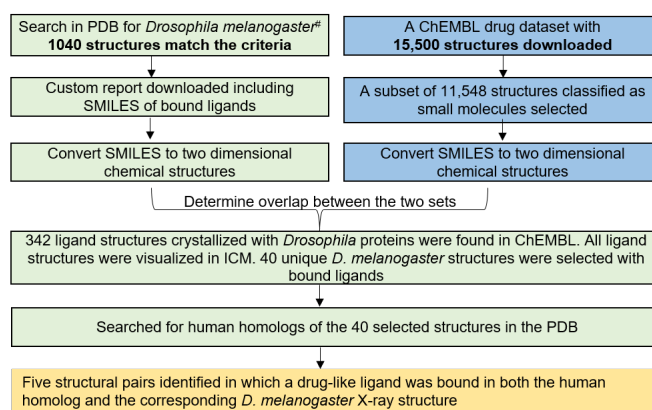
RESULTS

We initially queried the Protein Data Bank (PDB) for *D. melanogaster* crystal structures determined by X-ray diffraction, which yielded 1,040 entries (Figure 1) (6). We prioritized structures solved by X-ray crystallography because

this technique provides highly detailed, atom-level images of proteins and allows for precise identification of attached ligands (19). In parallel, we also obtained a subset of the ChEMBL drug set specifically categorized as small molecules to facilitate the prioritization of *D. melanogaster* structures based on the presence of a drug or potential drug candidates (20). ChEMBL is a database of drug-like bioactive molecules maintained by scientists at the European Molecular Biology Laboratory (EMBL) to support the development of new drugs (20). Using tools available in ICM browser version 3.9-4a, we identified *D. melanogaster* crystal structures with ligands also found in ChEMBL (21). We selected a subset of 40 unique *D. melanogaster* structures, each containing a ligand that was either an approved drug listed in ChEMBL or a chemical that was not a natural substrate (e.g., adenosine triphosphate) or a solvent (e.g., glycerol, polyethylene glycol, etc.). Five out of the 40 *D. melanogaster* structures had a human homolog bound to a ligand available in PDB, and hence we selected these five proteins (ACE, VPS34, acetylcholinesterase, NCS-1 and DNK) for detailed analysis. Reference sequences of the five proteins were downloaded from UniProt for analysis of the differences between the two species and comparison with X-ray sequences (22). In this paper we summarize the observations from the analysis of these structures.

ACE

Structures of *D. melanogaster* and human ACE, both complexed with the anti-hypertensive medicine lisinopril, were used for the detailed analysis. The human ACE structure with



*Search refined by structure determination methodology (experimental), scientific name of source organism (*Drosophila melanogaster*), experimental method (X-ray diffraction) and polymer entity type (protein)

Figure 1. Workflow used to identify and select protein structures for detailed analysis. Searches were conducted in the Protein Data Bank (PDB) for *Drosophila melanogaster* structures solved by X-ray crystallography. A custom report of the results was exported that contained the bound ligands in simplified molecular input line entry system (SMILES) format. An overlap with the ChEMBL drug set (categorized as small molecules) facilitated prioritizing *D. melanogaster* crystal structures based on the presence of approved drugs or potential drug candidates. Human and *D. melanogaster* structures of angiotensin-converting enzyme (ACE), phosphatidylinositol 3-kinase catalytic subunit type 3 (VPS34), acetylcholinesterase, neuronal calcium sensor 1 (NCS-1), deoxynucleoside kinase (DNK; known as deoxycytidine kinase [DCK] in humans) were identified crystallized with a drug or drug-like ligand. Structural analysis tools in ICM browser version 3.9-4a were used for processing the data, visualizing, and detailed analysis.

lisinopril (1O86.pdb) includes only ACE's extracellular domain, containing only the 493 C-terminal amino acids out of the 1,306 total amino acids in the complete sequence, while the *D. melanogaster* ACE structure (2X91.pdb) includes 595 out of 615 total residues (23, 24). The helical structures showed a good overlap between the two structures, but there were multiple flexible loops present in both structures, resulting in a root mean square deviation (RMSD) of 0.433 (**Figure 2A**). While the overall sequence identity between *D. melanogaster* and human ACE was 46%, which is not considered high, there was high conservation of amino acids in the catalytic site, with 24 out of the 33 amino acids (73%) within 5 Å of lisinopril being identical between the two structures (**Figure 2B**, **Figure 2C**).

VPS34

The human VPS34 structure (4UWF.pdb) contains 550 out of 887 amino acids, and the *D. melanogaster* structure (2X6F.pdb) contains 542 out of 949 amino acids, with an RMSD of 0.653 between the two structures (**Figure 3A**) (9, 25). The bound ligands in the *D. melanogaster* and human VPS34 X-ray structures are different. The overall sequence identity between the human and *D. melanogaster* VPS34 structural sequences was 62%. Seventeen out of 22 amino acids (77%) within 5 Å of the complexed inhibitors were identical, indicating high homology at the substrate binding pocket (**Figure 3B**, **Figure 3C**).

Acetylcholinesterase

The human X-ray structure (7E3H.pdb) contains 527 out of 614 amino acids, while the *D. melanogaster* structure (6XYX.pdb) contains 537 out of 649 amino acids, with an RMSD of 0.372 between the two structures (**Figure 4A**) (12, 26). The ligand complexed with human acetylcholinesterase was the drug donepezil, a medication used to treat dementia associated with Alzheimer's disease (12). The ligand complexed with *D. melanogaster* acetylcholinesterase is not an approved drug. The sequence identity between *D. melanogaster* and human acetylcholinesterase structural sequences was 39%, with 17 out of 31 amino acids (54.84%) being identical within 5 Å of the bound ligands (**Figure 4B**, **Figure 4C**).

NCS-1

The *D. melanogaster* structure (5G08.pdb) was complexed with chlorpromazine, a drug primarily used to treat serious mental illnesses like schizophrenia and bipolar disorder (15, 16). The ligand complexed with human NCS-1 is not an approved drug. Both the *D. melanogaster* and human NCS-1 crystal structures represent nearly the full-length sequences, with 180 out of 187 amino acids in the *D. melanogaster* structure and 187 out of 190 amino acids in the human structure (6QI4.pdb) (27). In both the *D. melanogaster* and human X-ray structures of NCS-1, three calcium ions were complexed that showed a perfect overlap when the two structures were superimposed (**Figure 5A**). The overall sequence identity between the two NCS-1 structures was 71%, and all 22 amino acids (100%) within 5 Å of the bound ligands were identical between *D. melanogaster* and human structures (**Figure 5B**, **Figure 5C**). One flexible loop does not overlap perfectly between the two structures, resulting in an RMSD of 0.503 (**Figure 5A**). In the *D. melanogaster* structure, one amino acid (isoleucine 178) was mutated to methionine (**Figure 5C**).

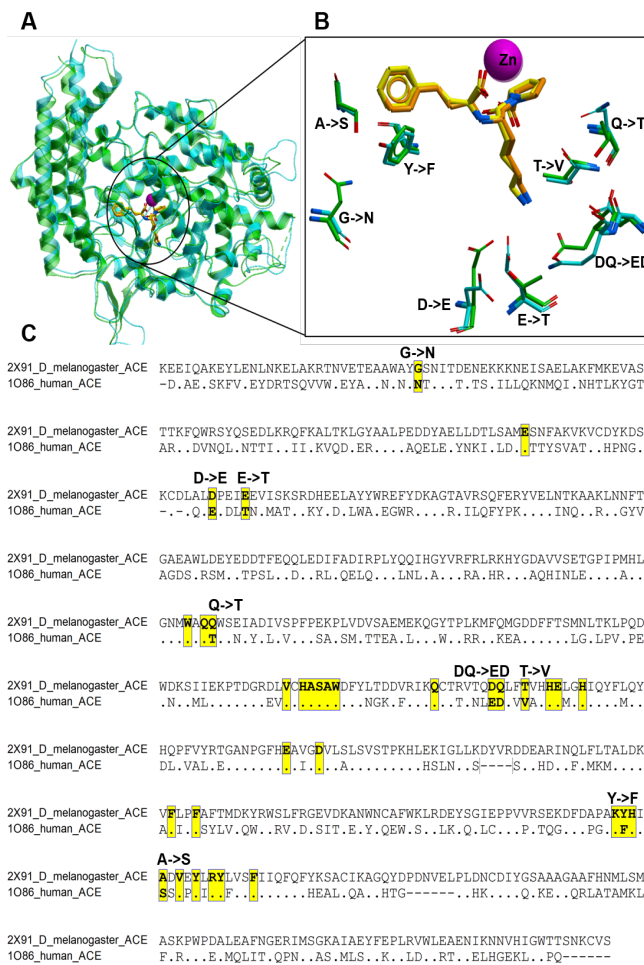


Figure 2. Comparison of *D. melanogaster* and human ACE protein sequences and crystal structures. (A) Overlay of the complete X-ray structure of *D. melanogaster* (cyan, 2X91.pdb) and human angiotensin-converting enzyme (ACE) (green, 1O86.pdb) is presented. (B) Close-up view of the nine amino acids that are within 5 Å of lisinopril and different between *D. melanogaster* and human ACE. Lisinopril carbon atoms are colored yellow (human) and orange (*D. melanogaster*), with zinc atoms depicted as magenta spheres. (C) Sequence alignment of *D. melanogaster* and human ACE protein sequences extracted from crystal structures 2X91.pdb (*D. melanogaster*) and 1O86.pdb (human). Identical amino acids are indicated by dots in the human sequence. Residues in bold with yellow background are within 5 Å of bound lisinopril. The nine differing amino acids between the two structures species are annotated.

DNK

DNK from *D. melanogaster* (2VPP.pdb) was compared with human DCK (2NO0.pdb), as both structures were complexed with the anti-cancer drug gemcitabine (**Figure 6A**, **Figure 6B**) (17, 18). The human X-ray structure contains 242 out of 260 amino acids, with 3 cysteine residues mutated to serine (**Figure 6C**). The *D. melanogaster* structure contains 189 out of 250 amino acids. The overall sequence identity between *D. melanogaster* DNK and human DCK was 38%. However, 15 out of the 20 amino acids (75%) within 5 Å of gemcitabine are identical between *D. melanogaster* and human structures (**Figure 6B**, **Figure 6C**). The structural differences between *D. melanogaster* DNK and human DCK are more evident in the loop region, resulting in an RMSD of 0.500.

and human in the hinge region were Tyr746_{D. melanogaster} - Phe684_{human}, Val747_{D. melanogaster} - Ile685_{human}, and Asp748_{D. melanogaster} - Gln686_{human}. In addition, the gatekeeper residue was leucine in *D. melanogaster* and methionine in human VPS34 (Leu744_{D. melanogaster} - Met682_{human}). Researchers have reported that the size of the gatekeeper residue affects the shape and depth of the catalytic site in kinases, which in turn influences the binding affinity and selectivity of ligands that bind the catalytic site (31). Since kinase inhibitors function by binding to this pocket, variations in the gatekeeper residue between *D. melanogaster* and human VPS34 may influence how well a drug accesses the pocket, how strongly it binds, and how well it functions across the two species (31). Individuals with mutations affecting the gatekeeper can become resistant to the medicines that function by binding in this pocket (31, 32). Because of the differences in the amino acids in the hinge and gatekeeper between *D. melanogaster* and human VPS34, the binding affinity of a drug binding in this pocket will likely be different between the two species. Hence, our findings suggest that *D. melanogaster* VPS34 can be used to study protein structure and function, as well as other basic biology concerns involving VPS34, but cannot be used for designing selective VPS34 inhibitors. To design selective human VPS34 inhibitors, researchers should use the human VPS34 protein. Alternatively, the hinge and gatekeeper amino acids in *D. melanogaster* could be mutated to match those in humans for early drug discovery efforts, such as testing new compounds for therapeutic use.

In the case of acetylcholinesterase from *D. melanogaster* and humans, the low overall sequence identity of 39% does not appear to influence the structural fold. We observed comparable secondary structure and good overlap between the structures of the two species except the multiple flexible loops resulting in low structural identity (RMSD of 0.372). The amino acid differences in the catalytic ligand binding pocket were substantial (only 17 out of the 31 amino acids were identical), resulting in varied sizes of the pocket. The amino acid differences may influence the size and shape of the pocket as well as the binding affinity of the ligands. Therefore, when designing ligands as inhibitors of human acetylcholinesterase, it will be more advantageous to use the human protein. The *D. melanogaster* protein will be useful for understanding general biology questions related to acetylcholinesterase including development of better insecticides (33).

NCS-1 exhibited high overall sequence identity of 71% between *D. melanogaster* and human structural sequences. In addition, all 22 amino acids (100%) within 5 Å of the bound ligand were identical between *D. melanogaster* and human structures. The mutated amino acid isoleucine 178 in *D. melanogaster* structure was not within the ligand binding pocket hence it will not impact ligand binding.

The overall sequence identity between *D. melanogaster* DNK and human DCK was 38%. However, 15 out of the 20 amino acids (75%) within 5 Å of gemcitabine were identical between *D. melanogaster* and human structures. The structural differences between *D. melanogaster* DNK and human DCK were more evident in the loop region, resulting in an RMSD of 0.500. Based on the structural similarity of *D. melanogaster* and human NCS-1 and DNK/DCK, *D. melanogaster* proteins may be used as a surrogate of human proteins for understanding structure and function, screening new therapeutic compounds, computational modeling, and

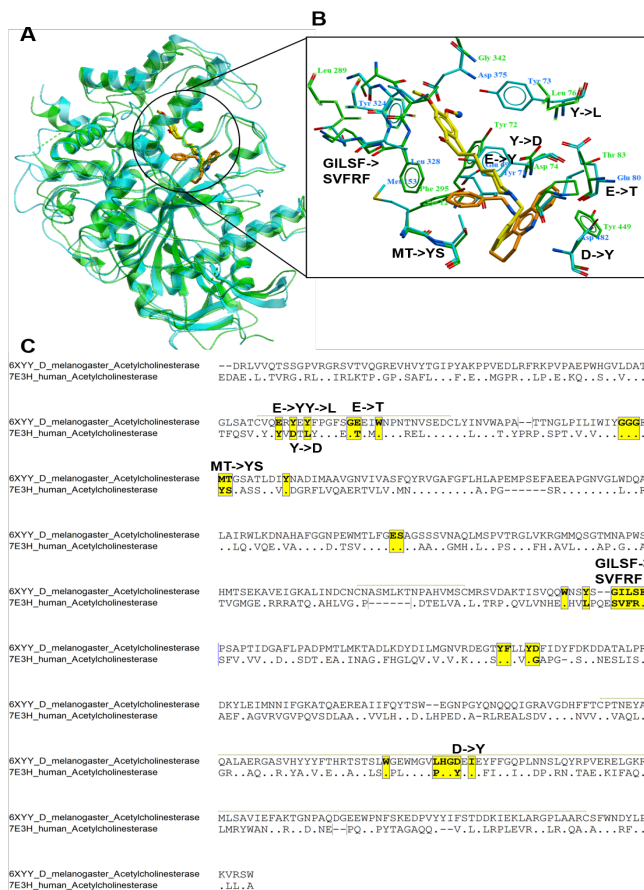


Figure 4. Comparison of *D. melanogaster* and human acetylcholinesterase protein sequences and crystal structures. (A) Overlay of the complete X-ray structure of *D. melanogaster* (cyan, 6XYY.pdb) and human acetylcholinesterase (green, 7E3H.pdb) is presented. (B) Close-up view of amino acids that are different within 5 Å of the bound inhibitors between *D. melanogaster* and human acetylcholinesterase. Inhibitor carbon atoms are colored yellow (human) and orange (*D. melanogaster*). (C) Sequence alignment of *D. melanogaster* and human acetylcholinesterase protein sequences extracted from crystal structures 6XYY.pdb (*D. melanogaster*) and 7E3H.pdb (human). Identical amino acids are indicated by dots in the human sequence. Residues in bold with yellow background are within 5 Å of bound inhibitors. In panels (B) and (C) selected amino acids differing between the two species are annotated for visualization purposes.

other human drug discovery efforts.

Sequence identity of *Drosophila* proteins ranged from 38 to 71% overall compared to their human homologs for the five proteins we investigated. Consistent with our hypothesis, the sequence identity in the drug binding pocket of *D. melanogaster* proteins was between 55 and 100% similar to their human homologs. However, we were surprised to find that subtle differences in the amino acids in functionally significant positions, such as the hinge and gatekeeper residues in VPS34, may influence decisions about their utility for drug discovery efforts. All five proteins examined in this report have contributed to advancing our understanding of their respective functions, catalytic mechanisms, and drug binding modes, even when the sequence identity is relatively low. We focused on sequence identity (amino acids are the

same between two species) rather than sequence similarity (amino acids are of the same type; e.g., glutamic acid in one sequence vs. aspartic acid in another would be considered similar, as they are both acidic) because of our interest in interactions in the drug binding sites where amino acids are engaged in drug binding. It should be noted that binding affinity and kinetics may vary between *D. melanogaster* and humans because of the differences in protein conformations, dynamics of ligand binding, assay conditions, and these differences warrant careful considerations (34).

In conclusion, even though fruit flies are useful for studying genes, proteins, and biological pathways, there are some fundamental limitations because human biology is more complex and dynamic compared to *D. melanogaster*. Proteins interact with other proteins in a cell, and subtle differences can influence these interactions and the biological processes that follow. Researchers have inferred that studying human diseases can present challenges when using *D. melanogaster* as a model (1). Proteins from *D. melanogaster* may be used in early stages of drug discovery when sequence identity in the functional or catalytic region of the protein is greater than 50% and when the functionally critical residues are also identical (35). *D. melanogaster* structures with lower sequence identity

to humans (30–50%) may be used for building computational models of human proteins (homology models) for structure-based design of drugs (35). Although *D. melanogaster* remains a valuable and widely used model organism in biomedical research, it is essential to complement and validate its findings with functional studies in mammalian models such as rats, mice, pigs, or monkeys (1). The multiple aspects considered in this study while making a conclusion on the utility of a *Drosophila* protein for early human drug discovery can be applied to determine the utility of other proteins.

Looking ahead, several directions can be pursued to further expand the impact of *D. melanogaster* in drug discovery and biomedical research. One exciting possibility is to use artificial intelligence-based approaches to mine huge collections of structural and sequence data to predict conserved regions of proteins across species that could be targeted for drug discovery. A limitation of this study was the availability of only five pairs of *D. melanogaster* and human X-ray structures complexed with a drug or drug-like ligand. As next steps, expanding the structural coverage of *D. melanogaster* proteins complexed with diverse drug-like molecules through experimental and computational methods will allow comparison with human homologs to generate additional insights and enhance the utility of this model organism in rational drug design. Establishing an open-access, integrated database linking *D. melanogaster* protein structures, sequences, genetic phenotypes, experimental ligand-binding data, and human disease associations would facilitate cross-species translational research. Collaborative efforts to validate key findings in human cell systems and mammalian models will remain essential to ensure clinical relevance. By embracing innovative technologies, future research can accelerate the discovery of next-generation therapeutics.

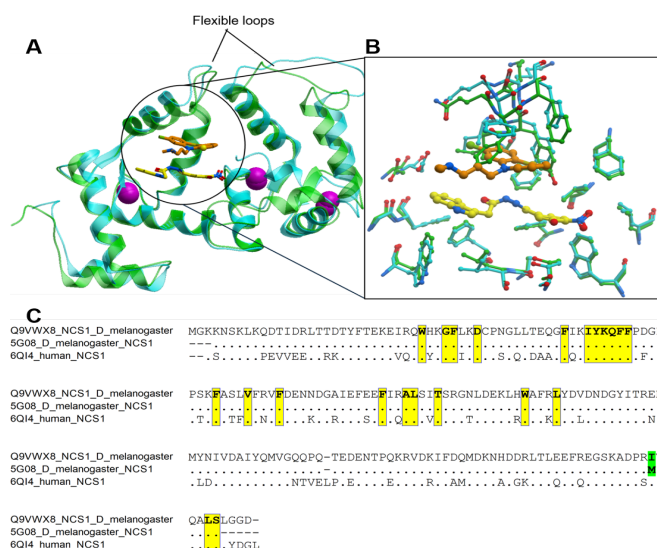


Figure 5. Comparison of *D. melanogaster* and human NCS-1 protein sequences and crystal structures. (A) Overlay of the complete X-ray structure of *D. melanogaster* (cyan, 5G08.pdb) and human neuronal calcium sensor 1 (NCS-1) (green, 6QI4.pdb) is presented. (B) Close-up view of amino acids within 5 Å of bound inhibitors is shown. Inhibitor carbon atoms are colored yellow (human) and orange (*D. melanogaster*), with calcium atoms depicted as magenta spheres. Inhibitors do not overlap, as they make different interactions in the binding pocket. All 22 amino acids in the binding pocket are identical, demonstrating strong overlap, except for those within the flexible loop region. (C) Sequence alignment of *D. melanogaster* and human NCS-1 protein sequences extracted from crystal structures 5G08.pdb (*D. melanogaster*) and 6QI4.pdb (human), with the third sequence being *D. melanogaster* NCS-1 from UniProt (accession id Q9VWX8). Identical amino acids are indicated by dots in the human sequence. Residues in bold with yellow background are within 5 Å of bound inhibitors. One amino acid (isoleucine 178), highlighted with a green background, is mutated to methionine in the *D. melanogaster* NCS-1 crystal structure.

MATERIALS AND METHODS

Data acquisition and processing

To select X-ray structures for detailed analysis, we first performed a full text search using *Drosophila melanogaster* as query text in the advanced search query feature in PDB (Figure 1) (6). The query matched 2,048 structures, which were refined in the PDB web portal by structure determination methodology (experimental), scientific name of source organism (*Drosophila melanogaster*), and polymer entity type (protein), resulting in 1,358 structural entries. Further filtering for experimental method as “X-ray diffraction” resulted in 1,040 entries (Figure 1). The remaining 318 structures were determined largely by electron microscopy and solution nuclear magnetic resonance (NMR) spectroscopy. The 1,040 structures determined by X-ray crystallography correspond to 385 unique proteins, with multiple structural records available for several proteins. A custom report was generated from the PDB to include comprehensive protein details including UniProt identification numbers, relevant citations, and information about associated bound ligands, which were represented using Simplified Molecular Input Line Entry System (SMILES) (22, 36). SMILES is a way to represent a chemical structure as a string of text in a linear format using regular keyboard characters, making it easier for computers to process and understand. The custom report was downloaded as a comma separated file.

We used the ChEMBL drug set specifically categorized as small molecules to facilitate the selection of *D. melanogaster*

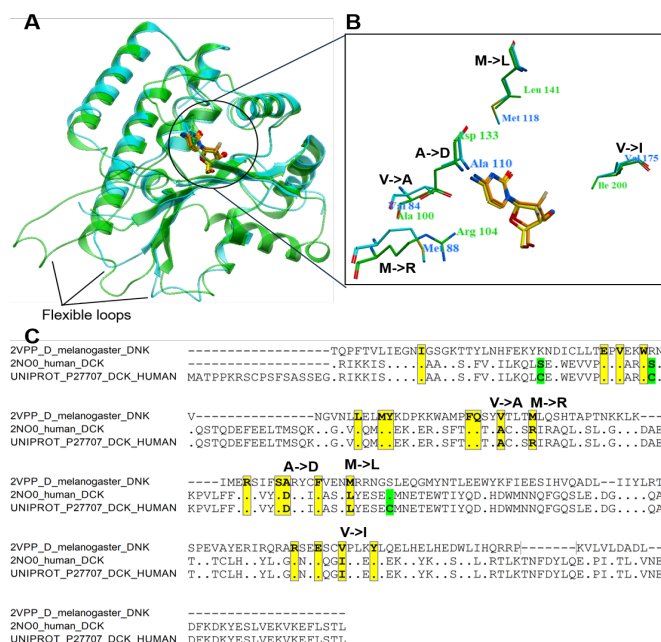


Figure 6. Comparison of *D. melanogaster* DNK and human DCK protein sequences and crystal structures. (A) Overlay of the complete X-ray structure of *D. melanogaster* deoxynucleoside kinase (DNK) (cyan, 2VPP.pdb) and human deoxycytidine kinase (DCK) (green, 2NO0.pdb) is presented. (B) Close-up view of the five amino acids that are within 5 Å of gemcitabine (bound ligand) and different between *D. melanogaster* DNK and human DCK is shown. Inhibitor carbon atoms are colored yellow (human) and orange (*D. melanogaster*). (C) Sequence alignment of *D. melanogaster* DNK and human DCK protein sequences extracted from crystal structures 2VPP.pdb (*D. melanogaster*) and 2NO0.pdb (human DCK), with the third sequence being human DCK from UniProt (accession id P27707). Identical amino acids are indicated by dots in the human sequence. Residues in bold with yellow background are within 5 Å of bound ligand (gemcitabine). Three amino acids (Cysteine 45, 59 and 146), highlighted with a green background, are mutated to serine in the human DCK crystal structure.

structures based on the presence of a drug or potential drug candidates (20). ChEMBL is a carefully maintained collection of drug-like bioactive molecules that integrates chemical information and biological activity data including genomic data from different sources to support the development of new drugs (20). Within the ChEMBL web portal, an initial set of 15,500 compounds was selected (37). This dataset was subsequently refined by filtering for small molecules, resulting in a final list of 11,548 compounds. This filtered set was downloaded as a comma separated file.

The files downloaded from both PDB and ChEMBL were imported as two separate tables in ICM version 3.9-4a (21). ICM was used to convert SMILES notation in each of the tables to 2D chemical structures. The data from the ChEMBL set was merged with the PDB table based on overlapping 2D ligand structures. This approach enabled tagging of ChEMBL IDs and other relevant information from the ChEMBL file to the PDB table only for ligands present in both tables. 342 ligands crystallized in 294 *D. melanogaster* structures were also found in the ChEMBL set, with multiple ligands present in some *D. melanogaster* structures.

The 2D structures of all ligands crystallized with *D.*

melanogaster were visualized using ICM. From this analysis, a subset of 40 unique protein structures of *D. melanogaster* was selected such that the structures contained bound ligands that were either approved drugs with available ChEMBL data, or a chemical ligand not classified as natural substrates (e.g., ATP) or solvents (e.g., glycerol, polyethylene glycol). UniProt identification numbers were used to retrieve information about each protein's function, the full-length native sequence, and disease relevance (22). Using protein names, additional searches were performed in PDB to identify human homologs for each of the 40 selected *D. melanogaster* structures (6). Ultimately, only five structural pairs were identified in which a drug or drug-like ligand was bound in both the human homolog and the corresponding *D. melanogaster* X-ray structure. Structural overlays, visualization, sequence alignment, analysis and image generation were conducted using ICM software (21).

Protein structure superimposition and analysis

Selected protein crystal structures were downloaded from the PDB for detailed analysis (6). High resolution crystal structures complexed with approved drugs were prioritized for analysis if available. Structures of *D. melanogaster* (2X91.pdb) and human ACE (1O86.pdb), both complexed with the anti-hypertensive medicine lisinopril, were used for the detailed analysis (23, 24). For VPS34, X-ray structures were not available bound with the same inhibitor, hence high-resolution structures of (2X6F.pdb) and human VPS34 (4UWF.pdb) were selected for analysis (9, 25). Crystal structure of *D. melanogaster* (6XYY.pdb) complexed with a drug-like ligand and human acetylcholinesterase (7E3H.pdb) complexed with the drug donepezil were downloaded from the PDB (12, 26). The *D. melanogaster* NCS-1 structure (5G08.pdb) complexed with a drug chlorpromazine and human NCS-1 (6QI4.pdb) complexed with a drug-like ligand represented nearly the full-length sequence, hence selected for analysis (16, 27). DNK from *D. melanogaster* (2VPP.pdb) was compared with human DCK (2NO0.pdb), as both structures were complexed with the anti-cancer drug gemcitabine (17, 18). All ten crystal structures were loaded in ICM. Protein sequences from each of the ten crystal structures were extracted for sequence alignment. Pairs of homologous protein structures were converted to ICM objects and superimposed. Using tools available in ICM, amino acids within 5 Å of each of the bound ligand were selected for visualization and analysis in each pair of homologous proteins. Differences in ligand did not influence the amino acid selection as both complexed ligands were considered for selecting amino acids in homologous protein pairs. It is common practice to analyze ligand-protein interactions within 5 Å of the ligand, as the interactions within this distance are the strongest and relevant for tight ligand binding (28).

Protein sequence alignment

The reference protein sequences from UniProt utilized for analysis were Q10714 (*D. melanogaster* ACE), P12821 (human ACE), Q9W1M7 (*D. melanogaster* VPS34), Q8NEB9 (human VPS34), P07140 (*D. melanogaster* acetylcholinesterase), P22303 (human acetylcholinesterase), Q9VWX8 (*D. melanogaster* NCS-1), P62166 (human NCS-1), Q9XZT6 (*D. melanogaster* DNK) and P27707 (human DCK) (22). In addition, protein sequences extracted from each of the 10 crystal structures were included for the sequence analysis.

Sequence alignment was automated in ICM with no manual intervention. Reference sequences downloaded from UniProt were compared with the sequence from the crystal structures to identify if amino acids were mutated for crystallization. Following this, sequence extracted from each pair of *D. melanogaster* and human X-ray structures were aligned and analyzed to assess sequence identity. The overall sequence identity values were automatically calculated by ICM after sequence alignment.

RMSD calculation

RMSD was determined to measure protein structure similarity between pairs of *D. melanogaster* and its human homolog. RMSD quantifies the average distance between corresponding atoms in protein structures after they are superimposed. To calculate RMSD, pairs of homologous protein structures were opened in ICM and superimposed, followed by RMSD calculation using a widget available in ICM. All atoms were selected to calculate RMSD between two homologous structures.

Received: August 17, 2025

Accepted: February 3, 2026

Published: October 11, 2026

REFERENCES

1. Giansanti, Maria Grazia, et al. "Drosophila melanogaster: How and Why It Became a Model Organism." *International Journal of Molecular Sciences*, vol. 26, no. 15, 2 Aug. 2025. <https://doi.org/10.3390/ijms26157485>.
2. Misra, Sima, et al. "Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review." *Genome Biology*, vol. 3, 31 Dec. 2002, <https://doi.org/10.1186/gb-2002-3-12-research0083>.
3. Ugur, Berrak, et al. "Drosophila tools and assays for the study of human diseases." *Disease Models & Mechanisms*, vol. 9, no. 3, Mar. 2016, pp. 235–244. <https://doi.org/10.1242/dmm.023762>.
4. Pandey, Udai Bhan, and Charles D Nichols. "Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery." *Pharmacological Reviews*, vol. 63, no. 2, June 2011, pp. 411–436. <https://doi.org/10.1124/pr.110.003293>.
5. Santos, Rita, et al. "A comprehensive map of molecular drug targets." *Nature Reviews Drug Discovery*, vol. 16, no. 1, Jan 2017, pp. 19–34. <https://doi.org/10.1038/nrd.2016.230>.
6. Berman, H. M., et al. "The Protein Data Bank." *Nucleic Acids Research*, vol. 28, no. 1, 1 Jan 2000, pp. 235–242. <https://doi.org/10.1093/nar/28.1.235>.
7. Naqvi, Nawazish, et al. "Molecular basis of exopeptidase activity in the C-terminal domain of human angiotensin I-converting enzyme: insights into the origins of its exopeptidase activity." *The Journal of Biological Chemistry*, vol. 280, no. 8, 21 Dec. 2004, pp. 6669–6675. <https://doi.org/10.1074/jbc.M412638200>.
8. Anthony, Colin S., et al. "The N domain of human angiotensin-I-converting enzyme: the role of N-glycosylation and the crystal structure in complex with an N domain-specific phosphinic inhibitor, RXP407." *The Journal of Biological Chemistry*, vol. 285, no. 46, 8 Sep. 2010, pp. 35685–35693. <https://doi.org/10.1074/jbc.M110.167866>.
9. Miller, Simon, et al. "Shaping development of autophagy inhibitors with the structure of the lipid kinase VPS34." *Science*, vol. 327, no. 5973, 26 Mar. 2010, pp. 1638–1642. <https://doi.org/10.1126/science.1184429>.
10. Liu, Yuan, et al. "Targeting VPS34 in autophagy: An update on pharmacological small-molecule compounds." *European Journal of Medicinal Chemistry*, vol. 256, 5 Aug. 2023. <https://doi.org/10.1016/j.ejmech.2023.115467>.
11. Shafferman, A., et al. "Mutagenesis of human acetylcholinesterase. Identification of residues involved in catalytic activity and in polypeptide folding." *The Journal of Biological Chemistry*, vol. 267, no. 25, 5 Sep. 1992, pp. 17640–17648.
12. Dileep, K. V., et al. "Crystal structure of human acetylcholinesterase in complex with tacrine: Implications for drug discovery." *International Journal of Biological Macromolecules*, vol. 210, 15 June 2022, pp. 172–181. <https://doi.org/10.1016/j.ijbiomac.2022.05.009>.
13. Romero-Pozuelo, Jesús, et al. "Chronic and acute alterations in the functional levels of Frequentins 1 and 2 reveal their roles in synaptic transmission and axon terminal morphology." *The European Journal of Neuroscience*, vol. 26, no. 9, 26 Oct. 2007, pp. 2428–2243. <https://doi.org/10.1111/j.1460-9568.2007.05877.x>.
14. Dason, Jeffrey S., et al. "Multiple roles for frequenin/NCS-1 in synaptic function and development." *Molecular Neurobiology*, vol. 45, no. 2, 7 Mar. 2012, pp. 388–402. <https://doi.org/10.1007/s12035-012-8250-4>.
15. Boyd-Kimball, Debra, et al. "Classics in Chemical Neuroscience: Chlorpromazine." *ACS Chemical Neuroscience*, vol. 10, no. 1, 16 Jan. 2019, pp. 79–88. <https://doi.org/10.1021/acschemneuro.8b00258>.
16. Mansilla, Alicia, et al. "Interference of the complex between NCS-1 and Ric8a with phenothiazines regulates synaptic function and is an approach for fragile X syndrome." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 6, 24 Jan. 2017, pp. E999–E1008. <https://doi.org/10.1073/pnas.1611089114>.
17. Knecht, Wolfgang, et al. "Drosophila melanogaster deoxyribonucleoside kinase activates gemcitabine." *Biochemical and Biophysical Research Communications*, vol. 382, no. 2, 1 May 2009, pp. 430–433. <https://doi.org/10.1016/j.bbrc.2009.03.041>.
18. Sabini, Elisabetta, et al. "Nonenantioselectivity property of human deoxycytidine kinase explained by structures of the enzyme in complex with L- and D-nucleosides." *Journal of Medicinal Chemistry*, vol. 50, no. 13, 27 May 2007, pp. 3004–3014. <https://doi.org/10.1021/jm0700215>.
19. Papageorgiou, Anastassios C., et al. "Protein Structure Analysis and Validation with X-Ray Crystallography." *Methods in Molecular Biology*, vol. 2178, 01 Nov. 2020, pp. 377–404. https://doi.org/10.1007/978-1-0716-0775-6_25.
20. Gaulton, Anna, et al. "ChEMBL: a large-scale bioactivity database for drug discovery." *Nucleic Acids Research*, vol. 40, Issue D1, 23 Sep 2011, pp. D1100–D1107. <https://doi.org/10.1093/nar/gkr777>.
21. Molsoft L.L.C. ICM-Browser. Version 3.9-4a. Copyright 1989-2025. Molsoft L.L.C.
22. UniProt Consortium. "UniProt: the Universal Protein Knowledgebase in 2025." *Nucleic Acids Research*,

- vol. 53, D1, 6 Jan 2025, pp. D609–D617. <https://doi.org/10.1093/nar/gkae1010>.
23. Natesh, Ramanathan, et al. "Crystal structure of the human angiotensin-converting enzyme-lisinopril complex." *Nature*, vol. 421, 19 Jan. 2003, pp. 551–554. <https://doi.org/10.1038/nature01370>.
24. Akif, Mohd, et al. "High-resolution crystal structures of *Drosophila melanogaster* angiotensin-converting enzyme in complex with novel inhibitors and antihypertensive drugs." *Journal of Molecular Biology*, vol. 400, no. 3, 16 Jul. 2010, pp. 502–517. <https://doi.org/10.1016/j.jmb.2010.05.024>.
25. Pasquier, Benoit, et al. "Discovery of (2S)-8-[(3R)-3-methylmorpholin-4-yl]-1-(3-methyl-2-oxobutyl)-2-(trifluoromethyl)-3,4-dihydro-2H-pyrimido[1,2-a]pyrimidin-6-one: a novel potent and selective inhibitor of VPS34 for the treatment of solid tumors." *Journal of Medicinal Chemistry*, vol. 58, no. 1, 17 Nov. 2014, pp. 376–400. <https://doi.org/10.1021/jm5013352>.
26. Nachon, Florian, et al. "A Second Look at the Crystal Structures of *Drosophila melanogaster* Acetylcholinesterase in Complex with Tacrine Derivatives Provides Insights Concerning Catalytic Intermediates and the Design of Specific Insecticides." *Molecules*, vol. 25, no. 5, 6 Mar. 2020. <https://doi.org/10.3390/molecules25051198>.
27. Canal-Martín, Andrea, et al. "Insights into real-time chemical processes in a calcium sensor protein-directed dynamic library." *Nature Communications*, vol. 10, no. 1, 26 Jun. 2019, pp. 2798. <https://doi.org/10.1038/s41467-019-10627-w>.
28. Bissantz, Caterina, et al. "A medicinal chemist's guide to molecular interactions." *Journal of medicinal chemistry*, vol. 53, no. 14, 26 Mar 2010, pp 5061-5084. <https://doi.org/10.1021/jm100112j>.
29. Acharya, K. Ravi, et al. "Ace revisited: a new target for structure-based drug design." *Nature Reviews Drug Discovery*, vol. 2, no. 11, Nov 2003, pp. 891–902. <https://doi.org/10.1038/nrd1227>.
30. Zhao, Zheng, and Bourne E. Philip. "How Ligands Interact with the Kinase Hinge." *ACS Medicinal Chemistry Letters*, Oct 2023 vol. 14, no. 11, 2023, pp. 1503–1508. <https://doi.org/10.1021/acsmmedchemlett.3c00212>.
31. Zhou, Yang, et al. "Targeting Gatekeeper Mutations for Kinase Drug Discovery." *Journal of Medicinal Chemistry*, vol. 65, no. 23, 17 Nov. 2022, pp. 15540–15558. <https://doi.org/10.1021/acs.jmedchem.2c01361>.
32. Alaimo, Peter J., et al. "Targeting the gatekeeper residue in phosphoinositide 3-kinases." *Bioorganic & Medicinal Chemistry*, vol. 13, no. 8, 15 April 2005, pp. 2825–2836. <https://doi.org/10.1016/j.bmc.2005.02.021>.
33. Menozzi, Philippe, et al. "Mutations of acetylcholinesterase which confer insecticide resistance in *Drosophila melanogaster* populations." *BMC Evolutionary Biology*, vol. 4, 5 Feb. 2004, <https://doi.org/10.1186/1471-2148-4-4>.
34. Spassov, Danislav S., "Binding Affinity Determination in Drug Design: Insights from Lock and Key, Induced Fit, Conformational Selection, and Inhibitor Trapping Models." *International Journal of Molecular Sciences*, vol. 25, 2004, 7124, <https://doi.org/10.3390/ijms25137124>.
35. Hillisch, Alexander, et al. "Utility of homology models in the drug discovery process." *Drug Discovery Today*, vol. 9, 2004, pp. 659-669. [https://doi.org/10.1016/S1359-6446\(04\)03196-4](https://doi.org/10.1016/S1359-6446(04)03196-4).
36. Weininger, David, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules", *Journal of Chemical Information and Computer Sciences*, vol. 28, no. 1, 1 Feb 1988, pp. 31–36. <https://doi.org/10.1021/ci00057a005>.
37. ChEMBL web portal: <https://www.ebi.ac.uk/chembl/>

Copyright: © 2026 Shankar and Karki. All JEI articles are distributed under the Creative Commons Attribution Noncommercial No Derivatives 4.0 International License. This means that you are free to share, copy, redistribute, remix, transform, or build upon the material for any purpose, provided that you credit the original author and source, include a link to the license, indicate any changes that were made, and make no representation that JEI or the original author(s) endorse you or your use of the work. The full details of the license are available at <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>.