



OPEN ACCESS

EDITED BY

Serdal Kirmizialtin,
New York University Abu Dhabi, United
Arab Emirates

REVIEWED BY

Jian Wang,
The Pennsylvania State University,
United States
Jose A. Gavira,
Spanish National Research Council
(CSIC), Spain

*CORRESPONDENCE

Georg Künze,
✉ georg.kuenze@uni-leipzig.de

RECEIVED 02 March 2023

ACCEPTED 26 April 2023

PUBLISHED 09 May 2023

CITATION

Engelberger F, Zakary JD and Künze G
(2023), Guiding protein design choices by
per-residue energy breakdown analysis
with an interactive web application.
Front. Mol. Biosci. 10:1178035.
doi: 10.3389/fmolb.2023.1178035

COPYRIGHT

© 2023 Engelberger, Zakary and Künze.
This is an open-access article distributed
under the terms of the [Creative
Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Guiding protein design choices by per-residue energy breakdown analysis with an interactive web application

Felipe Engelberger, Jonathan D. Zakary and Georg Künze*

Institute for Drug Discovery, Leipzig University, Leipzig, Germany

Recent developments in machine learning have greatly facilitated the design of proteins with improved properties. However, accurately assessing the contributions of an individual or multiple amino acid mutations to overall protein stability to select the most promising mutants remains a challenge. Knowing the specific types of amino acid interactions that improve energetic stability is crucial for finding favorable combinations of mutations and deciding which mutants to test experimentally. In this work, we present an interactive workflow for assessing the energetic contributions of single and multi-mutant designs of proteins. The energy breakdown guided protein design (ENDURE) workflow includes several key algorithms, including per-residue energy analysis and the sum of interaction energies calculations, which are performed using the Rosetta energy function, as well as a residue depth analysis, which enables tracking the energetic contributions of mutations occurring in different spatial layers of the protein structure. ENDURE is available as a web application that integrates easy-to-read summary reports and interactive visualizations of the automated energy calculations and helps users selecting protein mutants for further experimental characterization. We demonstrate the effectiveness of the tool in identifying the mutations in a designed polyethylene terephthalate (PET)-degrading enzyme that add up to an improved thermodynamic stability. We expect that ENDURE can be a valuable resource for researchers and practitioners working in the field of protein design and optimization. ENDURE is freely available for academic use at: <http://endure.kuenzelab.org>.

KEYWORDS

protein design, energy calculation, amino acid interaction, web application (app), machine learning

1 Introduction

The design of proteins with improved stability and activity is a critical aspect of research in biotechnology and related fields. It holds the potential to revolutionize a wide range of applications (Arnold, 2018), from developing enzymes for industrial processes (Chen and Arnold, 2020), antibodies and antivirals for medicine (Sevy and Meiler, 2014; Willis et al., 2015) to molecular switches and biosensors (Stein and Alexandrov, 2015; Quijano-Rubio et al., 2021). Protein design has been demonstrated to play a crucial role in many different areas of biotechnology (Castro et al., 2022; Habibi et al., 2022; Reetz, 2022).

Two widely used protein design approaches are directed evolution and computer-aided protein design. The former approach mimics the natural gene diversification and selection

process and involves iterative rounds of mutagenesis, which create a library of mutants, and selection of mutants with desired functions (Arnold, 2018). Computer-aided protein design typically involves algorithms that suggest mutations for experimental testing (Pan and Kortemme, 2021). These algorithms may be based on in-depth molecular modeling, e.g., with the Rosetta software suite (Leman et al., 2020), or machine learning predictions (Dauparas et al., 2022). The experimental testing of the designed proteins is time-, cost-, and labor-intensive. Thus, prioritizing the most probable candidates for experimental testing is necessary. To facilitate mutant selection, it can be informative to determine the specific types of amino acid interactions that contribute to protein stability and assess the energetic impact of mutations (Goldenzweig et al., 2016).

Machine learning algorithms have revolutionized the field of protein design, enabling researchers to generate novel proteins with improved properties more efficiently (Dauparas et al., 2022; Ferruz and Höcker, 2022). Two current state-of-the-art methods are the evolutionary and structure-based design method PROSS (Goldenzweig et al., 2016; Weinstein et al., 2021), and the deep learning method ProtMPNN (Dauparas et al., 2022). Both methods can yield tens to hundreds of candidate structures or sequences. Selecting the best candidates for experimental testing is a rather tricky task, particularly for designed proteins with multiple mutations, as the effect of each mutation is dependent upon the presence of other mutations—a phenomenon referred to as epistasis (Starr and Thornton, 2016). Yet the selection process determines the success of the overall design process. Thus, it is essential to have reliable, comprehensive, and easy-to-use methods for evaluating and selecting the most probable designs, based on the energetic magnitude and type of interactions (hydrogen bonds, salt bridges, etc.) introduced by the mutations.

Some existing tools for protein design and mutant selection include Rosetta (Leman et al., 2020), HotSpotWizard3.0 (Sumbalova et al., 2018), ProteinSolver (Strokach et al., 2021), and FoldX (Schymkowitz et al., 2005). Scoring functions and modeling algorithms from Rosetta were previously tested for the prediction of protein stability and affinity changes ($\Delta\Delta G$). Kellogg et al. (2011) investigated the role of conformational sampling in computing mutation-induced changes in protein stability and compared the predictions to experimental $\Delta\Delta G$ values. Barlow et al. (2018) developed the Flex ddG method using Rosetta ensemble-based estimation of changes in protein-protein binding affinity upon mutation. Frenz et al. (2020) focused on improving the prediction of protein mutational free energy using Cartesian coordinate minimization. However, all these tools often require extensive knowledge of the software, leaving non-expert users without easy access to analyze the outcomes of their protein design experiments. In the field of *de novo* protein design, DE-STRESS (Stam and Wood, 2021) has been developed to help non-expert users evaluate the plausibility of such designs. Unfortunately, equivalent tools for assessing the results of sequence design experiments operated on a provided structure are lacking. To address this gap, we have developed ENergy breakDown gUided pRotein dEsign (ENDURE), a modular web application that provides an interactive and user-friendly interface for analyzing the energetic contributions of protein designs. ENDURE integrates easy-to-read summary tables and interactive visualizations of automated energy calculations, helping users to explore and

reveal mutational hotspots—which confer stabilization or destabilization—and compare the specific types of interactions that a particular mutation is introducing. In that way, ENDURE helps selecting the best protein mutants for further experimental characterization.

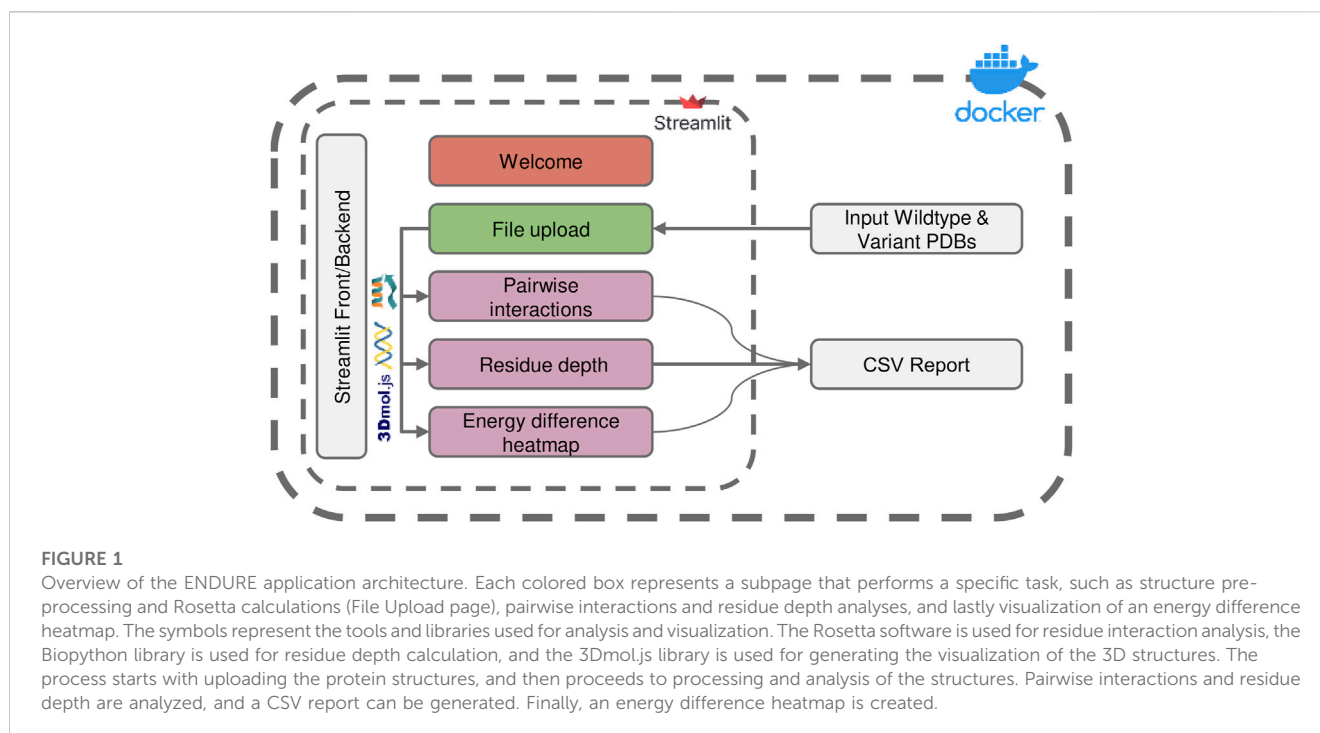
The application workflow (Figure 1) integrates several key algorithms, which analyze the protein structure using the Rosetta energy function, including per-residue energy breakdown and the sum of interaction energies calculations. Additionally, the tool provides a residue depth analysis, which enables users to track the energetic contributions of mutations occurring at different spatial layers of the protein structure, thus easily shedding light on the particular strategies that a design pipeline might have used and assessing its particular energetic impact. We demonstrate the use of ENDURE in assessing a previously designed version of a polyethylene terephthalate (PET)-hydrolyzing enzyme from *Ideonella sakaiensis* (IsPETase), called DuraPETase (Cui et al., 2021), carrying ten mutations compared to the wildtype IsPETase (Yoshida et al., 2016). We expect that ENDURE will be a valuable resource for protein designers, filling a crucial gap in assessing and explaining the outcomes of protein design calculations.

2 Methods

The architecture and interface of the ENDURE web-app are designed in such a way that users are guided through the process of analyzing the energetic contributions of their protein designs in an intuitive and user-friendly manner. As shown in Figure 1, the whole workflow consists of several pages implemented in Streamlit (<https://streamlit.io>), a web application framework for Python. Starting from the File Upload page, the user uploads the PDB files of the designed protein and the reference protein structure. Alternatively, users can choose to provide amino acid sequences of the reference and mutant protein, and ENDURE will automatically predict their structures using ESMFold (Rives et al., 2021). The app then guides the user through three analysis steps, including pairwise interaction analysis, residue depth analysis, and inspection of all pairwise interaction changes by means of an energy difference heatmap. The results of these analyses are presented to the user in an interactive way that can be easily adapted to suit their specific needs using the controls on the left sidebar. The resulting tables can also be exported and downloaded as CSV file. Streamlit abstracts both front-end and back-end programming, making the application easily extensible by users with Python programming knowledge. Finally, the source code of ENDURE is packaged in a Docker container for easy installation and replication of the tool across different compute environments such as servers or HPC clusters. More details on the technological implementation of the app and the computational algorithms are provided in the following paragraphs.

2.1 Front end/backend in streamlit

The application's front end consists of the user interface (UI), which includes the welcome page, file upload section, and interactive visualizations. The UI is designed to be user-friendly and easy to navigate, allowing users to upload their protein structure file,



prepare the structure, and run the energy breakdown and residue depth calculations. Since Streamlit is our web development framework, the front and backend are managed under the hood. We do include several analysis modules, which are either pure Python functions or Python functions that call other auxiliary executables. The names of important internal functions in the ENDURE app are written in typewriter font in the following sections.

2.2 Processing structures with ENDURE

The protocol consists of two main stages: 1) structure pre-processing actions and 2) structure analysis actions. Those actions are run on the File Upload page.

The pre-processing is an important step in the ENDURE web tool as it ensures that the uploaded protein structure is in the correct format and ready for analysis. The pre-processing actions include cleaning of the PDB files to remove any ligands, ions, or water molecules, relaxing of the protein structures to remove any steric clashes or unfavorable interactions, and determining the mutations for the designed protein relative to the reference protein. Additionally, the PDB file is renumbered in this step so that the first residue in the file is at position 1. Relaxing the protein structures is the second important action in preparing the protein structures for analysis. This ensures that the protein structure is in a low energy state according to the Rosetta energy function (Alford et al., 2017), which can help to minimize false positive results. The ENDURE web tool uses RosettaScripts (Fleishman et al., 2011) to perform a single iteration of FastRelax (Khatib et al., 2011) by default, taking a previously minimized protein structure and optimizing its energy landscape further. Overall, the input preparation protocol is crucial for ensuring that the protein structure is in the correct format and

ready for analysis. (See [Supplementary Material](#) for the specific relax and energy breakdown commands).

The analysis actions include the energy breakdown (EB) calculations, which provide information about the energetics of each residue's interaction in the two analyzed protein structures, and the residue depth (RD) calculations. These algorithms are powered by the Rosetta energy function (Alford et al., 2017) and the Biopython (Cock et al., 2009) library, respectively. The EB calculations are performed using the Rosetta EB executable. In short, EB determines the one-body and two-body energies for each residue and decomposes them further into individual score term contributions, thus allowing the simultaneous exploration of, e.g., sidechain and backbone interactions. By clicking on the Calculate Energy button on the File Upload page, the Rosetta EB calculation is run in a subprocess. Internally, the run function is launched with four parameters as input: the input PDB file name, the location to save the result file, the location to save the log file, and the file path of the Rosetta executable. The output of the protocol is converted to a downloadable CSV file using the `convert_outfile` function, which saves the CSV file as a dictionary in the current session state.

These actions are followed by the RD calculation, which uses the MSMS algorithm (Sanner et al., 1996) from Biopython to calculate the distance of each residue to the molecular surface. The MSMS software computes the solvent-excluded surface from a set of spheres, representing the atoms in a protein structure. The reduced surface is calculated, and an analytical description of the solvent-excluded surface is derived from it. The calculation is done in the `calculate_depth` function.

The processing and analysis actions launched from the File Upload page are run in the background to prevent the UI from freezing, and their results are integrated into the interactive visualizations in the front end.

2.3 Analysis of ENDURE outputs: pairwise interactions analysis

The Interaction Analysis page allows users carrying out a comprehensive analysis of the energetic changes in pairwise interactions of single- and multi-mutant protein designs. The user can select and analyze interactions from different categories (A to F explained below) and from different physical interaction types (salt bridges, etc.) through different control parameters on the side bar. Individual residue pairs, affected directly or indirectly by the mutations, can be selected and displayed in interactive 3D visualizations.

We defined six categories of residue pair interactions to identify the regions in the protein structures most affected by the mutations.

Category A: residue pairs that are interacting in the reference structure have different interaction energy in the mutant, even though neither of the two residues were mutated.

Category B: residue pairs that are interacting in the reference structure have one member replaced, resulting in a different interaction energy in the mutant.

Category C: residue pairs that are interacting in the reference structure no longer interact in the mutant, even though neither of the two residues was mutated.

Category D: residue pairs that are interacting in the reference structure no longer interact in the mutant because one member was mutated.

Category E: residue pairs that are not interacting in the reference structure interact in the mutant, even though neither of the two residues was mutated.

Category F: residue pairs that are not interacting in the reference structure interact in the mutant because one member was mutated.

In order for ENDURE to detect the different interaction categories A-F and different physical interaction types, there are several functions implemented on the page that users can execute by clicking the Start Calculations button (after having run all the pre-processing actions on the File Upload page). Such functions perform a post-processing and filtering of the scorefile generated by the Rosetta EB calculations.

The `energy_calc` function is used to identify the essential changes in interaction energies between the mutant and the reference protein structure. It takes the outputs of the residue EB computation performed for the mutant and reference structures, the list of mutations between the reference and mutant, and a streamlit progress bar. The function calls several sub-functions to perform various interaction energy comparisons for each interaction category and physical interaction type. The former is managed by the `interaction_analysis` function and the latter is managed by the following functions: `salt_bridges`, `disulfide_bonds`, and `hydrogen_bonds`. As the names suggest, these functions calculate the energy differences for different types of interactions, such as salt bridges, disulfide bonds, and hydrogen bonds.

The `interaction_analysis` function calculates the difference in interaction energies between a mutant and a reference protein structure for a given list of mutations. The `interaction_analysis` function processes the outputs of the per-residue EB performed on the reference and mutant structures. It calculates the differences in all single-body (i.e., within a single residue) and two-body (i.e., between two residues) energies for all categories (A to F) between the two proteins.

In analyzing the interaction energy changes between the wildtype and mutant protein, it is important to distinguish between total energy changes and significant energy changes. The former is the sum of all energy changes, including those with a small value. However, since thousands of small changes can occur, it is possible for insignificant changes to mask chemically important changes. To overcome this issue, the significant energy change is calculated, which is the sum of only those interaction energy changes that exceed a minimum magnitude. In ENDURE changes that are larger than +1.0 Rosetta Energy Units (REU) or smaller than -1.0 REU are considered significant. By focusing only on significant energy changes, the mutations that are likely to have a significant impact can be more easily detected.

For the REF2015 scoring function for soluble proteins, there is an approximate 1:1 correspondence of REU and kcal/mol (Alford et al., 2017). However, for other Rosetta scoring functions, which include statistically derived potentials, the correspondence of Rosetta score units to thermodynamic energies is convoluted. For compatibility with other Rosetta scoring functions, which we plan to add to ENDURE in the future, we decided to report energy values on the ENDURE web page in REU.

The total energy change for all interactions and the sum of the subset of significant changes, is calculated using the `total_energy_changes` and `significant_changes` functions, respectively. These functions operate on a dictionary returned by `interaction_analysis` and sum all as well as the significant energy changes for each interaction category specified in an interaction list. Interactions from categories A-F can be selected from a list and the change for a given interaction type (salt bridges, disulfide bonds, sidechain-sidechain hydrogen bonds, sidechain-backbone hydrogen bonds, backbone-backbone short-range hydrogen bonds, backbone-backbone long-range hydrogen bonds, and all interactions) can be further inspected.

2.4 Analysis of ENDURE outputs: residue depth analysis

Mutation-induced energetic changes can have different effects on different layers of the protein structure. Therefore, we implemented a residue depth analysis and combined it with the per-residue energy breakdown analysis to distinguish changes occurring on the protein surface from those occurring in buried regions of the protein structure.

The Residue Depth page of the ENDURE app allows the user to analyze and compare the effect of mutations on the energy and spatial location of residues. Specifically, the user can select a residue pair that displays a strong negative energetic contribution and visualize the interaction in the protein structure. By adjusting the threshold slider, the user can see which residues have a significant impact on stability and select the most promising mutations for further analysis. Once the user selects a particular mutation by clicking on a point in the scatter plot, the app displays a side-by-side 3D visual comparison of the residue in the mutant and reference structures, which allows for a direct comparison of the effect of mutations on the protein structure. This analysis provides valuable insights into the structural and energetic changes resulting from mutations.

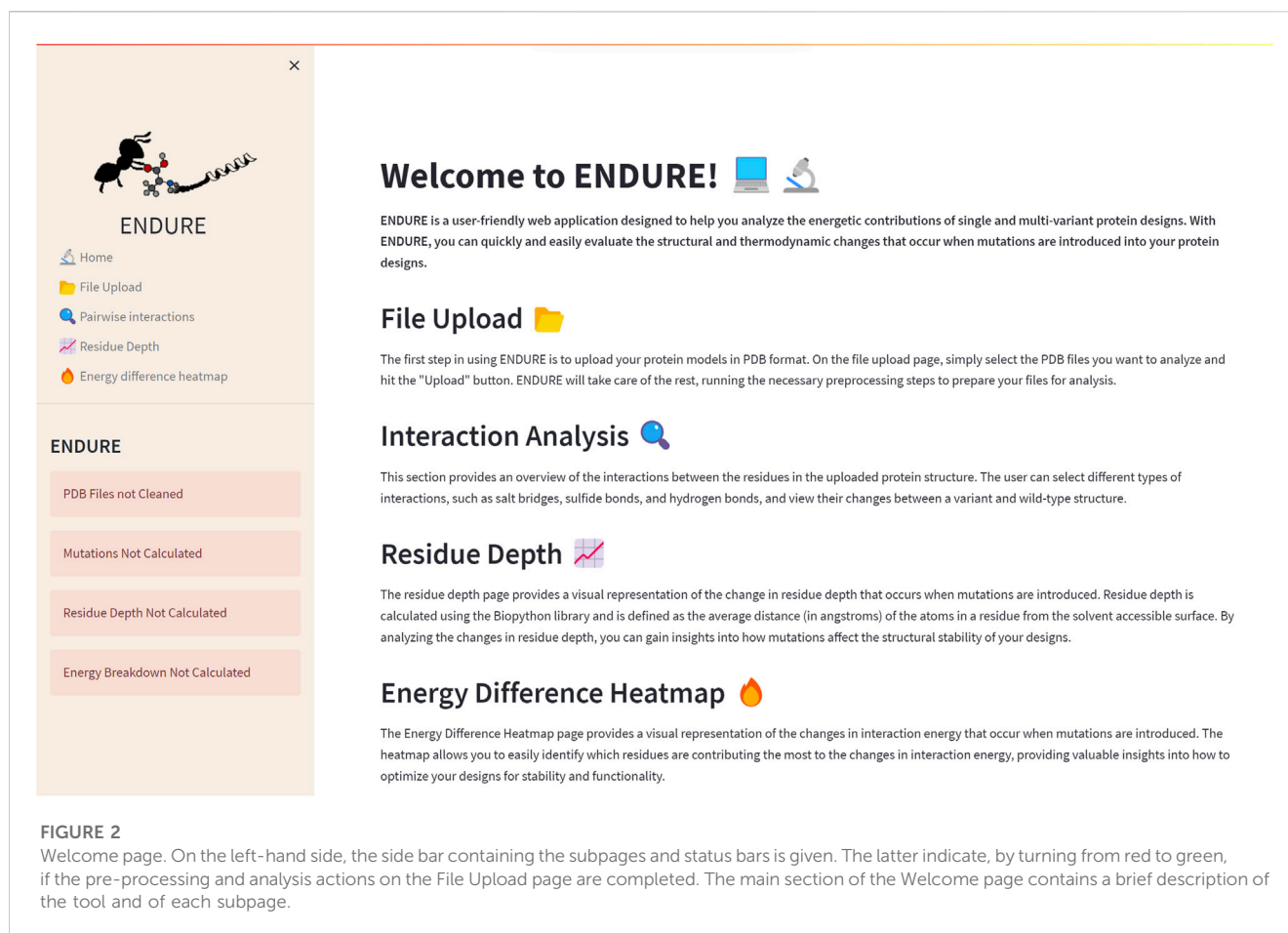


FIGURE 2

Welcome page. On the left-hand side, the side bar containing the subpages and status bars is given. The latter indicate, by turning from red to green, if the pre-processing and analysis actions on the File Upload page are completed. The main section of the Welcome page contains a brief description of the tool and of each subpage.

3 Results

The Welcome page (Figure 2) is the first page users will encounter when accessing the web application. This page provides an overview of the tool and its functionality, allowing new users to quickly familiarize themselves with the interface. The Welcome page includes a brief description of the purpose of ENDURE and the steps involved in using the tool to analyze protein designs. The Welcome page is designed to be user-friendly and intuitive, providing a clear and concise introduction to the tool and its capabilities, making it easier for users to get started and use the tool most effectively for their particular research questions.

Figure 3 presents the File Upload section, where users can upload their protein structure files in PDB format. The interface is designed to be intuitive and user-friendly, with options for either selecting the file from the local file system or dragging and dropping the file into the designated area. After uploading the file, the user is prompted to run five important pre-processing actions: 1) cleaning PDBs, which ensures that the files are correctly parsed, and the residues are renumbered so that the first residue is at position 1. 2) Relaxing PDB files, which prepares the files for the analysis. 3) Determining mutations by identifying the amino acid differences between the reference and mutant sequences. This information is crucial for many components of the analysis, as it allows tracking the position of mutations. 4) Calculating residue depth determines the

average distance of residues from the solvent-accessible surface. This is a key factor in understanding the energetic contributions of mutations. The calculation is performed using the Biopython library and is executed in a separate thread to avoid hanging the GUI. 5) Creating energy breakdown files, which provide a detailed breakdown of the energy contributions of individual residues. This calculation is performed using the Rosetta EB executable, as explained above, and is run in the background to prevent the GUI from being frozen.

3.1 ENDURE use case: analysis of an *in silico*-designed PET-degrading enzyme

To demonstrate the workflow and scope of application of ENDURE, we present here the results obtained for a previously reported designed PET-degrading enzyme, called DuraPETase (PDB ID: 6KY5) (Cui et al., 2021), which has higher thermostability than the wildtype *IsPETase* enzyme (PDB ID: 5XJH) (Yoshida et al., 2016; Joo et al., 2018). As can be seen in Figure 4A, ENDURE confirms that the particular mutant (carrying ten mutations) has favorable total and significant energy changes (−5.8 REU), indicating improved stability. This result is in line with the experimentally determined increase in the apparent melting temperature (T_m) of DuraPETase by 31°C compared the wildtype *IsPETase* (Cui et al., 2021). In addition, we calculated the significant

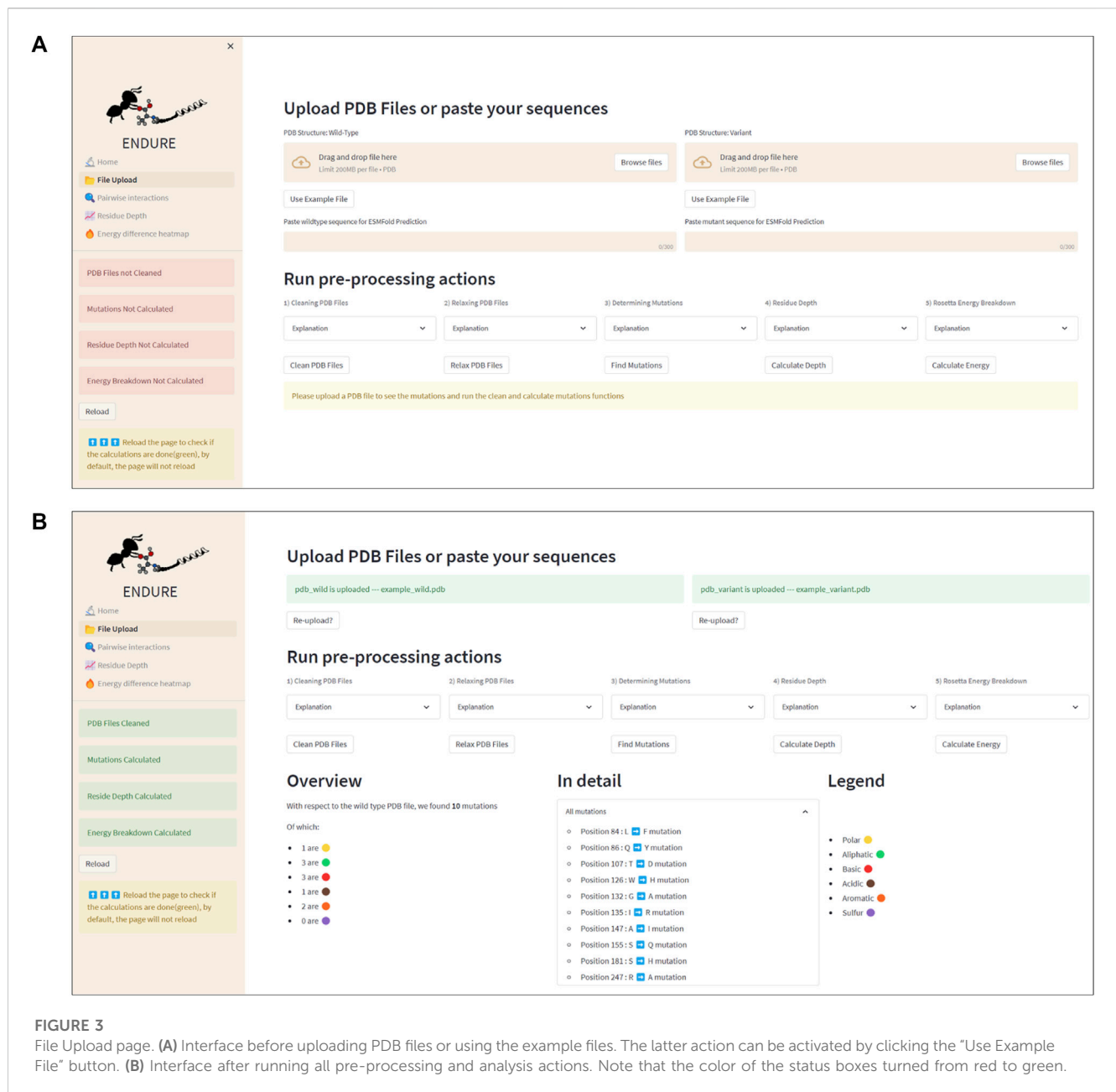


FIGURE 3

File Upload page. (A) Interface before uploading PDB files or using the example files. The latter action can be activated by clicking the “Use Example File” button. (B) Interface after running all pre-processing and analysis actions. Note that the color of the status boxes turned from red to green.

energy changes for two other designed PET hydrolases, FastPETase (Lu et al., 2022) and HotPETase (Bell et al., 2022), using ENDURE (Supplementary Table S1). The favorable energy changes for FastPETase (−10.2 REU) and HotPETase (−27.6 REU) confirmed their higher stability of $T_m = 67^\circ\text{C}$ and $T_m = 82^\circ\text{C}$, respectively, compared to wildtype *IsPETase*. This shows the utility of ENDURE in estimating overall stability changes.

In addition to the overall energy comparison, ENDURE also provides detailed information about the specific amino acid interactions that contribute to the improved stability of the selected mutant. The interaction analysis feature allows focusing on and visualizing specific residue interactions in the protein structure, which help to rationalize the underlying molecular mechanisms contributing to the improved stability of the selected mutant. These features will be described next.

3.1.1 Interaction analysis—Changes in pairwise interactions

The Interaction Analysis page of ENDURE enables performing a detailed examination of the energetic changes for residue pairs of different types (see Figure 4B). With this help, the user can identify the particular interactions contributing to improved or impaired energetic stability.

For example, Figure 4B shows a salt bridge interaction between residues Arg135 and Asp153, which causes a significant energy improvement of −3.32 REU in DuraPETase compared to wildtype *IsPETase*. This salt bridge isn’t present in the wildtype protein but is in DuraPETase—i.e., it belongs to interaction category F. *IsPETase* has an isoleucine (Ile135) at the same position, which cannot form a salt bridge. In addition, as shown in the left-hand table in the screenshot in Figure 4B, there are two more salt bridges detected by

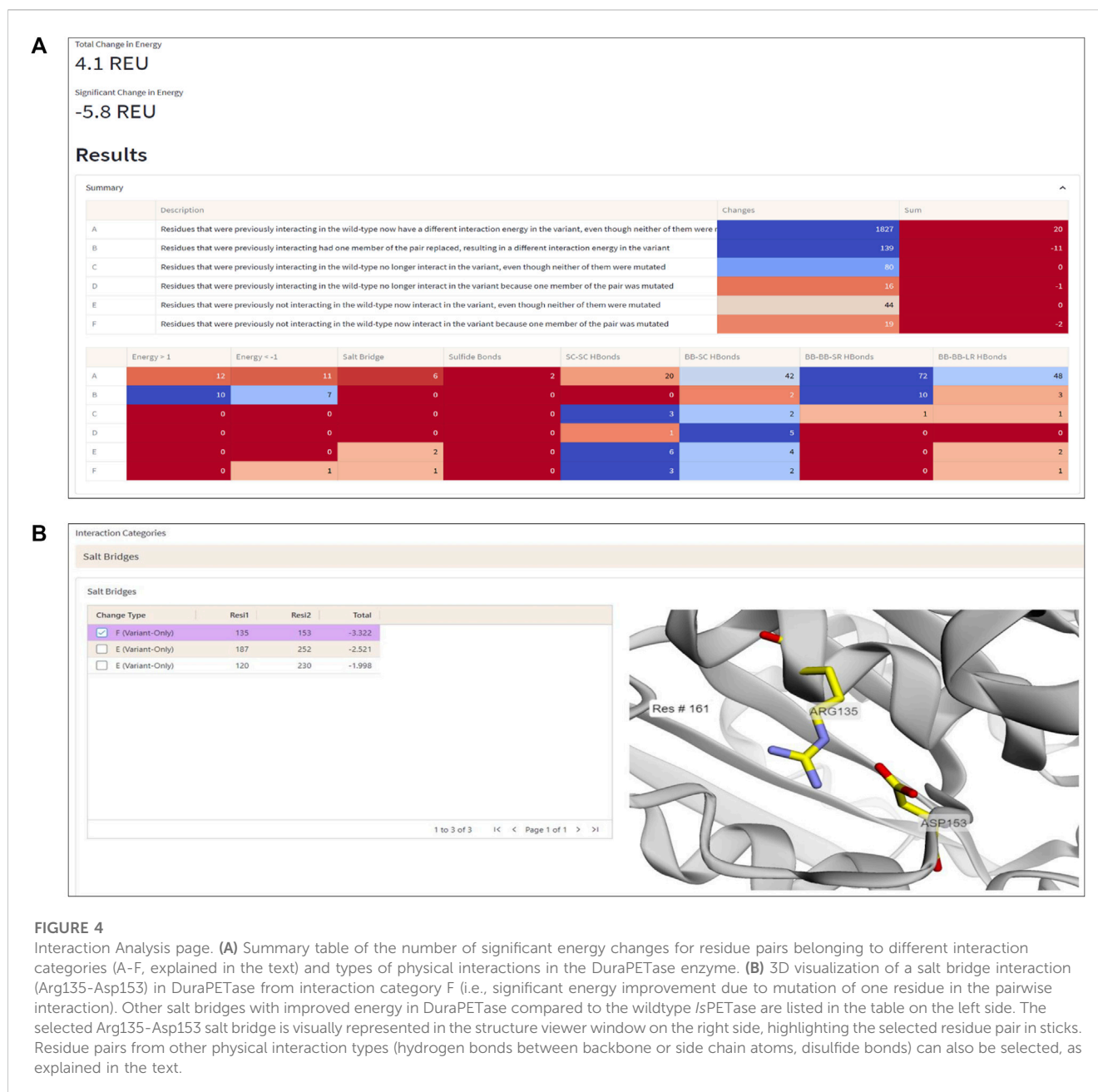


FIGURE 4

Interaction Analysis page. (A) Summary table of the number of significant energy changes for residue pairs belonging to different interaction categories (A-F, explained in the text) and types of physical interactions in the DuraPETase enzyme. (B) 3D visualization of a salt bridge interaction (Arg135-Asp153) in DuraPETase from interaction category F (i.e., significant energy improvement due to mutation of one residue in the pairwise interaction). Other salt bridges with improved energy in DuraPETase compared to the wildtype *IsPETase* are listed in the table on the left side. The selected Arg135-Asp153 salt bridge is visually represented in the structure viewer window on the right side, highlighting the selected residue pair in sticks. Residue pairs from other physical interaction types (hydrogen bonds between backbone or side chain atoms, disulfide bonds) can also be selected, as explained in the text.

ENDURE, which contribute significant energy changes in DuraPETase: Asp187-Arg252 (-2.52 REU) and Arg120-Asp230 (-1.99 REU). These two interactions belong to category E, i.e., they do not involve a mutated residue but are probably an indirect effect of nearby mutations in the environment of the four residues. Information like this provides valuable insight into protein structure and can help guide the user in their efforts to design a better-functioning mutant protein.

The different categories of interaction changes (A to F) allow the user to quickly identify the changes in interactions that have occurred and to focus their attention on the most important ones. For example, if a user observes that most changes are of type E or F, this might suggest that new interactions have formed, which could also significantly affect function. By providing this

information, the Interaction Analysis page helps the user to quickly understand and prioritize the changes in interactions that have occurred and guide their efforts in protein design.

3.1.2 Residue depth analysis

The residue depth analysis feature allows determining the depth of each residue in the protein structure, which reflects its accessibility to the solvent. By analyzing the energetic changes of mutations occurring in different spatial layers of the protein structure, the location of mutations that improve or impair stability can be determined. In Figure 5, the aforementioned mutation Ile135Arg is displayed as an example. In wildtype *IsPETase*, Ile135 is located on the protein surface, indicated by a low residue depth value of $\sim 2\text{\AA}$ (Figure 5A). This surface-exposed

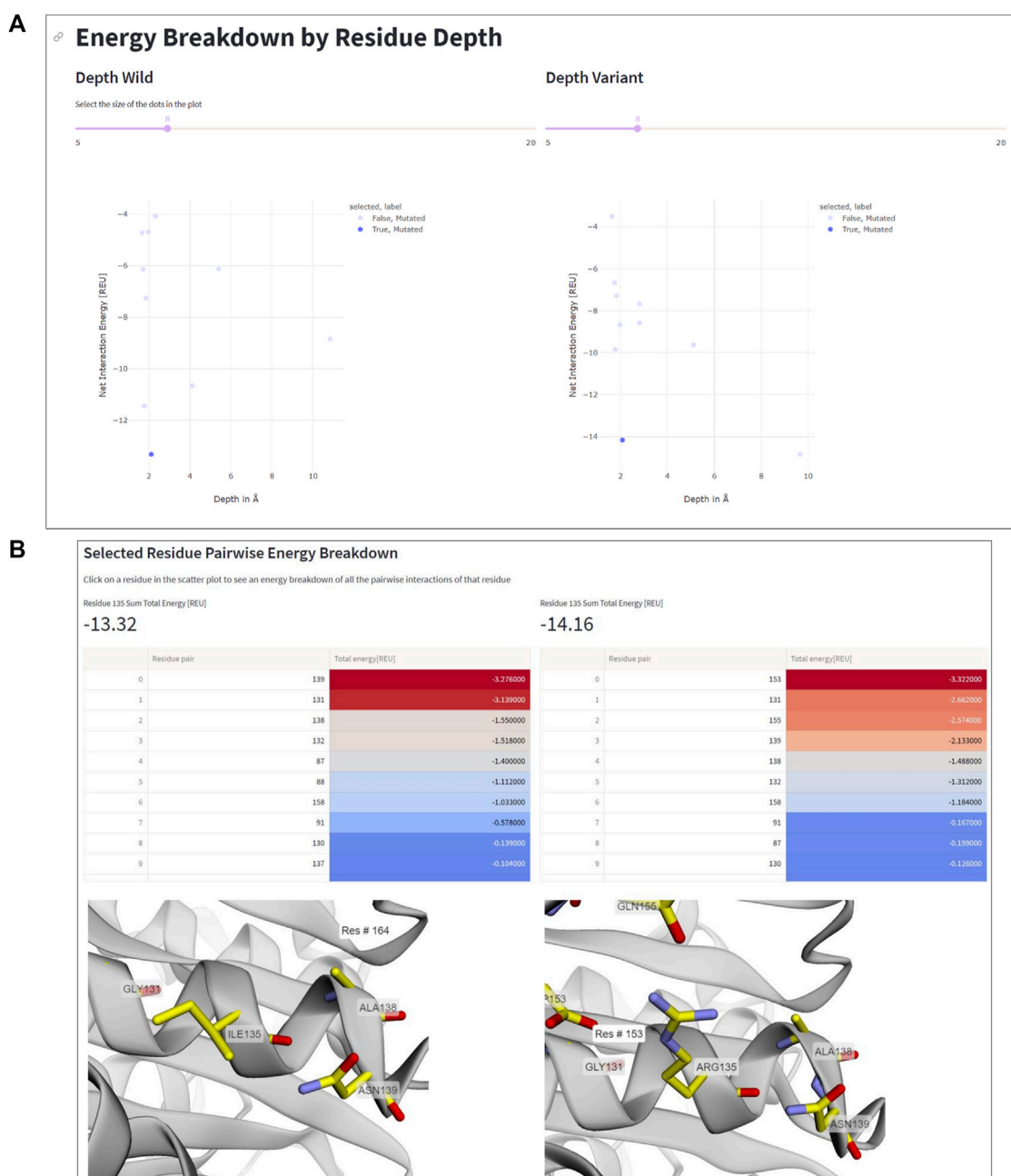
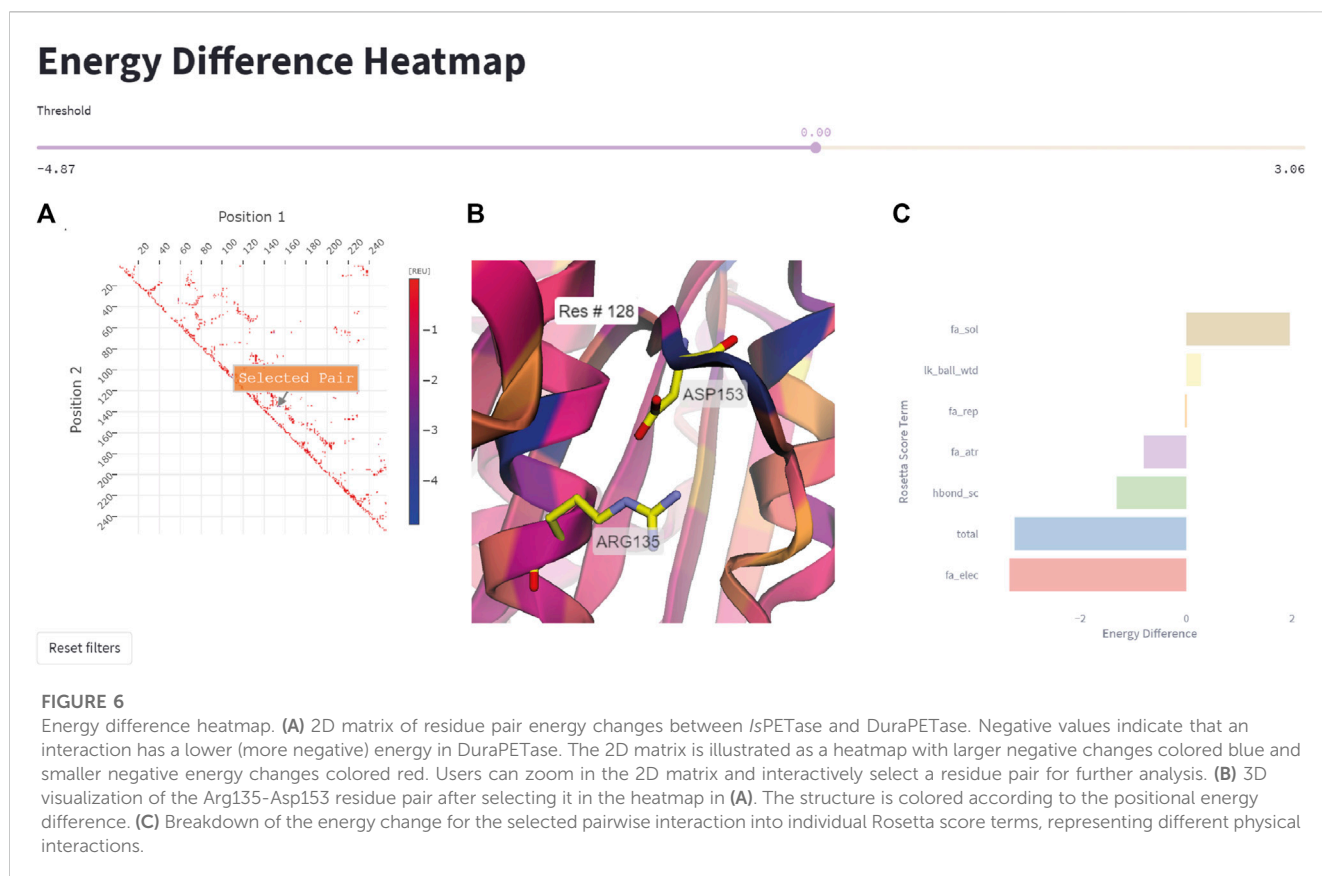


FIGURE 5 Residue depth analysis. **(A)** Net interaction energy versus residue depth plots for wildtype *IsPETase* (left) and *DuraPETase* (right). The blue data point in the lower left corner of the plot, corresponding to Ile135 in *IsPETase* or Arg135 in *DuraPETase*, respectively, has been selected. **(B)** Side-by-side comparison of the location and surrounding amino acids of Ile135 and Arg135, respectively. The interacting residues and their energy contributions are listed above the 3D viewer.

location is unfavorable for a hydrophobic amino acid. By contrast, the Arg135 residue in *DuraPETase* can form favorable interactions on the protein surface. In addition to the already mentioned Arg135-Asp153 salt bridge, the side chain of Arg135 interacts with the side chain of Gln155 through a hydrogen bond (Figure 5B). The table in Figure 5B, which lists the interacting residues for Ile135 or Arg135, respectively, and highlights their respective energy contributions (blue: high energy, red: low energy), confirms these visual observations. These extra interactions, can explain the lower net

energy of -14.16 REU for Arg135 in *DuraPETase* compared to -13.32 REU for Ile135 in *IsPETase* (Figure 5A).

Information like this is important because it can help decide which residues to mutate in order to improve protein stability. For example, mutations in the protein core may have a greater impact on stability than those located near the surface, and thus targeting core residues for mutagenesis can considerable impact the stability of the designed protein. Additionally, the residue depth analysis feature can be used to select the most promising mutants for further



characterization, by identifying mutations that have an improved net interaction energy and are located in favorable positions within the protein structure.

3.1.3 Energy difference heatmap

This feature allows the user to quickly represent all residue pairs that have a significantly changed interaction energy in the mutant protein compared to the wildtype, and selectively track specific pair interactions. The user can select interactively from the residue pair interaction matrix on the left side in [Figure 6](#) a pair of residues that display a large negative energetic contribution. The relevance threshold for the energy can be adjusted with the threshold slider present above. Once the user has selected a residue pair ([Figure 6A](#)), the corresponding pair (Arg135-Asp153 in this example) will be highlighted in the structure viewer in the middle of the page ([Figure 6B](#)), and a breakdown of the interaction energy change into individual score terms from the Rosetta energy function will appear ([Figure 6C](#)). For visual clarity only non-zero score terms are displayed to the user.

4 Discussion

The ENDURE web application provides a user-friendly interface for analyzing protein structures to facilitate mutant selection in protein design workflows. The study has shown that the application can accurately and efficiently process PDB files, clean and renumber them, determine mutations, calculate residue depth, and generate energy breakdown files. The interaction analysis section allows users

to view changes in pairwise interactions between residues in the wildtype and mutant structures by providing visual representations of the significant and total energy changes.

One of the key innovations of the ENDURE web application is its ability to group changes in residue pairwise interactions into different types and categories, ranging from residues that are interacting in the reference and mutant protein structure with different interaction energies (category A), to residues that aren't interacting in the reference structure but make interactions in the mutant due to a mutation (category F). These different categories of changes provide a useful way for the user to identify which mutations have the strongest impact on the protein structure, and consequently focus design efforts on specific areas of the protein structure.

Compared to previous research, our web application offers a user-friendly and accessible solution for analyzing protein structures and interaction changes. Prior research in this field has typically been focused on developing computational tools for protein structure prediction and analysis ([Schymkowitz et al., 2005](#); [Leman et al., 2020](#); [Stam and Wood, 2021](#)) or the analysis of the effects of single point mutations ([Yin et al., 2007](#)). For instance, the Eris web server ([Yin et al., 2007](#)) is an estimator of protein stability that primarily focuses on single mutations. In contrast, our ENDURE server is specifically designed to handle multiple mutations simultaneously while assessing the introduction of significant pairwise interactions. ENDURE provides a unique solution by incorporating advanced Rosetta modeling and analysis algorithms into a user-friendly interface and making them more broadly accessible.

It is important to note that the tool is limited by the accuracy of the underlying computational tools and algorithms used for protein structure prediction and analysis. Additionally, the interaction analysis section is based on a static protein structure analysis. Accuracy in predicting mutation-induced energy changes could benefit from a model ensemble approach (Peccati et al., 2023), in which protein dynamic changes like loop rearrangements can be considered. These limitations should be taken into consideration when interpreting the results of the analysis.

Despite these limitations, it represents a significant advancement in the field of protein structure and interaction analysis. The integration of computational tools into a user-friendly interface makes it possible for scientists outside the field of computational structural biology to quickly and efficiently analyze protein structures and identify potential areas for improvement. In future versions, ENDURE could be expanded to include additional features and improvements, such as the ability to examine other kinds of proteins, including membrane proteins and proteins with noncanonical amino acid modifications, through the incorporation of different energy functions. It could also be enhanced to consider ligand molecules in the design analysis, allowing the identification of designs with improved binding affinity. These features are currently planned for the next version of ENDURE, which will be released in the future. Additionally, the tool could be further developed to incorporate machine learning techniques to improve the accuracy of the analysis. With these advancements, ENDURE could become an even more powerful tool for protein design and analysis.

In conclusion, the ENDURE web application provides a unique and accessible solution for analyzing protein structure and interaction changes and, in that way, represents a significant advancement for the field of protein design. Categorizing changes in pairwise interactions for different interaction types provides a straightforward way for the user to guide their protein design strategies. Integrating computational tools into a user-friendly interface makes it possible for a broader audience to quickly and efficiently analyze protein structures. The future direction of the research will focus on further developing the application to incorporate analysis on protein dynamics, support for non-standard amino acid residues, and application of machine learning techniques. Furthermore, a command line interface integrated in the front end is planned, which will help further customize some of the analyses.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. The source code for ENDURE can be found on <https://github.com/kuenzelab/ENDURE>. Further inquiries can be directed to the corresponding author.

References

Alford, R. F., Leaver-Fay, A., Jeliaskov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., et al. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput* 13, 3031–3048. doi:10.1021/acs.jctc.7b00125

Author contributions

Conceptualization, FE and JZ; methodology, FE and JZ; software, FE and JZ; investigation, FE and JZ; data curation, FE; writing—original draft preparation, FE; writing—review and editing, FE, JZ, and GK; visualization, FE; supervision, GK; project administration, GK. All authors have read and agreed to the published version of the manuscript.

Funding

This work was funded by the German Research Foundation (CRC1052, project Z6) and the Open Access Publishing Fund of Leipzig University, supported by the German Research Foundation within the program Open Access Publication.

Acknowledgments

JZ acknowledges the Max Kade foundation for receiving a summer exchange scholarship. The authors thank Ivan Ivanikov for his technical support in installing the ENDURE webserver. The authors further acknowledge the Leipzig University Scientific Computing Center and the Institute for Drug Discovery for providing the computational resources.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1178035/full#supplementary-material>

Arnold, F. H. (2018). Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie International Edition* 57, 4143–4148. doi:10.1002/anie.201708408

- Barlow, K. A., S. O. C., Thompson, S., Suresh, P., Lucas, J. E., Heinonen, M., et al. (2018). Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein-Protein Binding Affinity upon Mutation. *J Phys Chem B* 122, 5389–5399. doi:10.1021/acs.jpbc.7b11367
- Bell, E. L., Smithson, R., Kilbride, S., Foster, J., Hardy, F. J., Ramachandran, S., et al. (2022). Directed evolution of an efficient and thermostable PET depolymerase. *Nat Catal* 5, 673–681. doi:10.1038/s41929-022-00821-3
- Castro, K. M., Scheck, A., Xiao, S., and Correia, B. E. (2022). Computational design of vaccine immunogens. *Current Opinion in Biotechnology* 78, 102821. doi:10.1016/j.copbio.2022.102821
- Chen, K., Arnold, F. H., Cheng, R., Fisher, M., Zhang, B. H., Di Maggio, M., et al. (2020). Facial Recognition Neural Networks Confirm Success of Facial Feminization Surgery. *Nat Catal* 3, 203–209. doi:10.1097/PRS.00000000000006342
- Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2020). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi:10.1093/bioinformatics/btp163
- Cui, Y., Chen, Y., Liu, X., Dong, S., Tian, Y., Qiao, Y., et al. (2021). Computational Redesign of a PETase for Plastic Biodegradation under Ambient Condition by the GRAPE Strategy. *ACS Catal* 11, 1340–1350. doi:10.1021/acscatal.0c05126
- Dauparas, J., Anishchenko, I., Bennett, N., Bai, H., Ragotte, R. J., Milles, L. F., et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. *Science* 378, 49–56. doi:10.1126/science.add2187
- Ferruz, N., and Höcker, B. (2022). Controllable protein design with language models. *Nat Mach Intell* 4, 521–532. doi:10.1038/s42256-022-00499-z
- Fleishman, S. J., Leaver-Fay, A., Corn, J. E., Strauch, E. M., Khare, S. D., Koga, N., et al. (2011). RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS one* 6, e20161. doi:10.1371/journal.pone.0020161
- Frenz, B., Lewis, S. M., King, I., DiMaio, F., Park, H., and Song, Y. (2020). Prediction of Protein Mutational Free Energy: Benchmark and Sampling Improvements Increase Classification Accuracy. *Front. Bioeng. Biotechnol.* 8, 558247. doi:10.3389/fbioe.2020.558247
- Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., et al. (2016). Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Molecular Cell* 63, 337–346. doi:10.1016/j.molcel.2016.06.012
- Habibi, N., Mauser, A., Ko, Y., and Lahann, J. (2022). Protein Nanoparticles: Uniting the Power of Proteins with Engineering Design Approaches. *Advanced Science* 9, 2104012. doi:10.1002/advs.202104012
- Joo, S., Cho, I. J., Seo, H., Son, H. F., Sagong, H.-Y., Shin, T. J., et al. (2018). Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nat Commun* 9, 382. doi:10.1038/s41467-018-02881-1
- Kellogg, E. H., Leaver-Fay, A., and Baker, D. (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* 79, 830–8. doi:10.1002/prot.22921
- Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popovic, Z., et al. (2011). Algorithm discovery by protein folding game players. *Proceedings of the National Academy of Sciences of the United States of America* 108, 18949, 18953. doi:10.1073/pnas.1115898108
- Leman, J. K., Weitzner, B. D., Lewis, S. M., Adolf-Bryfogle, J., Alam, N., Alford, R. F., et al. (2020). Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods* 17, 665–680. doi:10.1038/s41592-020-0848-2
- Lu, H., Diaz, D. J., Czarnecki, N. J., Zhu, C., Kim, W., Shroff, R., et al. (2022). Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature* 604, 662–667. doi:10.1038/s41586-022-04599-z
- Pan, X., and Kortemme, T. (2021). Recent advances in de novo protein design: Principles, methods, and applications. *Journal of Biological Chemistry* 296, 100558. doi:10.1016/j.jbc.2021.100558
- Peccati, F., Alunno-Rufini, S., and Jiménez-Osés, G. (2023). Accurate Prediction of Enzyme Thermostabilization with Rosetta Using AlphaFold Ensembles. *J. Chem. Inf. Model.* 63, 898–909. doi:10.1021/acs.jcim.2c01083
- Quijano-Rubio, A., Yeh, H.-W., Park, J., Lee, H., Langan, R. A., Boyken, S. E., et al. (2021). De novo design of modular and tunable protein biosensors. *Nature* 591, 482–487. doi:10.1038/s41586-021-03258-z
- Reetz, M. (2022). Making Enzymes Suitable for Organic Chemistry by Rational Protein Design. *ChemBioChem* 23, e202200049. doi:10.1002/cbic.202200049
- Rives, A., Meier, J., Sercu, T., Goyal, S., Lin, Z., Liu, J., et al. (2021). Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc Natl Acad Sci USA* 118, e2016239118. doi:10.1073/pnas.2016239118
- Sanner, M. F., Olson, A. J., and Spehner, J.-C. (1996). Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* 38, 305–320. doi:10.1002/(SICI)1097-0282(199603)38:3<305::AID-BIP4>3.0.CO;2-Y
- Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: an online force field. *Nucleic Acids Research* 33, W382–W388. doi:10.1093/nar/gki387
- Sevy, A. M., and Meiler, J. (2014). Antibodies: Computer-Aided Prediction of Structure and Design of Function. *Microbiology spectrum* 2. doi:10.1128/microbiolspec.AID-0024-2014
- Stam, M. J., and Wood, C. W. (2021). DE-STRESS: a user-friendly web application for the evaluation of protein designs. *Protein Engineering, Design and Selection* 34, gzab029. doi:10.1093/protein/gzab029
- Starr, T. N., and Thornton, J. W. (2016). Epistasis in protein evolution. *Protein Science* 25, 1204–1218. doi:10.1002/pro.2897
- Stein, V., and Alexandrov, K. (2015). Synthetic protein switches: design principles and applications. *Trends in Biotechnology* 33, 101–110. doi:10.1016/j.tibtech.2014.11.010
- Strokach, A., Becerra, D., Corbi-Verge, C., Perez-Riba, A., and Kim, P. M. (2021). Computational generation of proteins with predetermined three-dimensional shapes using ProteinSolver. *STAR Protocols* 2, 100505. doi:10.1016/j.xpro.2021.100505
- Sumbalova, L., Stourac, J., Martinek, T., Bednar, D., and Damborsky, J. (2018). HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. *Nucleic Acids Research* 46, W356–W362. doi:10.1093/nar/gky417
- Weinstein, J. J., Goldenzweig, A., Hoch, S., and Fleishman, S. J. (2021). PROSS 2: a new server for the design of stable and highly expressed protein variants. *Bioinformatics* 37, 123–125. doi:10.1093/bioinformatics/btaa1071
- Willis, J. R., Sapparapu, G., Murrell, S., Julien, J. P., Singh, V., King, H. G., et al. (2015). Redesigning HIV antibodies exhibit enhanced neutralizing potency and breadth. *The Journal of clinical investigation* 125, 2523–31. doi:10.1172/JCI80693
- Yin, S., Ding, F., and Dokholyan, N. V. (2007). Eris: an automated estimator of protein stability. *Nat Methods* 4, 466–467. doi:10.1038/nmeth0607-466
- Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., et al. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science* 351, 1196–1199. doi:10.1126/science.aad6359