The Shortest Path Method (SPM) webserver for computational enzyme design

Guillem Casadevall,[1] Jordi Casadevall,[2] Cristina Duran[1] and Sílvia Osuna[1,3]

[1]Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, c/Maria Aurèlia Capmany 69, 17003 Girona.
[2]Carrer Tancat 2, 17493 Vilajuïga.
[3]ICREA, Pg. Lluís Companys 23, 08010 Barcelona.

Abstract

SPMweb is the online webserver of the Shortest Path Map (SPM) tool for identifying the key conformationally-relevant positions of a given enzyme structure and dynamics. The server is built on top of the DynaComm.py code and enables the calculation and visualization of the SPM pathways. SPMweb is easy-to-use as it only requires three input files: the three-dimensional structure of the protein of interest, and the two matrices (distance and correlation) previously computed from a Molecular Dynamics simulation. We provide in this publication information on how to generate the files for SPM construction even for non-expert users and discuss the most relevant parameters that can be modified. The tool is extremely fast (it takes less than one minute per job), thus allowing the rapid identification of distal positions connected to the active site pocket of the enzyme. SPM applications expand from computational enzyme design, especially if combined with other tools to identify the preferred substitution at the identified position, but also to rationalizing allosteric regulation, and even cryptic pocket identification for drug discovery. The simple user interface and setup make the SPM tool accessible to the whole scientific community. SPMweb is freely available for academia at http://spmosuna.com/.

Introduction

Enzyme design aims to create novel biocatalysts with enhanced properties through the modification of their natural amino acid sequences or via generation of novel sequences and folds. The fascination with enzyme design and engineering is motivated by the advantageous features exhibited by these catalysts, including their capacity to function effectively under gentle biological conditions, achieving remarkable efficiency, selectivity, and specificity. Enzyme design is also an intellectual challenge, as it is a stringent examination of what we understand of enzyme stability, folding, evolution and catalysis.

Designing enzymes taking as starting point a natural or computationally reconstructed/generated scaffold involves selecting specific residues for mutagenesis, generating new variants, and employing screening protocols to assess improvements in targeted properties.(Bell *et al.*, 2021) Two main approaches exist: rational design(Damborsky and Brezovsky, 2014, Maria-Solano *et al.*, 2018, Romero-Rivera *et al.*, 2017) considering *de novo* and natural scaffolds, and Directed Evolution (DE),(Arnold, 2015, Currin *et al.*, 2015) which can be successfully combined to achieve higher levels of performance. Rational design focuses on predetermined hotspot positions, identified through multiple sequence alignments, structural analysis of active site pockets, potential substrate-binding tunnels, and comprehensive computational modeling (using techniques like Quantum Mechanics, Quantum Mechanics/Molecular Mechanics, Molecular Dynamics, and MonteCarlo simulations).(Romero-Rivera, Garcia-Borràs and Osuna, 2017, Sequeiros-Borja *et al.*, 2020) Rational design efforts often focus on the active site pocket or in the bottleneck regions of the computed substrate binding tunnels and gates. The user-friendly tools such as CAVER,(Stourac *et al.*, 2019) AQUA-DUCT,(Stourac, Vavra, Kokkonen, Filipovic, Pinto, Brezovsky, Damborsky and Bednar, 2019) and HotSpot Wizard,(Sumbalova *et al.*, 2018) among others can be used.(Sequeiros-Borja, Surpeta and Brezovsky, 2020) In contrast, DE,(Bornscheuer *et al.*, 2012, Francis and Hansche, 1972, Lutz and Bornscheuer, 2008, Packer and Liu, 2015) honored with the

2018 Nobel Prize in Chemistry, initially relied on iterative cycles of random mutagenesis. Recent advancements integrate bioinformatic tools, (Jiang *et al.*, 2008, Kourist *et al.*, 2010, Kuipers *et al.*, 2010, Kuipers *et al.*, 2009, Rothlisberger *et al.*, 2008, Siegel *et al.*, 2010) sequence analysis,(Addington *et al.*, 2013, Pavelka *et al.*, 2009) smarter libraries, protein engineering techniques,(Bornscheuer, Huisman, Kazlauskas, Lutz, Moore and Robins, 2012, Kazlauskas and Bornscheuer, 2009, Turner, 2009) gene synthesis,(Currin *et al.*, 2014) and high-throughput screening techniques.(Xiao *et al.*, 2015) Machine-learning sequence-function models can be used to guide DE.(Mazurenko *et al.*, 2020, Yang *et al.*, 2019) As mentioned above, the powerful DE strategy can be applied to boost the low activities of computational enzyme designs(Jaeckel *et al.*, 2008, Renata *et al.*, 2015, Romero and Arnold, 2009) and enhance promiscuous enzymatic side-activities.(Campbell *et al.*, 2016, Leveson-Gower *et al.*, 2019) Multiple laboratory-engineered enzymes have been reported in the literature, including enzymes for the production of drugs, biotherapeutics, potential bulk products, and fragrances.(Buller *et al.*, 2023)

A notable strength of DE lies in its capability to introduce mutations throughout the entire protein sequence. This contrasts with rational design approaches that are often restricted to alterations in the active site pocket or available tunnels and gates for promoting substrate binding/product release and altering the water content.(Gora *et al.*, 2013, Sequeiros-Borja, Surpeta and Brezovsky, 2020) As observed in numerous DE studies, the remarkable fold increases in catalytic activity achieved are accomplished thanks to mutations positioned far from the active site, which are computationally very challenging to predict.(Currin, Swainston, Day and Kell, 2015, Jiménez-Osés *et al.*, 2014, Obexer *et al.*, 2017, Osuna, 2021) This trend extends to diverse enzymes such as cytochrome P450, Diels-Alderase, phosphotriesterase, sitagliptinase, among many additional ones.(Osuna, 2021) Often laboratory-evolved enzymes present mutations introduced at an average distance of around 15 Å from the active site.(Currin, Swainston, Day and Kell, 2015) Intriguingly, there is no direct correlation between the impact of introduced mutations on enzyme turnover (k_{cat}) and their proximity to the active site, in contrast to the more deterministic role of active site mutations in specificity.(Currin, Swainston, Day and Kell, 2015) The coupling of distal residues affecting the enzyme catalytic activity suggests a substantial influence of long-range allostery, i.e., regulation of catalytic activity by effector and/or protein binding, in many proteins.(Gunasekaran *et al.*, 2004) Extensive MD simulations have successfully rationalized how distal mutations influence the multiple conformations enzymes can adopt thus impacting its catalytic activity.(Jiménez-Osés, Osuna, Gao, Sawaya, Gilson, Collier, Huisman, Yeates, Tang and Houk, 2014, Romero-Rivera *et al.*, 2017) Distal mutations often alter non-covalent interaction networks, which might favor some additional conformational states of the enzyme that are more optimal for the promiscuous activity to be enhanced and/or modify the flexibility of crucial structural elements such as loops and lids gating the active site pocket.(Campbell, Kaltenbach, Correy, Carr, Porebski, Livingstone, Afriat-Jurnou, Buckle, Weik, Hollfelder, Tokuriki and Jackson, 2016, Curado-Carballada *et al.*, 2019, Petrović *et al.*, 2018) While computational modeling can satisfactorily explain these changes in activity induced by distal alterations, the challenge remains in predicting which distal mutations can impact and regulate enzymatic activity.(Campitelli *et al.*, 2020, Jiménez-Osés, Osuna, Gao, Sawaya, Gilson, Collier, Huisman, Yeates, Tang and Houk, 2014, Romero-Rivera, Garcia-Borràs and Osuna, 2017) Given the insights from DE that distal mutations are essential for enhancing enzyme catalytic activity, the development of computational tools capable of predicting remote mutations holds great promise, potentially advancing our underdeveloped ability to computationally design efficient Nature-like enzymes.(Osuna, 2021)

The effect exerted by distal mutations in enzyme design reminds the allosteric regulation effect produced by effector binding in allosteric systems or within the active sites of heterocomplexes that present synchronised transportation of substrates. Distal mutations can induce a shift in the conformational landscape, thus favouring the catalytically competent arrangement of the catalytic residues for catalysis. Given the striking similarity between these two scenarios (enzyme design and allosteric regulation), we explored the potential development and application of correlation-based tools in enzyme design.(Maria-Solano, Serrano-Hervás, Romero-Rivera, Iglesias-Fernández and Osuna, 2018, Romero-Rivera, Garcia-Borràs and Osuna, 2017) We developed the

Shortest Path Map (SPM, DynaComm.py) tool by constructing a first complex graph based on mean distances and correlation values between the residues that compose the enzyme computed during MD simulations, similar to the protocol by Sethi et al.(Sethi *et al.*, 2009) for studying allosteric systems (see Figure 1).(Romero-Rivera, Garcia-Borràs and Osuna, 2017) In contrast to prior allosteric studies concentrating on identifying communities in the graph, (Sethi, Eargle, Black and Luthey-Schulten, 2009) our SPM approach involves computing shortest path lengths using the Dijkstra algorithm implemented in the igraph module.(Csárdi and Nepusz, 2006) Consequently, it identifies those pairs of residues that are more correlated and have a higher impact into the enzyme conformational dynamics. Unlike community analysis that highlights important regions of the enzyme, SPM directly identifies the most crucial residues rather than regions. This feature is particularly appealing for enzyme design, enabling the direct construction of small libraries of hotspot positions.

SPM narrows down the sequence space to a subset of conformationally relevant positions, with a notable capability to pinpoint challenging distal positions that enhance activity.(Osuna, 2021) The successful application of SPM in identifying DE mutations in retro-aldolase, monoamine oxidase, and tryptophan synthase enzymes suggests its potential utility in the rational design of enzyme variants.(Osuna, 2021) The Mulholland lab utilized our SPM tool to assess changes in dynamical networks during the transition-state ensemble along DE of a computationally designed Kemp eliminase. (Bunzel *et al.*, 2021) Additionally, we have used SPM to investigate allosteric communication within monomers, and in allosteric systems.(Calvó-Tusell *et al.*, 2022, Castelli *et al.*, 2024, Curado-Carballada, Feixas, Iglesias-Fernández and Osuna, 2019) More recently, we have also used SPM for rational enzyme design in combination with other tools to further reduce the number of identified positions and select the specific amino acid at each site, as described in the following examples. We combined SPM with ancestral sequence reconstruction for developing new stand-alone tryptophan synthase B (TrpB) variants.(Maria-Solano *et al.*, 2021) Focusing on including the ancestral amino acid in the non-conserved SPM positions, our approach increased the stand-alone activity of the new SPM6-TrpB variant by 7-fold (in terms of k_{cat}).(Maria-Solano, Kinateder, Iglesias-Fernández, Sterner and Osuna, 2021) It is worth noting that, while testing only a single variant, the fold increase in k_{cat} was comparable to the 9-fold obtained through DE, which required generating and screening over 3000 variants. In a recent pre-print paper, we showcased the efficacy of our SPM methodology in designing efficient Nature-like enzymes. Specifically, we achieved a more than 1300-fold increase in the esterase catalytic efficiency of a hydroxynitrile lyase (HNL), surpassing the esterase activity of the reference enzyme.(Casadevall *et al.*, 2023) Altogether, these studies provide compelling evidence for the potential of our SPM methodology in computational enzyme design.

In this study, we develop and describe the webserver version of the SPM tool for its application in enzyme design for academic use. First, we discuss the user-friendly webserver generated, the input files needed and the overview of the settings that the user can alter to generate different SPM maps. Second, we show with the tryptophan synthase example how information of inter and intramolecular SPM communications networks can be withdrawn. With this tool, we hope the academic community can benefit from the application of the SPM in the study of biomolecular systems and aim to expand the current area of application of the SPM methodology.

Results

WORKFLOW

The basic workflow for SPM construction is shown in Figure 1. As described in the introduction, the enzyme structure and dynamics is simplified using a weighted graph (step 1), which is then further processed to identify the shortest paths to generate the final SPM graph (step 2). SPM can then be plotted back into the 3D-dimensional structure to visualize how the active site pocket is connected to more remote sites.

Figure 1. Workflow and equations for Shortest Path Map (SPM) construction for computational enzyme design. The enzyme is simplified as a weighted graph as done for studying allostery,(Sethi, Eargle, Black and Luthey-Schulten, 2009) however, this complex graph is simplified to identify the shortest paths (pairs of residues) that have a higher contribution to the conformational dynamics. SPM can be drawn back on the 3D structure to directly assess how the active site pocket is connected to active site and distal sites. The key equations (1 and 2) for converting a protein into a graph are also displayed. Each node in the first complex graph represents a residue. The edges linking each pair of nodes (residues) are assigned weights in accordance with equation 1 and 2, where $C_{i,j}$ is the computed correlation value, Δr_i and Δr_j are the displacement of the C_α of the *i*, *j* residue observed in the MD trajectory with respect to a reference structure.

Generation of the first weighted graph

Initial attempts to apply graph theory to investigate allosteric regulation primarily focused on the static X-ray structure of the enzyme.(Guo and Zhou, 2016) In the constructed graph, two sets of nodes (residues) were linked by an edge if the distance between their representative atoms fell below a specific threshold. The significant advancement in graph construction came from Sethi et al.,(Sethi, Eargle, Black and Luthey-Schulten, 2009) who employed short MD simulations (lasting a few nanoseconds) to determine the connected nodes and their respective edge weights. An edge was established between a pair of residues (nodes) if the representative atoms of each residue remained below a defined distance threshold (see Figure 1) for a specified fraction of the MD simulation time. The edge connecting residues *(i,j)* was weighted based on their correlation values ($C_{i,j}$, as outlined in equations 1 and 2, Figure 1). Residues undergoing highly correlated conformational changes during the MD simulation (i.e., $C_{i,j} \rightarrow 1$) were linked by a relatively short edge ($l_{i,j} \rightarrow 0$). Conversely, a pair of residues with non-correlated movements ($C_{i,j} \rightarrow 0$) were connected by relatively long edges ($l_{i,j} \rightarrow \infty$).

In this protocol, the enzyme conformational dynamics is summarized through this first weighted graph (shown in Figure 1). Further subdivision of the graph into communities, utilizing the Girvan-Newman algorithm,(Girvan and Newman, 2002) results in the identification of what is called the optimal community network used in the study of allosterically-regulated enzymes.(Rivalta *et al.*, 2012, Schupfner *et al.*, 2020) However, for computational enzyme design it is more preferred to identify a subset of positions rather than regions or communities.

Generation of the SPM

For SPM generation instead of pinpointing communities within the initial graph, we use the Dijkstra algorithm, implemented in the igraph module,(Csárdi and Nepusz, 2006) to calculate the shortest path lengths. The algorithm considers all nodes of the graph and determines the shortest path from the first to the last protein residue. Consequently, the method identifies the edges in the graph that are shorter, thus indicating higher correlation and more frequently used in going through all protein residues. All edges are then normalized, and only those with the most significant contribution (a visualization/ significance threshold is applied, see Figure 2) are represented in the SPM. Drawing the SPM directly onto the 3D structure of the protein, rather than its 2D graph representation (see Figure 1), is more advantageous as one can directly see how the network expands through the 3D structure. The primary benefit of SPM over community analysis lies in directly identifying the most critical residues (as opposed to regions), making it more appealing for enzyme design, as small libraries of hotspot positions can be constructed directly. SPM enables the prediction of distal active site mutations that lead to enhanced enzymatic activity for the first time in a computational protocol.(Romero-Rivera, Garcia-Borràs and Osuna, 2017)

DESCRIPTION OF THE WEBSERVER

Input files

There are three mandatory files for SPM construction: the tertiary structure of the enzyme/protein in pdb format for visualizing the results, and the distance and correlation matrices obtained often through MD simulations (but not necessarily restricted to). Our recommendation is to generate the distance and correlation matrices using at least three replicates of MD simulations of 200-500 ns of simulation length in explicit solvent and considering either C_α or C_β positions. The calculation of the distance and correlation matrices can be done considering the whole MD trajectory, the last 100-200 nanoseconds of the MD simulations or using distinct sets of conformations in case of proteins undergoing large conformational changes. We, however, recommend using either the whole MD trajectory or the last 100-200 ns of the MD runs.(Duran *et al.*, 2024)

The distance and correlation matrix can be computed with different MD analysis software, but we provide as example the input file used for cpptraj included in AMBER tools:

Input files for cpptraj module for computing the correlation and proximity matrices:

```
# We recommend taking as reference the most populated cluster from the MD trajectory. This is especially relevant for proteins undergoing large conformational changes.
cpptraj <parm file>
reference structure.pdb
trajin MD_trajectory.nc  1 last 1
rms reference @CA
matrix dist @CA out dist_mat.dat
matrix correl @CA out corr_mat.dat
exit
```

The three mandatory files (structure.pdb, dist_mat.dat, corr_mat.dat) can then be uploaded in the corresponding boxes included in the main page of the webserver (see Figure 2). It should be also mentioned that the webserver also accepts the distance and correlation matrices as numpy binary files (.npy).

Figure 2. SPM main page of the webserver. The user needs to upload in the corresponding boxes the three mandatory files that are needed for SPM construction: the enzyme 3D structure, and the two matrices: distance and correlation previously computed from the MD simulations. Two important parameters can be modified for SPM construction: the distance threshold (bottom right panel), and the significance threshold (bottom left panel). The webserver link is: https://spmosuna.com.

SPM parameters

As discussed in the previous section, two thresholds need to be defined for SPM construction. The first one is related to the mean distance value between the user defined atoms along the MD simulation (often distances between either C_α or C_β). While we recommend the use of a distance threshold of 6 Å, in some cases, it might be useful to play with the distance matrix threshold. Increasing this value to higher numbers will of course consider a higher portion of the protein residues for each targeted site, and thus the computed SPM graph will contain a larger number of positions. In the opposite direction, rather small values for the distance matrix will only consider nearby residues thus being very local and restricted (see the distance threshold tests in the case example below).

The other important threshold is related to the number of positions represented in the final SPM graph. This visualization/significance threshold will restrict the number of edges and nodes displayed. We recommend a

threshold of 0.3, as it will reduce the number of positions and will only display the ones playing a higher role in the conformational dynamics. In any case, we believe it might be also useful to play with the visualization/ significance threshold as well to visualize a higher proportion of the identified edges and evaluate how the disconnected parts of the graph are actually connected. Therefore, this has been added as an extra parameter in the SPM webserver. In Figure 2, the two boxes related to distance and significance threshold are also displayed.

Output files

SPM visualization

After uploading the three requested input files, the SPM is built and visualized in the main screen panel (see Figure 3). SPM can be shown in the 3D structure of the uploaded protein structure, where the important residues are marked with spheres and labelled according to its ranked ID, and edges connecting the pairs of residues are highlighted in black. Those pairs of residues that have a higher contribution to the conformational dynamics present bigger spheres and thicker edges. However, the sizes of spheres and widths of edges are mostly qualitative. By default, a distance threshold of 6 Å and a visualization/significance threshold of 0.3 is used. However, as mentioned before, these two parameters can be modified using the threshold panels. The SPM is also displayed in 2D in an additional panel below the 3D representation, in which the residue labels and connections can be more easily seen.

PyMoL script for SPM visualization

Another interesting feature of the SPM webserver is that it generates a PyMoL script that can be executed in PyMoL software after loading the 3D structure of the enzyme. The visualization in PyMoL is rather simple, the user needs to first load the reference structure (load reference.pdb), and then in the command line execute the SPM pymol script defining the correct path where it is located (@$PATH/ pymol_shortest_path_reference). In pymol, the user can tune all parameters and also include some transparency into the cartoon of the protein structure to visualize better the SPM graph (for instance by typing the command in the command line: set cartoon_transparency, 0.6).

Figure 3. SPMweb main page displaying the output after running the SPM calculation. SPM is visualized on top of the 3D structure of the protein as shown in the left panel. The 2D representation of the SPM graph is also shown (right panel). The results can be visualized as full screen by clicking the “Toggle Fullscreen” button, and the labels of the atoms can also be added/removed by clicking the “Toggle Labels (It can take several seconds to process)” button.

CASE EXAMPLES

SPMweb can be used to address different relevant enzymatic properties. We provide some examples of how the SPM tool can be employed: (1) to identify the conformationally relevant distal positions connected to the enzyme active site for the generation of some mutational libraries, and (2) to rationalize the existing allosteric communication between the enzyme subunits in a dimeric structure.

Case example 1. Identification of the key conformationally relevant positions either at the active site or at distal sites connected to the catalytic pocket.

SPM can be applied for identifying mutational spots not restricted to the active site and neither to the tunnel regions targeted by DE. Along the years, we have shown how SPM allows, for the first time, the prediction of which distal active site positions might lead to enhanced enzymatic activity after mutation.(Romero-Rivera, Garcia-Borràs and Osuna, 2017) This has been tested in different unrelated enzyme families showcasing the

potential of SPM for the rational design of enzyme variants.(Osuna, 2021) In this case example, we applied the SPM in the case of tryptophan synthase B (TrpB) subunit, as we first realized that SPM was capturing some of the DE positions(Maria-Solano *et al.*, 2019) and subsequently applied it for designing a stand-alone TrpB.(Maria-Solano, Kinateder, Iglesias-Fernández, Sterner and Osuna, 2021) As shown in Figure 4, the computed SPM in the webserver shows how the graph connects the active site pocket that holds the catalytic lysine and the PLP cofactor with remote sites that interestingly contain many of the DE mutations. For constructing this main SPM, the default parameters for the distance and visualization/significance thresholds have been used (panel A in Figure 4). However, as shown in panel B in Figure 4 by changing the two threshold parameters the obtained SPM maps differ quite substantially. Despite *Pf*TrpB^{0B2} being dimeric in solution in the absence of its binding TrpA partner, we computed the distance and correlation matrices considering only one of the monomeric units. This computed SPM therefore identifies the intramolecular conformationally relevant positions with the monomeric structure connected to the active site pocket.

Figure 4. Case example of computed SPM for investigating the distal sites connected to the active site pocket of the enzyme tryptophan synthase B (*Pf*TrpB^{0B2}) considering only the monomeric structure. A. Visualization of the SPM using the default thresholds for significance and distance. B. Top panel: visualization of the effect of altering the SPM significance threshold using 0.1 (left) and 0.5 (right). Bottom panel: visualization of the effect of altering the distance threshold and using a value of 5 Å (left) and 8 Å (right). The active site of the enzyme that holds the PLP-cofactor and the catalytic residues is highlighted with a blue discontinuous cycle.

Case example 2. Rationalization of the allosteric pathway existing between monomers in a dimeric enzyme structure.

Another interesting feature to analyze in those enzymatic systems that are not monomeric in solution is the existing communication pathway within subunits. This is particularly relevant for allosterically regulated enzymes such as tryptophan synthase, but also in enzymes that require a higher order oligomeric structure for function like monoamine oxidase (MAO-N).(Curado-Carballada, Feixas, Iglesias-Fernández and Osuna, 2019, Osuna, 2021) We have again used the example of TrpB that adopts a dimeric structure to analyze the communication existing between the two subunits. In this case, the whole dimeric structure has been used for SPM construction: the distance and correlation matrices are computed considering the complete dimeric structure. As shown in Figure 5, the computed SPM pathway using the default parameters now expands from one subunit to the other and does not necessarily connect the respective active site pockets of both TrpB monomers. This analysis can be used to identify residues crucial for the intersubunit (allosteric) communication and can also be relevant for explaining cooperative effects.

Figure 5. Case example of computed SPM for studying the allosteric communication existing between monomers in a dimeric tryptophan synthase B (*Pf*TrpB^{0B2}) structure. A. Visualization of the SPM using the default thresholds for significance and distance. B. Top panel: visualization of the effect of altering the SPM significance threshold using 0.1 (left) and 0.5 (right). Bottom panel: visualization of the effect of altering the distance threshold and using a value of 5 Å (left) and 8 Å (right). The active site of the enzyme that holds the PLP-cofactor and the catalytic residues is highlighted with a blue discontinuous cycle.

Conclusions

SPMweb is a new webserver for identifying a subset of conformationally relevant positions located throughout the protein structure. This unique tool can be used for rationally identifying distal sites whose conformational dynamics is connected to the enzyme active site pocket. Although the tool was initially developed for computational enzyme design as discussed in the whole paper, the potential applications of this novel methodology are broad. SPM can be directly used for rationalizing the allosteric communication between enzyme subunits as shown in the case example discussed above. However, it could also be potentially applied for instance for identifying cryptic pockets for designing allosteric inhibitors in drug discovery. We hope that by releasing this webserver to the scientific community, the number of applications and successful cases in which SPM can be applied is expanded.

Acknowledgments:

We thank the Generalitat de Catalunya for the consolidated group TCBioSys (SGR 2021 00487), Spanish MICIN for grant projects PID2021-129034NB-I00 and PDC2022-133950-I00. S.O. is grateful to the funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC-2015-StG-679001, ERC-2022-POC-101112805, and ERC-2022-CoG-101088032), and the Human Frontier Science Program (HFSP) for project grant RGP0054/2020. C.D. was supported by the Spanish MINECO for a PhD fellowship (PRE2019-089147), and G. C. by a research grant from ERC-StG (ERC-2015-StG-679001) and ERC-POC (ERC-2022-POC-101112805).

References:

Addington T.A., Mertz R.W., Siegel J.B., Thompson J.M., Fisher A.J., Filkov V., Fleischman N.M., Suen A.A., Zhang C. and Toney M.D. (2013) Janus: Prediction and Ranking of Mutations Required for Functional Interconversion of Enzymes. *J Mol Biol,* **425**, 1378-1389.

Arnold F.H. (2015) The nature of chemical innovation: new enzymes by evolution. *Q Rev Biophys,* **48**, 404-410.

Bell E.L., Finnigan W., France S.P., Green A.P., Hayes M.A., Hepworth L.J., Lovelock S.L., Niikura H., Osuna S., Romero E. *et al.* (2021) Biocatalysis. *Nat Rev Methods Primers,* **1**, 46.

Bornscheuer U.T., Huisman G.W., Kazlauskas R.J., Lutz S., Moore J.C. and Robins K. (2012) Engineering the third wave of biocatalysis. *Nature,* **485**, 185-194.

Buller R., Lutz S., Kazlauskas R.J., Snajdrova R., Moore J.C. and Bornscheuer U.T. (2023) From nature to industry: Harnessing enzymes for biocatalysis. *Science,* **382**, eadh8615.

Bunzel H.A., Anderson J.L.R., Hilvert D., Arcus V.L., van der Kamp M.W. and Mulholland A.J. (2021) Evolution of dynamical networks enhances catalysis in a designer enzyme. *Nat Chem,* **13**, 1017-1022.

Calvó-Tusell C., Maria-Solano M.A., Osuna S. and Feixas F. (2022) Time Evolution of the Millisecond Allosteric Activation of Imidazole Glycerol Phosphate Synthase. *J Am Chem Soc,* **144**, 7146-7159.

Campbell E., Kaltenbach M., Correy G.J., Carr P.D., Porebski B.T., Livingstone E.K., Afriat-Jurnou L., Buckle A.M., Weik M., Hollfelder F. *et al.* (2016) The role of protein dynamics in the evolution of new enzyme function. *Nat Chem Biol,* **12**, 944-950.

Campitelli P., Modi T., Kumar S. and Ozkan S.B. (2020) The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu Rev Biophys,* **49**, 267-288.

Casadevall G., Pierce C., Guan B., Iglesias-Fernandez J., Lim H.-Y., Greenberg L.R., Walsh M.E., Shi K., Gordon W., Aihara H. *et al.* (2023) Designing Efficient Enzymes: Eight Predicted Mutations Convert a Hydroxynitrile Lyase into an Efficient Esterase. *bioRxiv*, 2023.2008.2023.554512, DOI: 554510.551101/552023.554508.554523.554512.

Castelli M., Marchetti F., Osuna S., F. Oliveira A.S., Mulholland A.J., Serapian S.A. and Colombo G. (2024) Decrypting Allostery in Membrane-Bound K-Ras4B Using Complementary In Silico Approaches Based on Unbiased Molecular Dynamics Simulations. *J Am Chem Soc,* **146**, 901-919.

Csárdi G. and Nepusz T. (2006) The igraph software package for complex network research. *InterJournal,* **Complex Systems**, 1695-1704.

Curado-Carballada C., Feixas F., Iglesias-Fernández J. and Osuna S. (2019) Hidden Conformations in Aspergillus niger Monoamine Oxidase are Key for Catalytic Efficiency. *Angew Chem Int Ed,* **58**, 3097-3101.

Currin A., Swainston N., Day P.J. and Kell D.B. (2014) SpeedyGenes: an improved gene synthesis method for the efficient production of error-corrected, synthetic protein libraries for directed evolution. *Protein Eng Des Sel,* **27**, 273-280.

Currin A., Swainston N., Day P.J. and Kell D.B. (2015) Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem Soc Rev,* **44**, 1172-1239.

Damborsky J. and Brezovsky J. (2014) Computational tools for designing and engineering enzymes. *Curr Opin Chem Biol,* **19**, 8-16.

Duran C., Casadevall G. and Osuna S. (2024) Harnessing Conformational Dynamics in Enzyme Catalysis to achieve Nature-like catalytic efficiencies: The Shortest Path Map tool for computational enzyme design. *submitted for publication*.

Francis J.C. and Hansche P.E. (1972) DE FIRST. *Genet,* **70**, 59-73.

Girvan M. and Newman M.E.J. (2002) Community structure in social and biological networks. *Proc Natl Acad Sci U S A,* **99**, 7821-7826.

Gora A., Brezovsky J. and Damborsky J. (2013) Gates of Enzymes. *Chem Rev,* **113**, 5871-5923.

Gunasekaran K., Ma B. and Nussinov R. (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins,* **57**, 433-443.

Guo J. and Zhou H.-X. (2016) Protein Allostery and Conformational Dynamics. *Chem Rev,* **116**, 6503-6515.

Jaeckel C., Kast P. and Hilvert D. (2008), *Annu Rev Biophys*, pp. 153-173.

Jiang L., Althoff E.A., Clemente F.R., Doyle L., Rothlisberger D., Zanghellini A., Gallaher J.L., Betker J.L., Tanaka F., Barbas C.F. *et al.* (2008) De novo computational design of retro-aldol enzymes. *Science,* **319**, 1387-1391.

Jiménez-Osés G., Osuna S., Gao X., Sawaya M.R., Gilson L., Collier S.J., Huisman G.W., Yeates T.O., Tang Y. and Houk K.N. (2014) The role of distant mutations and allosteric regulation on LovD active site dynamics. *Nat Chem Biol,* **10**, 431-436.

Kazlauskas R.J. and Bornscheuer U.T. (2009) Finding better protein engineering strategies. *Nature Chem Biol,* **5**, 526-529.

Kourist R., Jochens H., Bartsch S., Kuipers R., Padhi S.K., Gall M., Böttcher D., Joosten H.-J. and Bornscheuer U.T. (2010) The α/β-Hydrolase Fold 3DM Database (ABHDB) as a Tool for Protein Engineering. *ChemBioChem,* **11**, 1635-1643.

Kuipers R.K., Joosten H.-J., van Berkel W.J.H., Leferink N.G.H., Rooijen E., Ittmann E., van Zimmeren F., Jochens H., Bornscheuer U., Vriend G. *et al.* (2010) 3DM: Systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins,* **78**, 2101-2113.

Kuipers R.K.P., Joosten H.-J., Verwiel E., Paans S., Akerboom J., van der Oost J., Leferink N.G.H., van Berkel W.J.H., Vriend G. and Schaap P.J. (2009) Correlated mutation analyses on super-family alignments reveal functionally important residues. *Proteins,* **76**, 608-616.

Leveson-Gower R.B., Mayer C. and Roelfes G. (2019) The importance of catalytic promiscuity for enzyme design and evolution. *Nat Rev Chem,* **3**, 687-705.
Lutz S. and Bornscheuer U.T. (2008) *Protein Engineering Handbook*, Wiley-VCH Verlag GmbH & Co. KGaA.
Maria-Solano M.A., Iglesias-Fernández J. and Osuna S. (2019) Deciphering the Allosterically Driven Conformational Ensemble in Tryptophan Synthase Evolution. *J Am Chem Soc,* **141**, 13049-13056.
Maria-Solano M.A., Kinateder T., Iglesias-Fernández J., Sterner R. and Osuna S. (2021) In Silico Identification and Experimental Validation of Distal Activity-Enhancing Mutations in Tryptophan Synthase. *ACS Catal,* **11**, 13733-13743.
Maria-Solano M.A., Serrano-Hervás E., Romero-Rivera A., Iglesias-Fernández J. and Osuna S. (2018) Role of conformational dynamics in the evolution of novel enzyme function. *Chem Comm,* **54**, 6622-6634.
Mazurenko S., Prokop Z. and Damborsky J. (2020) Machine Learning in Enzyme Engineering. *ACS Catal,* **10**, 1210-1223.
Obexer R., Godina A., Garrabou X., Mittl P.R.E., Baker D., Griffiths A.D. and Hilvert D. (2017) Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat Chem,* **9**, 50-56.
Osuna S. (2021) The challenge of predicting distal active site mutations in computational enzyme design. *Wiley Interdiscip Rev Comput Mol Sci*, e1502.
Packer M.S. and Liu D.R. (2015) Methods for the directed evolution of proteins. *Nat Rev Genet,* **16**, 379-394.
Pavelka A., Chovancova E. and Damborsky J. (2009) HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res,* **37**, W376-W383.
Petrović D., Risso V.A., Kamerlin S.C.L. and Sanchez-Ruiz J.M. (2018) Conformational dynamics and enzyme evolution. *J R Soc Interface,* **15**.
Renata H., Wang Z.J. and Arnold F.H. (2015) Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew Chem Int Ed,* **54**, 3351-3367.
Rivalta I., Sultan M.M., Lee N.-S., Manley G.A., Loria J.P. and Batista V.S. (2012) Allosteric pathways in imidazole glycerol phosphate synthase. *Proc Natl Acad Sci U S A,* **109**, E1428-E1436.
Romero P.A. and Arnold F.H. (2009) Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol,* **10**, 866-876.
Romero-Rivera A., Garcia-Borràs M. and Osuna S. (2017) Computational tools for the evaluation of laboratory-engineered biocatalysts. *Chem Comm,* **53**, 284-297.
Romero-Rivera A., Garcia-Borràs M. and Osuna S. (2017) Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal,* **7**, 8524-8532.
Rothlisberger D., Khersonsky O., Wollacott A.M., Jiang L., DeChancie J., Betker J., Gallaher J.L., Althoff E.A., Zanghellini A., Dym O. *et al.* (2008) Kemp elimination catalysts by computational enzyme design. *Nature,* **453**, 190-U194.
Schupfner M., Straub K., Busch F., Merkl R. and Sterner R. (2020) Analysis of allosteric communication in a multienzyme complex by ancestral sequence reconstruction. *Proc Natl Acad Sci U S A,* **117**, 346-354.
Sequeiros-Borja C.E., Surpeta B. and Brezovsky J. (2020) Recent advances in user-friendly computational tools to engineer protein function. *Briefings in Bioinformatics,* **22**.

Sethi A., Eargle J., Black A.A. and Luthey-Schulten Z. (2009) Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A,* **106**, 6620-6625.
Siegel J.B., Zanghellini A., Lovick H.M., Kiss G., Lambert A.R., St.Clair J.L., Gallaher J.L., Hilvert D., Gelb M.H., Stoddard B.L. *et al.* (2010) Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science,* **329**, 309-313.
Stourac J., Vavra O., Kokkonen P., Filipovic J., Pinto G., Brezovsky J., Damborsky J. and Bednar D. (2019) Caver Web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport. *Nucleic Acids Res,* **47**, W414-W422.
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 Res,* **46**, W356-W362.
Turner N.J. (2009) Directed evolution drives the next generation of biocatalysts. *Nature Chem Biol,* **5**, 567-573.
Xiao H., Bao Z. and Zhao H. (2015) High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind Eng Chem Res,* **54**, 4011-4020.
Yang K.K., Wu Z. and Arnold F.H. (2019) Machine-learning-guided directed evolution for protein engineering. *Nat Methods,* **16**, 687-694.