Journal of Structural Biology 218 (2026) 108341

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/jsb

Research Article

Architecture and dynamics of a supramolecular oxygen transport system in human homogentisate 1,2-Dioxygenase

Alfonso Trezza [a,1], Rebecca Finetti [a,1], Bianca Roncaglia [a,*], Francesco Giuliani [a], Francesca Furlan [b], Annalisa Santucci [a,c,d]

[a] Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Via Aldo Moro, 2, Siena, 53100, Italy
[b] Centro clinico di riferimento regionale per lo screening metabolico allargato, SC Pediatria Pneumoinfettivologia, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Via Commenda 9, Milano, 20122, Italy
[c] Industry 4.0 Competence Center ARTES 4.0, Viale Rinaldo Piaggio, Pontedera, 56025, Italy
[d] MetabERN, Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Via Aldo Moro, 2, Siena, 53100, Italy

ARTICLE INFO

Edited by Rajan Sankaranarayanan

Keywords:
Classical and steered molecular dynamics simulations
Protein tunnel analysis
Oxygen diffusion pathways
Homogentisate 1,2-dioxygenase
Alkaptonuria
Multiple sequence alignment
Molecular modeling

ABSTRACT

Alkaptonuria (AKU) is an ultra-rare inherited metabolic disorder caused by impaired activity of homogentisate 1,2-dioxygenase (HGD), a Fe(II)-dependent enzyme that catalyzes the oxidative cleavage of homogentisic acid in the tyrosine degradation pathway. Although high-resolution structures of human HGD have been solved, a fundamental mechanistic question has remained unresolved: how molecular oxygen reaches the deeply buried catalytic iron required for catalysis. Here, we identify a previously unreported AKU-associated HGD variant, c.925G>A (p.G309R), and use it as a mechanistic perturbation reference to dissect the structural determinants of oxygen access. By integrating replicated classical and steered molecular dynamics simulations with transient pocket detection, tunnel mapping, O_2 spatial-occupancy analysis, residue-level tunnel composition, and PCA/tICA-based dynamic validation, we identify a structurally accessible and dynamically supported O_2-translocation architecture connecting the central pore of the hexameric enzyme to the non-heme Fe(II) active sites. This pathway is not intrinsic to a single subunit but emerges from a cooperative arrangement of residues contributed by three protomers, generating six symmetry-related O_2-access routes per hexamer. The G309R substitution perturbs the architecture and continuity of this tunnel system, providing a mechanistic explanation for enzyme dysfunction without evidence of active-site structural perturbation or global destabilization. Together, our findings support oxygen-tunnel integrity as a previously unrecognized mechanistic requirement for human HGD activity and introduce disruption of oxygen trafficking as an additional pathogenic mode in AKU, with implications for structure-guided variant interpretation and precision-medicine strategies.

1. Introduction

Homogentisate 1,2-dioxygenase (HGD) is a non-heme Fe(II)-dependent dioxygenase that catalyzes the oxidative cleavage of homogentisic acid (HGA) into maleylacetoacetic acid, thereby linking tyrosine degradation to central metabolic pathways such as the tricarboxylic acid cycle and ketone body catabolism (Titus et al., 2000). The reaction requires molecular oxygen as a co-substrate: the Fe^{2+} ion in the active site binds O_2 and promotes its conversion into a reactive species that enables incorporation of both oxygen atoms into the aromatic ring of HGA (Amaya et al., 2004). Loss or severe reduction of HGD activity leads to intracellular accumulation of HGA, which undergoes spontaneous oxidation and polymerization in the presence of oxygen to form a melanin-like pigment (Braconi et al., 2018; Milella et al., 2024). Progressive pigment deposition in collagen-rich tissues causes ochronosis (Bernardini et al., 2024; Braconi et al., 2015), the hallmark of alkaptonuria (AKU), and contributes to cartilage rigidity and brittleness (Bernardini et al., 2024; Braconi et al., 2015). HGD therefore operates at an oxygen-dependent interface between physiological HGA clearance and harmful non-enzymatic oxidation that drives tissue damage. HGD is a 445-amino acid polypeptide encoded by the HGD gene on chromosome 3q13.33, comprising 14 exons (Fernández-Cañón et al., 1996). Six protomers assemble into a functional homohexamer arranged as a dimer of trimers, with each monomer adopting a β-sandwich fold composed of an N-terminal (280 residues) and a C-terminal (140 residues) domain (Titus et al., 2000). The catalytic

* Corresponding author.
E-mail address: bianca.roncaglia@unisi.it (B. Roncaglia).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jsb.2026.108341
Received 10 February 2026; Received in revised form 22 May 2026; Accepted 3 June 2026
Available online 11 June 2026
1047-8477/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

site is located at the interface between adjacent subunits, underscoring the importance of quaternary organization for enzymatic activity (Borowski et al., 2005). This hexameric architecture confers low tolerance to amino-acid substitutions: even alterations that preserve the monomer fold can destabilize the assembly and compromise function (Titus et al., 2000). Consistently, extensive efforts have cataloged HGD variants; to date, 259 unique entries are reported in the HGD mutation database, including 37 classified as non-pathogenic. These variants are commonly grouped into three major functional categories: (i) active-site alterations, (ii) disruption of hexamer assembly, and (iii) destabilization of individual protomers (Zatkova et al., 2012). Despite differences in residual catalytic efficiency, clinical manifestations of AKU are broadly comparable among patients, and clear genotype-phenotype correlations remain limited. This is supported by studies of siblings carrying identical variants, who show similar HGA excretion profiles but variable symptom severity and disease progression (Zatkova et al., 2022). From a genetic standpoint, AKU patients carry either homozygous or compound heterozygous HGD mutations (Vilboux et al., 2009). In compound heterozygosity, the coexistence of two allelic forms could, in principle, generate hetero-oligomeric assemblies with unpredictable functional consequences, although mixed hexamers have not yet been demonstrated experimentally. In some cases, only a single detectable HGD variant is identified, likely reflecting undetected mutations in non-coding or regulatory regions or somatic mosaicism (Usher et al., 2014). Here, we report an AKU-associated HGD variant, c.925G>A (p.Gly309Arg), and use it as an entry point for a focused structural investigation of its local environment. Initial structural analyses showed no overt active-site alteration, protomer destabilization, or hexamer disruption caused by the substitution, leading us to hypothesize an additional mechanism of HGD dysfunction involving impaired O_2 access to the non-heme Fe(II) catalytic site. Building on this rationale, we define the molecular architecture of an O_2-access tunnel connecting the central pore of the hexamer to the active site and examine how G309R perturbs tunnel continuity. These observations expand the current framework of mutational effects in HGD by introducing disruption of oxygen trafficking as an additional, structurally grounded mechanism distinct from active-site impairment, protomer destabilization, or hexamer disruption.

2. Materials & methods

2.1. Genetic analysis and variant identification

The genetic screening revealed a heterozygous c.925G>A transition in exon 12 of the *HGD* gene, resulting in a missense p.Gly309Arg (p.G309R) substitution. This specific variant was confirmed absent from major population and disease-specific reference repositories, including gnomAD and the HGD mutation database (https://hgddatabase.cvtisr.sk). Given that AKU is an autosomal recessive disorder, this single heterozygous finding does not constitute a definitive genetic diagnosis for the patient; rather, it serves as the foundational entry point for investigating the mechanistic impact of a novel, likely pathogenic structural variant.

2.2. Structures retrieval, system preparation, and classical molecular dynamics (MD)

The canonical amino acid sequence of human HGD was retrieved from the UniProtKB/Swiss-Prot database (accession code: Q93099) (UniProt Consortium, 2025), and the p.G309R substitution was introduced *in silico*. To establish a reliable structural platform, both the available crystallographic structure (PDB ID: 1EYB) (Titus et al., 2000) and predictive models were critically evaluated. Due to incomplete sequence coverage, missing side-chain densities, and non-native sequence deviations inherent to the 1EYB template, complete full-length hexameric models for both the wild-type (WT) and G309R HGD systems were generated using AlphaFold 3 (v3.0.1) (Abramson et al., 2024), explicitly modeling the coordination of the catalytic Fe^{2+} ion within each of the six symmetry-related active sites.

Model reliability was validated using the predicted Local Distance Difference Test (pLDDT), yielding a global score of 96.75 for the WT and G309R assembly. Structural divergence from the experimental fold was quantified by C_α superposition onto 1EYB over the commonly resolved core, resulting in a sub-ångström root-mean-square deviation (RMSD) (C_α-RMSD < 1 Å). Geometry and stereochemical quality were confirmed using PROCHECK (Laskowski et al., 1993) and Ramachandran analysis (Ramachandran et al., 1963), showing excellent backbone dihedral distributions, in detail $R_\varphi = 0.964$; $R_\psi = 0.975$ for WT and $R_\varphi = 0.964$; $R_\psi = 0.975$ for G309R (Figure S1).

All molecular mechanics and MD simulations were carried out using the GROMACS 2019.3 suite (Abraham et al., 2015). System setups, including optimal protonation states at pH 7.4, topology generation, and initial solvation configurations, were managed via the CHARMM-GUI interface (Jo et al., 2008) using the CHARMM36-mar2019 all-atom force field (Huang et al., 2017). The active site iron atoms were treated using the standard non-bonded Lennard-Jones/Coulomb model optimized for coordination environments in CHARMM. Each hexameric system was centered within a cubic box, keeping a minimum distance of 1.2 nm to the boundaries, and solvated with explicit TIP3P water molecules. Electroneutrality was achieved by adding 6 K^+ counterions to the WT system and 7 K^+ counterions to the mutant system, resulting in a total size of approximately 258,000 atoms per box. Periodic boundary conditions (PBC) were applied along all spatial directions.

Energy minimization was performed using a steepest-descent algorithm for 10,000 steps or until the maximum force fell below 100 kJ mol^{-1} nm^{-1}. During minimization, position restraints were enforced on the protein backbone and side-chain heavy atoms with force constants of 400.0 kJ mol^{-1} nm^{-2} and 40.0 kJ mol^{-1} nm^{-2}, respectively. Long-range electrostatic interactions were resolved using the Particle Mesh Ewald (PME) method with a real-space cutoff of 1.2 nm. Van der Waals forces were calculated using a force-switching function spanning 1.0 to 1.2 nm. All covalent bonds involving hydrogen atoms were constrained via the LINCS algorithm (Hess et al., 1997), allowing an integration timestep of 2 fs. Thermal equilibration was conducted for 1 ns in the NVT ensemble at 300 K using the Nosé–Hoover thermostat (Nosé, 1984; Hoover, 1985) ($\tau_t = 1$ ps). Subsequently, pressure equilibration was performed for 5 ns in the NPT ensemble at 1 bar utilizing an isotropic Parrinello–Rahman barostat (Parrinello and Rahman, 1981) ($\tau_p = 2$ ps). Production runs were performed in independent triplicates for both systems for a duration of 500 ns per replica, aggregating 3 μs of unbiased cMD sampling. Trajectory coordinates were archived every 1 ns, while energy parameters were written every 10 ps. Structural inspections and rendering were done via PyMOL 3.3 (Schrödinger, 2020).

2.3. Trajectory analyses: stability, cavity detection, and free-energy landscapes

Conformational stability and local structural variations were assessed independently across all six 500 ns trajectories. Global structural deviations and structural compactness were tracked via backbone RMSD and radius of gyration (R_g) using the `gmx rms` and `gmx gyrate` modules in GROMACS.

Dynamic cavity detection was executed via CAVER 3.0 (Chovancova et al., 2012) to map transient channels capable of accommodating small gas molecules. For each of the three replicas of both WT and mutant setups, snapshots were extracted at 2 ns intervals (251 frames per trajectory). The coordinate of the catalytic Fe^{2+} ion within each protomer was designated as the tunnel origin point. Tunnel searches were conducted using a minimum probe radius of 0.9 Å, a shell radius of 3 Å, a shell depth of 4 Å, and a clustering threshold of 3.5 Å.

The multi-protomer tunnel data were post-processed using a custom Python script to filter pathways based on three interconnected structural criteria: (i) spatial and geometric equivalence across all six protomers, (ii) convergence of the outer tunnel mouth toward the shared solvent-exposed central pore of the HGD hexamer, and (iii) selective closure or physical obstruction of the path in the presence of the p.G309R mutation. Cross-protomer and cross-replica tunnel alignment was validated using an inter-protomer centerline spatial deviation threshold of ≤2.5 Å, which matches the physical dimensions of an O_2 transport pathway while preventing the artificial mixing of distinct topological cavities.

To map the baseline equilibrium conformations of the tunnel boundaries prior to biased pulling, Principal Component Analysis (PCA) was performed on the Cartesian coordinates of the tunnel-lining residues using the MDTraj library (McGibbon et al., 2015), following structural least-squares fitting against the core backbone. Free-energy landscapes (FEL) were subsequently built by projecting the tracking coordinates onto the primary essential subspaces (PC1/PC2), enabling the identification of stable low-energy conformational basins.

2.3.1. β-hairpin RMSD, spatial-displacement, and tunnel disruption analysis

The structural and positional rearrangements of the β-hairpin loop containing the mutation site (spanning residues Thr302–Ile312 in chain-A indexing) were evaluated via chain-wise C_α-RMSD and three-dimensional spatial density profiling. C_α-RMSD profiles for both WT and mutant trajectories were calculated against the initial WT energy-minimized structure using `gmx rms` to directly monitor deviations from the native, open-tunnel architecture.

2.3.2. Residue-level tunnel-composition analysis

The complete residue-level architecture of the pathways was mapped using the full ensemble of 18 Steered Molecular Dynamics (SMD) trajectories. Distance-based contact analysis was executed via a custom Python workflow utilizing the MDAnalysis framework (Michaud-Agrawal et al., 2011) together with NumPy and pandas.

Atomic distances between the traveling O_2 molecule (`resname O2`) and the surrounding protein residues (`protein`) were tracked frame by frame across each translocation event using the `capped_distance` utility. Intermolecular contacts were evaluated using a strict first-shell interaction cutoff of 3.5 Å. The resulting raw data were collapsed into single-residue profiles to calculate descriptive parameters, including total contact frame counts, interaction frequencies, minimum distances, and contact persistence metrics across replicas. Extended chemical and steric boundaries of the gas channels were mapped by running identical distance scripts at expanded cutoffs of 5.0 Å and 10.0 Å.

Consensus residue lists were derived based on strict (3/3 replicas shared among all tunnels) persistence scores. Channel presence–absence matrices and contact frequency heatmaps were plotted using Matplotlib. Supramolecular channels were indexed according to their inter-subunit composition: Tunnel A (chains A–C–E), Tunnel B (chains B–A–F), Tunnel C (chains C–B–D), Tunnel D (chains D–E–C), Tunnel E (chains E–F–A), and Tunnel F (chains F–D–B).

2.4. Steered molecular dynamics (SMD)

To verify the physical feasibility and structural accessibility of the six symmetry-related channels, constant-velocity SMD simulations (Israelewitz et al., 2001) were deployed as an active structural probe. These simulations were designed to test whether an O_2 molecule could successfully traverse the internal pathways without triggering unphysical distortions of the surrounding hexameric fold, and were not intended to estimate standard equilibrium permeation rates or absolute binding free energies.

A neutral O_2 molecule, parameterized using standard CHARMM compatible parameters, was explicitly placed at the center-of-mass of the solvent-filled inner pore of the HGD hexamer, where all six CAVER tunnels converge. Following placement, the system underwent brief energy minimization and a dual NVT/NPT equilibration step (1 ns each) keeping the gas molecule restrained to its initial position.

During production SMD, the O_2 center of mass was pulled toward the target catalytic Fe^{2+} ion along a one-dimensional reaction coordinate defined by the inter-atomic distance vector. For each of the six available channels, three independent SMD replicas were executed, generating 18 distinct translocation paths. Pulling was driven by a harmonic spring potential with a force constant of 2000 kJ mol^{-1} nm^{-2} moving at a constant rate of 0.01 nm ps^{-1} over a 500 ps window. No positional restraints were applied to the protein atoms during pulling, allowing full conformational flexibility of the channel walls. Pulling forces and coordinates were sampled every 0.1 ps.

2.5. Post-SMD analyses

2.5.1. O_2 spatial occupancy and pathway reproducibility across SMD replicas

The instantaneous linear distance between the migrating O_2 molecule and the target catalytic iron center was monitored across all trajectories using the `gmx pairdist` utility in GROMACS 2019.3. Volumetric occupancy maps for the traveling gas molecule were constructed via the `gmx spatial` tool. Prior to calculation, all frames were aligned against an invariant backbone selection to remove system drift.

Three-dimensional density matrices were processed using custom Python utilities based on SciPy and NumPy to extract spatial centerlines representing the highest-probability paths for each independent replica. Overlap checks among the three replica-specific centerlines were used to confirm path reproducibility. The 3D density fields were projected onto orthogonal planes (XY, XZ, and YZ matrices) to check path continuity. Longitudinal occupancy profiles were calculated along the pulling axis to locate and define stable intermediate landing sites, entry gates, and exit points along the channel.

2.6. Conformational and dynamic analysis via PCA and tICA

To evaluate whether the forced O_2 translocation was coupled with dominant geometric variances and slow, time-correlated kinetic modes, Principal Component Analysis (PCA) and Time-Lagged Independent Component Analysis (tICA) (Pérez-Hernández et al., 2013) were performed on the SMD-derived O_2 trajectories.

Prior to any dimensionality reduction analysis, raw trajectories were systematically processed to remove periodic boundary condition (PBC) artifacts by unwrapping the molecular coordinates. To isolate the local displacement of the gaseous co-substrate from global macroscopic movements, the hexameric structure was iteratively aligned to the protein backbone C_α atoms of the wild-type (WT) reference framework using a least-squares fitting procedure. PCA and tICA were subsequently executed on the un-rototranslated center-of-mass coordinates of the O_2 molecule.

Geometric dimension reduction was conducted via PCA using the *scikit-learn* package to project the coordinates onto the first two principal components (n_components = 2), thereby capturing the largest orthogonal spatial variances of the translocation pathways. Concurrently, microscopic kinetic modes were resolved via tICA using the *deeptime* library. To account for the accelerated transport dynamics induced by the external steering bias while retaining a physically meaningful description of the slowest time-correlated transitions, a lag time of $\tau = 5$ ps (corresponding to 5 simulation frames) was selected, retaining the first two independent components (dim = 2).

Two-dimensional free-energy landscapes (FELs) were reconstructed from the projected probability distributions along the PC1/PC2, TIC1/

TIC2, and PC1/TIC1 subspaces. The underlying probability densities were discretized using a uniform grid resolution of 40 bins per dimension. To prevent statistical grid artifacts, histograms were smoothed using a two-dimensional Gaussian filter with a uniform width of $\sigma = 1.0$. The smoothed distributions were normalized to the maximum probability density (P_{max}) and converted into relative free-energy surfaces (ΔG). The global energy minimum of each individual landscape was definitionally anchored to zero.

2.7. Evolutionary conservation and variant mapping

Homology searches and sequence sampling were performed via BLASTp (Altschul et al., 1990) using the human HGD canonical sequence as the query against the non-redundant (nr) protein database. To eliminate self-conservation bias, all hits matching the *Homo sapiens* taxon were excluded from the dataset. The alignment search was configured with a word size of 2, retrieving a maximum of 5000 homologous sequences.

Multiple Sequence Alignment (MSA) matrices were compiled and evaluated using the Skylign 1.8.2 tool. Positional conservation weights and evolutionary constraints were quantified using Shannon entropy scores, where values near 0 (100%) reflect total sequence conservation, while, values approaching 100 (0%) represent highly variable hypermutable loci. Experimentally validated pathogenic missense mutations linked to AKU were harvested from the HGD mutation database and structurally mapped onto the verified tunnel regions.

3. Results

3.1. Global conformational behavior of WT and G309R HGD

To evaluate the global conformational stability of WT and G309R HGD, backbone RMSD and total R_g were monitored over 500 ns for the three independent replicas of each system (Fig. 1). Backbone RMSD showed a comparable overall trend in WT and G309R. In both systems, RMSD increased during the initial phase of the simulation and then reached a stable plateau. WT replicas stabilized mainly around 4.8–5.2 Å after approximately 150–200 ns, whereas G309R replicas showed a similar stabilization profile, with average values generally around 5.2–5.6 Å. Individual trajectories displayed moderate replica-dependent fluctuations, but no progressive structural drift was observed after equilibration. The mean RMSD profile of G309R remained slightly higher than that of WT for most of the simulation, although the two systems showed broadly similar conformational behavior. The total R_g also showed comparable behavior between WT and G309R. Both systems displayed a gradual increase during the early simulation phase, followed by stabilization. WT R_g values increased from approximately 31.4 Å to 32.7 Å, whereas G309R reached values around 32.8–33.1 Å. The mutant system showed a slightly higher mean R_g than WT, but the overall temporal profiles and final values remained close between the two systems. Overall, the replica-based RMSD and R_g analyses indicate that WT and G309R HGD maintain comparable global structural stability and compactness over the 500 ns simulations.

3.2. Composition, architecture, and dynamics of O_2 tunnel conformational states

As previous results showed no evidence of progressive unfolding, global destabilization, or loss of hexameric compactness in either system, it suggests a potential involvement of the mutation in the alteration of HGD biological function rather than in the perturbation of its 3D structure. Such evidence led us to hypothesize that G309R could influence the O_2 transport mechanism toward the catalytic site; thus, further analyses were performed. CAVER analysis of replicated WT trajectories identified a set of six recurrent, symmetry-related cavities that connect the solvent-accessible central pore of the HGD hexamer to its six catalytic iron (Fe^{2+}) centers (Fig. 2).

Considering the unbiased WT HGD MD trajectories, we detected all residues within a radius of 10 Å from the center of mass of each tunnel probe identified by CAVER, the tunnel-associated residues were: PHE49, THR50, CYS51, PRO52, GLU209, LEU210, PRO211, ASP212, LEU213, GLY214, PRO215, ILE216, GLY217, PHE227, TYR275, PRO295, SER296, ILE297, PHE298, THR299, VAL300, LEU301, THR302, LYS304, SER305, GLY309 (chain A-numbering). We noted that the detected cavities were not confined to individual protomers, but were positioned at trimeric inter-subunit interfaces. The six pathways were assigned to the contributing chains as follows: tunnel A, chains A–C–E; tunnel B, chains B–A–F; tunnel C, chains C–B–D; tunnel D, chains D–E–C; tunnel E, chains E–F–A; and tunnel F, chains F–D–B. This arrangement was reproduced at the catalytic sites related to symmetry, defining a supramolecular tunnel organization centered on the hexameric pore.

To evaluate whether the CAVER-identified tunnels corresponded to recurrent conformational states of the WT enzyme, PCA and free-energy landscape analyses were performed on the tunnel-associated residues. For each of the six WT tunnels, the trajectory was projected onto the first two principal components, and the corresponding free-energy landscape was reconstructed (Fig. 3).

The six tunnels showed defined low-energy regions in the PC1/PC2 space, indicating that the tunnel-associated residues sampled recurrent conformational states rather than isolated geometries. In each pathway, the time-resolved projection showed a structured progression across the conformational landscape, with the CAVER-defined tunnel conformations localized within discrete regions of the PCA/FEL space. The PCA/FEL analysis therefore supported the presence of organized tunnel-associated conformational states in WT HGD before SMD-based O_2 translocation analysis.

3.3. Localization of G309 and tunnel-associated β-hairpin, structural stability, and tunnel disruption

Next, we investigated the structural factors underlying the absence of WT-like tunnels in the G309R mutant. Backbone RMSD and R_g analyses indicate that WT and G309R HGD maintain comparable global structural stability and compactness, suggesting a potential role of the mutation in selectively disrupting the tunnel architecture without affecting the overall fold. GLY309 is located at the center of a short loop region connecting two antiparallel β-sheets and represents the pivot of a β-hairpin (THR302–ILE312 segment) that contributes directly to the tunnel-associated boundary. Structural localization of this segment confirmed that residue 309 is positioned within the tunnel environment rather than outside the identified pathway. The effect of the G309R substitution on the tunnel-associated β-hairpin was quantified by chain-wise Cα-RMSD analysis. The analysis was performed separately for each of the six β-hairpins, and for each replica, to avoid averaging symmetry-related local rearrangements across the hexameric assembly. RMSD analyses were conducted following three complementary strategies designed to capture distinct layers of structural information: (i) structural stability of WT β-hairpins relative to their own initial reference; (ii) structural stability of G309R β-hairpins relative to their own initial reference; and (iii) structural displacement of G309R β-hairpins relative to the WT reference structure. WT and G309R β-hairpin Cα-RMSD profiles exhibited distinct structural stability trends over the course of the MD trajectories. Across all chains, the WT β-hairpins displayed RMSD values similar G309R counterparts. In detail, the β-hairpin self-reference Cα-RMSD profiles of WT and G309R were evaluated chain-by-chain over the 500 ns trajectories, and the resulting values are reported as mean ± SD across the three replica-averaged values. In chain A, the WT β-hairpin showed a mean self-reference Cα-RMSD of 0.208 ± 0.088 nm, compared with 0.322 ± 0.024 nm for G309R. In chain B, WT and G309R exhibited nearly identical mean values, with 0.287 ± 0.057 nm

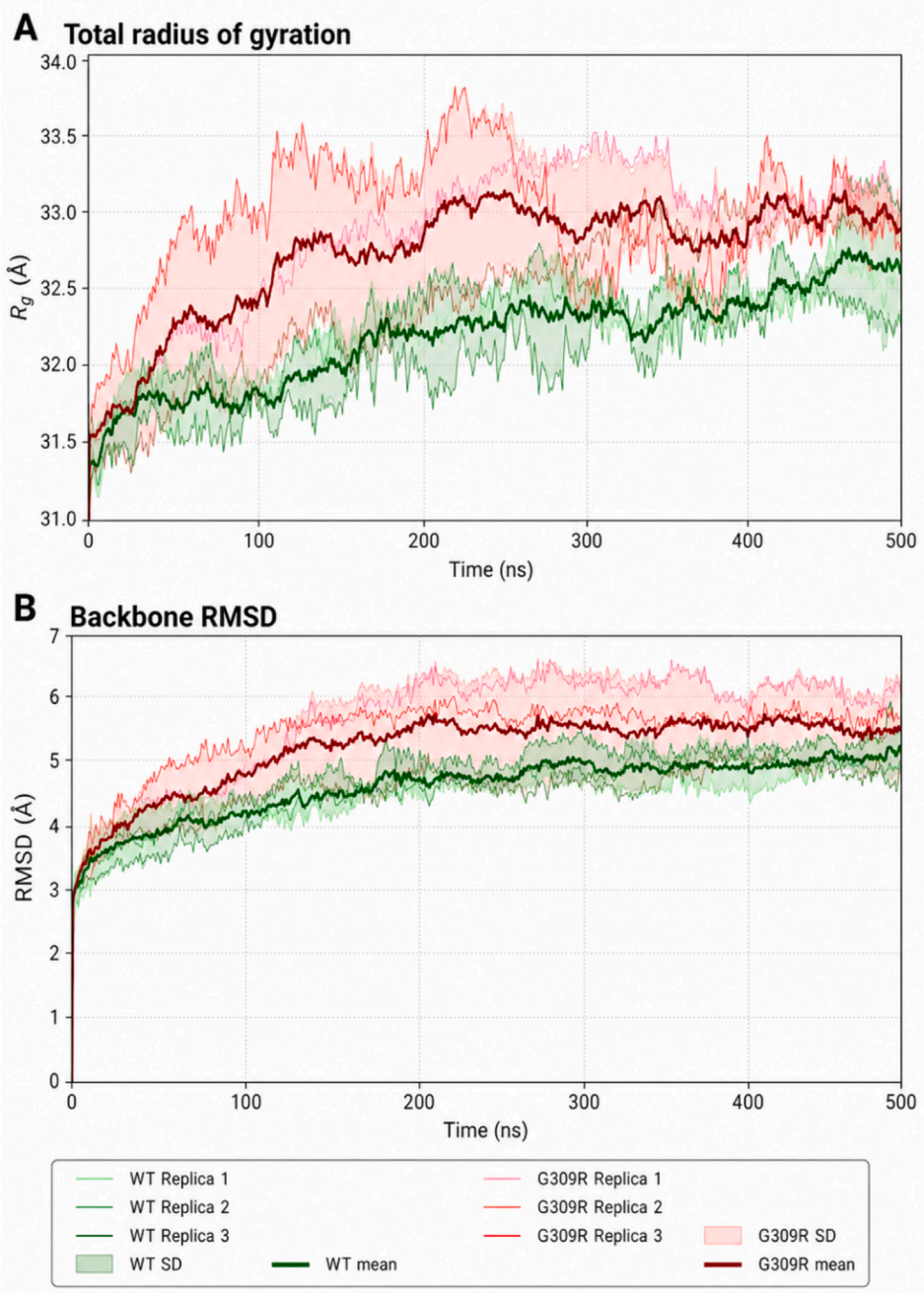


Fig. 1. Global conformational stability of WT and G309R HGD. (A) Total R_g calculated over 500 ns for three independent WT and G309R replicas. (B) Backbone RMSD calculated over the same simulation window. WT is shown in green and G309R in red. Both systems display comparable global stability and compactness, with no evidence of progressive unfolding, loss of hexameric compactness, or large-scale destabilization in the G309R mutant.

for WT and 0.283 ± 0.102 nm for G309R. In chain C, the WT profile reached a mean value of 0.341 ± 0.051 nm, while G309R reached 0.375 ± 0.094 nm. In chain D, the WT β-hairpin showed a mean Cα-RMSD of 0.224 ± 0.049 nm, whereas G309R reached 0.327 ± 0.060 nm. In chain E, the WT and G309R profiles were comparable in magnitude, with mean values of 0.226 ± 0.062 nm and 0.264 ± 0.037 nm, respectively. Finally, in chain F, the WT β-hairpin showed a mean Cα-RMSD of 0.153 ± 0.021 nm, while G309R reached 0.311 ± 0.021 nm (Fig. 4).

The self-reference RMSD profiles of both WT and G309R remained within the same range across all six chains, refuting the hypothesis that the tunnel disruption is caused by structural destabilization associated β-hairpin.

To further dissect the role of the G309R substitution in tunnel disruption, we examined potential spatial shifts of the β-hairpin within the tunnel region. Chain-wise analysis revealed distinct structural rearrangement patterns among the six HGD protomers. Representative structural snapshots at 0, 250, and 500 ns showed that the β-hairpin maintained chain-specific conformational trajectories throughout the simulations, with markedly different displacement amplitudes across chains A–F (Fig. 5a). The RMSD profiles calculated relative to the WT reference exhibited stable temporal trends over the 500 ns trajectories, with no abrupt RMSD collapse or discontinuous transitions. The three independent replicas displayed reproducible chain-dependent behavior, with higher RMSD amplitudes in selected chains and lower values in others. Chain A showed the smallest difference between WT and mutant β-hairpin positioning, with a mean displacement of approximately 0.14 nm. Chain B exhibited a higher mean difference of approximately 0.32 nm, comparable to chain C (~0.36 nm). Larger

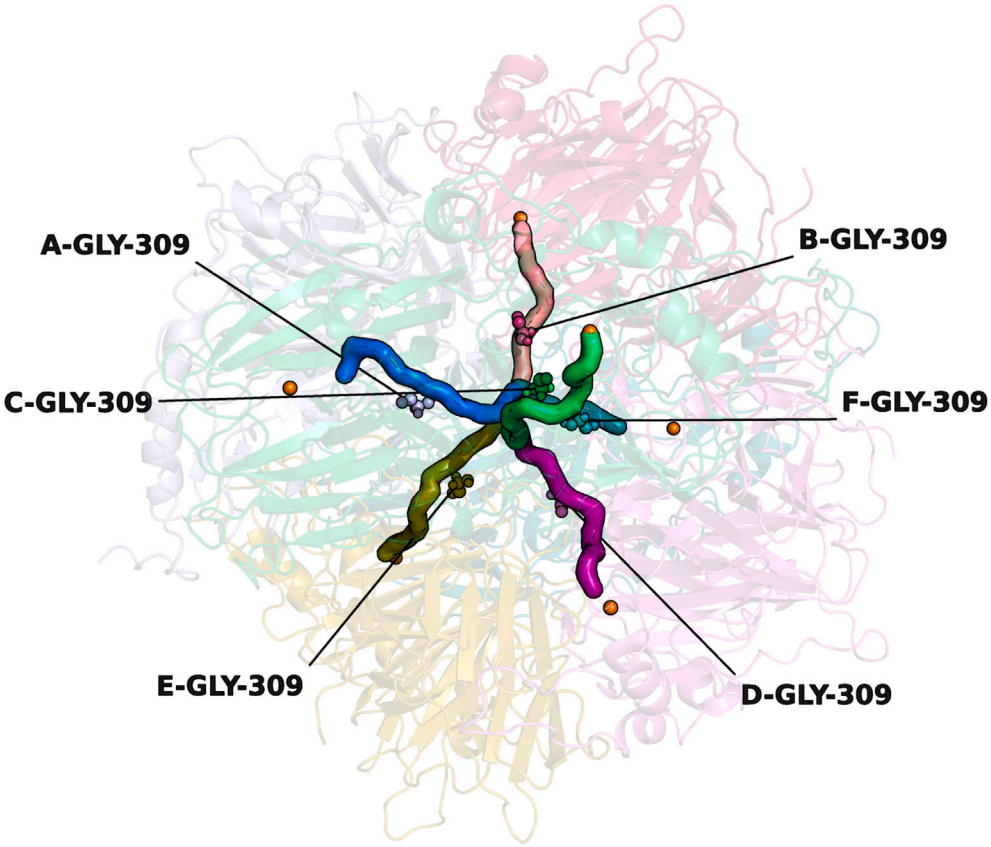


Fig. 2. CAVER-derived tunnel network in the WT HGD hexamer. The six protomers are shown in cartoon representation, Fe(II) catalytic centers as orange spheres, and Gly309 residues are labeled in each chain. Colored tunnel surfaces indicate the six symmetry-related pathways connecting the central pore to the catalytic sites through inter-subunit interfaces.

mean differences were observed in chain D and chain E, with values of approximately 0.74 nm and 0.61 nm, respectively. Chain F displayed the largest mean difference, reaching approximately 1.01 nm (Fig. 5b). Overall, the chain-wise mean RMSD comparison indicated a non-uniform β-hairpin rearrangement across the hexamer. The magnitude of displacement followed the approximate order chain A < chain B ≈ chain C < chain E < chain D < chain F, with chain F displaying the largest β-hairpin displacement relative to the WT reference (Fig. 5c). The structural snapshots and RMSD profiles consistently indicated that the β-hairpin rearrangement is chain-dependent and asymmetric across the HGD hexamer. Taken together, these observations support a direct role of the G309R mutation in disrupting the WT tunnel architecture, likely caused from the high side-chain flexibility of the ARG309 in the role of pivot of the β-hairpin. Accordingly, all subsequent analyses were performed on WT HGD to define the architecture and dynamics of the supramolecular oxygen transport system.

3.4. SMD: O_2 translocation through WT tunnel toward catalytic Fe(II) center

The O_2 path through the potential WT tunnel toward catalytic Fe(II) center was evaluated for each of the six WT tunnels performing SMD. O_2 was initially positioned at the solvent-accessible central pore and steered toward the corresponding catalytic Fe(II) center. Three independent SMD replicas were performed for each tunnel, resulting in 18 O_2 translocation trajectories. The O_2–Fe(II) distance progressively decreased along the pulling coordinate, indicating continuous translocation from the central pore toward the catalytic site. The final O_2 positions were located in proximity to the catalytic Fe(II) centers for all six tunnel pathways. No large-scale disruption of the WT hexameric architecture was observed during the SMD translocation process. The replicated SMD trajectories therefore supported the geometric accessibility of the six WT CAVER-identified tunnels to O_2, as shown by the time-dependent decrease in the O_2–Fe(II) distance in Fig. 6.

However, to overcome possible over-interpretation due to the bias applied during the steering and to define the composition, architecture, stability, and reproducibility of the path in the same tunnel and among them, further extensive structural, energy and evolutionary analyses were performed.

3.5. Residue-level mapping and architecture of the supramolecular oxygen tunnel

The tunnel forming residue composition was analyzed across the six pathways and across replicas using distance-based O_2–protein contact mapping. Direct O_2-contact residues were identified from the SMD trajectories through a contact criterion, identifing the spatial neighborhood surrounding the O_2 trajectory in a radius of 10 Å, considering only the residues in common in each chain for each replicate among all the chains. The extended tunnel architecture converged on a recurrent set of 12 equivalent residues per contributing chain (previously identified in WT HGD cMDs, see the section 3.4). These positions are organized into three main structural regions: an N-terminal tunnel-associated segment, corresponding to THR50, CYS51, PRO52, ARG53, SER54, and THR55; an intermediate interfacial segment comprising ASP212; and a β-hairpin tunnel-associated segment corresponding to LY304, SER305, VAL306, ARG307, PRO308, and GLY309 (chain-A numbering). Equivalent residue sets were recovered in all other chains, preserving the same local organization. A representative three-dimensional visualization of this supramolecular architecture is shown in Fig. 7 for tunnel A, assembled at the A–C–E trimeric interface.

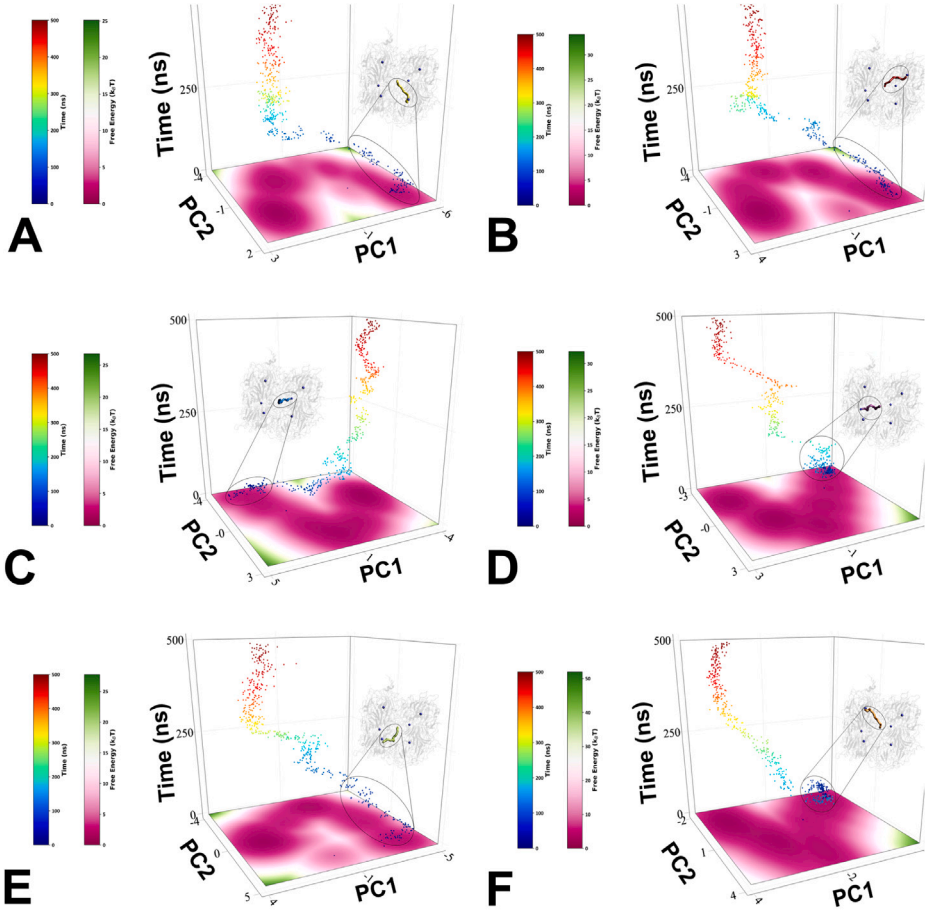


Fig. 3. PCA/FEL characterization of WT tunnel-associated conformational states. For each of the six CAVER-identified WT O_2-access tunnels, tunnel-associated residues were projected onto the first two principal components. Free-energy landscapes were reconstructed in PC1/PC2 space, with low-energy regions indicating recurrent tunnel-associated conformational states. Time-resolved trajectory projections and representative structural views support the presence of organized tunnel states in WT HGD.

Table 1
Chain-specific tunnel-lining residue contacts for the six supramolecular O_2 tunnels identified by CAVER analysis. For each tunnel-defining position (chain-A reference numbering), the corresponding residue in each chain is reported together with its mean O_2–residue distance (mean ± SD, Å) calculated across SMD trajectories.

Chain-A position	Chain A	Chain B	Chain C	Chain D	Chain E	Chain F
THR50	THR50 (6.48 ± 1.51)	THR495 (4.23 ± 2.37)	THR940 (4.48 ± 3.54)	THR1385 (6.29 ± 3.25)	THR1830 (3.68 ± 1.44)	THR2275 (5.96 ± 3.34)
CYS51	CYS51 (7.03 ± 0.04)	CYS496 (4.37 ± 1.80)	CYS941 (4.91 ± 2.85)	CYS1386 (4.92 ± 3.86)	CYS1831 (3.60 ± 1.29)	CYS2276 (5.70 ± 2.28)
PRO52	PRO52 (6.45 ± 0.41)	PRO497 (2.77 ± 0.21)	PRO942 (3.30 ± 0.66)	PRO1387 (3.92 ± 1.95)	PRO1832 (2.98 ± 0.73)	PRO2277 (5.52 ± 2.68)
SER54	SER54 (8.71 ± 0.00)	SER499 (6.81 ± 0.90)	SER944 (7.66 ± 2.71)	SER1389 (8.08 ± 1.60)	SER1834 (6.96 ± 2.64)	SER2279 (8.36 ± 0.33)
THR55	THR55 (8.92 ± 0.00)	THR500 (5.86 ± 1.27)	THR945 (5.09 ± 2.16)	THR1390 (5.01 ± 3.92)	THR1835 (5.90 ± 3.19)	THR2280 (8.36 ± 1.71)
ASP212	ASP212 (6.93 ± 2.58)	ASP657 (5.41 ± 1.04)	ASP1102 (3.49 ± 1.36)	ASP1547 (3.37 ± 1.19)	ASP1992 (5.02 ± 0.29)	ASP2437 (7.73 ± 1.90)
LYS304	LYS304 (5.35 ± 1.07)	LYS749 (5.94 ± 2.06)	LYS1194 (3.72 ± 1.62)	LYS1639 (4.84 ± 1.85)	LYS2084 (4.35 ± 1.34)	LYS2529 (6.36 ± 2.01)
SER305	SER305 (5.25 ± 1.85)	SER750 (6.75 ± 2.18)	SER1195 (4.06 ± 1.67)	SER1640 (4.01 ± 1.86)	SER2085 (4.42 ± 2.19)	SER2530 (6.17 ± 1.09)
VAL306	VAL306 (4.79 ± 1.23)	VAL751 (4.77 ± 2.24)	VAL1196 (3.46 ± 1.46)	VAL1641 (3.44 ± 0.81)	VAL2086 (3.67 ± 1.37)	VAL2531 (4.67 ± 2.27)
ARG307	ARG307 (4.23 ± 1.41)	ARG752 (5.73 ± 0.71)	ARG1197 (5.45 ± 1.40)	ARG1642 (2.81 ± 0.30)	ARG2087 (3.57 ± 0.70)	ARG2532 (3.44 ± 1.20)
PRO308	PRO308 (2.76 ± 0.25)	PRO753 (4.25 ± 0.81)	PRO1198 (3.23 ± 0.27)	PRO1643 (2.36 ± 0.15)	PRO2088 (2.54 ± 0.18)	PRO2533 (3.94 ± 1.01)
GLY309	GLY309 (5.90 ± 0.51)	GLY754 (7.68 ± 1.60)	GLY1199 (5.29 ± 2.34)	GLY1644 (2.35 ± 0.04)	GLY2089 (3.82 ± 0.53)	GLY2534 (5.98 ± 2.73)

The full chain-specific mapping (Table 1) showed that each tunnel was assembled from three equivalent residue sets contributed by the corresponding interface-forming chains. Therefore, tunnel A was composed of the tunnel-associated regions from chains A, C, and E; tunnel B from chains B, A, and F; tunnel C from chains C, B, and D; tunnel D from chains D, E, and C; tunnel E from chains E, F, and A; and tunnel F from chains F, D, and B (Fig. 8).

3.6. Chian-based O_2 spatial analysis identifies stable translocation tunnel

To evaluate whether O_2 translocation followed the same and reproducible spatial tunnels to achieve the catalytic Fe(II) centers rather than other paths due to the replica-specific deviations, the 18 SMD trajectories were analyzed by spatial occupancy mapping. For each tunnel, O_2 coordinates from the three independent replicas were mapped onto a common structural reference frame and used to generate consensus occupancy distributions. For all six WT tunnels, the three independent replicas produced overlapping O_2 occupancy regions for each tunnel. The highest-density regions defined continuous spatial corridors connecting the central pore with the corresponding Fe(II)-proximal region. Replica-specific centerlines remained embedded within the same consensus density region, indicating that O_2 followed a reproducible pathway within each tunnel. Orthogonal projections of the O_2 occupancy maps confirmed pathway continuity in the XY, XZ, and YZ

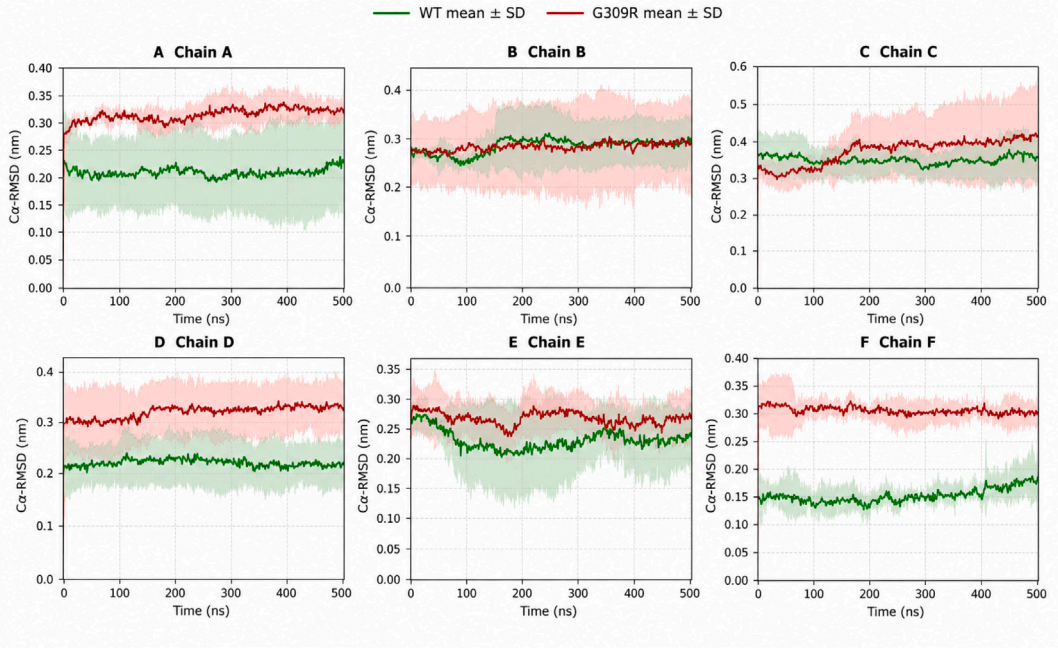


Fig. 4. Chain-resolved β-hairpin self-reference Cα -RMSD profiles of WT and G309R HGD. Self-reference Cα -RMSD values were calculated for the β-hairpin region of each HGD protomer over 500 ns MD simulations. Panels A–F report the profiles for chains A–F, respectively. WT trajectories are shown in green and G309R trajectories in red. Curves represent the mean Cα -RMSD calculated from three independent replicas, while shaded areas indicate the corresponding standard deviation at each time point. The analysis shows that both WT and G309R β-hairpin regions remain within the same sub-nanometer RMSD range, with chain-dependent but stable temporal profiles throughout the simulations.

planes exhibiting a very low standard deviation among the replicates. The reconstructed routes therefore did not depend on a single visual projection. Longitudinal density profiles further resolved the spatial organization of each pathway, identifying recurrently populated entry, central, and exit regions along the translocation coordinate. Across the six tunnels (and in all replicates), the O_2 occupancy analysis showed a consistent central-pore-to-active-site organization. Local fluctuations were observed among replicas, but the global pathway geometry remained conserved within each tunnel. These results indicate that O_2 translocation in WT HGD occurs through defined, spatially reproducible tunnels rather than through random displacement outside the tunnel region, as shown by the consensus density maps and replica-derived centerlines in Fig. 9.

3.7. PCA/tICA free-energy analysis supports the dynamic coherence of O_2 translocation

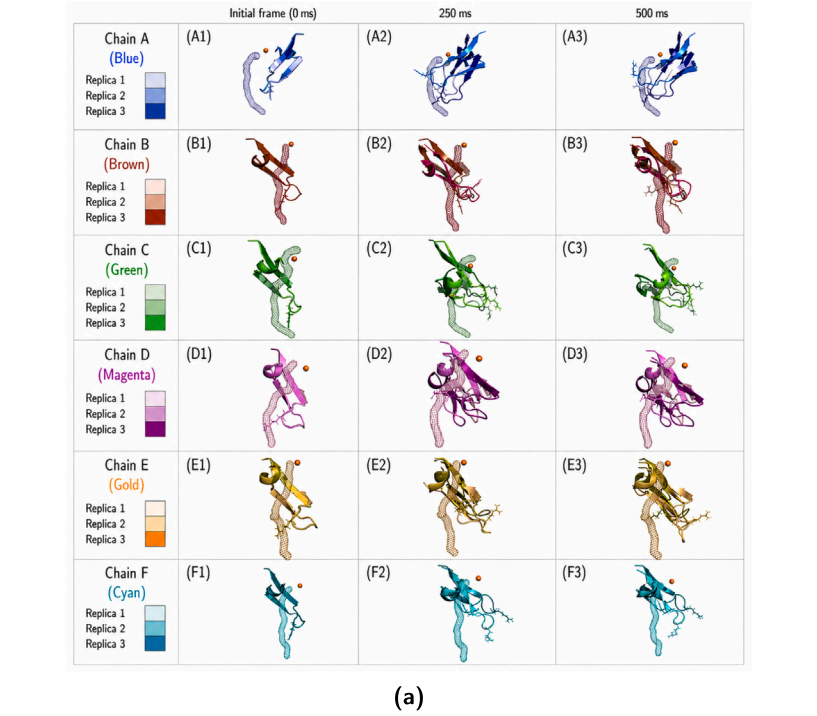
To assess whether the SMD-derived O_2 translocation coordinates were compatible with dominant geometric and slow kinetic modes, PCA and tICA analyses were performed on the O_2 coordinates extracted from the SMD trajectories. For each tunnel and replica, PCA-derived and tICA-derived free-energy landscapes were reconstructed and compared. The PCA free-energy landscapes described the dominant geometric organization of the O_2 translocation coordinate, whereas the tICA landscapes captured slow time-correlated components of the same process. Across the analyzed tunnels, PCA and tICA projections produced consistent low-energy regions associated with the O_2 translocation pathway. The agreement between PCA and tICA projections was quantified by Pearson correlation between PC1 and TIC1 (Fig. 10A-F). Most tunnel replicas showed high absolute correlation values, indicating that the dominant geometric coordinate and the slow kinetic coordinate described the same O_2 translocation process (Figure S2 A-F and Figure S3 A-F). The PCA/tICA/FEL analysis therefore provided a dynamic validation layer independent of visual trajectory inspection. The O_2 pathway detected was associated with coherent reduced-coordinate landscapes rather than with isolated trajectory-specific motion.

3.8. Evolutionary conservation and tunnel scaffold-lining mapping

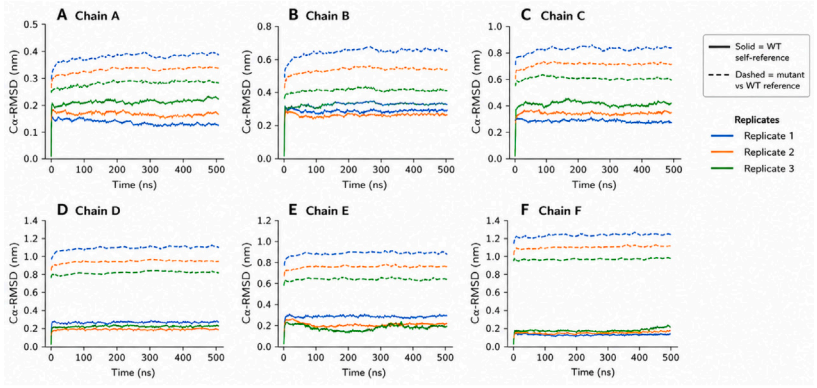
The evolutionary relevance of the tunnel-associated architecture was assessed through a BLASTp-based multiple sequence alignment of 1008 non-redundant HGD orthologs spanning a broad phylogenetic range. Twelve tunnel-lining positions distributed across the three structural regions of the O_2-access architecture were selected for residue-resolved conservation analysis: THR50, CYS51, ASP212, LYS304, SER305, and GLY309 are entirely invariant (100% conservation), suggesting an indispensable structural or catalytic role for enzyme function. Similarly, sites PRO52, ARG53, ARG307, and PRO308 show conservation rates exceeding 98%. Only moderate flexibility is observed in adjacent positions, specifically at SER54 (92.7% conservation) and THR55, where the predominant variant THR55S (14.0%) represents a highly predictable substitution between polar amino acids. At the opposite end of the spectrum, residue VAL306 distinctly emerges as the hypervariable hotspot of the dataset (conserved in only 57.3% of cases)(Fig. 11). However, it is crucial to emphasize the strictly conservative nature of these mutations: the predominant orthologous variants are Leucine (33.0%) and Isoleucine (2.3%). This demonstrates that, despite high permissivity in terms of primary sequence identity, the VAL in LEU and VAL in ILE transitions rigorously preserve the physicochemical properties, ensuring the maintenance of the required hydrophobic and aliphatic steric bulk.

4. Discussion

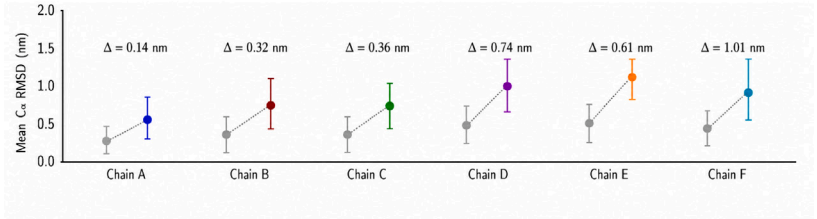
Efficient catalysis in non-heme Fe(II)-dependent dioxygenases requires not only a properly coordinated metal center and a correctly positioned organic substrate, but also a structural route through which molecular oxygen can access the buried catalytic environment (Tariq et al., 2022; Martinez and Hausinger, 2015; Solomon et al., 2000). In human HGD, the Fe(II)-dependent cleavage of HGA is well established, and the architecture of the active site has been extensively characterized. However, the structural basis by which O_2 reaches the deeply buried catalytic iron has remained unclear. This is a non-trivial



(a)



(b)



(c)

Fig. 5. G309R-induced β-hairpin displacement in HGD. (a) Representative snapshots of the β-hairpin (residues 309) in WT and G309R at 0, 250, and 500 ns, shown per chain (A–F) and replica. (b) Time-resolved Cα-RMSD profiles of the β-hairpin across the six chains and three replicas: WT relative to the WT reference (solid lines) and G309R relative to the WT reference (dashed lines). (c) Mean Cα-RMSD per chain for WT (gray) and G309R (colored); Δ values indicate the chain-wise WT–G309R displacement (nm).

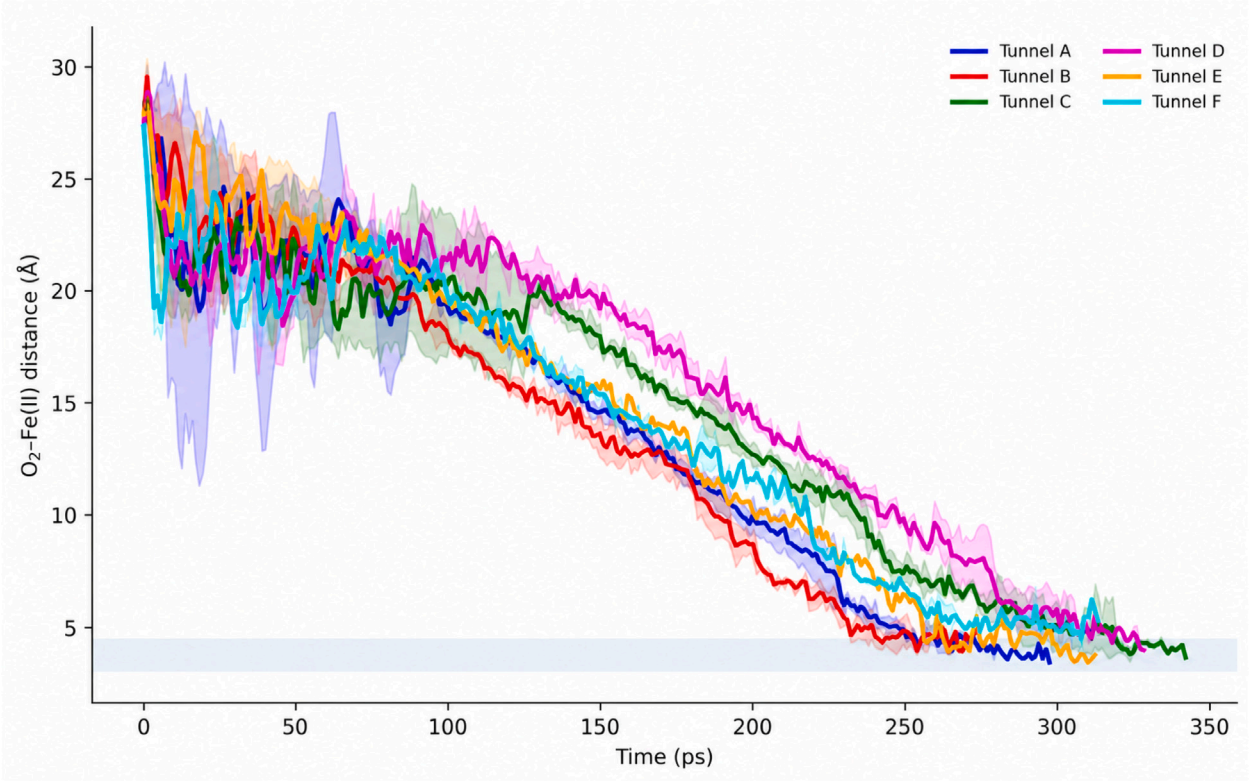


Fig. 6. SMD-based O_2 translocation toward the catalytic Fe(II) centers. Time evolution of the O_2–Fe(II) distance during SMD simulations along the six symmetry-related WT transport pathways. The profiles show progressive O_2 movement from the central pore toward the Fe(II)-proximal region. These distance profiles support metal-proximal translocation but are interpreted together with spatial-occupancy, centerline, and PCA/tICA analyses.

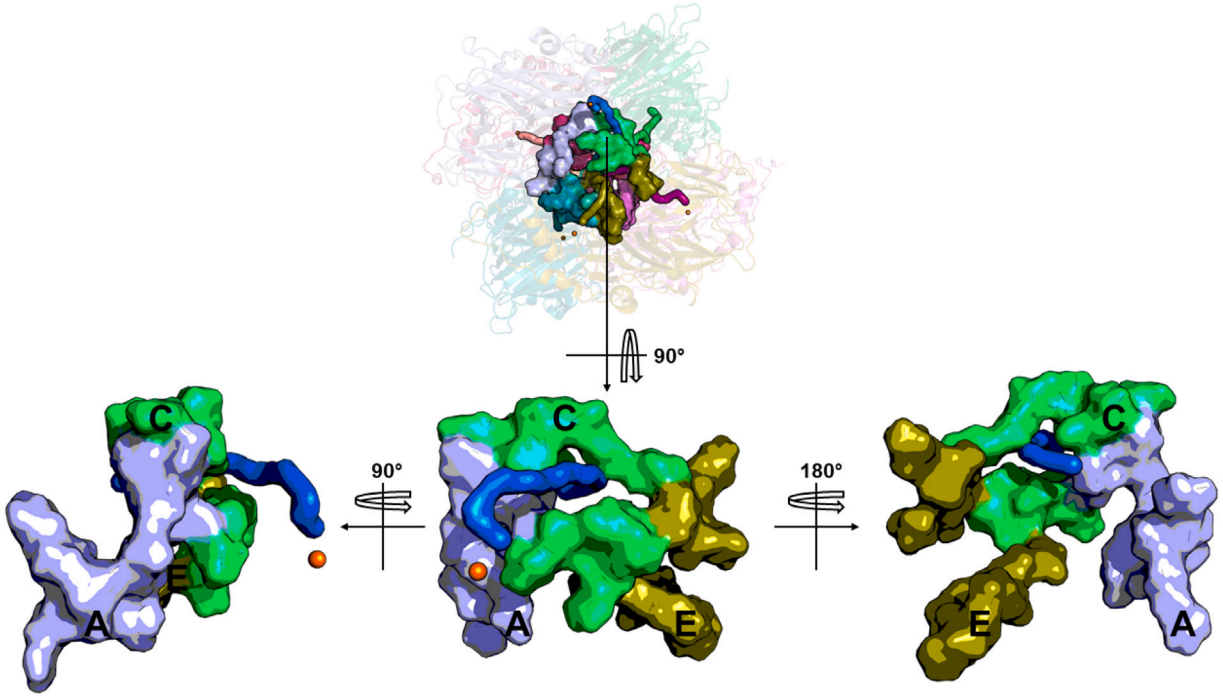
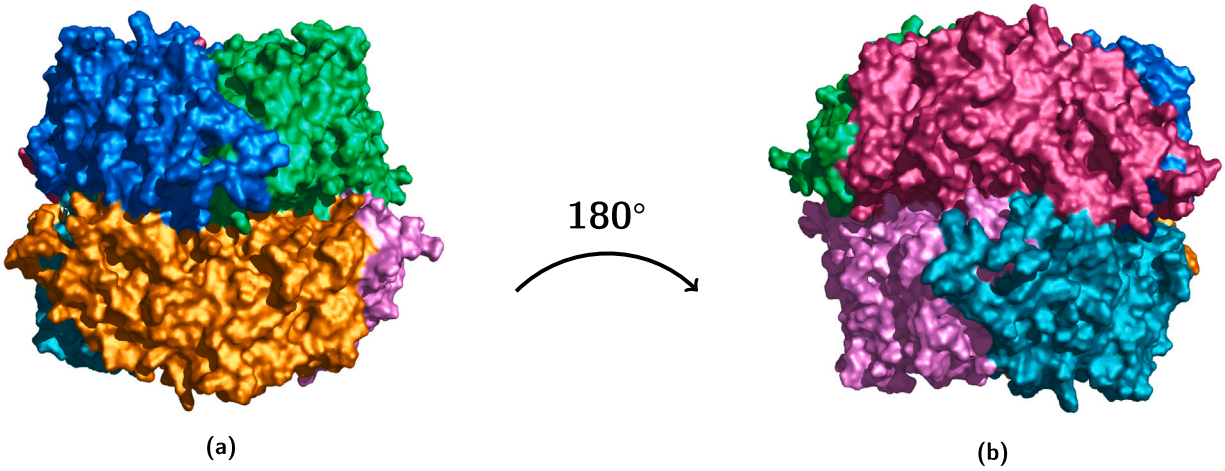


Fig. 7. Three-dimensional visualization of the tunnel A supramolecular architecture at the A–C–E representative trimeric interface. *Top:* overview of the HGD hexamer with the chain-A–C–E tunnel-lining surfaces highlighted. *Bottom, from left to right:* three rotated views (90° and 180° along the indicated axes) showing the cooperative assembly of tunnel-lining residues from chain A (ligth blue), chain C (green), and chain E (gold). The catalytic Fe(II) center is shown as an orange sphere. The continuity of the tunnel surfaces across the three chains highlights the supramolecular nature of the O_2-access cavity, which is not encoded by a single protomer but emerges from the spatial cooperation of three adjacent subunits.

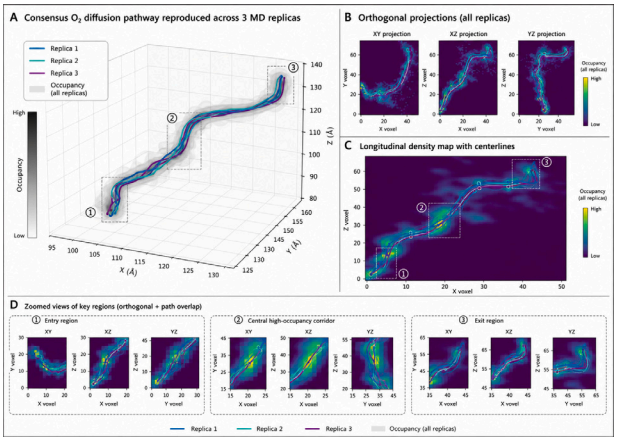


Tunnel	Position 50–55 (Chain)	Position 212 (Chain)	Position 304–308 (Chain)
Tunnel A	A	C	E
Tunnel B	B	A	F
Tunnel C	C	B	D
Tunnel D	D	E	C
Tunnel E	E	F	A
Tunnel F	F	D	B

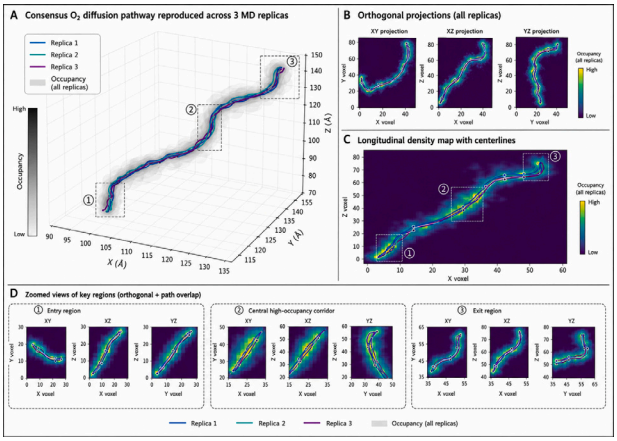
(c)

Fig. 8. Symmetry-related architecture and chain-specific composition of O_2-access tunnels in hexameric HGD.(a) Surface representation of the HGD hexamer in one orientation, showing the spatial arrangement of the six protomers.(b) Surface representation of the same HGD hexamer after a 180° rotation, providing the opposite view of the hexameric assembly and showing the relative position of the remaining protomers. (c) Chain-specific composition of the six symmetry-related O_2-access tunnels. For each tunnel, the table reports the protomer contributing the position 50–55, position 212, and the position 304–308. Chains are color-coded as follows: chain A, blue; chain B, red; chain C, green; chain D, magenta; chain E, orange; and chain F, cyan.

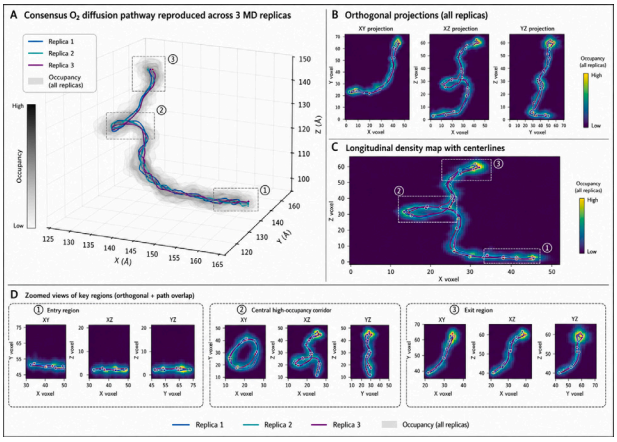
question because gas-access routes in enzymes are often transient, partially hydrated, conformationally gated, and poorly resolved by static structural models (Singh and Anand, 2021; Cohen et al., 2006). The present study addresses this unresolved aspect of HGD function by combining replicated unbiased MD simulations, CAVER-based tunnel detection, residue-level tunnel composition, β-hairpin displacement analysis, SMD path probing, O_2 spatial-occupancy mapping, centerline reproducibility, PCA/FEL, tICA/FEL, and evolutionary conservation analysis (Torabifard and Cisneros, 2017; Cohen et al., 2006). Together, these approaches support the presence of a supramolecular O_2-access architecture in WT HGD and show that the G309R substitution disrupts this architecture without inducing global destabilization of the hexamer. A first important result is that the G309R substitution does not cause major global structural disruption of HGD. Across three independent 500 ns replicas for both WT and mutant systems, backbone RMSD and total radius of gyration profiles showed broadly comparable global behavior. The mutant displayed only a modest upward shift in RMSD and Rg, without evidence of progressive unfolding, collapse, loss of compactness, or large-scale hexamer destabilization. This point is mechanistically important because HGD pathogenic variants have traditionally been interpreted within three main structural categories: active-site perturbation, protomer destabilization, or disruption of hexamer assembly (Titus et al., 2000; Zatkova et al., 2012). Our simulations indicate that G309R does not primarily act through these global mechanisms. Rather, the mutation preserves the overall structural framework while perturbing a localized architectural element involved in O_2 access. This distinction is central to the interpretation of the present work: the pathogenic effect proposed here does not require gross structural failure, but arises from selective disruption of a functional access pathway embedded in the quaternary structure of the enzyme. The tunnel architecture identified in WT HGD is not a simple monomeric cavity. CAVER analysis of the replicated WT trajectories identified six symmetry-related routes connecting the solvent-accessible central pore of the hexamer to the six catalytic Fe(II) centers (Chovancova et al., 2012). These routes are located at trimeric inter-subunit interfaces and are assembled from residues contributed by three neighboring protomers. Thus, the O_2-access pathway emerges as a supramolecular property of the HGD hexamer rather than as a feature encoded by a single subunit. This observation is consistent with the known requirement of HGD quaternary organization for activity and assigns a functional role to the central pore as a shared access region from which O_2 can be distributed toward individual catalytic sites (Titus et al., 2000). The six routes are equivalent in their global organization but chain-specific in their residue numbering, reflecting the symmetry of the hexameric assembly. This symmetry-derived organization provides a structural explanation for how a buried active site can remain accessible to a small gaseous co-substrate without requiring large opening events or complete solvent exposure of the catalytic iron. A key strength of the present analysis is that the tunnel was not defined solely as a geometric object generated by CAVER.



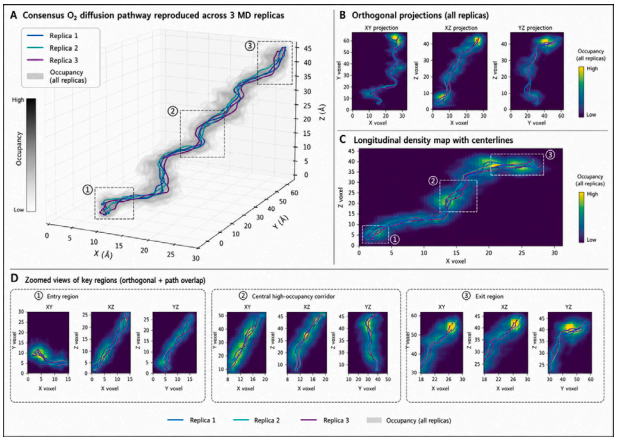
(a) Chain A



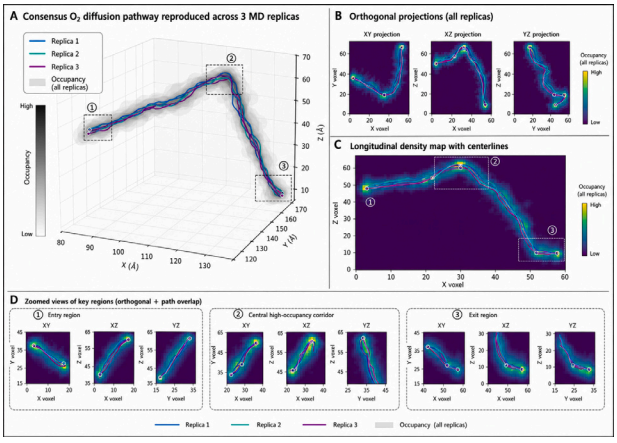
(b) Chain B



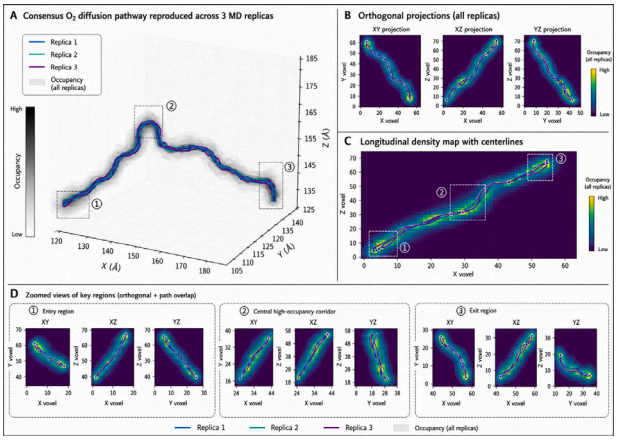
(c) Chain C



(d) Chain D



(e) Chain E



(f) Chain F

Fig. 9. Consensus O_2 diffusion pathways reproduced across 3 MD replicas for six independent chains (A–F). Each sub-panel shows: (A) the 3D consensus pathway with occupancy density and centerlines from the three replicas, (B) orthogonal projections (XY, XZ, YZ), (C) the longitudinal density map with overlaid centerlines, and (D) zoomed views of the entry region, central high-occupancy corridor, and exit region.

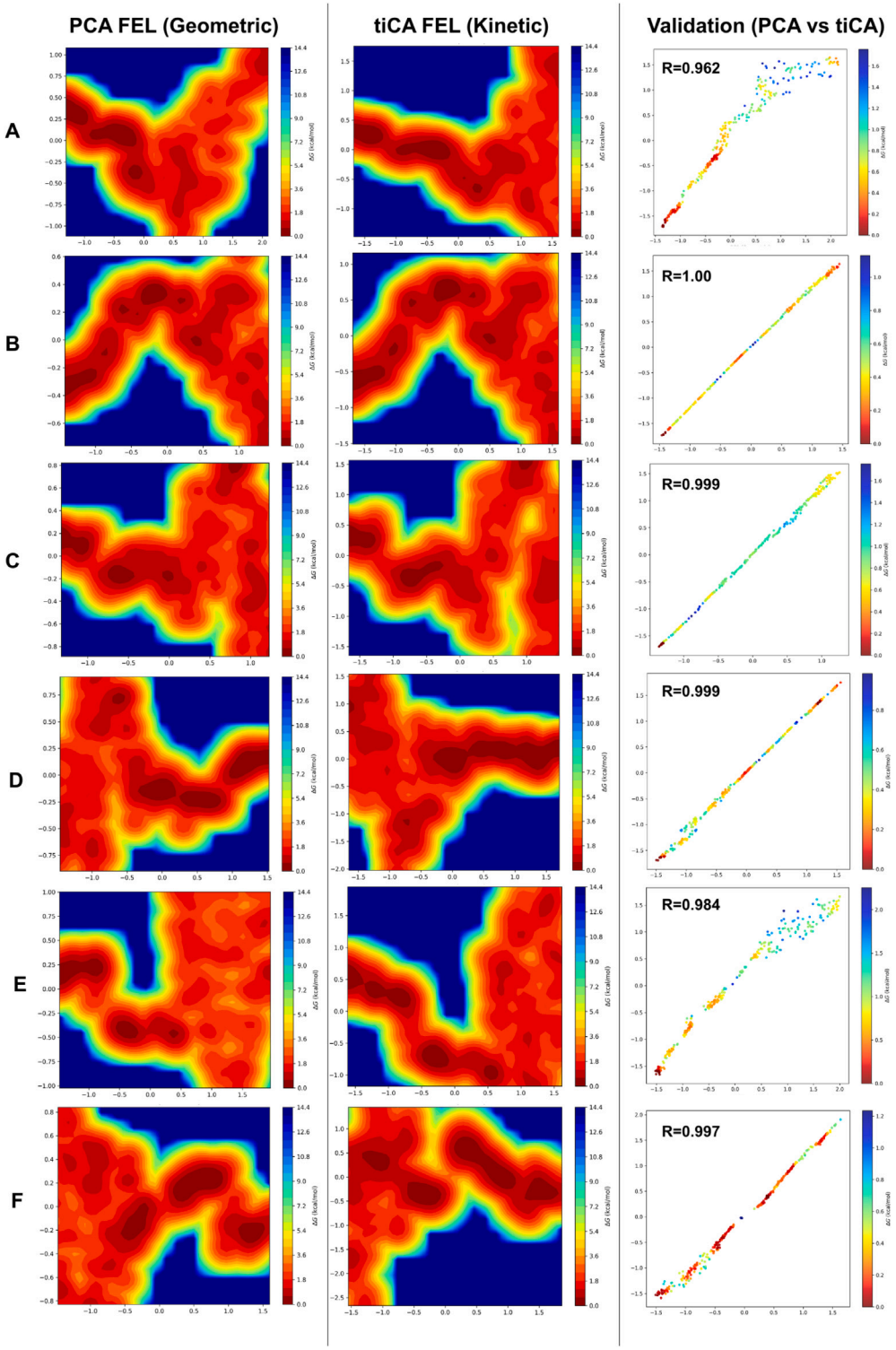


Fig. 10. PCA- and tICA-derived free-energy landscapes reconstructed from O_2 center-of-mass coordinates extracted from SMD trajectories. Each row (from A to F) reports horizontally (from left to right) the PCA-derived geometric FEL (PCA FEL geometric), the tICA-derived kinetic FEL (tICA FEL kinetic), and the PCA/tICA validation correlation (Validation PCA vs tICA). PCA describes the dominant geometric organization of O_2 translocation, whereas tICA captures slow time-correlated components of the same process. PC1/TIC1 correlation analysis shows high agreement in absolute value across most replicas, supporting the dynamic coherence of the O_2 translocation coordinate. To clarify, we reported the PCA- and tICA-derived free-energy landscapes picture of the replicate 1. The analyses of other replicates were shown in Figure S2-3.

The pathway was further resolved at residue level using O_2–protein contact mapping across the replicated SMD trajectories. This analysis identified both direct O_2-contact residues and the broader extended tunnel environment. The extended tunnel architecture converged on 12 equivalent residue positions per contributing chain, distributed across three structural regions: an N-terminal tunnel-associated segment, an intermediate interfacial segment, and a β-hairpin segment containing residue 309. In chain-A numbering, the tunnel-associated

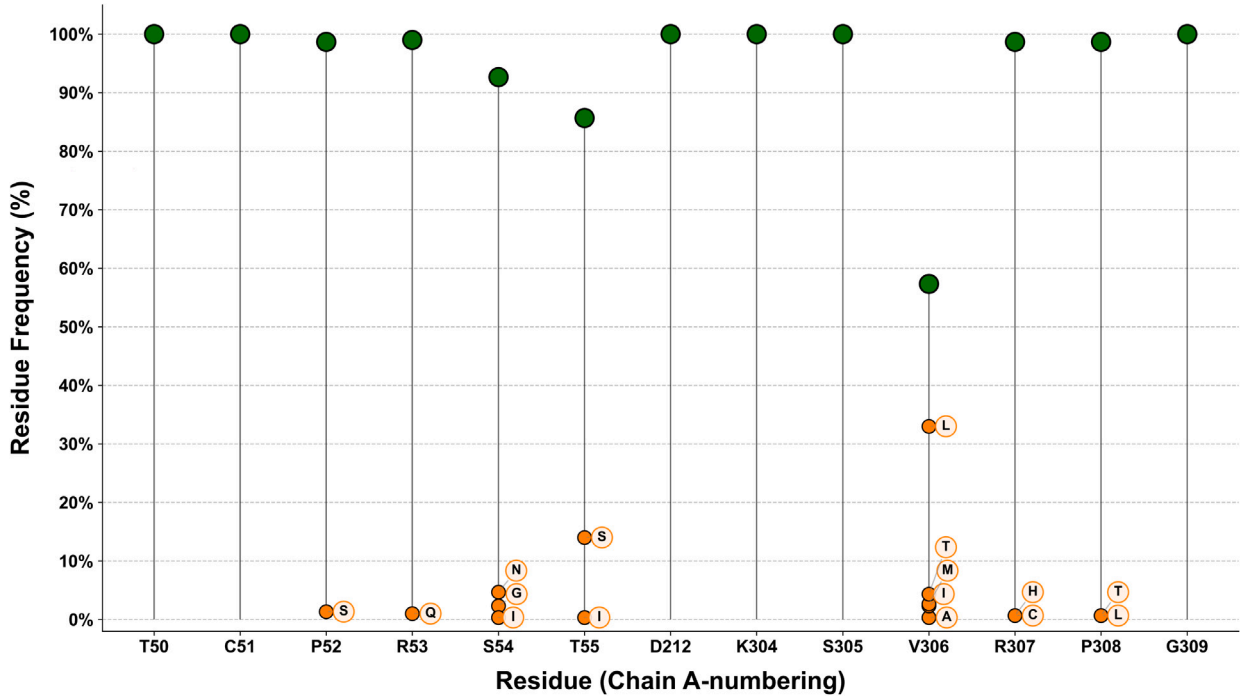


Fig. 11. Residue conservation and substitution profile at tunnel-lining positions of human HGD across 1008 non-redundant orthologs. Lollipops (green) report the conservation frequency (%) of the reference residue at each chain-A position. Orange circles indicate the most frequent non-reference substitutions, with the substitute amino acid identity reported within each circle.

environment includes THR50, CYS51, PRO52, ARG53, SER54, THR55, ASP212, LY304, SER305, VAL306, ARG307, PRO308, and GLY309. The recurrence of this positional architecture across the six pathways indicates that the tunnel is residue-defined and structurally organized, rather than an arbitrary void in the protein matrix. It lies within the ASP212 β-hairpin segment, which contributes directly to the tunnel-associated boundary. This localization is essential for interpreting the G309R effect. GLY309 is not peripheral to the proposed pathway; it is embedded within one of the recurrent tunnel-forming segments. The structural properties of glycine are particularly relevant in this context. Because glycine lacks a side chain beyond hydrogen, it imposes minimal steric restriction on the backbone and can populate local ϕ/ψ conformations that are less accessible to other amino acids. In tight turns and β-hairpin regions, this conformational permissiveness can be critical for maintaining local curvature and compact packing (DuPai et al., 2021a). In WT HGD, Gly309 therefore provides a small, flexible, and sterically unobtrusive element within the tunnel boundary. Its replacement by arginine introduces a long, bulky, positively charged side chain into a spatially constrained region directly adjacent to the O_2-access environment. The structural consequence of G309R was quantified by chain-wise β-hairpin RMSD and spatial-displacement analyses. These analyses were deliberately performed separately for each chain, because averaging the six symmetry-related β-hairpins would obscure chain-dependent local rearrangements within the hexamer. The mutant β-hairpins showed increased displacement relative to the WT reference in all chains, with the largest deviations observed in chains F, D, and E. Importantly, the signal is not interpreted as complete unfolding of the β-hairpin. Rather, it indicates positional displacement of an otherwise recognizable structural element relative to the WT tunnel-compatible geometry. This distinction is critical: disruption of a narrow gas-access route does not require loss of secondary structure. Small shifts in a loop or β-hairpin that defines the tunnel boundary can be sufficient to narrow, distort, or interrupt a continuous O_2 corridor. The spatial-density and structural snapshot analyses support this interpretation by showing that the β-hairpin containing residue 309 adopts a shifted spatial distribution in the mutant relative to WT. Thus, G309R perturbs the O_2 pathway through a local structural-dynamical mechanism involving loss of glycine-mediated backbone permissiveness, introduction of a bulky charged side chain, altered local packing, and positional displacement of the tunnel-associated β-hairpin. The effect of arginine at position 309 should not be viewed only as a static steric clash. Arginine has a long flexible side chain ending in a guanidinium group and can sample multiple rotameric states. In the local structural environment of residue 309, the neighboring internal cavity offers space in which R309 rotamers can be accommodated. During the simulations, the mutant side chain is therefore able to occupy the tunnel-proximal region and alter the local steric and electrostatic environment. This provides a plausible molecular basis for the observed β-hairpin displacement and loss of WT-like tunnel continuity. In WT, glycine does not impose such side-chain accommodation constraints, allowing the β-hairpin to preserve a compact geometry compatible with an open tunnel boundary. In the mutant, the side chain of R309 and the shifted β-hairpin jointly reduce the continuity of the access route. In a previous work, where analyses of the β-hairpin motif were extensively explored, it was shown that there is a strong preference for glycine in flexible turn regions; conversely, arginine emerged as one of the least prevalent amino acids within these motifs (DuPai et al., 2021b), strongly supporting our evidence. The resulting model explains how a single residue substitution can impair a functional pathway without causing overt active-site deformation or global hexamer destabilization. PCA and free-energy landscape analyses further supported the existence of organized WT tunnel conformational states (Papaleo et al., 2009; Schultze and Grubmüller, 2021). The CAVER-identified tunnels were associated with defined low-energy regions in PC1/PC2 space, indicating that the tunnel-associated residues sample recurrent conformational states rather than isolated or trajectory-specific geometries. This result is important because tunnel formation in proteins is inherently dynamic. A pathway visible in one frame or detected by a purely geometric algorithm may not necessarily correspond to a recurrent conformational feature. By projecting tunnel-associated residues onto PCA-derived landscapes, the present analysis shows that the WT tunnel architecture is embedded within sampled

conformational basins. These landscapes indicate that the tunnel is not a static structural artifact but part of the intrinsic conformational ensemble of the WT enzyme. At the same time, this analysis should not be interpreted as a quantitative ranking of pathway favourability. It supports the recurrence and organization of tunnel-associated states, while absolute thermodynamic barriers would require dedicated enhanced-sampling calculations. The SMD simulations were therefore used in a specific and controlled manner. They were not interpreted as direct proof of spontaneous O_2 diffusion, nor as a calculation of permeation kinetics or absolute free-energy barriers. Instead, SMD was used as a path-probing method to test whether O_2 can be translocated through the structurally identified WT tunnels without large-scale disruption of the protein architecture (Singh and Anand, 2021; Torabifard and Cisneros, 2017). This distinction directly addresses a common limitation of steered simulations: an externally applied bias can force movement along a coordinate that may not be frequently sampled under equilibrium conditions. For this reason, SMD alone was not used as the sole evidence for the proposed pathway. Rather, it was embedded within a broader validation framework combining tunnel mapping, replicated trajectories, O_2 occupancy, centerline reproducibility, residue-level contacts, PCA/FEL, and tICA/FEL analyses. Within this framework, the replicated SMD trajectories showed that O_2 can be guided from the central pore to the Fe(II)-proximal region through each of the six WT pathways. The spatial analysis of O_2 trajectories provided an additional level of validation beyond simple distance-to-metal plots. Distance profiles can show whether O_2 approaches the catalytic iron, but they do not define the three-dimensional route followed by the molecule, nor do they determine whether different replicas reproduce the same spatial corridor. For this reason, the O_2 coordinates from all 18 SMD trajectories were analyzed by spatial occupancy mapping, consensus density reconstruction, centerline comparison, orthogonal projections, and longitudinal density profiling. Across the six WT pathways, the three independent replicas per tunnel produced overlapping O_2 occupancy regions. Replica-specific centerlines remained embedded within the same high-density corridor, and orthogonal XY, XZ, and YZ projections confirmed that the reconstructed routes were continuous in three dimensions. Longitudinal density maps further resolved entry, central, and exit regions along the translocation coordinate. These analyses show that O_2 translocation did not occur through random displacement outside the protein tunnel environment, but through defined and reproducible spatial corridors constrained by the WT HGD architecture. The PCA/tICA/FEL analysis of O_2 coordinates provided a complementary dynamic validation layer. PCA describes dominant geometric variance, whereas tICA identifies slow time-correlated modes that are more closely related to kinetically persistent motions. By applying both methods to the SMD-derived O_2 coordinates, we assessed whether the O_2 translocation coordinate was compatible with both geometric and slow dynamic descriptions of the process. The high absolute PC1/TIC1 correlations observed across most tunnel replicas indicate that the same translocation coordinate is captured by both approaches. Negative correlations in selected cases do not represent disagreement, because the sign of PCA and tICA eigenvectors is arbitrary; rather, they indicate opposite orientation of an equivalent reduced coordinate. The corresponding PCA- and tICA-derived free-energy landscapes displayed coherent low-energy regions associated with O_2 translocation. These results do not replace a full potential-of-mean-force calculation, but they provide evidence that the SMD-derived route is dynamically coherent and supported by reduced-coordinate free-energy organization rather than by visual inspection alone (Borowski et al., 2005; Martinez and Hausinger, 2015; Solomon et al., 2000). Taken together, the computational evidence supports a coherent model of O_2 access in HGD. In WT, the hexamer contains six symmetry-related, residue-defined O_2-access routes that connect the central pore to the catalytic Fe(II) centers. These routes are supramolecular in nature, being assembled at trimeric interfaces from residues contributed by neighboring protomers. They are detected by CAVER, resolved by residue-level mapping, supported by PCA/FEL conformational analysis, traversed by O_2 in replicated SMD simulations, reproduced as spatial occupancy corridors across replicas, and dynamically supported by PCA/tICA/FEL analysis. In G309R, the same continuous architecture is locally disrupted through displacement of the β-hairpin segment containing residue 309. This provides a mechanistic link between a single AKU-associated substitution and selective impairment of O_2 accessibility. The evolutionary conservation analysis further supports the functional relevance of the tunnel-associated architecture (Amaya et al., 2004; Borowski et al., 2005). Residues defining the O_2-access environment showed high conservation across a broad set of HGD orthologs, indicating that the structural elements contributing to the tunnel are under evolutionary constraint. This observation is consistent with the proposed functional importance of maintaining a defined O_2-access route in an enzyme whose active site is buried within a multimeric assembly. Conservation alone cannot prove O_2 transport, but in combination with the structural, dynamical, and residue-level analyses, it strengthens the interpretation that the tunnel-associated residues are not incidental components of the fold. Several limitations should be explicitly acknowledged. First, the present work does not provide direct experimental measurement of O_2 flux through the proposed pathways. Experimental validation would require a dedicated study, potentially involving site-directed mutagenesis of tunnel-lining residues, kinetic assays under controlled oxygen availability, gas-analog trapping, xenon-derivatization experiments, or time-resolved spectroscopic approaches. Second, the simulations do not provide absolute permeation rates, residence times, or pathway-specific free-energy barriers. Such quantities would require enhanced-sampling approaches specifically designed for quantitative free-energy estimation, such as umbrella sampling, metadynamics, or related path-based methods. Third, SMD imposes an external bias and is therefore interpreted here only as a controlled pathway-probing method. These limitations are important, but they do not undermine the main conclusion of the study, which is not a quantitative kinetic ranking of O_2 pathways. The central conclusion is structural and mechanistic: WT HGD contains a conserved, residue-defined, dynamically supported O_2-access architecture, and G309R disrupts this architecture through local β-hairpin displacement. This work expands the mechanistic framework for interpreting pathogenic HGD variants. Beyond mutations that directly affect the catalytic site, destabilize individual protomers, or disrupt hexamer assembly (Zatkova et al., 2012; Vilboux et al., 2009; Usher et al., 2014), the present study identifies disruption of oxygen trafficking as an additional pathogenic mechanism. This category is particularly relevant because it can explain loss of enzymatic competence in the absence of global destabilization or overt active-site deformation. The G309R variant illustrates this principle: the enzyme remains globally compact, but a localized structural rearrangement compromises the architecture required for O_2 access. More broadly, the findings suggest that gas-access pathways in non-heme dioxygenases (Singh and Anand, 2021; Tariq et al., 2022; Torabifard and Cisneros, 2017; Cohen et al., 2006) may represent mutation-sensitive structural determinants of enzyme function.

Beyond explaining the local effect of the G309R mutation, the present findings provide a functional explanation for the peculiar quaternary organization of HGD. Indeed, the catalytic mechanism would not depend only on the individual protomer, but on a supramolecular architecture generated by the hexameric assembly. Since HGD is not known to form alternative functional oligomeric states, this would support the idea that the native hexamer is required to create the proper structural framework for O_2 access and catalysis.

Recognition of such pathways may improve variant interpretation, guide targeted mutagenesis, and support structure-informed therapeutic or enzyme-engineering strategies in AKU and related metabolic disorders.

5. Conclusion

This study identifies oxygen trafficking as a critical and previously unrecognized structural determinant of catalytic competence in human homogentisate dioxygenase, providing a new mechanistic framework for interpreting enzymatic failure in AKU. By integrating replicated molecular simulations with tunnel mapping, residue-level pathway characterization, spatial-occupancy analysis, PCA/tICA-based dynamic validation, and evolutionary profiling, we show that HGD function depends not only on preservation of the Fe(II) active site and hexameric assembly, but also on maintenance of a supramolecular O_2-delivery architecture embedded within the enzyme scaffold. Using the AKU-associated G309R variant as a mechanistic probe, our results indicate that loss of function can arise from selective disruption of O_2 access rather than from global destabilization or overt active-site impairment. The mutation locally displaces the tunnel-associated β-hairpin containing residue 309 and disrupts the continuity of the WT O_2-access route. The identification of six symmetry-related, residue-defined, and dynamically supported O_2 pathways provides a structural link between single-residue substitutions and impaired oxygen accessibility in HGD. Beyond advancing the molecular understanding of HGD dysfunction, these findings delineate oxygen-access pathways as mutation-sensitive functional elements and support their consideration in variant stratification, structure-guided interpretation, and mechanism-informed therapeutic strategies for AKU and related metabolic disorders.

CRediT authorship contribution statement

Alfonso Trezza: Writing – original draft, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Rebecca Finetti:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Bianca Roncaglia:** Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Francesco Giuliani:** Methodology, Investigation, Formal analysis. **Francesca Furlan:** Methodology, Investigation, Formal analysis, Data curation. **Annalisa Santucci:** Supervision, Resources, Funding acquisition.

Ethics approval and consent to participate

Informed consent was obtained from all subjects involved in the study. The informed consent and protocols conformed to the standards set by the latest revision of the Declaration of Helsinki.

Funding sources

This research received no external funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank the PRIN 2022 PNRR programme (Project code: P2022RYR5W), entitled “A precision medicine ecosystem for Alkaptonuria through clinical and experimental knowledge management and sharing system for an integrated approach”. The authors would also like to express their gratitude to aimAKU (Associazione Italiana Malati di Alcaptonuria) for their invaluable support. We would like to extend a special, heartfelt thank you to A.C.D., E.D. and A.A.C., whose extraordinary cooperation, strength, and dedication were fundamental to this work. This study is dedicated to them.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jsb.2026.108341.

Data availability

No data was used for the research described in the article.

References

Abraham, M.J., et al., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2, 19–25. http://dx.doi.org/10.1016/j.softx.2015.06.001.

Abramson, J., et al., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630 (8016), 493–500. http://dx.doi.org/10.1038/s41586-024-07487-w.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

Amaya, A.A., Brzezinski, K.T., Farrington, N., Moran, G.R., 2004. Kinetic analysis of human homogentisate 1,2-dioxygenase. Arch. Biochem. Biophys. 421 (1), 135–142. http://dx.doi.org/10.1016/j.abb.2003.10.014.

Bernardini, G., et al., 2024. Alkaptonuria. Nat. Rev. Dis. Prim. 10 (1), 16. http://dx.doi.org/10.1038/s41572-024-00498-x.

Borowski, T., Georgiev, V., Siegbahn, P.E.M., 2005. Catalytic reaction mechanism of homogentisate dioxygenase: A hybrid DFT study. J. Am. Chem. Soc. 127 (49), 17303–17314. http://dx.doi.org/10.1021/ja054433j.

Braconi, D., Millucci, L., Bernardini, G., Santucci, A., 2015. Oxidative stress and mechanisms of ochronosis in alkaptonuria. Free Radic. Biol. Med. 88, 70–80. http://dx.doi.org/10.1016/j.freeradbiomed.2015.02.021.

Braconi, D., et al., 2018. Inflammatory and oxidative stress biomarkers in alkaptonuria: data from the DevelopAKUre project. Osteoarthr. Cartil. 26 (8), 1078–1086. http://dx.doi.org/10.1016/j.joca.2018.05.017.

Chovancova, E., et al., 2012. CAVER 3.0: A tool for the analysis of transport pathways in dynamic protein structures. PLoS Comput. Biol. 8 (10), e1002708. http://dx.doi.org/10.1371/journal.pcbi.1002708.

Cohen, J., Arkhipov, A., Braun, R., Schulten, K., 2006. Imaging the migration pathways for O_2, CO, NO, and Xe inside myoglobin. Biophys. J. 91 (5), 1844–1857. http://dx.doi.org/10.1529/biophysj.106.085746.

DuPai, C.D., Davies, B.W., Wilke, C.O., 2021a. A systematic analysis of the beta hairpin motif in the protein data bank. Prot. Sci. 30 (3), 613–623. http://dx.doi.org/10.1002/pro.4020.

DuPai, C.D., Davies, B.W., Wilke, C.O., 2021b. A systematic analysis of the beta hairpin motif in the protein data bank. Prot. Sci. 30 (3), 613–623. http://dx.doi.org/10.1002/pro.4020.

Fernández-Cañón, J.M., et al., 1996. The molecular basis of alkaptonuria. Nature Genet. 14 (1), 19–24. http://dx.doi.org/10.1038/ng0996-19.

Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 18 (12), 1463–1472. http://dx.doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

Hoover, W.G., 1985. Canonical dynamics: Equilibrium phase-space distributions. Phys. Rev. A 31 (3), 1695–1697. http://dx.doi.org/10.1103/PhysRevA.31.1695.

Huang, J., et al., 2017. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nature Methods 14 (1), 71–73. http://dx.doi.org/10.1038/nmeth.4067.

Isralewitz, B., Gao, M., Schulten, K., 2001. Steered molecular dynamics and mechanical functions of proteins. Curr. Opin. Struct. Biol. 11 (2), 224–230. http://dx.doi.org/10.1016/S0959-440X(00)00194-9.

Jo, S., Kim, T., Iyer, V.G., Im, W., 2008. CHARMM-GUI: A web-based graphical user interface for CHARMM. J. Comput. Chem. 29 (11), 1859–1865. http://dx.doi.org/10.1002/jcc.20945.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26 (2), 283–291. http://dx.doi.org/10.1107/S0021889892009944.

Martinez, S., Hausinger, R.P., 2015. Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. J. Biol. Chem. 290 (34), 20702–20711. http://dx.doi.org/10.1074/jbc.R115.648691.

McGibbon, R.T., et al., 2015. MDTraj: A modern open library for the analysis of molecular dynamics trajectories. Biophys. J. 109 (8), 1528–1532. http://dx.doi.org/10.1016/j.bpj.2015.08.015.

Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., Beckstein, O., 2011. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. 32 (10), 2319–2327. http://dx.doi.org/10.1002/jcc.21787.

Milella, M.S., Geminiani, M., Trezza, A., Visibelli, A., Braconi, D., Santucci, A., 2024. Alkaptonuria: From molecular insights to a dedicated digital platform. Cells 13 (12), 1072. http://dx.doi.org/10.3390/cells13121072.

Nosé, S., 1984. A molecular dynamics method for simulations in the canonical ensemble. Mol. Phys. 52 (2), 255–268. http://dx.doi.org/10.1080/00268978400101201.

Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R., De Gioia, L., 2009. Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case. J. Mol. Graph. Model. 27 (8), 889–899. http://dx.doi.org/10.1016/j.jmgm.2009.01.006.

Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 52 (12), 7182–7190. http://dx.doi.org/10.1063/1.328693.

Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G., Noé, F., 2013. Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys. 139 (1), 015102. http://dx.doi.org/10.1063/1.4811489.

Ramachandran, G.N., Ramakrishnan, C., Sasisekharan, V., 1963. Stereochemistry of polypeptide chain configurations. J. Mol. Biol. 7, 95–99. http://dx.doi.org/10.1016/S0022-2836(63)80023-6.

Schrödinger, L., 2020. The PyMOL molecular graphics system. Version 2.5.

Schultze, S., Grubmüller, H., 2021. Time-lagged independent component analysis of random walks and protein dynamics. J. Chem. Theory Comput. 17 (9), 5766–5776. http://dx.doi.org/10.1021/acs.jctc.1c00273.

Singh, S., Anand, R., 2021. Tunnel architectures in enzyme systems that transport gaseous substrates. ACS Omega 6 (49), 33274–33283. http://dx.doi.org/10.1021/acsomega.1c05430.

Solomon, E.I., Brunold, T.C., Davis, M.I., Kemsley, J.N., Lee, S.-K., Lehnert, N., Neese, F., Skulan, A.J., Yang, Y.-S., Zhou, J., 2000. Geometric and electronic structure/function correlations in non-heme iron enzymes. Chem. Rev. 100 (1), 235–350. http://dx.doi.org/10.1021/cr9900275.

Tariq, M., Özbek, P., Moin, S.T., 2022. Hydration modulates oxygen channel residues for oxygenation of cysteine dioxygenase: Perspectives from molecular dynamics simulations. J. Mol. Graph. Model. 110, 108060. http://dx.doi.org/10.1016/j.jmgm.2021.108060.

Titus, G.P., Mueller, H.A., Burgner, J., Rodríguez De Córdoba, S., Peñalva, M.A., Timm, D.E., 2000. Crystal structure of human homogentisate dioxygenase. Nature Struct. Biol. http://dx.doi.org/10.1038/76756.

Torabifard, H., Cisneros, G.A., 2017. Computational investigation of O_2 diffusion through an intra-molecular tunnel in AlkB; influence of polarization on O_2 transport. Chem. Sci. 8 (9), 6230–6238. http://dx.doi.org/10.1039/C7SC00997F.

UniProt Consortium, 2025. UniProtKB entry Q93099 (HGD_HUMAN). (Accessed 11 December 2025).

Usher, J.L., Ascher, D.B., Pires, D.E.V., Milan, A.M., Blundell, T.L., Ranganath, L.R., 2014. Analysis of HGD gene mutations in patients with alkaptonuria from the United Kingdom: Identification of novel mutations. In: Zschocke, J., Baumgartner, M., Morava, E., Patterson, M., Rahman, S., Peters, V. (Eds.), JIMD Reports, Volume 24. In: JIMD Reports, vol. 24, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 3–11. http://dx.doi.org/10.1007/8904_2014_380.

Vilboux, T., et al., 2009. Mutation spectrum of homogentisic acid oxidase (HGD) in alkaptonuria. Hum. Mutat. 30 (12), 1611–1619. http://dx.doi.org/10.1002/humu.21120.

Zatkova, A., Olsson, B., Ranganath, L.R., Imrich, R., 2022. Analysis of the phenotype differences in siblings with alkaptonuria. Metabolites 12 (10), 990. http://dx.doi.org/10.3390/metabo12100990.

Zatkova, A., et al., 2012. Identification of 11 novel homogentisate 1,2 dioxygenase variants in alkaptonuria patients and establishment of a novel LOVD-based HGD mutation database. In: JIMD Reports - Case and Research Reports, 2012/1. In: JIMD Reports, vol. 4, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 55–65. http://dx.doi.org/10.1007/8904_2011_68.