Article

Computational Study of Network and Type-I Functional Divergence in Alcohol Dehydrogenase Enzymes Across Species Using Molecular Dynamics Simulation

Suhyun Park [1] , Petrina Jebamani [2] , Yeon Gyo Seo [1] and Sangwook Wu [1,3,]*

[1] Department of Physics, Pukyong National University, Busan 48513, Republic of Korea; suhyun@pukyong.ac.kr (S.P.); yeongs0709@pukyong.ac.kr (Y.G.S.)
[2] Department of Chemical Engineering, Pusan National University, Busan 46241, Republic of Korea; petrina@pusan.ac.kr
[3] PharmCADD, 1102-ho, 60, Centum Jungang-ro, Haeundae-gu, Busan 48059, Republic of Korea
[*] Correspondence: sangwoow@pknu.ac.kr; Tel.: +82-051-629-5549

Abstract: Alcohol dehydrogenases (ADHs) are critical enzymes involved in the oxidation of alcohols, contributing to various metabolic pathways across organisms. This study investigates type I functional divergence within three ADH1 families: *Saccharomyces cerevisiae* (PDB ID: 4W6Z), *Gadus morhua* (PDB ID: 1CDO), and *Homo sapiens* (PDB ID: 1HDX). Understanding the molecular evolution and mechanisms underlying functional divergence of ADHs is essential for comprehending their adaptive significance. For this purpose, we performed a computational analysis that included structural characterization of ADHs through three-dimensional modeling, site-specific analysis to evaluate selective pressures and evolutionary constraints, and network analysis to elucidate relationships between structural features and functional divergence. Our findings indicate substantial variations in evolutionary and structural adaptations among the ADH families.

1. Introduction

Alcohol dehydrogenases (ADH) are a group of enzymes that have attracted significant scientific interest due to their fundamental role in the metabolism of alcohol across a diverse range of organisms, including bacteria, algae, yeasts, plants, and vertebrates [1,2]. These enzymes catalyze the transformation of alcohol into aldehyde or ketone with the concurrent reduction of nicotinamide adenine dinucleotide (NAD+) (Figure 1a). While this reaction yields potentially toxic intermediates, it also generates aldehyde and alcohol groups that serve various physiological functions [3]. ADH enzymes are particularly vital in yeasts, plants, and many bacterial species, where they drive fermentation processes critical for replenishing NAD+ (Figure 1b,c) and sustaining key metabolic pathways, thereby supporting survival and growth in low-oxygen environments [2,4]. In addition to their metabolic roles, ADH enzymes are a subject of evolutionary study due to their involvement in local adaptation, a process that promotes genetic variation and functional specialization across organisms. Investigating these adaptations provides valuable insights into how ADH genes evolve and acquire distinct roles relevant to organismal health and environmental adaptation.
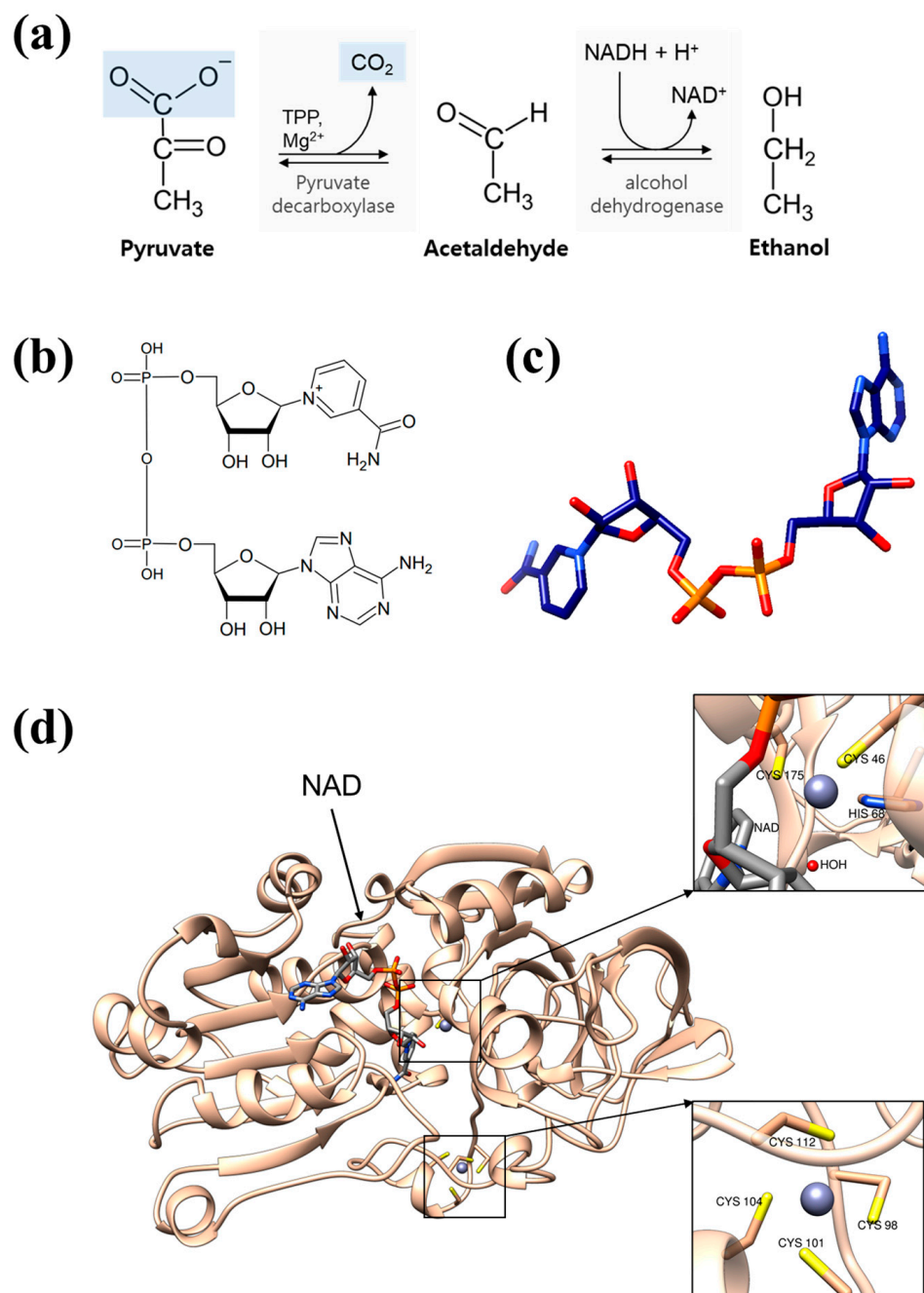
Figure 1. Alcohol dehydrogenase is an enzyme used together with pyruvate decarboxylase in alcoholic fermentation to convert NADH to NAD+ while converting acetaldehyde to ethanol. (**a**) The conversion process. (**b**) 2D and (**c**) 3D structures of NAD. (**d**) Structure of human ADH in complex with NAD (PDB id: 1HDX). The zinc ions are shown in the boxes.

Alcohol dehydrogenase (ADH) enzymes form a diverse gene family with several distinct classes. In yeasts and bacteria, ADH enzymes—typically functioning as tetramers with two zinc ions per monomer—catalyze reactions crucial for energy production through the conversion of acetaldehyde to ethanol. These functions are particularly important under conditions of low oxygen or low temperatures, enabling these organisms to adapt to environmental stressors [2,3]. ADH enzymes in plants similarly contribute to the maintenance of cellular redox balance and NAD+ regeneration [4,5]. In humans, ADH enzymes are structurally diverse, with at least seven unique genes encoding different isoforms. Human ADH is dimeric and incorporates two zinc ions (Zn^{2+}), reflecting its intricate molecular

architecture [6]. Among these, Class I ADHs (ADH1A, ADH1B, ADH1C) are liver-specific and central to ethanol metabolism [7], while Class II (ADH4), Class III (ADH5), and Class IV (ADH7) enzymes differ in substrate specificity and tissue distribution, highlighting the diversity within the human ADH family. Class II ADHs (ADH4) are also liver-based but are less efficient at metabolizing ethanol at low concentrations. Class III ADHs (ADH5) are found in nearly all tissues and exhibit broad substrate specificity. Class IV ADHs (ADH7), located in the stomach and esophagus, are effective at metabolizing other alcohols, such as retinol [8–10].

Gene duplication plays a key role in the evolution and diversification of ADH enzymes, as duplicated genes can evolve at different rates, a process known as type-I functional divergence. This divergence allows for the development of novel functions that enable organisms to adapt to their environments [11]. Function changes with evolution through processes like gene duplication, mutation, and selection, which allow proteins to develop new roles. For example, in the study by Carrijo de Oliveira et al., an ancestral enzyme (HI-Uase) involved in uric acid breakdown was duplicated, creating a "spare" gene copy [12]. Mutations in this duplicate allowed it to gradually lose its original enzymatic function and adopt a new role as transthyretin (TTR), a thyroid-hormone-binding protein. This transformation illustrates how structural modifications after duplication can lead to significant shifts in protein function, providing new biological capabilities over evolutionary time.

Thompson et al. explored the evolutionary functional diversification in plant ADHs and developed the first three-dimensional model of a plant ADH. Their study revealed site-specific amino acid changes in three crucial regions: the zinc-binding loop, the monomer interaction site, and the active site [13]. They also examined functional divergence in ADH enzymes across *Poaceae*, *Brassicaceae*, *Fabaceae*, and *Pinaceae* [14] identifying divergent amino acids in three critical regions: the zinc-binding loop, the monomer interaction site, and the active site. Additionally, they investigated ADH classes in animals, fungi, and plants, using evolutionary and modeling techniques to pinpoint key residues involved in different types of functional divergence between duplicated genes, providing insights into the process of ADH diversification [15].

The complex interplay between molecular evolution, structural diversity, and functional adaptation makes ADH enzymes an interesting subject for computational analysis. Techniques such as molecular dynamics simulations, integrated with network analysis, provide a powerful framework to investigate these relationships [16].

In this study, we investigate the molecular evolution (type-I functional divergence) of three ADH1 families (*Saccharomyces cerevisiae* (PDB id: 4W6Z, yeast alcohol dehydrogenase I) [17], *Gadus morhua* (PDB ID: 1CDO, cod(fish) alcohol dehydrogenase I) [18], and *Homo sapiens* (PDB ID: 1HDX, human alcohol dehydrogenase I) [19]) using network analysis and functional divergence methods. The sequence and structural alignments of these three PDB structures are shown in Supplementary Figure S1. The computational methods used in this study integrate molecular evolution techniques and structural bioinformatics to assess functional divergence [20]. Functional divergence analysis was employed to identify amino acid residues likely subjected to selective pressures post-gene duplication, potentially influencing unique functional roles among ADH1 family members [11,21]. Network analysis, focusing on centrality metrics such as betweenness, closeness, and degree, enabled the examination of residue interactions within the three-dimensional ADH1 structures. These centrality metrics reveal residues critical to structural integrity and dynamic communication pathways within the enzyme, further elucidating their roles in substrate binding or catalysis. The integration of network centrality with functional divergence scores allows for a refined understanding of the specific evolutionary constraints that shape each ADH1 enzyme's structural and functional landscape [22,23]. Computational methods help in understanding the molecular architecture and evolutionary forces driving functional differentiation in ADH1 homologs [24,25]

To investigate the structural stability and evolutionary patterns of alcohol dehydrogenase (ADH) enzymes, we employed several analytical methods. We conducted a phyloge-

netic analysis to discern the evolutionary relationships among ADH subtypes, emphasizing the conservation and divergence of key enzymes, particularly ADH1 and ADH3, across various taxa. Molecular dynamics (MD) simulations were then performed to explore the conformational dynamics and stability of ADH proteins over time. By applying Newtonian physics to model atomistic movements, these simulations provided detailed insights into structural fluctuations at high temporal resolution. Key measurements, such as root mean square deviation (RMSD), quantified the extent of structural deviations from the initial configuration, indicating protein stability. This approach enabled us to observe dynamic behaviors critical to understanding protein function and flexibility in a biological context. Additionally, we used dynamic cross-correlation maps (DCCM) to examine inter-residue dynamics, identifying correlated and anti-correlated motions that suggest functionally relevant domains. Lastly, network centrality analysis, focusing on betweenness, degree, and closeness measures, helped pinpoint residues critical for structural integrity and functional connectivity, allowing us to map essential nodes within the protein interaction network.

By leveraging these computational methods, our study aims to dissect the molecular architecture and evolutionary forces that have driven functional differentiation across ADH1 homologs, offering insights into both conserved and lineage-specific adaptations.

2. Materials and Methods

2.1. Phylogenetic Analysis

The phylogenetic tree was generated using the Diverge version 3.0 program [26,27], with the NJ (Neighbor-Joining) algorithm [28] employing the p-Distance method for distance estimation. Visualization was performed using the iTOL (Interactive Tree Of Life) program [29]. The NJ algorithm constructs a distance matrix by measuring the evolutionary distances between sequences, then iteratively combines the closest sequences into new clusters. The distances between these new clusters and the remaining sequences are recalculated, and this process is repeated until all sequences are represented in the form of a phylogenetic tree. A total of 139 ADH gene sequences were utilized to construct phylogenetic trees for each species and ADH type.

2.2. Molecular Dynamics Simulation

We performed MD simulations using the NAMD package—version 2.14 [30] with the CHARMM36 force field [31] and with protein parameters incorporating the CMAP corrections [32] for three PDB structures (1CDO, 1HDX, and 4W6Z). The TIP3P water model [33] was employed. The particle mesh Ewald (PME) method was used with a direct space cut-off of 12 Å [34]. The damping coefficient for the Langevin dynamics simulation was 5 ps−1. The Nosé–Hoover method was used to maintain constant pressure (1 atm) [35]. MD simulation was performed in the NPT ensemble at 310 K. We performed MD simulation for 100ns for each of the three PDB structures.

2.3. Network Analysis

The dynamic cross-correlation map (DCCM) method was used to analyze the correlation between residues in trajectories of MD simulations [36–38].

$$C_{ij} = \frac{<(r_i(t)-<r_i(t)>)(r_j(t)-<r_j(t)>)>}{\sqrt{(<r_i^2(t)>-<r_i(t)>^2)(<r_j^2(t)>-<r_j(t)>^2)}} \quad (1)$$

where $r_i(t)$ and $r_j(t)$ are the atomic positions of the i-th and j-th Cα atoms at time t. The quantity $r_i(t) - <r_i(t)>$ corresponds to the fluctuation of the i-th atom and $r_j(t) - <r_j(t)>$ to that of the j-th atom. A correlation map for the conformational changes for all the Cα atoms during the 100 ns MD simulation was obtained. The C_{ij} value (Equation (1)) in the DCCM is an adjacency matrix. In the constructed network, each node corresponds to a Cα atom, and each edge is an information transfer probability (i.e., cross-correlation). The weight w_{ij} of the correlation edge between the nodes i and j was defined as [39,40]

$$w_{ij} = -log|C_{ij}| \tag{2}$$

To identify and quantify the nodes that occupy critical positions in a network, several centrality measures were proposed—including the betweenness, closeness, and degree centralities [41–43]—and were used in our analysis. The closeness centrality is defined as the average length of the shortest paths between a node and all the other nodes in a network. This is a measure of how long it will take information to spread from one given node to the other nodes. The closeness centrality is calculated as

$$C_C(v_i) = \frac{n-1}{\sum_{j \neq i} g(v_i, v_j)} \tag{3}$$

where $g(v_i, v_j)$ is the shortest path with a weight between nodes i and j. n is the number of nodes in the graph. The betweenness centrality is a measure of how many information pathways flow through a node in a network. The betweenness of node i is the fraction of the shortest paths between pairs of nodes that pass through this node. The betweenness centrality is calculated as

$$b_i = \frac{\sum_{s<t} g_i^{st}/n_{st}}{\frac{1}{2}n(n-1)} \tag{4}$$

where g_i^{st} is the number of shortest paths from s to t with a weight that passes through node i, and n_{st} is the total number of shortest paths from s to t.

The degree centrality measures the number of edges incident on a node in a network and is calculated as

$$C_D(v_i) = \sum_j A_{ij} \tag{5}$$

where A_{ij} is the adjacency matrix: if $w_{ij} > 0$ then $A_{ij} = 1$, otherwise $A_{ij} = 0$.

2.4. Posterior Probability

Gene duplication is an important event of evolution for diversification. The divergence of ancestral genes by gene duplication leads to the generation of several multigene families (Figure 2a). Type-I functional divergence refers to the process in which duplicated genes evolve to have different functional divergences, often resulting in a change in the rate of evolution between the duplicates. This divergence is a key mechanism by which organisms gain new capabilities and adapt to their environment over time. A statistical method for the estimation of type-I functional divergence after gene duplication is well-defined [44]. After generating multiple alignments of amino acid sequences, the altered functional constraint between two gene clusters can be given by

$$r_\lambda = \frac{Cov(\lambda_1, \lambda_2)}{\sqrt{Var(\lambda_1)Var(\lambda_2)}} \tag{6}$$

where $Var(\lambda_1)$, $Var(\lambda_2)$, and $Cov(\lambda_1, \lambda_2)$ are the variances and covariance of λ_1 and λ_2 [45]. $\lambda 1$ and $\lambda 2$ are evolutionary rates for class 1 and class 2. For no functional divergence after gene duplication, $r_\lambda = 1$; otherwise, $r_\lambda < 1$. The measure of type-I functional divergence is defined by

$$\theta_\lambda = 1 - r_\lambda v \tag{7}$$

As functional divergence increases, θ_λ increases from 0 (functional constraint) to 1 (functional divergence). For the two-state model, all sites are categorized into two groups: functional constraint (F_0) and functional divergence (F_1) (Figure 2b,c). The prior probability for the site in the amino acid sequence is given by

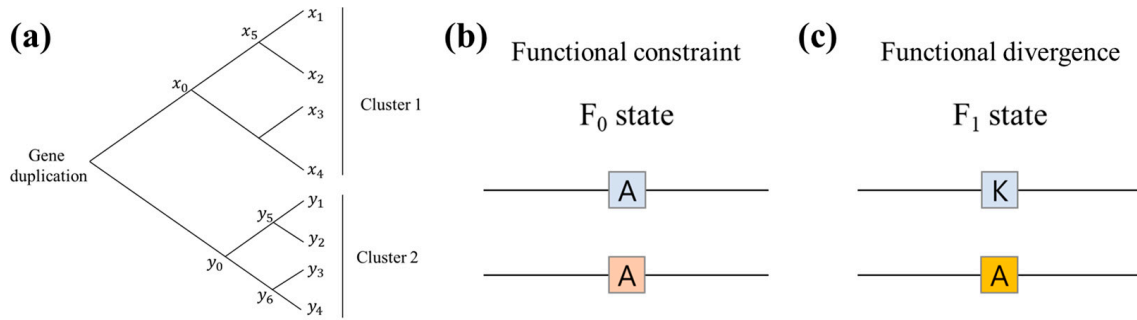$$P(F_1) = 1 - r_\lambda = \theta_\lambda, \quad P(F_0) = 1 - \theta_\lambda \tag{8}$$

Figure 2. (**a**) Emergence of gene families after gene duplication and formation of clusters. (**b**) A scheme of the F0 state (functional constraint). (**c**) A scheme of the F1 state (functional divergence).

Amino acids showing significant rate differences between two gene clusters are very important for understanding the functional and structural evidence of molecular evolution. The posterior probability of each site is a critical measure for selecting amino acids showing significant rate differences after the gene duplication event. Gu [45] estimated the coefficient of functional divergence based on the Poisson model. When the number of amino acid changes is X_i (i = 1,2 for gene cluster) at a given site, the probability of $X_i = k$ is given by

$$p_i(k) = \frac{(\lambda_i T_i)^k}{k!} e^{-\lambda_i T_i} \tag{9}$$

where T_1 and T_2 are the total evolutionary times of clusters 1 and 2, respectively [45].

By the assumption that the evolutionary rate changes among sites according to gamma distribution, the distribution of evolution rate, $\phi(\lambda)$, is defined by

$$\phi(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \tag{10}$$

where $\lambda = \lambda_1$ or λ_2.

Then, the joint distribution of the number of changes, $P(X_1 = i, X_2 = j)$, is given by

$$P(X_1 = i,\ X_2 = j\,|F_1) = Q_1\,(i)Q_2\,(j),\ P(X_1 = i,\ X_2 = j\,|F_0) = K_{12}\,(i,j) \tag{11}$$

where 1 and 2 are gene clusters. Q_1 and Q_2 are given as

$$Q_1(i) = \int_0^\infty p_1(i)\phi(\lambda_1)d\lambda_1,\ Q_2(j) = \int_0^\infty p_2(j)\phi(\lambda_2)d\lambda_2 \tag{12}$$

The posterior probability of state F_1 at the specific site with X_1 and X_2 changes in clusters 1 and 2 is given by [44]

$$P(F_1|X_1,\ X_2) = \frac{\theta_\lambda Q_1\,Q_2}{(1 - \theta_\lambda)K_{12} + \theta_\lambda Q_1\,Q_2} \tag{13}$$

Q values are calculated using Diverge [27] for the type-I functional divergence in this study.

3. Results

3.1. Phylogenetic Analysis

The generated phylogenetic tree provides a detailed overview of the evolutionary relationships among various alcohol dehydrogenase (ADH) enzymes across a wide range of species. In the fish lineage, sequences diverged into ADH1 and ADH3, while the mammalian and fungal lineages branched into ADH1, ADH2, ADH3, ADH4, and ADH5. The mammalian ADH lineage reveals a more complex structure, with ADH1 further subdivided into several subclusters: ADH1A, ADH1B (PDB ID:1HDX), ADH1C, along with

ADH2, ADH3, ADH4, and ADH5 forming a single cluster. Notably, ADH4 is positioned adjacent to ADH1, suggesting its origin from a gene duplication event of ADH1. In fungi, ADH1 (PDB ID: 4W6Z), ADH2, and ADH5 appear to cluster within the *Saccharomyces cerevisiae*, indicating a common ancestor that likely resulted from gene duplication of ADH1. Importantly, ADH2 is more closely related to ADH1 than to ADH5, implying a greater functional or structural similarity.

The ADH1 and ADH2 sequences from each species of *Lachancea* and *Kluyveromyces* were found to cluster together, whereas *Kluyveromyces* ADH3 and *Lachancea* ADH3 formed a separate cluster. This indicates that the genetic phylogenetic relationships among fungal ADH sequences differ from those of mammals, suggesting that each ADH type does not exist as a single cluster. This finding implies that although different species may share the same ADH type during the evolutionary process, they have undergone distinct genetic changes, resulting in functional differentiation and diverse characteristics.

Notably, ADH3 appears to be highly conserved, indicating its ancient origin and critical role across diverse taxa. It is present in a broad spectrum of species, from vertebrates like mammals, amphibians, and birds to invertebrates such as mollusks and nematodes. This conservation suggests that ADH3 plays a fundamental and possibly essential role in the metabolic processes of these organisms.

In contrast, ADH1 and ADH2 show significant evolutionary divergence, particularly within mammalian lineages, where multiple subtypes have emerged. This divergence suggests that these enzymes have undergone specialization to fulfill specific metabolic functions, such as alcohol metabolism, retinoid metabolism, or detoxification pathways, which are more complex in higher vertebrates. The presence of *Viridiplantae* ADH and fungi ADH as outgroups helps root the tree and provides a broader evolutionary context, allowing for a more comprehensive understanding of the evolutionary history and diversification of ADH enzymes (Figure 3). The branch lengths in the tree, which represent evolutionary distances, indicate varying levels of divergence, with longer branches reflecting greater evolutionary separation. This variation emphasizes the genetic diversity among the ADH enzymes and highlights the different evolutionary pressures that have shaped their development across species.

ADH1 is particularly significant due to its greater evolutionary divergence and specialization compared to other ADH types. While ADH3 is highly conserved across a wide range of species, ADH1 has evolved distinctively, particularly within mammals, amphibians, and birds. This evolutionary divergence suggests that ADH1 has adapted to meet specific metabolic requirements, such as alcohol metabolism, detoxification, and other specialized biochemical processes that are crucial for the survival and adaptation of these organisms. The evolution of ADH1 into different subtypes within mammals further underscores its importance, indicating a role in species-specific metabolic pathways. This makes ADH1 a critical enzyme for studying the evolutionary adaptations and functional specialization of metabolic processes across different species.

In summary, these results indicate that a single ADH type can diversify into subgroups with varying genetic mutations or structural differences rather than remaining as a singular cluster within the same classification.

3.2. RMSD Analysis

Root mean square deviation (RMSD) values are used to evaluate the structural stability and conformational changes of proteins during molecular dynamics (MD) simulations. Figure 4 shows the RMSD data for three pdb structures 1CDO, 1HDX, and 4W6Z. The thick line indicates the averages of the RMSD values. These RMSD values were calculated for a period of 100ns MD simulation.
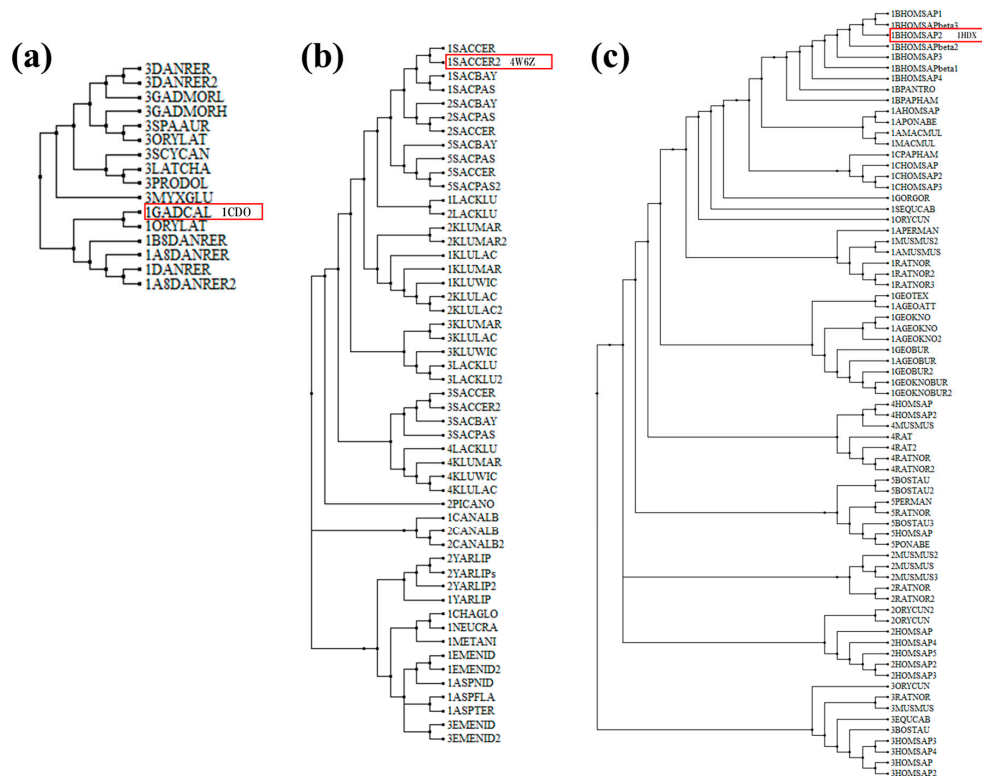
Figure 3. Phylogenetic tree of alcohol dehydrogenase (ADH) enzymes of (**a**) ADH1/ADH3, (**b**) ADH1/ADH2/ADH3/ADH4/ADH5, (**c**) ADH1/ADH2/ADH3/ADH4/ADH5. The entities in the phylogenic tree crystalized are shown in the red box.
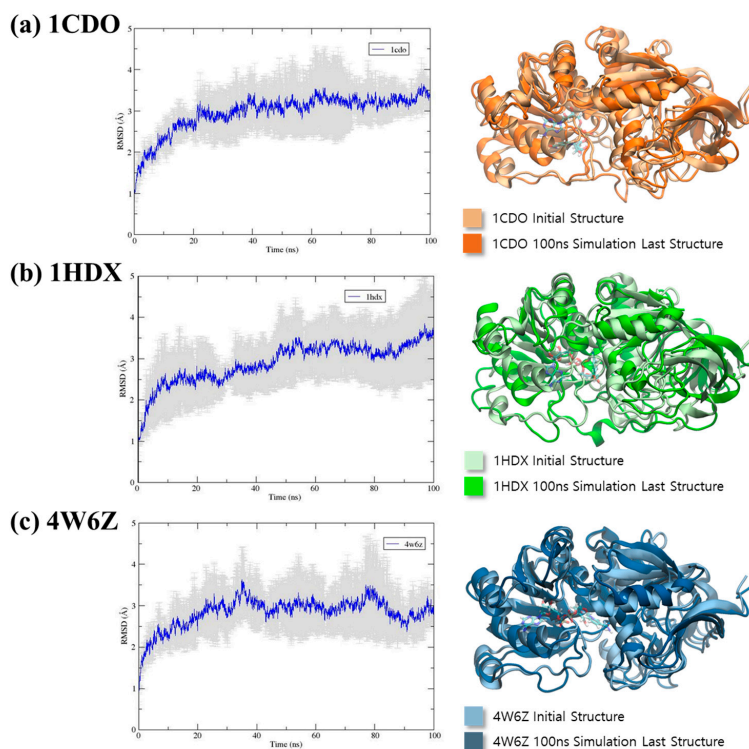

Figure 4. RMSD values and standard deviations of 100 ns MD simulations, repeated three times for three PDB structures and the alignment of the initial structures and final snapshot (100 ns) structures, respectively. (**a**) RMSD values and standard deviations of 1CDO. (**b**) RMSD values and standard deviations of 1HDX. (**c**) RMSD values and standard deviations of 4W6Z.

For 1CDO, the RMSD starts at a low value, showing that the protein initially stays close to its reference structure. During the first 20 ns, the RMSD increases quickly and then stabilizes around 3.0 Å. This indicates that 1CDO undergoes early conformational changes but eventually reaches a stable structure, with RMSD values consistently within the 5 Å range. Similarly, 1HDX shows an initial rise in RMSD, with values increasing until about 20–30 ns, where it stabilizes around 3.0 Å. However, 1HDX exhibits slightly more variability than 1CDO, suggesting that while the protein also reaches a stable conformation, it still experiences minor fluctuations during the simulation. The larger fluctuations in 1HDX's RMSD compared to 1CDO indicate more significant conformational changes before stabilization, though overall stability is comparable.

In contrast, 4W6Z shows greater fluctuations throughout the simulation, particularly between 50 and 100 ns, with RMSD values varying between 2.5 and 3.5 Å. This suggests that 4W6Z is less stable or more flexible than the other two proteins. The continuous fluctuations imply that 4W6Z does not settle into a single stable conformation but instead transitions between multiple conformations or remains dynamic throughout the simulation. The RMSD profile for 4W6Z also shows distinct phases, with an initial increase followed by periods of stability and then another increase towards the end, indicating multiple stages of conformational changes. The root mean square fluctuation (RMSF) values during 100 ns MD simulations repeated three times for three PDB structures shown in Supplementary Figure S2. We found that two loop regions for all three proteins show relatively higher fluctuation: for 1CDO (PHE 94–PRO 121 and LEU 298–GLY 311), 1HDX (GLY 98–ARG 128 and SER 298–ARG 312), and 4W6Z (SER 96–PRO112 and MET 270–ILE 288).

In summary, the MD simulations reveal that all three proteins undergo conformational changes, but 1CDO and 1HDX reach stable conformations after an initial period of adjustment. In contrast, 4W6Z remains more dynamic, suggesting greater flexibility or lower stability. These differences may be related to the proteins' functional roles, interactions with their environment, or inherent structural properties. The dynamic behavior of 4W6Z might imply a need for flexibility in its function, while the stability of 1CDO and 1HDX suggests more rigid biological roles.

3.3. Dynamic Cross-Correlation Analysis

Dynamic cross-correlation analysis was performed for the holo-proteins of the PDB structures 1CDO (*Gadus morhua*), 1HDX (*Homo sapiens*), and 4W6Z (*Saccharomyces cerevisiae*) (Figure 5a–c).

For 1CDO, the DCCM showed a complex pattern of correlations, with very sparse red patterns. This suggests that the movements of the residues within this protein are intricately linked, with only some residues moving in a highly correlated manner (as indicated by the red points), while others show anti-correlation in their movements (blue points). The DCCM for 1HDX showed prominent red areas, suggesting a strong correlation between residues. This could indicate a dynamic behavior different from that of 1CDO, with certain parts of the protein moving in a more coordinated manner, especially among residues 1–100 and near the central diagonal line. The DCCM of 4W6Z had patterns of residue correlations distinct from those in the other maps. This suggests that despite the structural similarities with the other two proteins, 4W6Z has unique dynamic properties. The red areas were more prominent in the 4W6Z structure than in the other two structures. The distinct shape of the protein might also influence these dynamic correlations. The strong correlations could be indicative of functional domains within the protein that work together and move in a coordinated fashion to achieve the protein's function.
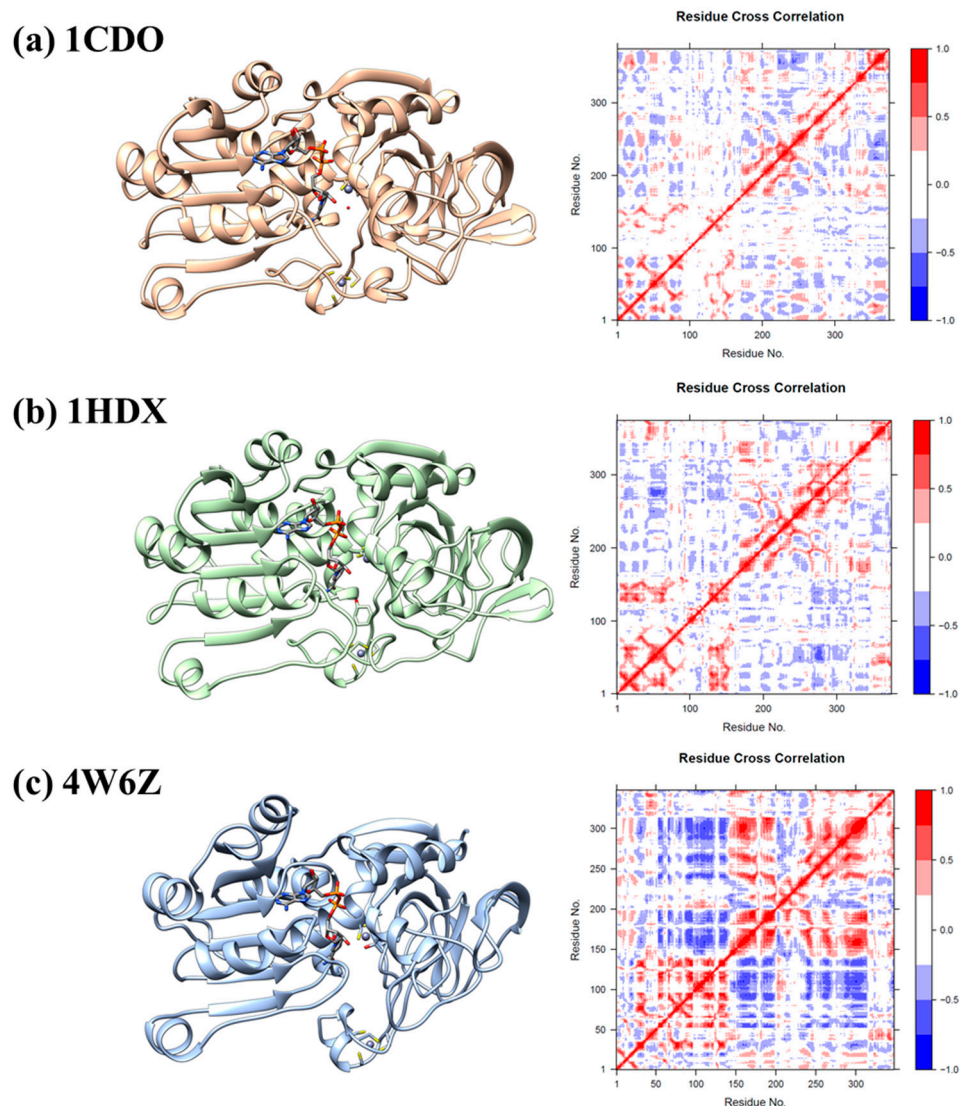
Figure 5. Dynamic cross-correlation map (DCCM) calculations throughout 100 ns molecular dynamics simulation of a protein. The maps provide information about the dynamic correlation between the Cα atoms of amino acids. Correlation between residue fluctuations: blue, negative; red, positive correlation. The color coding from blue to red indicates the degree of correlation, with red representing areas of high correlation which often implies coordinated movement or interaction between those residues. The diagonal line of the DCCM represents the correlation of each residue with itself, which is always perfect (hence the line is red). The points not on the diagonal line represent the correlations between different residues. (**a**) Dynamic cross-correlation map of 1CDO. (**b**) Dynamic cross-correlation map of 1HDX. (**c**) Dynamic cross-correlation map of 4W6Z.

3.4. Principal Component Analysis

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components [46]. The PCA plots of 1CDO, 1HDX, and 4W6Z are shown in Figure 6. For 1CDO, the first principal component (PC1) explained 28.15% of the variance, while the second principal component (PC2) explained 12.22%. For 1HDX, PC1 explained 29.56% of the variance, while PC2 explained 13.75%. For 4W6Z, PC1 explained a significant 49.68% of the variance, while PC2 explained 11.65%. The PCA results suggest a complex interplay of movements within the protein structure, with certain residues showing high correlation, implying coordinated movement or interaction.
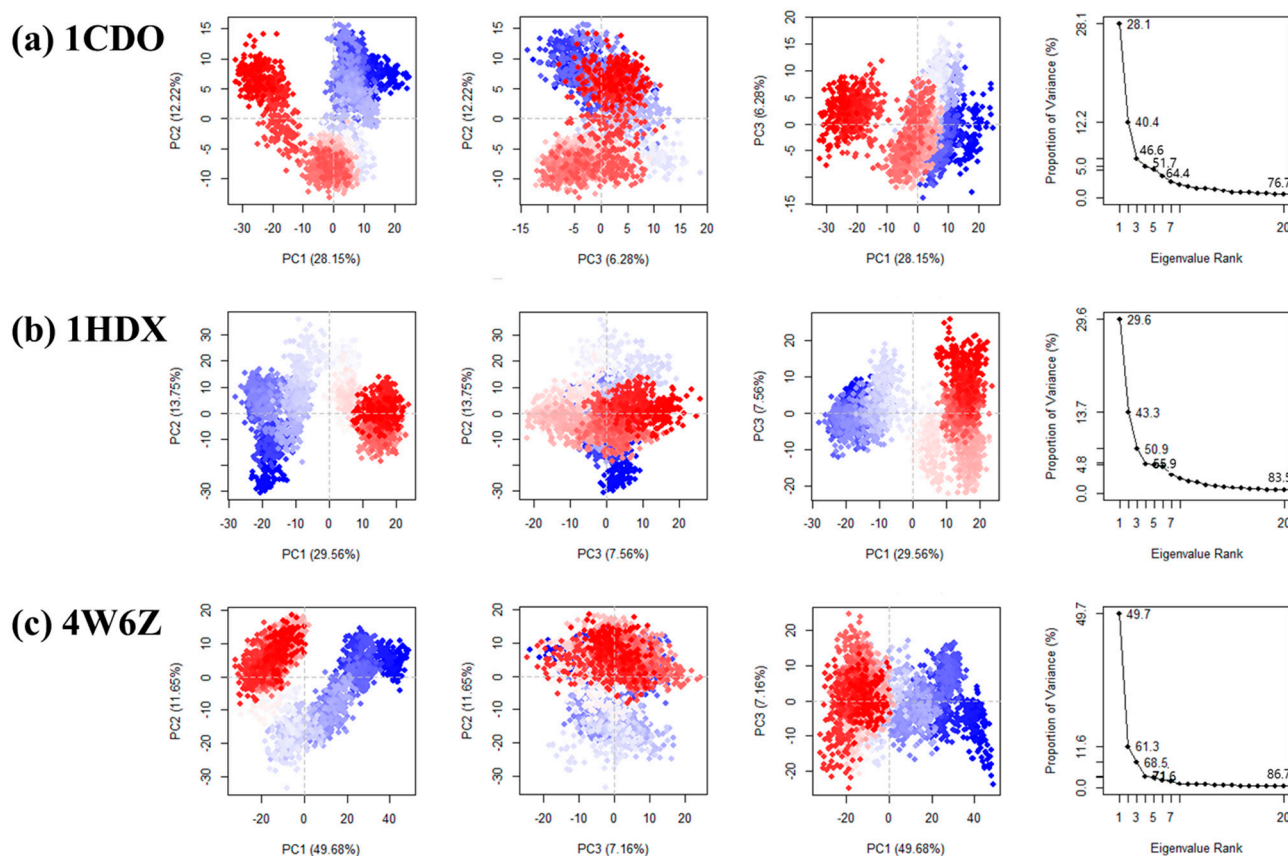
Figure 6. Principal component analysis throughout the 100 ns molecular dynamics simulation of each protein. The plots of PC2 vs. PC1 show the correlation between the first and second principal components, etc. Each point in the scatter plot represents a different conformation of the protein during the simulation. If two points are close together in the plot, it means that those two conformations are similar in terms of the movements of the residues. The percentage values given in the scatter plots represent the variance explained by each principal component. (**a**) Principal component analysis of 1CDO. (**b**) Principal component analysis of 1HDX. (**c**) Principal component analysis of 4W6Z.

When we look at how the path of movement is shown on a graph, we see a shape like half a circle or the letter U for all three proteins (Figure 6). Such a pattern could imply that the proteins are sampling a range of conformations. The half-circle or U shape might represent a conformational space where the proteins are transitioning between different conformational states. This pattern also likely suggests random diffusion within the simulation, which can be interpreted as thermal motion across a flat, gently sloping landscape of free energy [47–49]. Figure 6 also shows the resulting PC analysis scree plot, indicating the proportion of variance against its eigenvalue rank. This implies how much of the data's overall structure and variation can be captured by considering just the first few principal components. For 1CDO, the first three PCs together explained 47% of the variance. For 1HDX, they explained 50% of the variance. For 4W6Z, they explained 68% of the variance. For 4W6Z, a significant majority (68%) of the data's behavior can be understood by analyzing the first three PCs, indicating that these components are very informative. In contrast, for 1CDO and 1HDX, the first three PCs are less informative, capturing less than half of the total variance [50].

3.5. Centrality Measures

In the context of network analysis, the concept of centrality is instrumental in discerning the importance of a node (or residue) and an edge's (or interaction's) role in network

connectivity and the propagation of information within it. Two primary metrics in this study were closeness centrality and betweenness centrality. Closeness centrality serves as an estimate of the speed at which information can disseminate through a node to other nodes, taking into account the shortest paths. It essentially measures the average length of the shortest paths from a node to all other nodes in the network. Betweenness centrality is used to pinpoint nodes that lie on the most direct paths of communication, thereby controlling the flow of information. It quantifies the frequency at which a node appears on all shortest paths between two other nodes, thus highlighting its role as a "bridge" in the network. By integrating centrality measures, we can deepen our understanding of the dynamics of the network. This integrated approach aids in identifying critical elements that are integral to maintaining the network's optimal functioning and integrity.

The top 15 residues with the highest betweenness, closeness, and degree centralities are listed in Tables 1–3. ALA 207, PHE 230, LEU 206, and ASN 278 in 1CDO (Table 1); LEU 279, SER 289, GLU 267, and MET 276 in 1HDX (Table 2); and LEU 116, ASN 94, HIS 240, LEU 93, LYS 160, and TRP 92 in 4W6Z (Table 3) are the residues that have both high betweenness and closeness. The high values suggest that these residues play an important role in the protein's structural network.
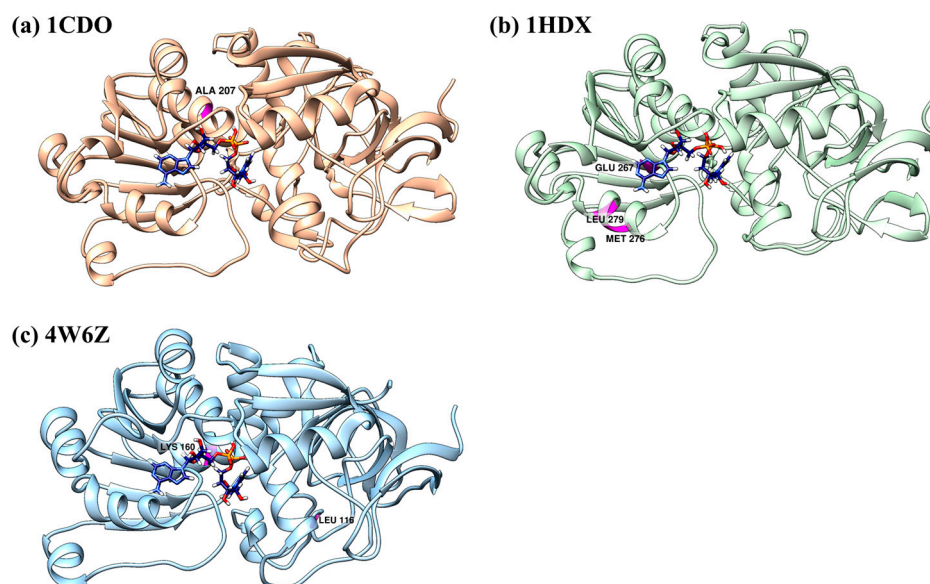
Table 1. Top 15 residues with high betweenness, closeness, and degree in 1CDO. The residues that overlap in both betweenness, closeness, and degree are shown in bold.

	Betweenness	Closeness	Degree
1	**ALA 207**	**ALA 207**	SER 76
2	PHE 230	PHE 181	PRO 250
3	LEU 206	LYS 366	LEU 361
4	CYS 104	LEU 361	ILE 251
5	ASP 337	LEU 206	VAL 77
6	LYS 285	PHE 230	PHE 181
7	THR 347	MET 362	ASN 243
8	GLU 27	MET 231	GLY 182
9	ASN 278	ASN 243	GLU 249
10	GLU 35	ASN 278	**ALA 207**
11	LEU 172	GLY 182	ILE 359
12	ARG 130	ALA 358	ALA 358
13	LEU 298	LYS 232	VAL 254
14	ILE 156	GLY 365	LYS 366
15	GLU 281	ASP 357	VAL 30

Table 2. Top 15 residues with high betweenness, closeness, and degree in 1HDX. The residues that overlap in both betweenness, closeness, and degree are shown in bold.

	Betweenness	Closeness	Degree
1	**LEU 279**	MET 275	**GLU 267**
2	SER 206	**MET 276**	THR 274
3	PHE 335	THR 274	**MET 276**
4	SER 289	**LEU 279**	ARG 129
5	ARG 37	SER 278	MET 275
6	ASN 118	ALA 277	ALA 12
7	**GLU 267**	**GLU 267**	GLY 66
8	GLY 66	LEU 280	GLY 270
9	THR 131	ASP 273	SER 278
10	ARG 129	LEU 272	**LEU 279**
11	HIS 51	ARG 271	ASP 49
12	GLY 117	PHE 266	PHE 130
13	ILE 137	SER 289	PHE 146
14	ARG 47	VAL 13	PHE 266
15	**MET 276**	CYS 282	THR 131

Table 3. Top 15 residues with high betweenness, closeness, and degree in 4W6Z. The residues that overlap in both betweenness, closeness, and degree are shown in bold.

	Betweenness	Closeness	Degree
1	ASP 132	LEU 93	ASN 297
2	**LEU 116**	**LEU 116**	THR 301
3	ASN 94	TRP 92	**LYS 160**
4	HIS 240	ASN 94	VAL 266
5	LEU 93	LYS 163	GLY 296
6	ARG 196	LEU 162	ASP 300
7	ILE 288	HIS 240	ARG 302
8	CYS 111	**LYS 160**	CYS 111
9	TYR 189	SER 117	ILE 156
10	**LYS 160**	LEU 167	TYR 294
11	VAL 266	TYR 159	ARG 298
12	GLU 67	ASP 115	ALA 161
13	ASN 110	ALA 161	ALA 299
14	TYR 195	HIS 171	TYR 159
15	TRP 92	VAL 173	**LEU 116**

Residues that have high betweenness, closeness, and degree centrality measures in a protein structure are often considered important due to their potential role in the protein's function, stability, and interactions: 207 ALA for 1CDO, 279 LEU, 267 GLU/276 MET for 1HDX, and 116 LEU /160 LYS for 4W6Z, positioned close to the ligand. They often reside in key positions that connect different parts of the protein, acting as "hubs" or "bridges" in the protein network. High-centrality residues are often found in or near the active site of enzymes or the binding site of receptors (Figure 7). They may be directly involved in the protein's function or mediate catalytic reactions with other molecules. Mutations at these positions are more likely to disrupt the protein's structure and function than other mutations [51].

(a) 1CDO

(b) 1HDX

(c) 4W6Z

Figure 7. Positions of residues with high betweenness centrality, closeness centrality, and degree centrality in the 3D structure of (**a**) 1CDO, (**b**) 1HDX, and (**c**) 4W6Z.

3.6. Type-I Functional Divergence

To assess the Type-I functional divergence in the three proteins, we analyzed the relationship between Q-values and three key centrality measures: betweenness, closeness, and degree. The maps presented (left to middle panels) display the distribution

of normalized Q-values against these centrality measures for three structures (Figure 8 (a) 1CDO, (b) 1HDX, and (c) 4W6Z). The results show a stronger correlation between closeness centrality and Q-values, as seen in the dense clustering of high-Q-value regions in the closeness maps. This suggests that residues with higher closeness centrality, which are more efficiently connected to other residues, play a key role in functional divergence. In contrast, betweenness centrality shows a sparse distribution of Q-values, indicating that residues on key paths between other residues may not contribute as strongly to divergence. The degree centrality maps also show some correlation with Q-values, but not as significantly as closeness centrality.
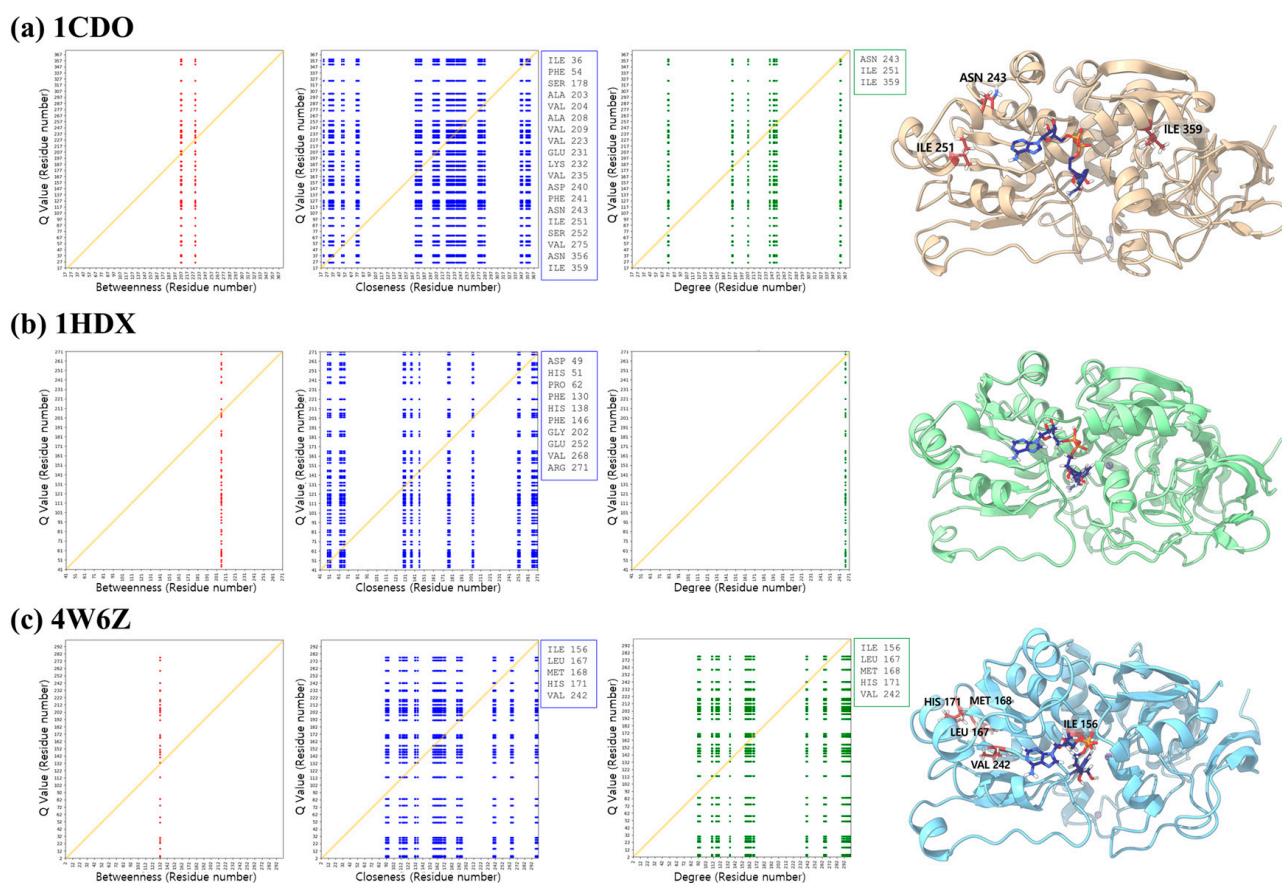
Figure 8. Correlation between Q values and centrality measures betweenness, closeness, and degree in (**a**) 1CDO, (**b**) 1HDX, and (**c**) 4W6Z. The amino acid residues on the diagonal lines (orange color) are shown in the boxes.

Several key residues that exhibit both high closeness centrality and Q-values were identified and highlighted in the 3D structural representations. For the fish ADH1/ADH3 divergence (ADH1: 1CDO), these include ILE 36, PHE 54, SER 178, ALA 203, VAL 204, ALA 208, VAL 209, VAL 223, GLU 231, LYS 232, VAL 235, ASP 240, PHE 241, ASN 243, ILE 251, SER 252, VAL 275, ASN 356, and ILE 359 (Supplementary Figure S3). In the case of fungi ADH1/ADH5 divergence (ADH1: 4W6Z), the key residues are ILE 156, LEU 167, MET 168, HIS 171, and VAL 242 (Supplementary Figure S4). For the Mammalia ADH1/ADH4 divergence (ADH1: 1HDX), important residues include ASP 49, HIS 51, PRO 62, PHE 130, HIS 138, PHE 146, GLY 202, GLU 252, VAL 268, and ARG 271 (Supplementary Figure S5). These residues are central to the protein structure and likely crucial for maintaining protein function, making them hotspots for functional divergence. This analysis suggests that closeness centrality is strongly correlated with type-I functional divergence, helping to identify evolutionary important sites within protein structures.

4. Discussion

In this study, we investigated the type-I functional divergence of the ADH1 family in three species—*Saccharomyces cerevisiae* (PDB id: 4W6Z), *Gadus morhua* (PDB id: 1CDO), and *Homo sapiens* (PDB id: 1HDX)—using a combination of phylogenetic analysis, molecular dynamics (MD) simulations, and network analysis. Through posterior probability estimation, we identified residues showing significant functional divergence, underscoring the evolutionary adaptations that differentiate these proteins across species.

Our network analysis, based on centrality measures derived from MD simulations, revealed distinct evolutionary patterns in the ADH1 proteins. Specifically, in 1CDO, residues Leu298 and Lys232 exhibited both high betweenness and closeness centrality, correlating strongly with high posterior probabilities for functional divergence. Similarly, His51 in 1HDX demonstrated a significant correlation between high centrality measures and functional divergence, emphasizing its role in catalysis, likely due to its proximity to NAD+ binding sites. These findings suggest that key catalytic and binding residues, which act as structural hubs, may undergo functional diversification during evolution, facilitating novel functional capabilities.

Interestingly, the absence of residues with both high centrality and functional divergence in 4W6Z points to a more conserved evolutionary pathway in *S. cerevisiae*. This indicates that while structural and functional divergence occurs in the ADH1 family, it may be species-specific, shaped by differing evolutionary pressures. This conserved nature in 4W6Z could imply that functional divergence in *S. cerevisiae* is limited to non-catalytic regions or that it is driven by factors not captured by centrality measures in our analysis.

The correlation observed between functional divergence and network centrality measures such as closeness highlights the evolutionary importance of structural "hub" residues. Residues central to the network tend to play pivotal roles in maintaining protein stability and function. Presumably, evolutionary changes at these sites may drive significant functional shifts. These insights provide a framework for further exploring the relationship between protein structure, functional divergence, and evolution across species.

5. Conclusions

This study offers an analysis of type-I functional divergence in the ADH1 families of *Saccharomyces cerevisiae*, *Gadus morhua*, and *Homo sapiens*. By integrating phylogenetic analysis, MD simulations, and network theory, we identified critical residues that have undergone functional divergence. The strong association between centrality measures (closeness) and functionally divergent residues, particularly in *Gadus morhua* and *Homo sapiens*, underscores the importance of structural hubs in protein evolution.

Our results suggest that evolutionary divergence at these "hub" residues has likely contributed to the diversification of ADH1 functions, especially in catalysis and NAD+ binding. The conserved nature of the *S. cerevisiae* ADH1 structure, as evidenced by the lack of functionally divergent residues despite high centrality, suggests that different evolutionary pressures may guide functional divergence in this species.

These findings not only enhance our understanding of the molecular evolution of the ADH1 family but also provide valuable insights for future studies investigating evolutionary changes in other enzyme families. The identification of key structural residues that drive functional shifts offers potential applications in enzyme engineering, drug design, and the broader study of protein evolution.

value and centrality measure for fish ADH1/ADH3 divergence (ADH1: 1CDO). (a) Closeness. (b) Key residues with matching closeness centrality and Q-value are shown; Figure S4: Correlation between Q-value and centrality measure for fungi ADH1/ADH5 divergence (ADH1: 4W6Z). (a) Closeness. (b) Key residues with matching closeness centrality and Q-value are shown; Figure S5: Correlation between Q-value and centrality measure for the Mammalia ADH1/ADH4 divergence (ADH1: 1HDX). (a) Closeness. (b) Key residues with matching closeness centrality and Q-value are shown.

References

1. Yin, S.-J.; Han, C.-L.; Lee, A.-I.; Wu, C.-W. *Enzymology and Molecular Biology of Carbonyl Metabolism 7*; Weiner, H., Maser, E., Crabb, D.W., Lindahl, R., Eds.; Springer: New York, NY, USA, 1999; p. 265.
2. Di, L.; Balesano, A.; Jordan, S.; Shi, S.M. The Role of Alcohol Dehydrogenase in Drug Metabolism: Beyond Ethanol Oxidation. *AAPS J.* **2021**, *23*, 1–21. [CrossRef] [PubMed]
3. Zakhari, S. Alcohol metabolism and epigenetics changes. *Alcohol Res. Curr. Rev.* **2013**, *35*, 6–16.
4. McQuillan, M.A.; Ranciaro, A.; Hansen, M.E.; Fan, S.; Beggs, W.; Belay, G.; Woldemeskel, D.; Tishkoff, S.A. Signatures of convergent evolution and natural selection at the alcohol dehydrogenase gene region are correlated with agriculture in ethnically diverse Africans. *Mol. Biol. Evol.* **2022**, *39*, msac183. [CrossRef] [PubMed]
5. Plapp, B.V.; Green, D.W.; Sun, H.-W.; Park, D.-H.; Kim, K. Substrate Specificity of Alcohol Dehydrogenases. In *Enzymology and Molecular Biology of Carbonyl Metabolism 4*; Weiner, H., Crabb, D.W., Flynn, T.G., Eds.; Springer: Boston, MA, USA, 1993; pp. 391–400.
6. Persson, B.; Hedlund, J.; Jornvall, H. Medium–and short-chain dehydrogenase/reductase gene and protein families: The MDR superfamily. *Cell. Mol. Life Sci.* **2008**, *65*, 3879–3894. [CrossRef]
7. Kimura, M.; Yokoyama, A.; Matsushita, S.; Higuchi, S. Enzymatic Aspects of Alcoholism-ADH and ALDH. In *Textbook of Addiction Treatment: International Perspectives*; el-Guebaly, N., Carrà, G., Galanter, M., Eds.; Springer: Milan, Italy, 2015; pp. 333–342.
8. Karabec, M.; Łyskowski, A.; Tauber, K.C.; Steinkellner, G.; Kroutil, W.; Grogan, G.; Gruber, K. Structural insights into substrate specificity and solvent tolerance in alcohol dehydrogenase ADH-'A'from *Rhodococcus ruber* DSM 44541. *Chem. Commun.* **2010**, *46*, 6314–6316. [CrossRef]
9. Liu, X.; Li, T.; Kong, D.; You, H.; Kong, F.; Tang, R. Prognostic implications of alcohol dehydrogenases in hepatocellular carcinoma. *BMC Cancer* **2020**, *20*, 1204. [CrossRef]
10. Szalai, G.; Veres, M.; Duester, G.; Lawther, R.; Lockhart, M.; Felder, M.R. Tissue expression pattern of class II and class V genes found in the Adh complex on mouse chromosome 3. *Biochem. Genet.* **2008**, *46*, 685–695. [CrossRef]
11. Nguyen Ba, A.N.; Strome, B.; Hua, J.J.; Desmond, J.; Gagnon-Arsenault, I.; Weiss, E.L.; Landry, C.R.; Moses, A.M. Detecting functional divergence after gene duplication through evolutionary changes in posttranslational regulatory sequences. *PLoS Comput. Biol* **2014**, *10*, e1003977. [CrossRef]
12. Carrijo de Oliveira, L.; Figueiredo Costa, M.A.; Gonçalves Pedersolli, N.; Heleno Batista, F.A.; Migliorini Figueira, A.C.; Salgado Ferreira, R.; Alves Pinto Nagem, R.; Alves Nahum, L.; Bleicher, L. Reenacting the Birth of a Function: Functional Divergence of HIUases and Transthyretins as Inferred by Evolutionary and Biophysical Studies. *J. Mol. Evol.* **2021**, *89*, 370–383. [CrossRef]
13. Thompson, C.E.; Salzano, F.M.; de Souza, O.N.; Freitas, L.B. Sequence and structural aspects of the functional diversification of plant alcohol dehydrogenases. *Gene* **2007**, *396*, 108–115. [CrossRef]

14. Thompson, C.E.; Fernandes, C.L.; de Souza, O.N.; de Freitas, L.B.; Salzano, F.M. Evaluation of the impact of functional diversification on Poaceae, Brassicaceae, Fabaceae, and Pinaceae alcohol dehydrogenase enzymes. *J. Mol. Model.* **2010**, *16*, 919–928. [CrossRef] [PubMed]

15. Thompson, C.E.; Freitas, L.B.; Salzano, F.M. Molecular evolution and functional divergence of alcohol dehydrogenases in animals, fungi and plants. *Genet. Mol. Biol.* **2018**, *41*, 341–354. [CrossRef] [PubMed]

16. Wollenberg Valero, K.C. Aligning functional network constraint to evolutionary outcomes. *BMC Evol. Biol.* **2020**, *20*, 58. [CrossRef] [PubMed]

17. Raj, S.B.; Ramaswamy, S.; Plapp, B.V. Yeast alcohol dehydrogenase structure and catalysis. *Biochemistry* **2014**, *53*, 5791–5803. [CrossRef]

18. Ramaswamy, S.; Ahmad, M.E.; Danielsson, O.; Jörnvall, H.; Eklund, H. Crystal structure of cod liver class I alcohol dehydrogenase: Substrate pocket and structurally variable segments. *Protein Sci.* **1996**, *5*, 663–671. [CrossRef]

19. Hurley, T.D.; Bosron, W.F.; Stone, C.L.; Amzel, L.M. Structures of three human β alcohol dehydrogenase variants: Correlations with their functional differences. *J. Mol. Biol.* **1994**, *239*, 415–429. [CrossRef]

20. Israr, J.; Alam, S.; Siddiqui, S.; Misra, S.; Singh, I.; Kumar, A. Advances in Structural Bioinformatics. In *Advances in Bioinformatics*; Singh, V., Kumar, A., Eds.; Springer: Singapore, 2024; pp. 35–70.

21. Gu, X. Functional divergence in protein (family) sequence evolution. *Genetica* **2003**, *118*, 133–141. [CrossRef]

22. Lei, T.; Liao, X.; Liang, X.; Sun, L.; Xia, M.; Xia, Y.; Zhao, T.; Chen, X.; Men, W.; Wang, Y.; et al. Functional network modules overlap and are linked to interindividual connectome differences during human brain development. *PLoS Biol.* **2024**, *22*, e3002653. [CrossRef]

23. Chiem, B.; Crevecoeur, F.; Delvenne, J.C. Structure-informed functional connectivity driven by identifiable and state-specific control regions. *Netw. Neurosci.* **2021**, *5*, 591–613. [CrossRef]

24. Zheng, X.; Hu, C.; Spooner, D.; Liu, J.; Cao, J.; Teng, Y. Molecular evolution of Adh and LEAFY and the phylogenetic utility of their introns in Pyrus (Rosaceae). *BMC Evol. Biol.* **2011**, *11*, 255. [CrossRef]

25. Fukuda, T.; Yokoyama, J.; Nakamura, T.; Song, I.J.; Ito, T.; Ochiai, T.; Kanno, A.; Kameya, T.; Maki, M. Molecular phylogeny and evolution of alcohol dehydrogenase (Adh) genes in legumes. *BMC Plant Biol.* **2005**, *5*, 6. [CrossRef] [PubMed]

26. Gu, X.; Vander Velden, K. Diverge: Phylogeny-based analysis for functional–structural divergence of a protein family. *Bioinformatics* **2002**, *18*, 500–501. [CrossRef] [PubMed]

27. Gu, X.; Zou, Y.; Su, Z.; Huang, W.; Zhou, Z.; Arendsee, Z.; Zeng, Y. An update of DIVERGE software for functional divergence analysis of protein family. *Mol. Biol. Evol.* **2013**, *30*, 1713–1719. [CrossRef] [PubMed]

28. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425.

29. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]

30. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [CrossRef]

31. MacKerell Jr, A.D.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolym. Orig. Res. Biomol.* **2000**, *56*, 257–265. [CrossRef]

32. Buck, M.; Bouguet-Bonnet, S.; Pastor, R.W.; MacKerell, A.D. Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. *Biophys. J.* **2006**, *90*, L36–L38. [CrossRef]

33. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]

34. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]

35. Feller, S.E.; Zhang, Y.; Pastor, R.W.; Brooks, B.R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613–4621. [CrossRef]

36. Hünenberger, P.; Mark, A.; Van Gunsteren, W. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J. Mol. Biol.* **1995**, *252*, 492–503. [CrossRef] [PubMed]

37. Ichiye, T.; Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins Struct. Funct. Bioinform.* **1991**, *11*, 205–217. [CrossRef] [PubMed]

38. Kasahara, K.; Fukuda, I.; Nakamura, H. A novel approach of dynamic cross correlation analysis on molecular dynamics simulations and its application to Ets1 dimer–DNA complex. *PLoS ONE* **2014**, *9*, e112419. [CrossRef] [PubMed]

39. Sethi, A.; Eargle, J.; Black, A.A.; Luthey-Schulten, Z. Dynamical networks in tRNA: Protein complexes. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 6620–6625. [CrossRef]

40. VanWart, A.T.; Eargle, J.; Luthey-Schulten, Z.; Amaro, R.E. Exploring residue component contributions to dynamical network models of allostery. *J. Chem. Theory Comput.* **2012**, *8*, 2949–2961. [CrossRef]

41. Newman, M.E. A measure of betweenness centrality based on random walks. *Soc. Netw.* **2005**, *27*, 39–54. [CrossRef]

42. Freeman, L.C. Centrality in social networks: Conceptual clarification. In *Social Network: Critical Concepts in Sociology*; Routledge: London, UK, 2002; Volume 1, pp. 238–263.

43. Mason, O.; Verwoerd, M. Graph theory and networks in biology. *IET Syst. Biol.* **2007**, *1*, 89–119. [CrossRef]

44. Gu, X. Maximum-likelihood approach for gene family evolution under functional divergence. *Mol. Biol. Evol.* **2001**, *18*, 453–464. [CrossRef]
45. Gu, X. Statistical methods for testing functional divergence after gene duplication. *Mol. Biol. Evol.* **1999**, *16*, 1664–1674. [CrossRef]
46. Kitao, A. Principal component analysis and related methods for investigating the dynamics of biological macromolecules. *J* **2022**, *5*, 298–317. [CrossRef]
47. Hess, B. Similarities between principal components of protein dynamics and random diffusion. *Phys. Rev. E* **2000**, *62*, 8438. [CrossRef] [PubMed]
48. Hess, B. Convergence of sampling in protein simulations. *Phys. Rev. E* **2002**, *65*, 031910. [CrossRef] [PubMed]
49. Wolf, A.; Kirschner, K.N. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11· 23S subdomain. *J. Mol. Model.* **2013**, *19*, 539–549. [CrossRef] [PubMed]
50. Cozzolino, D.; Power, A.; Chapman, J. Interpreting and reporting principal component analysis in food science analysis and beyond. *Food Anal. Methods* **2019**, *12*, 2469–2473. [CrossRef]
51. Saxena, A.; Iyengar, S. Centrality measures in complex networks: A survey. *arXiv* **2020**, arXiv:2011.07190.