eLife

Reviewed Preprint
v2 • March 12, 2025
Revised by authors

Reviewed Preprint
v1 • January 16, 2025

Structural Biology and Molecular Biophysics

Functionally Important Residues from Graph Analysis of Coevolved Dynamic couplings

Manming Xu, Sarath Chandra Dantu, James A Garnett, Robert A Bonomo, Alessandro Pandini, Shozeb Haider

UCL School of Pharmacy, London, United Kingdom • Department of Computer Science, Brunel University London, London, United Kingdom • Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom • Research Service, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, United States • Department of Molecular Biology and Microbiology, Case Western Reserve University School of Medicine, Cleveland, United States • Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, United States • Clinician Scientist Investigator, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, United States • Departments of Pharmacology, Biochemistry, and Proteomics and Bioinformatics, Case Western Reserve University School of Medicine, Cleveland, United States • CWRU-Cleveland VAMC Center for Antimicrobial Resistance and Epidemiology (Case VA CARES), Cleveland, United States • University of Tabuk (PFSCBR), Tabuk, Saudi Arabia • UCL Center for Advanced Research Computing, University College London, London, United Kingdom

eLife Assessment

This paper reports the analysis of coevolutionary patterns and dynamical information for identifying functionally relevant sites. These findings are considered **important** due to the broad utility of the unified framework and network analysis capable of revealing communities of key residues that go beyond the residue-pair concept. The data are **solid** and the results are clearly presented.

https://doi.org/10.7554/eLife.105005.2.sa4

Abstract

The relationship between protein dynamics and function is essential for understanding biological processes and developing effective therapeutics. Functional sites within proteins are critical for activities such as substrate binding, catalysis, and structural changes. Existing computational methods for the predictions of functional residues are trained on sequence, structural and experimental data, but they do not explicitly model the influence of evolution on protein dynamics. This overlooked contribution is essential as it is known that evolution can fine tune protein dynamics through compensatory mutations, either to improve the proteins' performance or diversify its function while maintaining the same structural scaffold. To model this critical contribution, we introduce DyNoPy, a computational method that combines residue coevolution analysis with molecular dynamics (MD) simulations, revealing hidden correlations between functional sites. DyNoPy constructs a graph model of residue-residue interactions, identifies communities of key residue groups and annotates

critical sites based on their roles. By leveraging the concept of coevolved dynamical couplings —residue pairs with critical dynamical interactions that have been preserved during evolution—DyNoPy offers a powerful method for predicting and analysing protein evolution and dynamics. We demonstrate the effectiveness of DyNoPy on SHV-1 and PDC-3, chromosomally encoded β-lactamases linked to antibiotic resistance, highlighting its potential to inform drug design and address pressing healthcare challenges.

Introduction

Quantifying the contribution of individual residues or residue groups to protein function is important to estimate the pathogenic effect of mutations (1). Identifying the functional roles of individual residues has primarily been done through mutagenesis experiments (2). Bioinformatics methods have complemented these approaches through analysis of multiple sequence alignments (MSA) of homologous proteins and structural data (3–8). Among these methods, computational techniques that can decode inter-residue evolutionary relationships from MSAs have paved the way for machine learning (ML) based strategies that can predict protein structure (9–12), stability (13), and function (7) and extend the scope of computational protein design (14–16). A most recent approach has combined experimental data from three proteins, NUDT15, PTEN and CYP2C9, on stability and function with sequence and structural features to train a ML model to predict functional sites (17).

Functional sites are often regulated by both, local and global interactions. Changes in these interactions are instrumental for functional events like substrate binding, catalysis, and conformational changes (18). The development of physical models of protein dynamics and the increase in available computational power has stimulated the adoption of computational techniques (19, 20) to investigate the conformational dynamics of proteins, an essential component of the many biological functions (21, 22). Different models have been proposed to describe the interactions between residues during simulations and network models have been particularly popular, including methods on single structures and MD simulations data built by analysing the response to external forces on residue networks (23), by estimating the prevalence of non-covalent energy interaction networks in homologous proteins (24), or by analysing linear or non-linear correlation in atomic fluctuations (25, 26). These techniques have demonstrated their usefulness in extracting allosteric networks from structural data with applications in enzyme design (26).

However, none of these techniques incorporate information on residue evolution into the computational approach, While it has been established that evolution through compensatory mutations in dynamic regions, like hinges and loops, can fine tune protein structural dynamics and introduce promiscuity, thereby diversifying biological function. Assuming that protein functional dynamics is conserved during evolution, significant information on dynamic regions and substrate recognition sites should be recoverable using inter residue coevolution scores extracted from MSAs (27, 28). Coevolution analysis and Molecular Dynamics (MD) simulations have independently (29) and synergistically been combined in the past to identify important residues for function (30–34). Yet a method that combines hidden information on dynamics from evolution with direct information on local and global dynamics from conformational ensembles from MD is not yet available.

Here, we present DyNoPy, a computational method that can extract hidden information on functional sites from the combination of pairwise residue coevolution data and powerful descriptors of dynamics extracted from the analysis of MD ensembles. The method can detect coevolved dynamic couplings, i.e. residue pairs with critical dynamical interactions that have been preserved during evolution. These pairs are extracted from a graph model of residue-residue interactions. Communities of important residue groups are detected, and critical sites are

identified by their eigenvector centrality in the graph (**Fig. 1**). We demonstrate the power of this approach on SHV-1 and PDC-3 β-lactamases of major clinical importance (35, 36). DyNoPy successfully detects residue couplings that align with previous studies, guide in the explanations of mutation sites with previously unexplained mechanisms and provide predictions on plausible important sites for the emergence of clinically relevant variants.

Results and Discussion

β-lactamases are a group of enzymes capable of hydrolysing β-lactams, conferring resistance to β-lactam antibiotics (37). These enzymes are evolving rapidly, as single amino acid substitutions are sufficient to drive their evolution and increase their catalytic spectrum and inhibitor resistance profile (38). The widespread dissemination of β-lactamases across different bacterial species and their extensive emergence highlight their global impact on antibiotic resistance (39). The rapid evolution of β-lactamases and their clinical significance (38) makes them an ideal target for evaluating the robustness of DyNoPy.

In this study, we applied DyNoPy to two model enzymes from different β-lactamase families: class A β-lactamase SHV-1 (a chromosomally encoded enzyme in *Klebsiella pneumoniae*) and class C β-lactamase PDC-3 (a chromosomally encoded enzyme in *Pseudomonas aeruginosa*) (35, 36) (Supplementary Fig. S1 and S2). Both class A and class C β-lactamases comprise an α/β domain and an α helical domain, with the active site situated in between (40, 41). Moreover, both enzymes target the carbonyl carbon of the β-lactams using a highly conserved serine residue (42, 43). Despite these similarities, the structures of class A and class C β-lactamases are remarkably different (**Fig. 2**).

SHV-1 is a very well characterised enzyme with wealth of information on mutations and their corresponding effects on protein function. In contrast, the information available on PDC-3 remains limited. Detailed structural information on these enzymes can be found in the supplementary materials. Essential catalytic residues in SHV-1 are: S_{70}, K_{73}, S_{130}, E_{166}, N_{170}, K_{234}, G_{236}, and A_{237} (44) and conserved catalytic residues in PDC-3 include S_{64}, K_{67}, Y_{150}, N_{152}, K_{315}, T_{316}, and G_{317}. Highly conserved stretches of 3-9 hydrophobic residues, annotated as hydrophobic nodes, exists in class A β-lactamases and have been proven to be essential for protein stability (45). Residues defined as belonging to hydrophobic nodes within SHV-1 are listed in Supplementary Table S1.

In SHV-1, the predominant extended spectrum β-lactamase (ESBL) substitutions occur at L_{35}, G_{238}, and E_{240}, while R_{43}, E_{64}, D_{104}, A_{146}, G_{156}, D_{179}, R_{202}, and R_{205} appear in ESBLs with lower frequency (46). Mutations at M_{69}, S_{130}, A_{187}, T_{235}, and R_{244} are known to induce inhibitor resistance in the enzyme (47). In PDC-3, substitutions primarily occur on the Ω-loop, enhancing its flexibility to accommodate the bulky side chains of antibiotics, while deletions are more common in the R2-loop (43). The predominant Ω-loop mutations isolated from clinics are found at positions V_{211}, G_{214}, E_{219}, and Y_{221} (48).

Emergence of highly conserved dynamic couplings

DyNoPy builds a pairwise model of conserved dynamic couplings detected by combining coevolution scores and information on functional motions into a score J_{ij} (see Methods and **Fig. 1**). To this end a dynamics descriptor should be selected. When the descriptor is associated with functional conformational changes, it is expected that functionally relevant couplings will report higher scores. Dynamics descriptors can be selected from commonly used geometrical collective variables (CVs) for the analysis of MD trajectories (see Methods). As expected, the average $\boldsymbol{J}$ matrix score varies across the different CVs, with some of them showing no signal of dynamic coupling (Supplementary Fig. S4C).

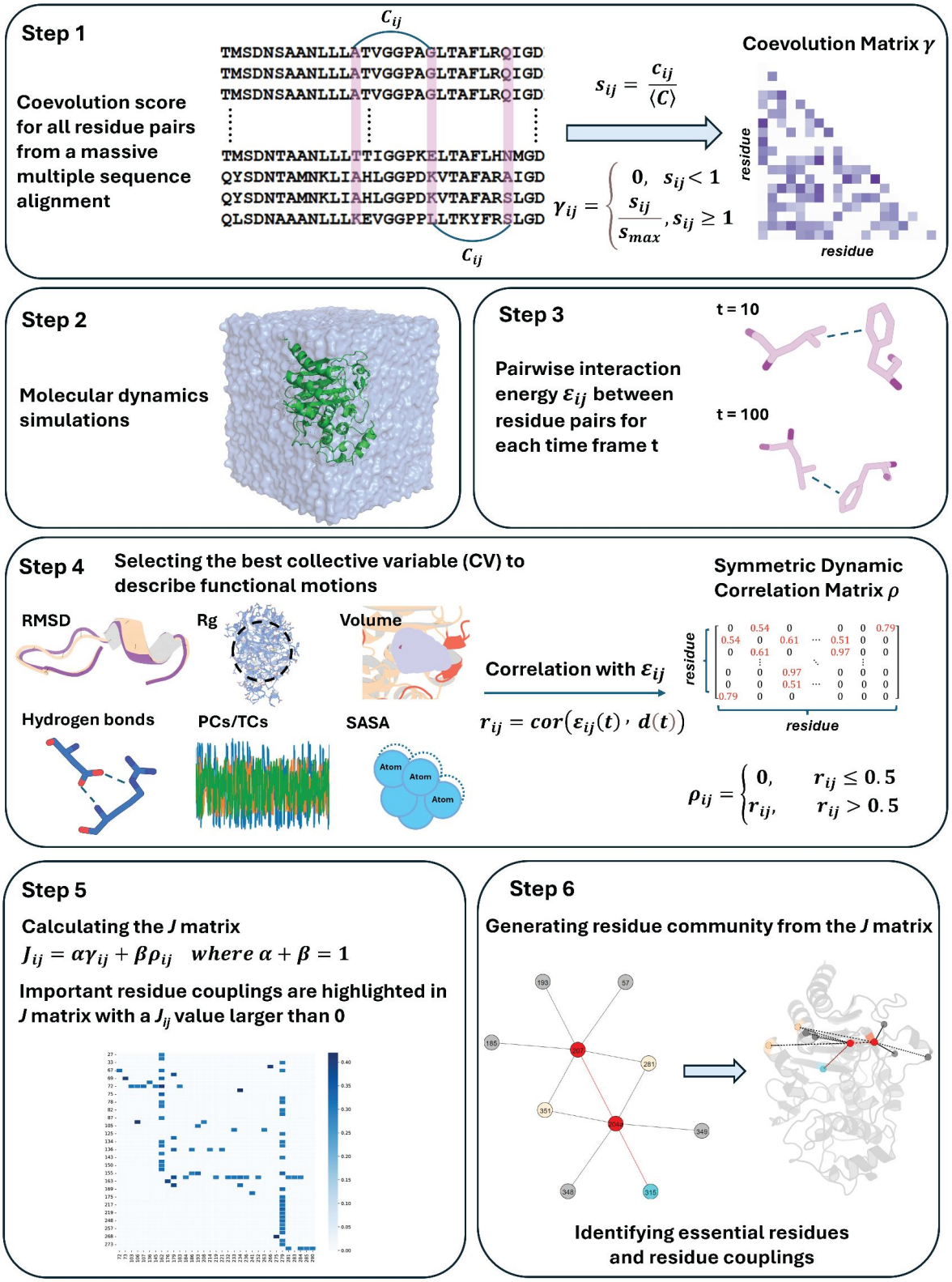

Fig. 1 –

Overview of the DyNoPy workflow

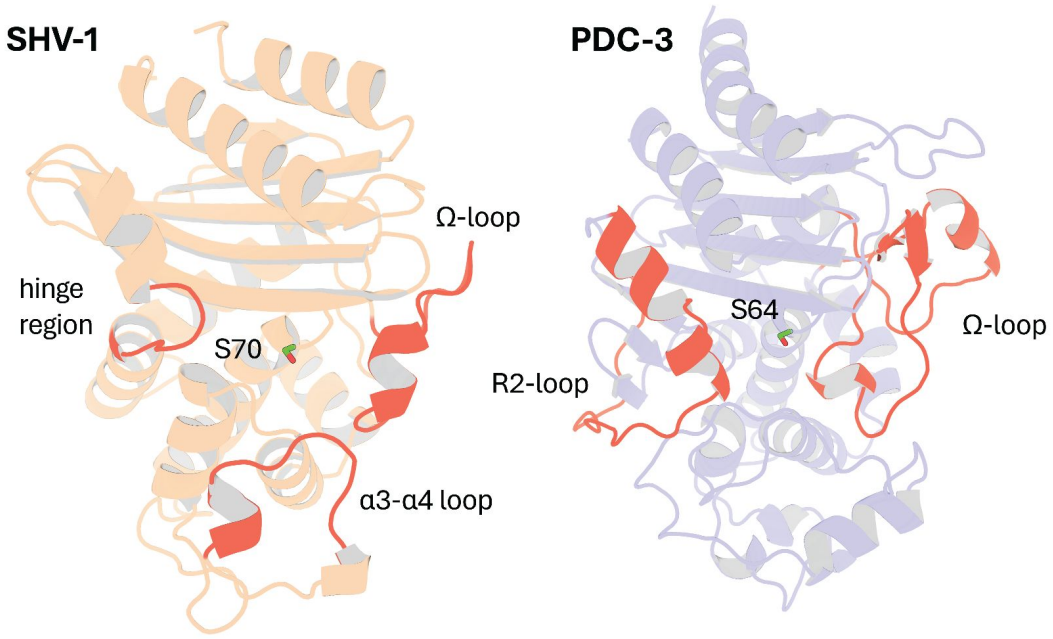

Fig. 2 –

Structural Comparison of SHV-1 (PDB ID: 3N4I) and PDC-3 (PDB ID: 4HEF) β-Lactamases. Catalytic serine S_{70} (SHV-1) and S_{64} (PDC3) are highlighted using stick representation. Important loops surrounding the active site are highlighted in red. In SHV-1, highlighted loops are the α3-α4 loop (residues 101-111), the Ω-loop (residues 164-179), and the hinge region (residues 213-218). In PDC-3, highlighted loops are the Ω-loop (residues 183-226) and the R2-loop (residues 280-310).

SHV-1 and PDC-3 exhibit distinct dynamics, requiring a different choice of the CV that best captures the functional dynamics. For SHV-1, the global first principal component (PC1) proved to be the most effective feature, identifying 571 residue pairs with a J_{ij} value greater than 0. Conversely, PDC-3 requires selection of more localized features that can extract the Ω-loop dynamics from the overall protein motion. Among the dynamic descriptors, the partial first time-lagged component (TC1_partial) performed best for PDC-3, detecting 216 residue pairs with a J_{ij} value greater than 0. Consequently, PC1 and TC1_partial were selected to build the $\boldsymbol{J}$ matrix for SHV-1 and PDC-3, respectively. The performance of all 12 CVs for each protein was assessed and listed in the Supplementary Table S2.

The importance of dynamical information is evident when coevolution couplings (γ_{ij}) and conserved dynamic couplings (J_{ij}) are compared: the number of non-zero couplings decrease from 40% to <2% of total residue pairs in the protein (Supplementary Fig. S4D) when information from the dynamics descriptor is added. Thus, the inclusion of protein dynamics in coevolution studies acts as an effective filter that rules out residue pairs that do not have significant correlations with functional motions. Moreover, when relying only on γ_{ij}, all the residues in SHV-1 and PDC-3 are included within four identified communities (Supplementary Table S3), suggesting that coevolution scores (γ_{ij}) alone do not effectively discriminate residues relevant for protein functions. Furthermore, it would be hard to distinguish critical core residues for each community using only γ_{ij}, as the eigenvector centrality (EVC) values for the residues do not show remarkable differences (Supplementary Fig. S5A and S5B). This means that detailed dynamic investigation of the top residues is needed to determine which pairs should be picked up and further analysed. On the other hand, it is much easier to identify essential residues based on J scores calculated, as clear outliers with significantly higher EVC values could be seen for almost all communities (Supplementary Fig. S5C and S5D) (29, 49). In conclusion, the lack of specificity in the statistically based coevolution analysis supports the choice of incorporating a score for the correlation between residue interactions and dynamic behaviours that enables deconvolution of community information.

DyNoPy reveals critical residues and predicts evolutionary pathways in SHV-1

DyNoPy identified eight meaningful communities, each consisting of at least three strongly coupled residues within SHV-1 (Supplementary Fig. S4A). All crucial catalytic residues and critical substitution sites previously mentioned participating in one of these communities with the exceptions of R_{43}, R_{202}, and S_{130}. Residues previously known to have critical role in function or conferring ESBLs/IRBLs phenotype are either directly coupled to protein dynamics or act as a central hub. The hubs interact with residues with either a role in catalysis or structural stability through their membership of hydrophobic nodes (35). Furthermore, DyNoPy identified key positions (L_{162} and N_{136}) within some communities that are known to undergo substitutions, conferring an ESBL phenotype in other class A β-lactamases. These substitutions have not yet emerged in the SHV family, providing insightful predictions about the potential future evolution of the enzyme. Detailed description of communities with secondary importance for protein function (community 3, 8, and 9) is provided in the supplementary information (Supplementary Fig. S6).

DyNoPy predicts mutation hotspots in SHV-1

DyNoPy detects critical mutation sites (L_{162} and N_{136}) that are known to extend the range of substrates in other class A β-lactamases but have not yet emerged as variants in the SHV family. These sites have not been modified in SHV family because of their plausible central role within the communities as they are mediating couplings with key functional residues essential for catalytic activity and structural stability, indicating their critical role in protein function and the potential lower mutation rate. These findings provide insightful predictions about the potential future evolution of the enzyme, as well as plausible explanations for why these mutations have not yet appeared.

$L_{162,}$ positioned at the start of the Ω-loop and adjacent to the crucial catalytic residue $E_{166,}$ is assigned as the core residue for community 1 (**Fig. 3A**). While it remains conserved in SHV family, variants of L_{162} have been isolated in other class A β-lactamase and are known to expand the enzyme catalytic spectrum. Single amino acid substitution at L_{162} can intensify antibiotic resistance in BEL-1 (50), a class A ESBL clinical variant, exhibiting robust resistance to ticarcillin and ceftazidime (51). BEL-2 diverges from BEL-1 by single amino acid substitution ($L_{162}F$) which alters the kinetic properties of the enzyme significantly and increases its affinity towards expanded-spectrum cephalosporins (52). The relationship between L_{162} and protein catalytic functions can be explained using DyNoPy model, as there are couplings with catalytic important residues M_{69}, K_{73}, E_{166}, and K_{234}. Moreover, the BEL case has confirmed that $L_{162}F$ mutation significantly destabilizes the overall protein structure, highlighting the crucial role of L_{162} in maintaining protein stability (50). DyNoPy accurately identifies the centrality of L_{162} by reporting its connections with 28 backbone residues, including nine hydrophobic node residues critical for protein stability. Among these, five hydrophobic residues are part of the α2 node: V_{75}, L_{76}, G_{78}, V_{80}, and L_{81}, highlighting the contribution of L_{162} to the stability of the α2 helix (35).

Just like L_{162}, N_{136} undergoes advantageous mutations in other class A β-lactamases while remains highly conserved within the SHV family. It is the core residue for community 7 (**Fig. 4B**). This residue forms a hydrogen bond with E_{166}, stabilizing the Ω-loop (53). Although DyNoPy did not detect this direct interaction between N_{136} and E_{166}, the established relationship between N_{136} and N_{170} highlights the role of N_{136} in influencing E_{166}. N_{170}, an essential catalytic residue located on the Ω-loop, contributes to priming the water molecule for the deacylation step with E_{166} (54) and is directly coupled with N_{136}. Due to the essential contribution of N_{136} in facilitating E_{166} to maintain its proper orientation, it was previously thought to be intolerant to mutations as substitution of Asparagine to Alanine at this position would make the enzyme lose its function completely (55). However, $N_{136}D$ substitution has emerged as a new clinical variant very recently in PenL, a class A β-lactamase, by increasing its ability in hydrolysing ceftazidime (55), suggesting that this site has potential to mutate. This gain of function is mainly triggered by the increased flexibility of the Ω-loop (55). DyNoPy correctly detect a dynamical relationship between N_{136} and the Ω-loop (residues 164-179). Six residues present in the Ω-loop participate within this community, including R_{164} and D_{179}. These two residues are critical as they are forming the 'bottleneck' of the Ω-loop which is essential for the correct position of E_{166} (56). D_{179} is also a critical mutation site for SHV-1. Single amino acid substitutions like $D_{179}A$, $D_{179}N$, and $D_{179}G$ are enough for the extended spectrum phenotype (46).

DyNoPy detects residue couplings essential for protein stability

DyNoPy identifies residue couplings critical for protein functional motions, particularly associated with protein stability. These residue pairs exhibit strong relationships as they are not only directly coupled with each other, but also forms various indirect couplings via other residues. As a result, both residues are considered as core residues inside these communities. It is expected that disruption of these couplings through mutation could compromise collective motions essential for enzyme activity.

As the secondary core residues in community 1 (**Fig. 3A**) F_{72} is showing a strong coupling with the primary core residue L_{162} and also forms nine indirect couplings with L_{162}, including via the catalytic K_{234}. This network of direct and indirect relationships reveals the importance of F_{72} and L_{162} coupling in maintaining protein functional motions. Interestingly, previous studies identified a small hydrophobic cavity formed by L_{162} and F_{72}, together with L_{139}, and L_{148}, which is essential for the stability of the active site (50). Notably, DyNoPy successfully recovers the key residues of this local hydrophobic cavity (L_{162}, F_{72}, and L_{148}).

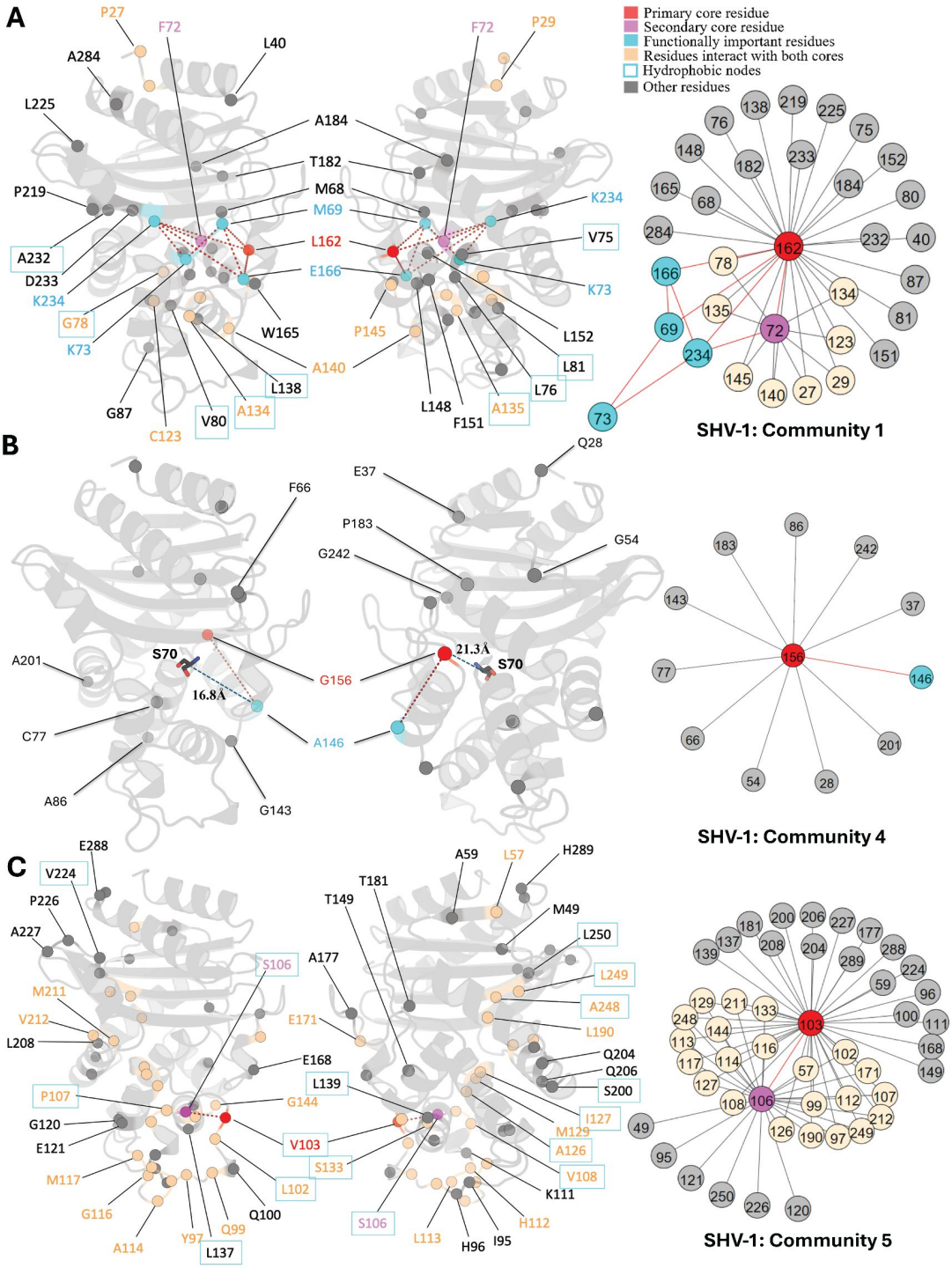

Fig. 3 –

Community 1, 4, and 5 of SHV-1 β-Lactamase. All the residues are depicted as spheres on the protein structure. The core residue for each community is highlighted in red, while purple is used to emphasize the secondary core residue. Residues that interact with both cores are coloured in light yellow. Functional important residues are marked in cyan. Hydrophobic nodes are enclosed with cyan boxes. A. Community 1 of SHV-1, comprising 33 residues with L_{162} being the primary core residue. B. Community 4 of SHV-1, containing 12 residues and is centred by G_{156}. G_{156} and A_{146} are two functional important residues distant from the active site. G_{156} is 21.3L away from the catalytic S_{70}. A_{146} is 16.8L away from S_{70}. C. Community 5 of SHV-1, embracing 48 residues and showing a strong correlation between V_{103} and S_{106}.

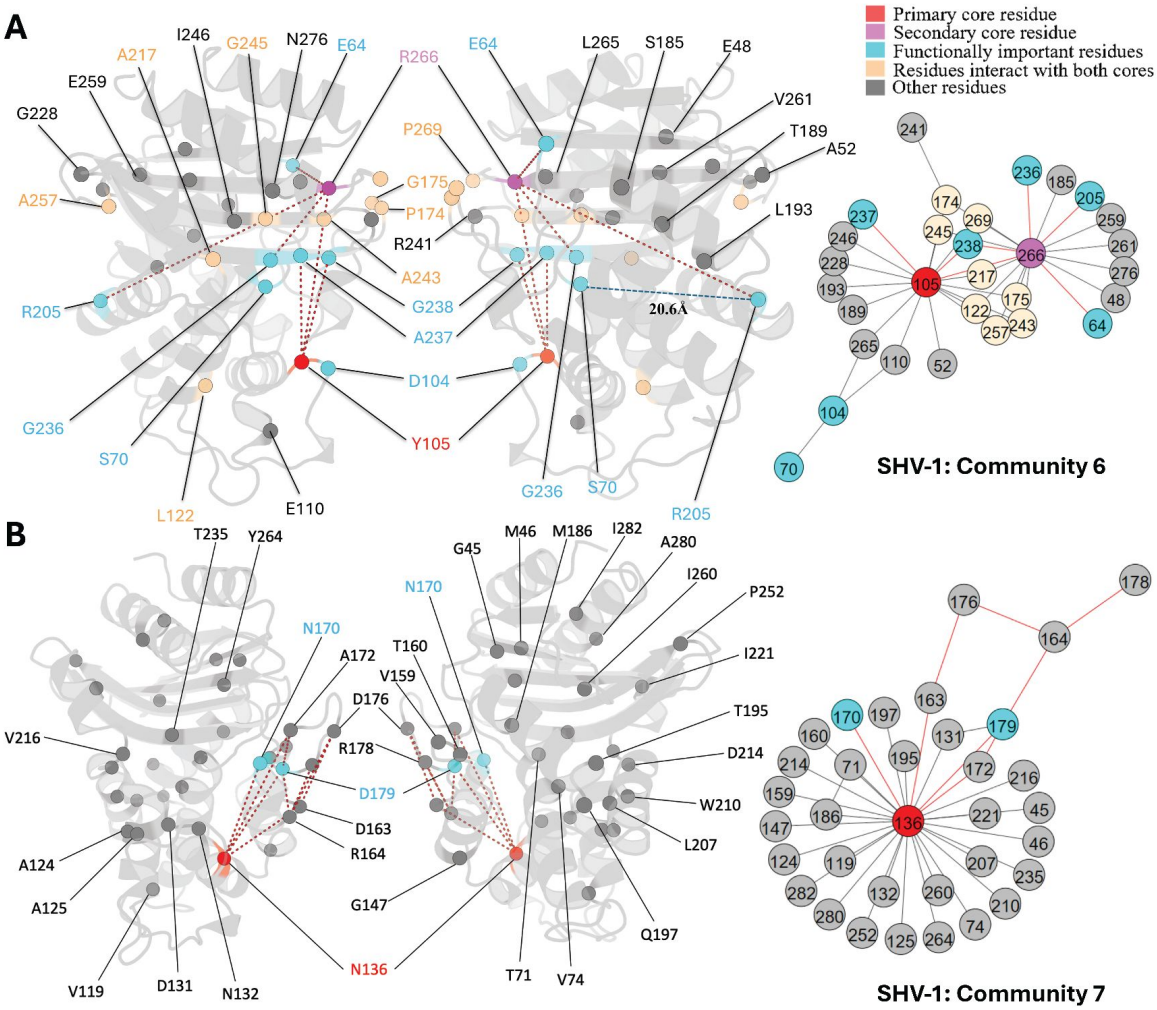

Fig. 4 –

Community 6 and 7 of SHV-1 β-Lactamase. All the residues are depicted as spheres on the protein structure. The core residue for each community is highlighted in red, while purple is used to emphasize the secondary core residue. Residues that interact with both cores are coloured in light yellow. Functional important residues are marked in cyan. A. Community 6 of SHV-1, comprising 30 residues with Y_{105} being the primary core residue. R_{205} is a functional important residue that is 20.6L away from the active site S_{70}. B. Community 7 of SHV-1, containing 34 residues and is centred by N_{136}.

The strong interplay between V_{103} and S_{106}, which are both residues on the α3-α4 loop, is seen in community 5 (**Fig. 3C**). These residues not only interact with each other directly but are also indirectly coupled via 22 other residues. This community emphasizes the significance of hydrophobic nodes in SHV stability and dynamics. Within the analysed 48 residues, 27 are hydrophobic, out of which 15 residues act as nodes critical for enzyme stabilization. Hydrophobic nodes stabilize their own secondary structures and interconnect to stabilize the overall protein (57). V_{103} and S_{106} themselves are hydrophobic nodes, stabilizing α3 helix and α4 helix respectively, and are strongly coupled with each other. In CTX-M, another class A enzyme, $N_{106}S$ is a common substitution that results in improved thermodynamic stability and compensate for the loss in stability of the variants (58). Interestingly, this residue is already a Serine in SHV, but still implies its pivotal role in protein stability.

DyNoPy provides valid explanations for mutation sites

During the evolution of β-lactamases, single mutations on specific sites that are distant from the functional sites have been observed to significantly alter protein catalytic functions. Additionally, single mutations on some surface exposed residues can dramatically increase protein stability. Understanding how these distant mutations impact function and stability becomes a major challenge in understanding protein evolutionary pathways. Communities extracted by DyNoPy show these residues linked with functional important residues, providing a rational for these mutation sites with unknown functions.

Mutations of G_{156} are limited but they lead to ESBL phenotype in the SHV family (46). G_{156} is the central residue for community 4 (**Fig. 3B**), but it is distant from the active site, over 20L away from the catalytic serine S_{70}. Clinical variant SHV-27, has extended resistance ability towards cefotaxime, ceftazidime, and aztreonam (59). It differs from SHV-1 by single amino acid substitution $G_{156}D$, suggesting that it has directly evolved from SHV-1 (59). Limited research has been done on position G_{156}, and the understanding of how it affects the enzyme catalytic properties given that it is far away from the active site is still unclear. Based on our results, we suggest that this residue is essential for the overall protein function because of its 11 coevolved dynamic couplings with protein dynamics, including A_{146}, another ESBL substitution site.

SHV-38, another ESBL that is capable of hydrolysing carbapenems, harbours a single $A_{146}V$ substitution compared to SHV-1 (60). Like G_{156}, A_{146} is 16.8 L away from S_{70} but shows an ability in altering protein catalytic function. The A_{146}-G_{156} residue pair shows a strong coevolutionary signal and strong correlation with protein overall dynamics, implying that there may compensatory mutations at these sites with potential to emerge in the SHV family in the future. These two residues are not connected to any catalytic residues but their coupling to functional dynamics can offer plausible explanation to ESBL activity of these two mutations.

Unlike other substitution sites that are adjacent to the active site, R_{205} is situated more than 20 L away from catalytic serine S_{70}. Its side chain points outward from the protein, exposing to the solvent. The $R_{205}L$ substitution often co-occurs with other ESBL mutations and is thought to indirectly contribute to the ESBL phenotype by compensating for stability loss induced by other mutations (61). SHV-3 is an ESBL that exhibits significant resistance to cefotaxime and ceftriaxone (62). Two substitutions in this enzyme, $R_{205}L$ and $G_{238}S$, extend its resistance profile (62). Thus, it is promising to see that DyNoPy detected these two mutation sites together within community 6 (**Fig. 4A**).

Y_{105} and R_{266} are the core residues for community 6. Y_{105} is situated on the α3-α4 loop positioned at the left side of the binding pocket. It is an important catalytic residue that recognizes and binds to the thiazolidine ring of penicillins or β-lactamase inhibitors (63). There is very limited information on the role of R_{266}, except that it may stabilize the Ω-loop in the SHV family similar to the analogous T_{266} in TEM (64). G_{238} is coupled with an essential catalytic residue Y_{105}, which further links with other catalytic functional residues: S_{70} and A_{237}, and R_{266}, a residue that known

to stabilize the Ω-loop. This indicates that mutations on G_{238} would result in an alteration on protein catalytic function, as well as an increased flexibility of the protein, which strongly aligns with previous finding (62). Its linked mutation site R_{205} does not showing direct coupling with any catalytic residues. Instead, it is directly coupled with $R_{266,}$ which we mentioned as an Ω-loop stabilizer. Thus, it is not surprising that R_{205} substitution alone is never observed in nature (65), as it would not give significant evolutionary advantage to the protein.

Insights into unexplained functional sites of PDC-3

Unlike the extensively studied SHV-1, the functional roles of individual amino acids in PDC-3 remains largely unexplored. This gap in understanding serves as welcome challenge for interpreting the effects of mutations and the dynamic behaviour of PDC-3 from our results. Although several mutation hotspots, such as those on the Ω-loop (48), have been identified, very little is known about the specific contributions of individual amino acids on the functionality of PDC-3.

In PDC-3, mutations have primarily been reported in the Ω-loop. They enhance its flexibility to accommodate the bulky side chains of antibiotics, while deletions are more common in the R2-loop (43). DyNoPy detected five communities in total (Supplementary Fig. S4B) with all the four predominant Ω-loop mutations appeared in these communities. Community 3, 4 and 5 are explained in the Supplementary information (Fig. S7). Furthermore, DyNoPy also detected several previously unexplored Ω-loop residues.

G_{214}, a known mutation site in PDC-3, is the core residue in community 1. Another two essential mutation sites: E_{219} and Y_{221} also participate in this community, directly coupled with G_{214} (**Fig. 5A**). G_{214} also has direct couplings with four other Ω-loop residues: A_{195}, A_{197}, G_{212}, and L_{216}. Previous results have demonstrated that substitutions of Glycine to Alanine or Arginine at 214 significantly destabilizes the Ω-loop (36). The strong correlation between G_{214} and these Ω-loop residues emphasizes the significant contribution of G_{214} towards the stability of the Ω-loop, which corroborates with previous results (36). Moreover, substitutions such as G_{214}A and G_{214}R and mutations on E_{219} and Y_{221} do not affect R2 loop flexibility, resulting in the smaller active site volume among variants (36) because none of the residues from the R2 loop are detected in this community offering plausible explanation to previously unexplained phenomenon.

G_{204} is the core residue of community 2, coupled with 73 other residues, most of which are distant from the catalytic site, suggesting plausible crucial role in overall protein stability like L_{162} in SHV-1 (**Fig. 5B**). G_{204}, a newly emerged mutation site in the PDC family (66), is located on the short β-sheet β5a within the Ω-loop, near the hinge region between β8 and β9 just above the active site. The only known variant of G_{204} is PDC-466, which was derived from PDC-462 (A_{89}V, Q_{120}K, V_{211}A, N_{320}S), with an addition of G_{204}D (66). Coupling of G_{204} to several catalytically important residues, including K_{67}, K_{315}, and T_{316} can suggest that mutations at this site can negatively impact catalytic power. This offers a plausible explanation of seeing fewer variants at this site and mutations at this site could have impact on hydrolysing capabilities of PDC variants. This should be confirmed by further experimental studies of variants of G_{204}. Unlike G_{214}, E_{219} and Y_{221} mutations which do not influence the dynamics of the R2 loop, substitutions on $V_{211,}$ a member of Ω-loop, has impact on dynamics of R2 loop because of its indirect couplings, through G_{204} to R2-loop residues (36). Two less critical substitution sites, H_{188} and V_{329}, were also observed in community 2.

Conclusions

DyNoPy offers two distinct advantages over existing computational tools (24, 26): a) information on residue-residue coevolution can be directly used to detect the components of protein dynamics that have been preserved during evolution b) dynamic descriptors extracted from the MD ensembles can be used to identify the function-specific conserved dynamic couplings.

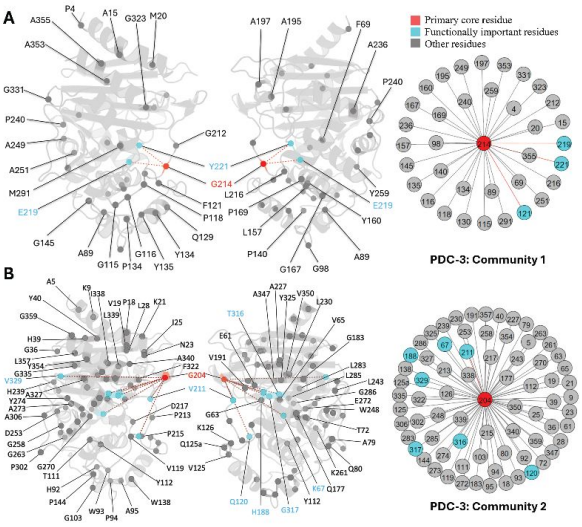

Fig. 5 –

Community 1 and 2 of PDC-3 β-Lactamase. All the residues are depicted as spheres on the protein structure. The core residue for each community is highlighted in red. Functional important residues are marked in cyan. A. Community 1 of PDC-3, comprising 36 residues with G_{214} being the primary core residue. B. Community 2 of PDC-3, containing 74 residues and is centred by G_{204}.

These couplings are then easily modelled as a graph and network analysis is used to extract epistatic communities and assign roles to residues based on their importance in the graph model. The choice of a relevant descriptor of functional dynamics has an impact on the ability to detect couplings that are involved in functional dynamics. In systems where there is limited role of dynamics in the function, the analysis done with DyNoPy is equivalent to conventional coevolution analysis, which can be consider one limitation of our method.

Here we demonstrated how the choice of relevant global and local descriptors returns a higher number of effective couplings (greater than 0), and in turn leads to interpretable graph models and communities. In other systems, when multiple descriptors can be used to quantify functional conformational change, it is expected that they will differently modulate the effect of coevolution coupling, which will be reflected in a different structure of the associated graph models. This suggests the use of DyNoPy to generate comparative models in proteins with multiple functions associated to distinct dynamical changes.

Mutations of L_{162} and N_{136} have not yet emerged in SHV-1, but they are detected by DyNoPy as core residues for communities. These residues are strongly coupled with other functional important residues, which play critical roles in protein stability and catalytic activity. The identification of these couplings shows high consistency with previous studies and highlights the importance of L_{162} and N_{136} in SHV-1 functional dynamics. Given their central role in these communities, mutations in L_{162} and N_{136} can significantly alter protein function, suggesting their potential for future evolutionary changes. However, their strong relationships with these critical functional residues also suggest that mutation at these sites would need to be balanced to maintain protein function, providing an explanation for why such mutations have not yet emerged in SHV-1 (67). The ability of DyNoPy in detecting functionally important mutation sites was demonstrated via well-characterized mutation sites including R_{205} and G_{238} from SHV-1. Moreover, DyNoPy shows predictive ability on less-studied mutation sites such as G_{156} and A_{146}, by detecting critical residue couplings that coevolved with functional motions.

Based on the knowledge we have gained from analysis of SHV-1 functional protein dynamics we suggest that in PDC-3, mutations at G_{204} because of its significant conserved dynamic couplings can lead to new ESBL/IRBL clinical variants. We suggest that DyNoPy can be used as a predictive tool to identify potential functional residues within this enzyme and guide future mutagenesis studies.

In summary, by integrating hidden evolutionary information with direct dynamic interactions, DyNoPy provides a powerful framework for identifying and analysing functional sites in proteins. The tool not only identifies key residues involved in local and global interactions, but also improves our ability to predict silent residues with previously unknown roles for future experimental testing. Our application of DyNoPy to broad-spectrum β-lactamases ESBLs and IRBLs demonstrates its potential to address key medical challenges such as antibiotic resistance by providing valid predictions on protein evolution.

Methodology

DyNoPy generates a graph representation of the protein structure that captures the couplings between amino acid residues contributing to the functional dynamics of the protein. Residues are represented as graph nodes, and conserved dynamic couplings are recorded as edges. Edge weights quantify the strength of these couplings. The model is built on two assumptions: residue pairs should have i) coevolved and their ii) time-dependent interactions correlate with a functional conformational change.

Therefore, edge weights (J_{ij}) for residue i and j are calculated as:

$$J_{ij} = \alpha\gamma_{ij} + \beta\rho_{ij} \quad where\ \alpha + \beta = 1 \tag{1}$$

where γ_{ij} is the scaled coevolution score and ρ_{ij} is the degree of correlation with the selected functional conformational change. α and β are weights assigned to γ_{ij} and ρ_{ij} that have a sum of one. The relative weight of the scaled coevolution score (α) is set to 0.5 in this study. When either of the assumptions listed above is not met, J_{ij} is set to zero.

Scaled coevolution scores

The occurrence of residue-residue coevolution can be estimated and quantified using probabilistic models of correlated mutations from deep multiple sequence alignments (MSA). DyNoPy supports generation of the MSA using the HH-Suite package (68) and calculation of scaled coevolution score (γ_{ij}) using CCMpred (69) as per the protocol described in Bibik et al. (70). For SHV-1 and PDC-3 hhblits returned 18,174 sequences (N_{eff}: 11.082) and 27,892 sequences (N_{eff}: 9.951). Sequences were detected from the UniRef30 (v2022_02) database (71). First a pairwise residue coevolution matrix (***C***) is calculated, then these raw scores (C_{ij}) are divided by the matrix mean (Equation 2). All scores (S_{ij}) smaller than 1 are set to zero, and the remaining values are normalised by the maximum value (Equation 3):

$$s_{ij} = \frac{c_{ij}}{\langle C \rangle} \tag{2}$$

$$\gamma_{ij} = \begin{cases} 0, & s_{ij} < 1 \\ \frac{s_{ij}}{s_{max}}, & s_{ij} \geq 1 \end{cases} \tag{3}$$

Correlation with functional motions

The contribution of a residue pair to a selected functional motion is estimated by how much the change in interaction energy between the two residues over time is correlated with a collective variable (CV) describing the functional motion:

$$r_{ij} = cor(\varepsilon_{ij}(t)\,, d(t)) \tag{4}$$

$$\rho_{ij} = \begin{cases} 0, & r_{ij} \leq 0.5 \\ r_{ij}, & r_{ij} > 0.5 \end{cases} \tag{5}$$

where $\varepsilon_{ij}(t)$ is the pairwise non-bonded interaction energy (see details in Supplementary Information) and $d(t)$ is the time-dependent value of the CV. Examples of CV and a discussion on the choice of the most relevant CV is presented in the results section. Correlation values smaller than 0.5 are set to 0. In absence of detectable contributions to the functional dynamics of the system, the couplings extracted by DyNoPy will describe a pure evolutionary model, and the community detection method presented below will be equivalent to a direct decomposition of the residue coevolution network into units.

Graph representation and analysis of conserved dynamic couplings

All pairwise conserved dynamic couplings (Equation 1) are collected into a square matrix ***J***. A graph is built from ***J***, using python-igraph v0.11 library (72). Nodes represent residues, and edges are drawn between nodes with positive J_{ij}. Edge weights are set to J_{ij}. The relative importance of the residues in this model of protein dynamics is calculated as eigenvector centrality of the nodes (73). Residues involved in extensive correlated dynamics with other highly connected residues have higher eigenvector centrality (EVC) scores. Groups of residues contributing to important collective motions are detected by community analysis of the graph structure. The Girvan-Newman algorithm is used to extract the community structure (74). A

meaningful community should contain at least three residues. Applying network analysis on the combined dynamics-coevolution matrix helps us extracting higher-order interactions beyond pairwise coupling and detecting critical residues, which show multiple interactions with each other. Moreover, indirect long-range relationships, which would be hard to identify from numerical data, could be detected through community clustering. Community-based analysis offers a more comprehensive understanding of residue relationships and enables the visualization of residue couplings on the protein structure.

Adaptive Sampling Molecular Dynamics Simulations

MD simulation data was sourced from our previous studies (35, 36). To summarise, SHV-1 structural coordinates (PDB ID: 3N4I) were obtained from the Protein Data Bank and modified to the wild type by introducing the E104D mutation. Similarly, the PDC-3 structure was derived from PDC-1 (PDB ID: 4HEF) by a T105A substitution. Both enzymes were protonated at pH 7.0 using PropKa from the PlayMolecule platform (75). One disulfide bond between C_{77} and C_{123} was specified in SHV-1. Both structures were solvated with TIP3P water molecules in a periodic box with a box size of 10 Å. Ions were added to neutralize the overall charge of each system at 150mM KCl. Amber force field ff14SB was used for all MD simulations (76). After an initial minimisation of 1000 steps, both the enzymes were equilibrated for 5 ns in the NPT ensemble at 1 atmospheric pressure using the Berendsen barostat (77). The initial velocities for each simulation were sampled from the Boltzmann distribution at 300 K. Multiple Markov State Model (MSM)-based adaptively sampled simulations were performed for both proteins based on the ACEMD engine (78, 79). A canonical (NVT) ensemble with a Langevin thermostat (80) (damping coefficient of 0.1 ps−1) and a hydrogen mass repartitioning scheme were employed to achieve time steps of 4 fs. For SHV-1, each trajectory spanned 60 ns with a time step of 0.1 ns, with a total of 593 trajectories. In the case of PDC-3, 100 trajectories were collected, each containing 3000 frames, lasting 300 ns. To manage the extensive datasets efficiently, trajectories were strategically stridden to ensure that a minimum of 30,000 frames were preserved for each system. The resulting trajectories are summarized in Supplementary Table S4.

Calculation and Selection of Collective Variables

DyNoPy works on the assumption that time-dependent interactions between critical residues, either having significant structural change or not will correlate with functional conformational motions. Since MD simulation data is high-dimensional, a time-dependent collected variable (CV) is required to extract the most relevant information for the process under study. The usefulness of DyNoPy is dependent on the choice of the CVs. To guide the selection of CVs, we selected 12 distinct features: radius of gyration (R_g), the first principal component (PC1), partial PC1 (PC1_partial), the first time-lagged independent component (TC1), partial TC1 (TC1_partial), global root mean square deviation (gRMSD), partial RMSD (pRMSD), dynamical RMSD (dRMSD), global solvent accessible surface area (gSASA), partial SASA (pSASA), active site pocket volume, and the number of hydrogen bonds (hbond). A description of the CVs, including the calculation methods and the residues used to calculate the partial variables, is detailed in the Supplementary information. CVs were subsequently used as input features for DyNoPy. A good collective variable (CV) should appropriately describe protein functional motions. Thus, a CV that detects the highest number of residue couplings is expected to be the most suitable descriptor. The length of the MD simulations should be appropriate to effectively sample the desired functional process as described by the selected CV.

Data Availability

All files required to run the simulations (topology, coordinates, input), processed trajectories (xtc), corresponding coordinates (pdb), can be downloaded from the DOI *https://doi.org/10.57760/sciencedb.15876* (PDC-3) and 10.5281/zenodo.13693144 (SHV-1). DyNoPy is available at *https://github.com/alepandini/DyNoPy*.

Acknowledgements

SCD was supported by Leverhulme Trust grant RPG-2017-222 awarded to AP and JAG. The authors would like to thank Arianna Fornili for insightful suggestions on the design of DyNoPy methodology.

Additional files

Supplementary Information

References

1. Stenson P. D., et al. 2017) **The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies** *Hum Genet* **136**:665–677

2. Matreyek K. A., et al. 2018) **Multiplex assessment of protein variant abundance by massively parallel sequencing** *Nat Genet* **50**:874–882

3. Poelwijk F. J., Krishna V., Ranganathan R. 2016) **The Context-Dependence of Mutations: A Linkage of Formalisms** *PLoS Comput Biol* **12**:e1004771

4. Høie M. H., Cagiada M., Beck Frederiksen A. H., Stein A., Lindorff-Larsen K. 2022) **Predicting and interpreting large-scale mutagenesis data using analyses of protein stability and conservation** *Cell Rep* **38**:110207

5. Blaabjerg L. M., et al. 2023) **Rapid protein stability prediction using deep learning representations** *eLife* **12**

6. Dunham A. S., Beltrao P. 2021) **Exploring amino acid functions in a deep mutational landscape** *Mol Syst Biol* **17**:e10305

7. Hopf T. A., et al. 2017) **Mutation effects predicted from sequence co-variation** *Nat Biotechnol* **35**:128–135

8. Radivojac P., et al. 2013) **A large-scale evaluation of computational protein function prediction** *Nat Methods* **10**:221–227

9. Jumper J., et al. 2021) **Highly accurate protein structure prediction with AlphaFold** *Nature* **596**:583–589

10. Lin Z., et al. 2023) **Evolutionary-scale prediction of atomic-level protein structure with a language model** *Science* **379**:1123–1130

11. Baek M., et al. 2021) **Accurate prediction of protein structures and interactions using a three-track neural network** *Science* **373**:871–876

12. Marks D. S., Hopf T. A., Sander C. 2012) **Protein structure prediction from sequence variation** *Nat Biotechnol* **30**:1072–1080

13. Broom A., Trainor K., Jacobi Z., Meiering E. M. 2020) **Computational Modeling of Protein Stability: Quantitative Analysis Reveals Solutions to Pervasive Problems** *Structure* **28**:717–726

14. Ding D., et al. 2024) **Protein design using structure-based residue preferences** *Nat Commun* **15**:1639

15. Wu Z., Kan S. B. J., Lewis R. D., Wittmann B. J., Arnold F. H. 2019) **Machine learning-assisted directed protein evolution with combinatorial libraries** *Proc Natl Acad Sci U S A* **116**:8852–8858

16. Russ W. P., et al. 2020) **An evolution-based model for designing chorismate mutase enzymes** *Science* **369**:440–445

17. Cagiada M., et al. 2023) **Discovering functionally important sites in proteins** *Nat Commun* **14**:4175

18. Wodak S. J., et al. 2019) **Allostery in Its Many Disguises: From Theory to Applications** *Structure* **27**:566–578

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

20. Rodrigues C. H. M., Pires D. E. V., Ascher D. B. 2021) **DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations** *Protein Sci* **30**:60–69

21. Henzler-Wildman K., Kern D. 2007) **Dynamic personalities of proteins** *Nature* **450**:964–972

22. James L. C., Tawfik D. S. 2003) **Conformational diversity and protein evolution--a 60-year-old hypothesis revisited** *Trends Biochem Sci* **28**:361–368

23. Nevin Gerek Z., Kumar S., Banu Ozkan S. 2013) **Structural dynamics flexibility informs function and evolution at a proteome scale** *Evol Appl* **6**:423–433

24. Yehorova D., Crean R. M., Kasson P. M., Kamerlin S. C. L. 2024) **Key interaction networks: Identifying evolutionarily conserved non-covalent interaction networks across protein families** *Protein Sci* **33**:e4911

25. Lange O. F., Grubmuller H. 2006) **Generalized correlation for biomolecular dynamics** *Proteins* **62**:1053–1061

26. Osuna S. 2020) **The challenge of predicting distal active site mutations in computational enzyme design** *WIREs Computational Molecular Science* **11**

27. Granata D., Ponzoni L., Micheletti C., Carnevale V. 2017) **Patterns of coevolving amino acids unveil structural and dynamical domains** *Proc Natl Acad Sci U S A* **114**:E10612–E10621

28. Liu Y., Bahar I. 2012) **Sequence evolution correlates with structural dynamics** *Mol Biol Evol* **29**:2253–2263

29. Parente D. J., Ray J. C., Swint-Kruse L. 2015) **Amino acid positions subject to multiple coevolutionary constraints can be robustly identified by their eigenvector network centrality scores** *Proteins* **83**:2293–2306

30. Ponzoni L., Polles G., Carnevale V., Micheletti C. 2015) **SPECTRUS: A Dimensionality Reduction Approach for Identifying Dynamical Domains in Protein Complexes from Limited Structural Datasets** *Structure* **23**:1516–1525

31. Sutto L., Marsili S., Valencia A., Gervasio F. L. 2015) **From residue coevolution to protein conformational ensembles and functional dynamics** *Proc Natl Acad Sci U S A* **112**:13567–13572

32. Estabrook R. A., et al. 2005) **Statistical coevolution analysis and molecular dynamics: identification of amino acid pairs essential for catalysis** *Proc Natl Acad Sci U S A* **102**:994–999

33. Chen Z., Rappert S., Sun J., Zeng A. P. 2011) **Integrating molecular dynamics and co-evolutionary analysis for reliable target prediction and deregulation of the allosteric inhibition of aspartokinase for amino acid production** *J Biotechnol* **154**:248–254

34. Wang J., Zhao Y., Wang Y., Huang J. 2013) **Molecular dynamics simulations and statistical coupling analysis reveal functional coevolution network of oncogenic mutations in the CDKN2A-CDK6 complex** *FEBS Lett* **587**:136–141

35. Olehnovics E., et al. 2021) **The Role of Hydrophobic Nodes in the Dynamics of Class A beta-Lactamases** *Front Microbiol* **12**:720991

36. Chen S., et al. 2024) **Omega-Loop mutations control the dynamics of the active site by modulating a network of hydrogen bonds in PDC-3 beta-lactamase** *bioRxiv* https://doi.org/10.1101/2024.02.04.578824

37. Poole K. 2004) **Resistance to beta-lactam antibiotics** *Cell Mol Life Sci* **61**:2200–2223

38. Bush K. 2018) **Past and Present Perspectives on beta-Lactamases** *Antimicrob Agents Chemother* **62**

39. Bush K. 2013) **Proliferation and significance of clinically relevant beta-lactamases** *Ann N Y Acad Sci* **1277**:84–90

40. Matagne A., Lamotte-Brasseur J., Frere J. M. 1998) **Catalytic properties of class A beta-lactamases: efficiency and diversity** *Biochem J* **330**:581–598

41. Philippon A., Arlet G., Labia R., Iorga B. I. 2022) **Class C beta-Lactamases: Molecular Characteristics** *Clin Microbiol Rev* **35**:e0015021

42. Palzkill T. 2018) **Structural and Mechanistic Basis for Extended-Spectrum Drug-Resistance Mutations in Altering the Specificity of TEM, CTX-M, and KPC beta-lactamases** *Front Mol Biosci* **5**:16

43. Jacoby G. A. 2009) **AmpC beta-lactamases** *Clin Microbiol Rev* **22**:161–182

44. Ambler R. P., et al. 1991) **A standard numbering scheme for the class A beta-lactamases** *Biochem J* **276**:269–270

45. Galdadas I., et al. 2018) **Defining the architecture of KPC-2 Carbapenemase: identifying allosteric networks to fight antibiotics resistance** *Sci Rep* **8**:12916

46. Liakopoulos A., Mevius D., Ceccarelli D. 2016) **A Review of SHV Extended-Spectrum beta-Lactamases: Neglected Yet Ubiquitous** *Front Microbiol* **7**:1374

47. Pagan-Rodriguez D., et al. 2004) **Tazobactam inactivation of SHV-1 and the inhibitor-resistant Ser130 -->Gly SHV-1 beta-lactamase: insights into the mechanism of inhibition** *J Biol Chem* **279**:19494–19501

48. Barnes M. D., et al. 2018) **Deciphering the Evolution of Cephalosporin Resistance to Ceftolozane-Tazobactam in Pseudomonas aeruginosa** *mBio* **9**

49. Negre C. F. A., et al. 2018) **Eigenvector centrality for characterization of protein allosteric pathways** *Proc Natl Acad Sci U S A* **115**:E12201–E12208

50. Pozzi C., et al. 2016) **Crystal Structure of the Pseudomonas aeruginosa BEL-1 Extended-Spectrum beta-Lactamase and Its Complexes with Moxalactam and Imipenem** *Antimicrob Agents Chemother* **60**:7189–7199

51. Bogaerts P., Bauraing C., Deplano A., Glupczynski Y. 2007) **Emergence and dissemination of BEL-1-producing Pseudomonas aeruginosa isolates in Belgium** *Antimicrob Agents Chemother* **51**:1584–1585

52. Poirel L., et al. 2010) **BEL-2, an extended-spectrum beta-lactamase with increased activity toward expanded-spectrum cephalosporins in Pseudomonas aeruginosa** *Antimicrob Agents Chemother* **54**:533–535

53. BolJs F., Pleiss J. 2008) **Conserved water molecules stabilize the Omega-loop in class A beta-lactamases** *Antimicrob Agents Chemother* **52**:1072–1079

54. Agarwal V., Yadav T. C., Tiwari A., Varadwaj P. 2023) **Detailed investigation of catalytically important residues of class A beta-lactamase** *J Biomol Struct Dyn* **41**:2046–2073

55. Cao T. P., et al. 2020) **Non-catalytic-Region Mutations Conferring Transition of Class A beta-Lactamases Into ESBLs** *Front Mol Biosci* **7**:598998

56. Parwana D., et al. 2024) **The Structural Role of N170 in Substrate-Assisted Deacylation in KPC-2 beta-Lactamase** *Angew Chem Int Ed Engl* **63**:e202317315

57. Galdadas I., et al. 2021) **Allosteric communication in class A beta-lactamases occurs via cooperative coupling of loop dynamics** *eLife* **10**

58. Lu S., et al. 2022) **An active site loop toggles between conformations to control antibiotic hydrolysis and inhibition potency for CTX-M beta-lactamase drug-resistance enzymes** *Nat Commun* **13**:6726

59. Corkill J. E., Cuevas L. E., Gurgel R. Q., Greensill J., Hart C. A. 2001) **SHV-27, a novel cefotaxime-hydrolysing beta-lactamase, identified in Klebsiella pneumoniae isolates from a Brazilian hospital** *J Antimicrob Chemother* **47**:463–465

60. Poirel L., et al. 2003) **Emergence in Klebsiella pneumoniae of a chromosome-encoded SHV beta-lactamase that compromises the efficacy of imipenem** *Antimicrob Agents Chemother* **47**:755–758

61. Ben Achour N., Mercuri P. S., Ben Moussa M., Galleni M., Belhadj O. 2009) **Characterization of a novel extended-spectrum TEM-type beta-lactamase, TEM-164, in a clinical strain of Klebsiella pneumoniae in Tunisia** *Microb Drug Resist* **15**:195–199

62. Nicolas M. H., Jarlier V., Honore N., Philippon A., Cole S. T. 1989) **Molecular characterization of the gene encoding SHV-3 beta-lactamase responsible for transferable cefotaxime resistance in clinical isolates of Klebsiella pneumoniae** *Antimicrob Agents Chemother* **33**:2096–2100

63. Bethel C. R., et al. 2006) **Role of Asp104 in the SHV beta-lactamase** *Antimicrob Agents Chemother* **50**:4124–4131

64. Kuzin A. P., et al. 1999) **Structure of the SHV-1 beta-lactamase** *Biochemistry* **38**:5720–5727

65. Neubauer S., et al. 2020) **A Genotype-Phenotype Correlation Study of SHV beta-Lactamases Offers New Insight into SHV Resistance Profiles** *Antimicrob Agents Chemother* **64**

66. Colque C. A., et al. 2021) **Development of antibiotic resistance reveals diverse evolutionary pathways to face the complex and dynamic environment of a long-term treated patient** *bioRxiv* https://doi.org/10.1101/2021.05.14.444257

67. Soskine M., Tawfik D. S. 2010) **Mutational effects and the evolution of new protein functions** *Nat Rev Genet* **11**:572–582

68. Remmert M., Biegert A., Hauser A., Soding J. 2011) **HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment** *Nat Methods* **9**:173–175

69. Seemayer S., Gruber M., Soding J. 2014) **CCMpred--fast and precise prediction of protein residue-residue contacts from correlated mutations** *Bioinformatics* **30**:3128–3130

70. Bibik P., Alibai S., Pandini A., Dantu S. C. 2024) **PyCoM: a python library for large-scale analysis of residue-residue coevolution data** *Bioinformatics* **40**

71. Mirdita M., et al. 2017) **Uniclust databases of clustered and deeply annotated protein sequences and alignments** *Nucleic Acids Res* **45**:D170–D176

72. Csárdi G., Nepusz T. 2006) **The igraph software package for complex network research**

73. Newman M. E. J. 2004) **Detecting community structure in networks** *The European Physical Journal B - Condensed Matter* **38**:321–330

74. Newman M. E. 2006) **Finding community structure in networks using the eigenvectors of matrices** *Phys Rev E Stat Nonlin Soft Matter Phys* **74**:036104

75. Martinez-Rosell G., Giorgino T., De Fabritiis G. 2017) **PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations** *J Chem Inf Model* **57**:1511–1516

76. Maier J. A., et al. 2015) **ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB** *J Chem Theory Comput* **11**:3696–3713

77. Berendsen H. J. C., Postma J. P. M., van Gunsteren W. F., DiNola A., Haak J. R. 1984) **Molecular dynamics with coupling to an external bath** *The Journal of Chemical Physics* **81**:3684–3690

78. Doerr S., Harvey M. J., Noe F., De Fabritiis G. 2016) **HTMD: High-Throughput Molecular Dynamics for Molecular Discovery** *J Chem Theory Comput* **12**:1845–1852

79. Harvey M. J., Giupponi G., Fabritiis G. D. 2009) **ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale** *J Chem Theory Comput* **5**:1632–1639

80. Davidchack R. L., Handel R., Tretyakov M. V. 2009) **Langevin thermostat for rigid body dynamics** *J Chem Phys* **130**:234101

Author information

Manming Xu*
UCL School of Pharmacy, London, United Kingdom

*Joint first authors

Sarath Chandra Dantu*
Department of Computer Science, Brunel University London, London, United Kingdom
ORCID iD: 0000-0003-2019-5311

*Joint first authors

James A Garnett
Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom

Robert A Bonomo
Research Service, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, United States, Department of Molecular Biology and Microbiology, Case Western Reserve University School of Medicine, Cleveland, United States, Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, United States, Clinician Scientist Investigator, Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, United States, Departments of Pharmacology, Biochemistry, and Proteomics and Bioinformatics, Case Western Reserve University School of Medicine, Cleveland, United States, CWRU-Cleveland VAMC Center for Antimicrobial Resistance and Epidemiology (Case VA CARES), Cleveland, United States

Alessandro Pandini
Department of Computer Science, Brunel University London, London, United Kingdom, University of Tabuk (PFSCBR), Tabuk, Saudi Arabia
ORCID iD: 0000-0002-4158-233X

For correspondence: alessandro.pandini@brunel.ac.uk

Shozeb Haider
UCL School of Pharmacy, London, United Kingdom, University of Tabuk (PFSCBR), Tabuk, Saudi Arabia, UCL Center for Advanced Research Computing, University College London, London, United Kingdom
ORCID iD: 0000-0003-2650-2925

For correspondence: Shozeb.haider@ucl.ac.uk

Editors

Reviewing Editor
Yogesh Gupta
The University of Texas Health Science Center at San Antonio, San Antonio, United States of America

Senior Editor
Qiang Cui
Boston University, Boston, United States of America

Reviewer #1 (Public review):

Summary:

As reported above, this paper by Xu et al reports on a new method to combine the analysis of coevolutionary patterns with dynamic profiles to identify functionally important residues and reveal correlations between binding sites.

Strengths:

In general, coevolutionary analysis and MD analysis are carried out separately and while there have been attempts to compare the information provided by the two, no unified framework exists. Here, the authors convincingly demonstrate that integrating signals from Dynamics and coevolution gives information that substantially overcomes the one provided by either method in isolation. While other methods are useful, they do not capture how dynamics is fundamental to define function and thus sculpts coevolution, via the 3D structure of the protein. At the same time, the authors demonstrate how coevolution in turn also influences internal dynamics. The Networks they rebuild unveil information at an even higher level: the model starts pairwise but through network representation the authors arrive to community analysis, reporting on interaction patterns that are larger than simple couples.

Comments on latest version:

I have nothing to add to this revision. The paper looks excellent and very interesting.

https://doi.org/10.7554/eLife.105005.2.sa3

Reviewer #2 (Public review):

Summary:

The authors introduced a computational framework, DyNoPy, that integrates residue coevolution analysis with molecular dynamics (MD) simulations to identify functionally important residues in proteins. DyNoPy identifies key residues and residue-residue coupling to generate an interaction graph and attempts to validate using two clinically relevant β-lactamases (SHV-1 and PDC-3).

Strengths:

DyNoPy could not only show clinically relevance of mutations but also predict new potential evolutionary mutations. Authors have provided biologically relevant insights into protein dynamics which can have potential applications in drug discovery and understanding molecular evolution.

Comments on latest version:

I appreciate the efforts of the authors to address my comments.

https://doi.org/10.7554/eLife.105005.2.sa2

Reviewer #3 (Public review):

Summary:

In this paper, Xu, Dantu and coworkers report a protocol for analyzing coevolutionary and dynamical information to identify a subset of communities that capture functionally relevant sites in beta-lactamases.

Strengths:

The combination of coevolutionary information and metrics from MD simulations is interesting for capturing functionally relevant sites, which can have implications in the fields of drug discovery but also in protein design.

Comments on latest version:

The authors have successfully addressed all my previous comments/concerns. I am happy with the current version of the manuscript.

https://doi.org/10.7554/eLife.105005.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *As reported above, this paper by Xu et al reports on a new method to combine the analysis of coevolutionary patterns with dynamic profiles to identify functionally important residues and reveal correlations between binding sites.*
>
> *Strengths:*
>
> *In general, coevolutionary analysis and MD analysis are carried out separately and while there have been attempts to compare the information provided by the two, no unified framework exists. Here, the authors convincingly demonstrate that integrating signals from Dynamics and coevolution gives information that substantially overcomes the one provided by either method in isolation. While other methods are useful, they do not capture how dynamics is fundamental to define function and thus sculpts coevolution, via the 3D structure of the protein. At the same time, the authors demonstrate how coevolution in turn also influences internal dynamics. The Networks they rebuild unveil information at an even higher level: the model starts pairwise but through network representation the authors arrive to community analysis, reporting on interaction patterns that are larger than simple couples.*
>
> *Weaknesses:*
>
> *The authors should*
>
> *- Make an effort in suggesting/commenting the limits of applicability of their method;*

We have added a sentence on Page 17, line 15 that describes the limitation of our method.

> *- Expand discussion on how DyNoPy compares to other methods;*

A paragraph has been added to explain the comparison with other models (Page 3, line 18)

> *- Dynamic is not essential in all systems (structural proteins): The authors may want to comment on possible strategies they would use for other systems where their framework may not be suitable/applicable.*

We agree with the reviewer that dynamics is not essential in all systems. In systems where there is limited role of dynamics in the function, the analysis done with DyNoPy is equivalent to conventional coevolution analysis, which can be consider one limitation of our method. Conversely, for dynamic proteins, combining functional dynamics descriptors with coevolution analysis using DyNoPy, helps in denoising information by deconvolution of communities. We have included this in the manuscript to highlight the suitability/applicability of the method.

Further, we have added a paragraph in the Introduction and conclusions highlighting the main difference between DyNoPy and existing computational tools like DCCM, KIN, and SPM and for your convenience it is provided below:

“Functional sites are often regulated by both, local and global interactions. Changes in these interactions are instrumental for functional events like substrate binding, catalysis, and conformational changes (18). The development of physical models of protein dynamics and the increase in available computational power has stimulated the adoption of computational techniques (19, 20) to investigate the conformational dynamics of proteins, an essential component of the many biological functions (21, 22). Different models have been proposed to describe the interactions between residues during simulations and network models have been particularly popular, including methods on single structures and MD simulations data built by analysing the response to external forces on residue networks (23), by estimating the prevalence of non-covalent energy interaction networks in homologous proteins (24), or by analysing linear or non-linear correlation in atomic fluctuations (25, 26). These techniques have demonstrated their usefulness in extracting allosteric networks from structural data with applications in enzyme design (26).”

> ***Reviewer #2 (Public review):***
>
> *Summary:*
>
> *Authors introduced a computational framework, DyNoPy, that integrates residue coevolution analysis with molecular dynamics (MD) simulations to identify functionally important residues in proteins. DyNoPy identifies key residues and residue-residue coupling to generate an interaction graph and attempts to validate using two clinically relevant β-lactamases (SHV-1 and PDC-3).*
>
> *Strengths:*
>
> *DyNoPy could not only show clinically relevance of mutations but also predict new potential evolutionary mutations. Authors have provided biologically relevant insights into protein dynamics which can have potential applications in drug discovery and understanding molecular evolution.*
>
> *Weaknesses:*
>
> *Although DyNoPy could show the relevance of key residues in active and non-active site residues, no experiments have been performed to validate their predictions.*

We thank the reviewer for highlighting this point. We acknowledge that direct experimental validation of our predictions for DyNoPy has not yet been performed. However, we have

provided explanations and evidence from experiments conducted on closely related homologs to support the relevance of key residues. These homologs share significant structural and functional similarity, which strengthens the reliability of our predictions.

> *In addition, they should compare their method with conventional techniques and show how their method could be different.*

We thank all the reviewers for highlighting this oversight on our behalf. In Introduction and conclusion, we have added the following paragraphs:

“Functional sites are often regulated by both, local and global interactions. Changes in these interactions are instrumental for functional events like substrate binding, catalysis, and conformational changes (18). The development of physical models of protein dynamics and the increase in available computational power has stimulated the adoption of computational techniques (19, 20) to investigate the conformational dynamics of proteins, an essential component of the many biological functions (21, 22). Different models have been proposed to describe the interactions between residues during simulations and network models have been particularly popular, including methods on single structures and MD simulations data built by analysing the response to external forces on residue networks (23), by estimating the prevalence of non-covalent energy interaction networks in homologous proteins (24), or by analysing linear or non-linear correlation in atomic fluctuations (25, 26). These techniques have demonstrated their usefulness in extracting allosteric networks from structural data with applications in enzyme design (26). ”

> *An explanation of "communities" divided in the work and how these communities are relevant to the article should be provided. In addition, choice of collective variables and their relevance in residue coupling movement is also not very well explained. Dynamics cross correlation map can also be a good method for understanding the residue movements and can explain the residue-residue coupling, it is not explained how DyNoPy is different from the conventional methods or can perform better.*

The following sentences have been included in the manuscript to address the questions raised by the reviewer:

On Community Definition and relevance

DyNoPy identified coevolving residue pairs (scaled coevolution score >1) with interactions strongly correlated with protein functional motions (i.e., J values larger than zero). Applying network analysis on the combined dynamics-coevolution matrix helps us extracting higher-order interactions beyond pairwise coupling and detecting critical residues, which show multiple interactions with each other. Moreover, indirect long-range relationships, which would be hard to identify from numerical data, could be detected through community clustering. Community-based analysis offers a more comprehensive understanding of residue relationships and enables the visualization of residue couplings on the protein structure.

On Choice of collective variables:

DyNoPy works on the assumption that time-dependent interactions between critical residues, either having significant structural change or not will correlate with functional conformational motions. Since MD simulation data is high-dimensional, a time-dependent dynamic descriptor is required to extract the most relevant information for the process under study. A good collective variable (CV) should appropriately describe protein functional motions. Thus, a CV that detects the highest number of residue couplings is expected to be the most suitable descriptor (Mentioned in Page 22 Line 14). In our study, we tested 12 CVs, either focusing on the entire protein or on selected regions. And the best performed CV (the one identified the most residue couplings) was selected for further analysis. In practical

applications, users can decide whether to focus on the most relevant global or local dynamics descriptor depending on the dynamics of their specific system.

We have added a paragraph in the Introduction differentiating DyNoPy with other methods including DCCM. DCCM differs from DyNoPy in two aspects 1) it does not account for inter-residue coevolution 2) the correlation matrix captures correlations of atomic/residue movements associated with the whole intrinsic dynamics of the system, without filtering for the contributions to the important motions involved in the biological function. Additionally, any residue pair contributing to functional motion without itself undergoing any structural change will not be visible in this approach.

> *In the sentence "DyNoPy identified eight significant communities of strongly coupled residues within SHV-1 (Supporting Fig. S4A)" I could not find a clear description of eight significant communities.*

The following sentences have been included in the results, methods and figure legends that define 'significant community':

'DyNoPy identified eight meaningful communities, each consisting of at least three strongly coupled residues within SHV-1 (Supplementary Fig. S4A). All crucial catalytic residues and critical substitution sites previously mentioned participating in one of these communities with the exceptions of R_{43}, R_{202}, and S_{130}.' (Page 8 Line 28)

'A meaningful community should contain at least three residues.' (Page 21 Line 2)

'A reasonable residue community should contain at least three residues.' (SI Page 11)

> *Again the description of communities is not clear to me in the following sentence "Detailed description of the other three communities is provided in the supporting information (Fig. S6)."*

This following sentence has been rewritten.

'Detailed description of communities with secondary importance for protein function (community 3, 8, and 9) is provided in the supplementary information (Supplementary Fig. S6).' (Page 9, line 8)

> *In the sentence "N170 acts as an intermediary between N136 and E166". Kindly cite the reference figure to show N179 as intermediate residue.*

This sentence has been rewritten to avoid any confusion.

'Although DyNoPy did not detect this direct interaction between N136 and E166, the established relationship between N136 and N170 highlights the role of N136 in influencing E166.' (Page 10 Line 8)

> *Please be careful with the numbers. In the sentence "These residues not only interact with each other directly but are also indirectly coupled via 21 other residues." I could count 22 other residues and not 21.*

We thank the reviewer for spotting this error. This has now been corrected. All the communities are counted again.

'These residues not only interact with each other directly but are also indirectly coupled via 22 other residues.' (Page 12 Line 14)

> *In the sentence "Unlike other substitution sites that are adjacent to the active site, R_{205} is situated more than 16 Å away from catalytic serine S_{70}". Please add this label somewhere in the figure.*

The figure legends have been updated to include this. Distances have been added to community 4 Fig. 3 and community 6 Fig. 4. Residue index in the legend of Fig.3 has been included as subscript. Distance in the main text has been changed to be more accurate.

'G_{156} and A_{146} are two functional important residues distant from the active site. G_{156} is 21.3Å away from the catalytic S_{70}. A_{146} is 16.8Å away from S_{70}.' (Page 12 Line 2)

'R_{205} is a functional important residue that is 20.6Å away from the active site S_{70}.' (Page 13 Line 10)

> *Please cite a reference in the sentence "This indicates that mutations on G238 would result in an alteration on protein catalytic function, as well as an increased flexibility of the protein, which strongly aligns with previous finding."*

The citation has been added

'This indicates that mutations on G238 would result in an alteration on protein catalytic function, as well as an increased flexibility of the protein, which strongly aligns with previous finding (62).' (Page 15 Line 2)

> ***Reviewer #3 (Public review):***
>
> *Summary:*
>
> *In this paper, Xu, Dantu and coworkers report a protocol for analyzing coevolutionary and dynamical information to identify a subset of communities that capture functionally relevant sites in beta-lactamases.*
>
> *Strengths:*
>
> *The combination of coevolutionary information and metrics from MD simulations is interesting for capturing functionally relevant sites, which can have implications in the fields of drug discovery but also in protein design.*
>
> *Weaknesses:*
>
> *The combination of coevolutionary information and metrics from MD simulations is not new as other protocols have been proposed along the years (the current version of the paper neglects some of them, see below), and there are a few parameters of the protocol that, in my opinion, should be better analyzed and discussed.*
>
> *(1) As mentioned, the introduction of the paper lacks some important publications in the field of using graph theory to represent important interaction networks extracted from MD simulations (DOI: 10.1002/pro.4911), and also combining MD data with MSA to identify functionally relevant sites for enzyme design (doi: 10.1021/acscatal.4c04587, 10.1093/protein/gzae005).*

We are very grateful for pointing us to these references. We have added a paragraph in the Introduction mentioning these and other computational tools similar to DyNoPy. Further, in conclusion we have highlighted the differences between DyNoPy and existing tools.

> *(2) The matrix used to apply graph theory (J_ij) is built from summing the scaled coevolution and degree of correlation values. The alpha and beta weights are defined, and the authors mention that alpha is set to 0.5, thus beta as well to fulfil with the alpha + beta = 1. Why a value of 0.5 has been selected? How this affects the overall results and conclusions extracted? The finding that many catalytically relevant residues are identified in the communities is not surprising given that such sites usually present a high conservation score.*

This is an excellent question. Our present formulation allows the user to easily assess the influence of coevolution and dynamic couplings on the output. Setting alpha to 0.5, weights both evolutionary and dynamics information equally and has shown promising results in SHV-1 and PDC-3. As it has been presented in the manuscript, setting alpha to 1, i.e., purely utilising coevolution data does not let us identify critical residues effectively as all residues are included in the set (Supplementary Fig. S4 and S5). In future work, we would like to investigate the effect of scanning alpha from 0 to 1 on the final residue list, possibly on a larger set of proteins and protein families.

We would also like to point out that some of the residue pairs with coevolution scores in the top 1% have J-scores set to 0, as they lacked significant coupling to the functional dynamics.

> *(3) Another important point that needs further explanation is the selection of the relevant descriptor of protein dynamics. In this study two different strategies have been used (one more global the other more local), but more details should be provided regarding their choice. What is the best strategy according to the authors? Why not using the same strategy for both related systems? The obtained results using one methodology or the other will have a large impact on the dynamical score. Another related point is: what is the impact of the MD simulation length, how the MSA is generated and number of sequences used for MSA construction?*

As in the case of many complex proteins, the flow of information occurs in β-lactamases via structural interactions (https://doi.org/10.7554/eLife.66567). These interactions occur both on a local level, as in the case of binding site residues or residues immediately surrounding the binding site; however, there are interactions far away (>20Å) from the binding site that have the ability to alter function. We have obtained this information from extensive surveys of clinical isolates and experimental data. To account for such interactions, a more global approach has to be taken. To answer the reviewer's question: each system is unique and there is no one-fixed strategy. In short, the method used should be able to denoise information and the user is advised to fine-tune their findings by corroborating with experimental and clinical information.

The length of MD simulations is also system specific. Some systems effectively sample the functional dynamics within a shorter simulation time, while others take a long timescale MD simulation to converge. The results won't change as long as the simulation has effectively sampled the functional dynamics associated with biological function.

The MSA is generated by the HH-Suite package as mentioned on Page 19 Line 19. More specifically, the MSA is constructed based on the UniRef30 database, where sequences are clustered, and each cluster contains sequences with at least 30% sequence identity. This provides a non-redundant set of protein sequences. Our package allows the automatic generation of MSAs from the database. For SHV-1, the alignment contains 18,175 protein sequences and for PDC-3, the alignment consists of 27,892 protein sequences. Full details of this protocol are published in Bibik et al. (https://doi.org/10.1093/bioinformatics/btae166). We have revised the methods section to include these details.

Other Minor Alterations

‘Fig. S1 and S2’ has been changed to ‘Supplementary Fig. S1 and S2’ for consistency (Page 6 Line 12)

(1) ‘Figure 5B’ has been changed to ‘Fig. 5B’ for consistency (Page 16 Line 11)

(2) All the ‘Figure’ has been changed to ‘Fig.’ in the SI for consistency

(3) Just as the suggestion, an alteration has been made on the Step 1 of Fig.1.

https://doi.org/10.7554/eLife.105005.2.sa0