From BioEmu Monomer Ensembles to AMBER-Ready Holo Models: Subunit-Aware Quality Control of a Cofactor-Rich Aldehyde Oxidoreductase

Piotr Dusza[1(✉)], Marek Wodziński[1,2], and Maciej Szaleniec[3]

[1] Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, AGH University of Krakow, Kraków, Poland
pdusza@student.agh.edu.pl
[2] Sano Centre for Computational Medicine, Kraków, Poland
[3] Jerzy Haber Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, Kraków, Poland

Abstract. Generative artificial intelligence (AI) models are emerging as tools for sampling protein conformational ensembles and providing starting structures for downstream molecular simulation. In cofactor-rich metalloenzymes, however, a plausible peptide chain is not enough, as the structure must also support successful reconstruction of cofactors, metal centers, protonation states, and solvent. Here, we present a subunit-aware workflow that converts BioEmu monomer ensembles into solvated, minimized holo models for the three subunits of the tungsten-containing aldehyde oxidoreductase from *Aromatoleum aromaticum* (AorA, AorB, and AorC). The workflow combines Cα-based alignment to the experimental reference with subunit-specific holo reconstruction, solvent preparation, staged minimization, outlier detection, principal component analysis (PCA), and HDBSCAN clustering.

The workflow started with requesting 100 BioEmu conformations per subunit, of which 96 AorA, 91 AorB, and 100 AorC structures remained after BioEmu's built-in filtering of steric clashes and chain discontinuities. At the end of the workflow, final no-flag retention was 79%, 73%, and 89% relative to the requested ensembles (82.3%, 80.2%, and 89.0% of the BioEmu-delivered sets). After minimization, the retained structures remained distinct from the cryo-EM reference, with median final Cα root-mean-square deviation (RMSD) values of 6.60, 1.23, and 2.72 Å for AorA, AorB, and AorC, respectively. At the same time, chemistry-centered metrics remained close to the reference-holo controls. PCA/HDBSCAN analysis showed the clearest density-defined ensemble structure for AorA, intermediate behavior for AorC, and the most diffuse retained ensemble for AorB. Together, these results support a practical workflow in which AI-derived conformations are screened after reconstruction and minimization to identify chemically plausible starting structures for subsequent simulation.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16765, pp. 193–211, 2027.
https://doi.org/10.1007/978-3-032-30536-7_13

Keywords: BioEmu · metalloenzymes · aldehyde oxidoreductase · holo reconstruction · molecular simulation · computational workflow

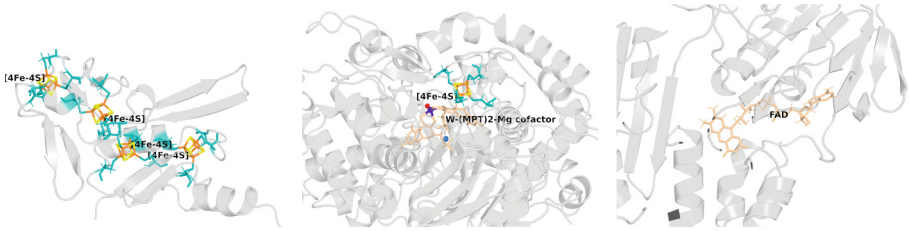
1 Introduction

In recent years, the focal point of structural modeling has shifted into a new direction. For many proteins, obtaining a single folded structure is no longer the primary challenge; instead, it is equally crucial to define a physically realistic ensemble and translate that ensemble into chemical models suitable for simulation [5,12,21]. BioEmu is emblematic of this transition, as it was developed to emulate protein equilibrium ensembles and to generate many statistically independent monomer conformations at a much lower cost compared to large-scale molecular dynamics (MD) simulations [13]. Such models are especially attractive as upstream generators of initial conditions because they can place later atomistic simulations in multiple regions of conformational space instead of forcing every trajectory to start from one experimental conformation [13,17].

However, that promise becomes harder to realize in cofactor-rich metalloenzymes, where a plausible protein backbone is only the starting point. In such systems, chemically important interactions between cofactors and the surrounding protein environment must be preserved during reconstruction. Protonation assignment, local cofactor geometry, explicit metal coordination, structurally retained water molecules, and topology-building choices such as bond definitions, residue templates, atom typing, and protonation-state mapping all influence whether the reconstructed system can be minimized without severe nonphysical local strain and whether it remains chemically interpretable as a starting point for MD [14,15]. These challenges are especially important in systems that combine multiple prosthetic groups and metal-centered chemistry, because even relatively small local distortions can alter coordination geometry, electron-transfer pathways, and access to the catalytic region [20,24].

As a demanding example of such a system, we selected the aldehyde oxidoreductase from *Aromatoleum aromaticum* (AOR_{Aa}). Its experimentally determined structure comprises a chemically heterogeneous assembly in which AorA forms an Fe–S-rich electron-transfer subunit, AorB contains the tungsten bis-metallopterin-magnesium catalytic cofactor together with a proximal [4Fe–4S] cluster, and AorC contains a bound flavin adenine dinucleotide (FAD) cofactor and a nicotinamide adenine dinucleotide (NAD^+) binding site, as shown on Fig. 1 [3,24]. The enzyme therefore combines metal centers and organic cofactors linked to the protein scaffold through both coordination-dependent and non-covalent interactions within one three-dimensional architecture. It is therefore well suited for examining the gap between a plausible AI-generated starting chain and a chemically explicit holo model that remains suitable for downstream simulation.

This study develops a practical methodological workflow for screening AI-derived monomer conformations as chemically acceptable starting points for simulation in a cofactor-rich oxidoreductase system. Using the three AOR subunits AorA, AorB, and AorC as a benchmark, we evaluate how subunit-specific



(a) AorA (b) AorB (c) AorC

Fig. 1. Structural overview of the cofactor architecture in the three AOR subunits used as the benchmark system in this study. (a) AorA, showing the Fe–S relay and the coordinating cysteine environment. (b) AorB, showing the tungsten-bis-MPT active site together with the associated Mg ion and the proximal [4Fe–4S] cluster. (c) AorC, showing the bound FAD cofactor in the flavin-containing subunit. Together, these panels illustrate the chemically heterogeneous architecture that makes AOR_{Aa} a demanding test case for reconstruction, solvent preparation, and staged minimization.

holo reconstruction and staged energy minimization affect the survival of AI-initialized conformations under conservative quality-control criteria. The analysis integrates subunit-specific reconstruction, reference-holo control models, Quality Control (QC) filtering, and density-based ensemble analysis to identify structures that remain chemically plausible after preparation. In this way, the study defines a practical screening framework for AI-derived starting structures in chemically complex oxidoreductase systems.

2 Materials and Methods

2.1 System Definition and Reference Structures

All models were based on the cryo-EM structure of AOR_{Aa} deposited as PDB 8C0Z [24]. Because BioEmu currently samples monomeric protein chains rather than whole multimeric assemblies, and returns structures in a backbone-frame representation that requires downstream side-chain reconstruction [13], the three AOR subunits were modeled separately. This was also the most sensible chemical choice, because each subunit required a different procedure for rebuilding its holo state before AMBER preparation.

BioEmu sampling was performed independently for the FASTA sequence of each subunit. For each subunit, 100 conformations were requested, yielding 96 exported structures for AorA, 91 for AorB, and 100 for AorC after BioEmu's built-in physicality filtering [13]. Detailed sampling settings are given in Appendix A.1. These structures were then converted into per-frame PDB files and rigidly aligned to the corresponding chains from the 8C0Z reference model.

2.2 Alignment and Subunit-Aware Reconstruction

Structural alignment was carried out in Python using MDAnalysis, following a Cα-based fit for both individual structures and the corresponding ensemble

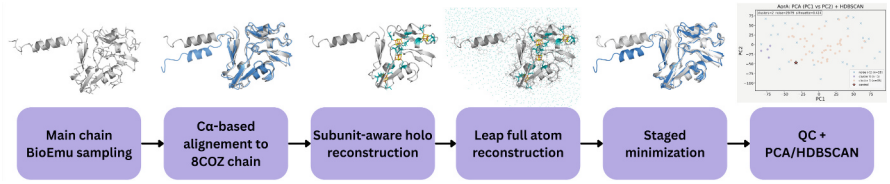


Fig. 2. Overview of the subunit-aware BioEmu-to-AMBER workflow used in this study. Presented above is the AorA subunit (white - BioEmu conformation; blue - Reference), while all three AOR subunits share the same high-level preparation logic, they differ in holo-restoration chemistry and in the need for local pre-relaxation before full minimization. (Color figure online)

sets [10,18]. In the final QC stage, the common alignment maps comprised 142 positions for AorA, 612 positions for AorB, and 424 positions for AorC. Positions absent from the mobile chain relative to the reference numbering were excluded from the common coordinate space.

Once aligned, each subunit underwent a separate holo-restoration procedure (Table 1). This stage not only restored the subunit-specific cofactors and metal centers, but also converted the BioEmu backbone-frame representation into a full-atom protein model during LEaP-based preparation, including side-chain reconstruction for the protein scaffold. AorA required rebuilding of the Fe–S electron transfer chain, including restoration of the [4Fe–4S] clusters and explicit definition of the coordinating cysteines so that the intended Fe coordination pattern was preserved during topology generation. AorB was the most demanding subunit from the chemical point of view. Besides the proximal [4Fe–4S] cluster, its preparation also required reconstruction of the W-bis-MPT-Mg center and structurally retained water molecules near the catalytic site. AorC was treated as a flavoprotein reconstruction problem, with the bound FAD cofactor transferred from the reference structure into the aligned subunit and then prepared with AMBER-compatible parameters.

Detailed protonation, force-field, solvation, and cofactor-parameterization settings are given in Appendix A.2.

2.3 Staged Minimization and Reference-Holo Controls

Energy minimization was carried out in explicit solvent using a staged, subunit-aware AMBER protocol. AorA and AorB were minimized in seven stages (`min0`–`min6`), beginning with a local Fe–S-region pre-relaxation step, whereas AorC followed the same overall logic but without `min0`. Detailed engine settings, restraint definitions, stage parameters, and software versions are given in Appendix A.3.

Reference-holo controls were prepared separately for each subunit using the corresponding experimental coordinates rather than BioEmu-derived backbones. These controls matter for analysis because they separate the accuracy of the pipeline from the variability of the input states. If the same reconstruction and minimization logic preserves the experimental structure but the AI-derived states

Table 1. Subunit scope and preparation logic used in the present AOR benchmark. "BioEmu-delivered" denotes the number exported after the model's built-in physicality filtering. "Retained" denotes the final no-flag set after downstream preparation, minimization, and QC, corresponding to 82.3%, 80.2%, and 89.0% of the BioEmu-delivered sets.

Subunit	Restored chemistry	Minimization schedule	Requested	BioEmu-delivered	Retained
AorA	Four Fe–S clusters and ligating cysteine network	min0–min6	100	96	79
AorB	W-bis-MPT, Mg, proximal Fe–S cluster, catalytic waters	min0–min6	100	91	73
AorC	Bound FAD cofactor and NAD-binding site	min1–min6	100	100	89

remain displaced, the displacement is attributable to alternative starting conformations rather than to a broken preparation protocol.

2.4 Quality-Control Metrics

The QC stage evaluated both global structural behavior and subunit-specific local chemistry. For each minimized system after alignment to the reference chain, we extracted final Cα RMSD, change in RMSD during minimization (ΔRMSD), radius of gyration (R_g) at the start and end of minimization, and final AMBER potential energy normalized by atom count. Here, R_g served as a coarse measure of overall compactness or mass distribution, and ΔR_g denotes the end-minus-start change during minimization. Outliers in energy-per-atom and final Cα RMSD were identified using a median absolute deviation (MAD)-based modified z-score threshold of 3.5 [11].

Additional chemistry-specific filters were applied in a subunit-dependent manner. For AorA and AorB, the reconstructed [4Fe–4S] regions were evaluated using the minimum and maximum Fe–S distances observed after minimization, and a frame was rejected if any Fe–S distance lay outside 2.05–2.45 Å. For AorB, the tungsten site was further assessed using the minimum bond-like W–S distance, with an admissible range of 2.20–2.60 Å. These bounds were derived from separate calibration analyses of metal–sulfur bonding, based on bond-energy calculations and the corresponding bond-length histograms. For AorC, FAD pose RMSD relative to the reference was tracked as a descriptive measure of cofactor placement but was not used as a hard rejection criterion.

The QC thresholds were chosen based on reference-holo baselines, empirical inspection of the minimized distributions, and conservative tolerance windows intended to exclude obvious geometric failures without forcing close agreement

with the control medians. For example, the reference-holo controls gave final minimum Fe–S distances of 2.18 Å for AorA and 2.18 Å for AorB, while retained structures clustered closely around similar values. Likewise, the AorB control gave a minimum-near W–S distance of 2.45 Å, and retained AorB structures remained centered near that value. The adopted Fe–S and W–S windows were therefore broad enough to permit physically reasonable relaxation while excluding clear over-compression or loss of near-contact coordination.

A structure was classified as a *keeper* if it had no active QC flags. The final retained set is therefore intentionally conservative and consists of systems that remained both energetically unremarkable and locally chemically plausible under the chosen criteria.

2.5 PCA and HDBSCAN Analysis

To assess the geometry of the retained ensembles, the common-reference Cα coordinates of all keeper structures in each subunit were projected into principal-component space. The first three PCs were used as input to Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) with a minimum cluster size defined as $\max(4, \lceil 0.03n \rceil)$, where n is the number of keeper structures in the subunit. HDBSCAN was chosen because it does not force all points into clusters and can therefore represent a mixture of dense basins and diffuse background occupancy [16]. Cluster quality was evaluated using the silhouette score on non-noise points and the density-based clustering validity index (DBCV), for which larger positive values indicate stronger density-defined separation [19]. Because HDBSCAN labels noise points explicitly, the silhouette score should be interpreted here as a measure of separation within the clustered subset rather than as a global quality measure when noise fractions differ strongly between subunits. Reference-holo controls were projected into the same PC space to allow direct comparison between experimentally anchored and AI-derived states. Additional implementation details are provided in Appendix A.4.

3 Results

3.1 QC Retention and Rejection Profiles Across Subunits

For each subunit, 100 conformations were requested from BioEmu. After BioEmu's built-in filtering of physically implausible samples, 96 AorA, 91 AorB, and 100 AorC structures entered the downstream preparation workflow (Fig. 3). The final no-flag retention rates therefore corresponded to 79%, 73%, and 89% relative to the originally requested ensembles, or 82.3% (79/96), 80.2% (73/91), and 89.0% (89/100) relative to the BioEmu-delivered minimized sets. Although the initial reduction at the BioEmu stage was limited, 4% for AorA and 9% for AorB, with no losses for AorC, it is important to distinguish this internal physicality filter from the subsequent QC procedure. The former excludes clearly problematic outputs (such as chain-break or clash-prone outputs) before export, whereas the latter evaluates whether a delivered backbone remains chemically

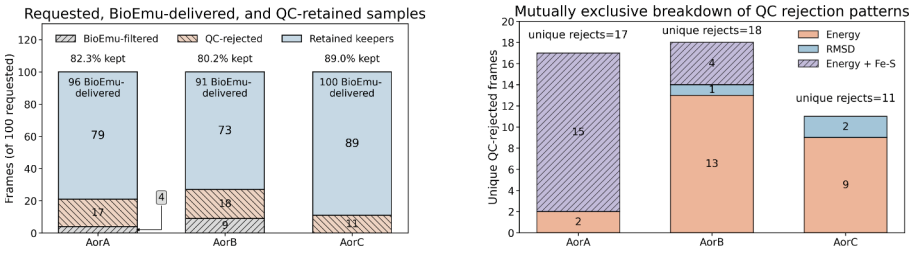


Fig. 3. QC outcomes across the three AOR subunits. Left: bars are normalized to 100 requested BioEmu samples per subunit and show BioEmu-side filtering, downstream QC rejections after preparation/minimization, and final retained keepers. Right: mutually exclusive breakdown of unique QC rejections, with each rejected frame counted once only.

and energetically acceptable after subunit-specific holo reconstruction, solvation, and minimization [13].

Within the downstream QC stage, energy-based outlier detection was the dominant rejection mode in all three subunits. AorA and AorB each showed 17 energy flags, whereas AorC showed 9. However, because some rejected structures carried more than one flag, per-category counts are not equivalent to numbers of rejected structures. In AorA, 15 of the 17 rejected structures combined energy and Fe–S-related abnormalities, whereas 2 were rejected on the basis of energy alone. In AorB, 13 of the 18 rejected structures were energy-only outliers, 1 was rejected on RMSD alone, and 4 showed a combined energy-plus-Fe–S pattern. In AorC, the rejected set consisted of 9 energy-only and 2 RMSD-only cases. Thus, the mutually exclusive rejection profiles differed across subunits, with AorA dominated by combined energy/Fe–S failures, AorB mainly by energy-only outliers, and AorC by energy-only plus a small number of RMSD-only cases.

The subunits also differed in the main features associated with rejection. In AorA, rejected structures showed clear abnormalities in the local Fe–S region, as the median final minimum Fe–S distance in the rejected subset was 1.52 Å, compared with 2.17 Å for the retained structures. In AorB, rejected structures remained very similar to the retained set in both Fe–S and W–S metrics. AorC showed a third pattern, where the rejected structures had a lower median FAD pose RMSD than the retained set. Thus, the three subunits differed not only in retention, but also in the combination of local and global features associated with rejection.

3.2 Structural and Energetic Effects of Minimization

After minimization, the retained structures remained distinct from the experimental reference at the backbone level (Table 2, Fig. 4). The median final Cα RMSD values for the retained sets were 6.60 Å for AorA, 1.23 Å for AorB, and 2.72 Å for AorC. By contrast, the corresponding reference-holo controls remained

essentially identical to the reference in all three cases, with final RMSD values of approximately 5×10^{-4} Å.

The energy distributions support the same general distinction between retained and rejected models. Within each subunit, the retained structures occupied relatively narrow ranges of favorable per-atom energies, whereas the rejected sets included the high-energy tail associated with unresolved strain or otherwise unfavorable geometry (Fig. 4). This separation between retained and rejected energy ranges was clearest in AorA and AorC.

Table 2. Summary of the retained (*keeper*) ensembles and the subunit-specific reference-holo controls. Reported keeper values are medians with interquartile ranges in brackets. Chemistry-aware values are stated only for the metrics relevant to a given subunit.

(a) Retained ensembles

Subunit	n	$\mathrm{RMSD}_{\mathrm{end}}$ (Å)	ΔRMSD (Å)	ΔR_g (Å)	E/N (kcal/mol/atom)	Chemistry-aware metric
AorA	79	6.60 [4.94, 9.00]	0.12 [0.05, 0.16]	0.18 [0.12, 0.24]	−4.15 [−4.16, −4.13]	Fe–S_{min}: 2.17 [2.17, 2.18]
AorB	73	1.23 [1.17, 1.29]	0.19 [0.17, 0.22]	0.39 [0.36, 0.41]	−3.84 [−3.84, −3.83]	Fe–S_{min}: 2.19 [2.18, 2.19] W–S_{min}: 2.45 [2.43, 2.45]
AorC	89	2.72 [2.48, 2.94]	0.15 [0.12, 0.17]	0.11 [0.06, 0.17]	−3.99 [−4.00, −3.95]	FAD pose RMSD: 1.85 [1.58, 2.02]

(b) Reference-holo controls

Subunit	n	$\mathrm{RMSD}_{\mathrm{end}}$ (Å)	ΔRMSD (Å)	ΔR_g (Å)	E/N (kcal/mol/atom)	Chemistry-aware metric
AorA control	1	0.0005	−0.79	−0.33	−4.15	Fe–S_{min}: 2.18
AorB control	1	0.0005	−0.91	−0.62	−3.91	Fe–S_{min}: 2.18 W–S_{min}: 2.45
AorC control	1	0.0005	−0.82	−0.45	−4.05	FAD pose RMSD: 0.0005

The same general picture is supported by the signs of ΔRMSD and ΔR_g. For retained structures, the median ΔRMSD was positive in all three subunits: +0.12 Å for AorA, +0.19 Å for AorB, and +0.15 Å for AorC. The median change in radius of gyration was likewise positive, at +0.18 Å, +0.39 Å, and +0.11 Å, respectively. These positive median shifts indicate that minimization did not simply return the retained models toward the experimental backbone. The largest median ΔR_g was observed for AorB (+0.39 Å), indicating the clearest shift in overall compactness between the pre-minimization and post-minimization structures.

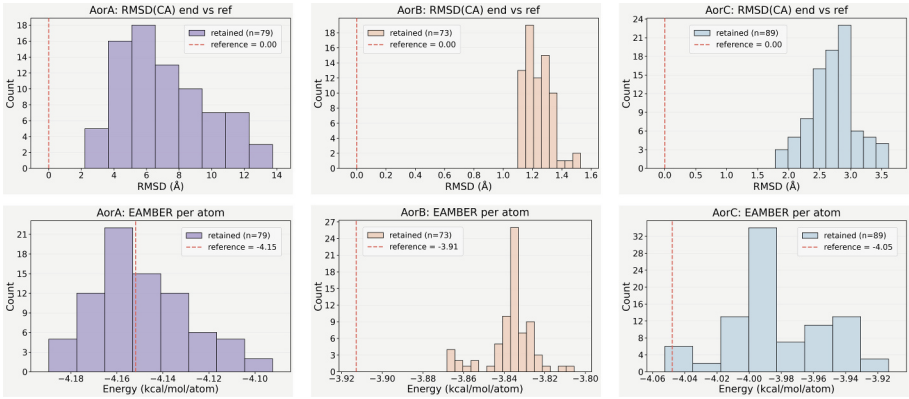


Fig. 4. Top row: distributions of final Cα RMSD to the reference. Bottom row: distributions of final AMBER potential energy per atom. Vertical lines correspond to the subunit-specific reference-holo controls.

3.3 Local Cofactor and Metal-Center Geometry Relative to the Reference-Holo Baseline

The chemistry-centered metrics help to interpret whether the retained structures remained locally plausible after minimization (Fig. 5). In AorA, the median final minimum Fe–S distance in the retained set was 2.17 Å, very close to the control value of 2.18 Å. In AorB, the corresponding median Fe–S minimum distance was 2.19 Å, while the median W–S minimum-near distance was 2.45 Å, again close to the control values of 2.18 and 2.45 Å, respectively. These results indicate that, for the retained structures, the local metal-centered geometry remained close to the reference-holo baseline after minimization, even when the protein backbone did not necessarily return to the experimental conformation.

AorC showed a different but equally informative behavior. The retained structures had a median FAD pose RMSD of 1.85 Å, with an interquartile range of 1.58 to 2.02 Å (Fig. 5). In protein-ligand benchmarking, heavy-atom pose RMSD values around 2 Å are commonly used as an indicator of near-native or chemically useful poses [1,23]. Although FAD is not a docked drug-like ligand and the present workflow involves protein relaxation rather than docking refinement, this scale remains useful as a point of reference. Thus, the retained AorC structures remained centered around a narrow range of FAD pose RMSD values.

3.4 Subunit-Specific Differences in Density-Based Ensemble Structure

The retained ensembles differed clearly in their geometric organization in PC space (Table 3, Fig. 6). AorA showed the strongest density-defined clustering, with two HDBSCAN clusters, 36.7% noise, a non-noise silhouette score of 0.424, and a full DBCV of 0.257. AorC also yielded two clusters, but with a substantially

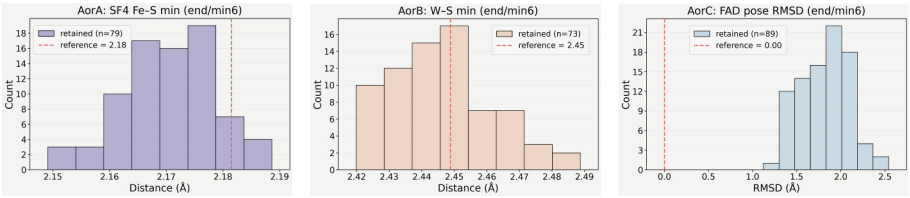


Fig. 5. Subunit-specific chemistry-aware metrics for the retained ensembles. Left: AorA minimum Fe–S distance in the reconstructed [4Fe–4S] relay. Middle: AorB minimum-near W–S distance in the tungsten site. Right: AorC FAD pose RMSD after alignment. In all panels, the vertical line marks the reference-holo control.

higher noise fraction (53.9%) and a lower full DBCV of 0.071. AorB showed the weakest global clustering structure, with 52.1% noise and a full DBCV of 0.043. Together, these metrics indicate a progression from the most compact retained ensemble in AorA to the most diffuse one in AorB, consistent with weaker density-defined separation as quantified by DBCV [19].

The variance explained by the first three principal components followed the same ordering. In AorA, the first three PCs accounted for 88.2% of the variance in the retained set. In AorC, the corresponding value was 60.4%, whereas in AorB the first three PCs explained 18.8% of the variance. This pattern is consistent with progressively more distributed structural heterogeneity from AorA to AorB.

Table 3. PCA/HDBSCAN summary for retained structures. Larger positive DBCV values indicate stronger density-defined cluster structure. The silhouette score was computed only on non-noise points. Cluster sizes and cluster-wise DBCV values are reported for the two clusters detected in each subunit.

Subunit	Kept	Clusters	Noise (%)	Silhouette (non-noise)	DBCV $_{\text{full}}$	Cluster 1 (n, DBCV)	Cluster 2 (n, DBCV)	PC1 / PC2 / PC3 variance (%)
AorA	79	2	36.7	0.424	0.257	5, 0.634	45, 0.380	43.1 / 31.9 / 13.2
AorB	73	2	52.1	0.332	0.043	6, 0.250	29, 0.056	8.2 / 6.0 / 4.6
AorC	89	2	53.9	0.452	0.071	13, 0.168	28, 0.146	28.6 / 21.2 / 10.6

Cluster-medoid representatives of the retained ensembles are shown in Fig. 7, providing structural exemplars of the density-defined regions identified by PCA/HDBSCAN and allowing visual comparison of retained conformational spread across the three subunits.

4 Discussion

4.1 MD-Readiness Beyond Structural Similarity

MD-readiness was defined here in operational rather than purely structural terms. Although RMSD to the experimental reference was tracked throughout

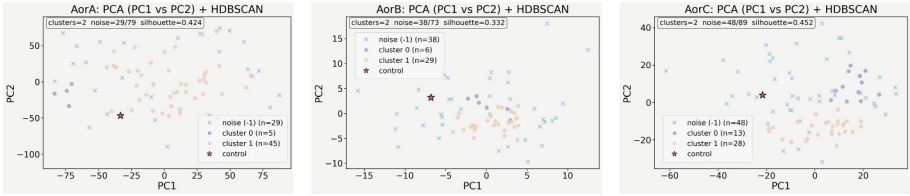
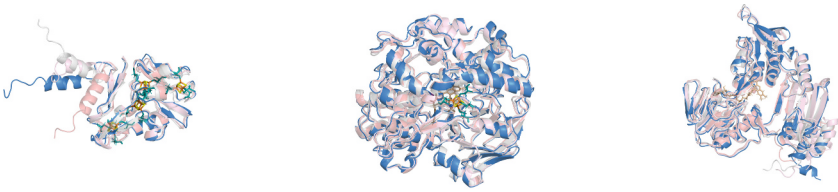


Fig. 6. PCA (PC1 vs. PC2) and HDBSCAN assignments for retained structures. Crosses indicate points labeled as noise, colored circles indicate cluster members, and the star denotes the projected reference-holo control. AorA shows the clearest density-structured separation, AorC is intermediate, and AorB shows the most diffuse retained ensemble with a large noise component.



(a) AorA (b) AorB (c) AorC

Fig. 7. Cluster-medoid representatives for AorA, AorB, and AorC, overlaid with the corresponding minimized reference structures. Colors are consistent across subunits: blue, reference; light pink, noise medoid; salmon, cluster 0 medoid; white, cluster 1 medoid. Cofactors and metal centers are shown from the reference structures. (Color figure online)

the analysis, a low RMSD was not required for a structure to be considered suitable for further simulation. Instead, a structure was classified as MD-ready if it could be reconstructed as a chemically complete holo model, minimized in explicit solvent without evident energetic or geometric abnormalities, and retained by the subunit-specific QC criteria. Under this definition, minimization functions primarily as a physical screening step rather than as a procedure expected to recover the experimental backbone.

This distinction is important for interpretation of the retained ensembles. As shown in the Results, minimization did not simply collapse the accepted models toward the cryo-EM reference. The consistently positive median ΔRMSD and ΔR_g values indicate instead that minimization relieved unfavorable contacts introduced during reconstruction and solvent preparation while preserving much of the conformational character of the starting models. This effect was especially evident in AorB, which showed the largest median ΔR_g, suggesting the clearest systematic shift in overall compactness between the pre-minimization and post-minimization states. At the same time, the reference-holo controls remained essentially coincident with the experimental structure after the same preparation and minimization procedure. This indicates that the protocol itself does not

impose substantial structural drift when applied to a reference-consistent starting state. The deviations observed in the retained BioEmu-derived structures should therefore not be interpreted as preparation artifacts alone.

Importantly, the final minimized coordinates reflect not only the conformational variation introduced by BioEmu, but also downstream side-chain reconstruction, protonation assignment, cofactor and ion transfer, solvation, and local relaxation during minimization. The retained sets therefore represent conformations whose backbone-level diversity survives chemically explicit reconstruction and minimization rather than being erased by the preparation workflow. In this sense, MD-readiness in the present study refers not to close structural agreement with the experimental reference, but to the ability of an alternative starting conformation to remain chemically plausible after full holo restoration and solvent relaxation.

As an additional practical check, selected representative retained structures were propagated for 51 ns in explicit-solvent MD and remained stable, without evident cofactor disruption, structural collapse, or loss of metal-center geometry. These runs support the practical MD compatibility of selected representatives, while broader replicate simulations would be needed for ensemble-wide dynamical validation.

4.2 Subunit-Specific Differences in Physicochemical Behavior After QC

The three subunits define a clear hierarchy in terms of preparation difficulty and post-QC ensemble behavior, where AorC was the most straightforward system to prepare. It showed the highest retention rate, the simplest rejection profile, and a cofactor-restoration problem that remained relatively local. After FAD placement and minimization, the retained ensemble preserved a broad range of conformations while remaining chemically reasonable. Although FAD pose RMSD was not used as a hard rejection criterion, the retained AorC structures remained centered around a narrow range of values, indicating that the cofactor occupied a reproducible binding mode related to the reference geometry even when the backbone did not return to the cryo-EM state. In practical terms, this makes AorC the most accessible starting point for subsequent representative-structure selection and trajectory-level stability checks.

AorA on the other hand, required stricter local control. Although many retained AorA structures remained substantially displaced from the reference at the backbone level, the acceptable Fe–S geometry was confined to a narrow range close to the control baseline. This behavior is consistent with the strongly constrained bonding pattern between the coordinating cysteines and the Fe atoms. In contrast, rejected AorA structures were characterized by unrealistically short minimum Fe–S distances and, in many cases, unfavorable energies. This indicates that the main sensitivity of AorA lies in the local metal-cluster region rather than in global fold variation. Therefore, the dedicated `min0` stage appears methodologically justified, as it allows local relaxation of the Fe–S environment before the broader minimization schedule is applied.

AorB was the most demanding subunit overall, as it combines the proximal Fe–S cluster with the W-bis-MPT center, associated magnesium, and structurally retained waters within a comparatively large catalytic framework. In this case, rejection was not explained primarily by the local metal-center metrics, since the retained and rejected structures remained similar in both Fe–S and W–S measures. Instead, AorB was characterized mainly by a diffuse post-QC ensemble, weak global clustering, and rejection driven largely by energy outliers together with a small number of RMSD outliers. The clustering analysis refines this picture further, as although the full DBCV value was low and just over half of the retained structures were labeled as noise, two local clusters were still detected, indicating that broad structural diffuseness coexists with limited pockets of more coherent organization. AorB should therefore be interpreted not as a failed system, but as the broadest and least compact of the retained ensembles, requiring more careful representative selection than AorA or AorC.

Taken together, these contrasts show that local cofactor- or metal-centered geometry and whole-structure energetic plausibility are related, but not interchangeable, indicators of model quality. A structure can remain locally acceptable yet still belong to a diffuse or energetically problematic ensemble, and conversely, global deviation from the experimental reference does not by itself imply failure of local chemical reconstruction.

4.3 Role of the Reference-Holo Controls

The reference-holo controls provide an essential baseline for interpreting the retained BioEmu-derived ensembles. Because they remain stable under the same preparation and minimization protocol, the deviations observed in the retained BioEmu-derived structures cannot be attributed solely to artifacts of LEaP assembly or minimization. A more plausible interpretation is that these structures represent alternative conformations that remain chemically viable after holo restoration and solvent relaxation.

Furthermore, the controls are also important for interpreting local chemistry. In AorA and AorB, the accepted structures retained Fe–S and W-site metrics that remained close to the corresponding control values despite larger backbone-level deviations. In AorC, the same general pattern was observed for the FAD cofactor, whose pose remained confined to a relatively narrow range even when the protein backbone did not recover the reference state exactly.

More broadly, the controls help distinguish between protocol reliability and input-state variability. They show that the workflow is capable of preserving a reference-consistent structure, while also providing a chemically anchored baseline against which AI-derived alternative states can be judged. This makes them more than a technical sanity check: they are part of the interpretive framework needed to decide whether a prepared model is merely different from the reference or chemically problematic.

4.4 Implications for Simulation Workflow Design

The present results suggest a practical workflow for preparing AI-derived starting structures for simulation in cofactor-rich enzymes. Generative models are useful for introducing backbone-level conformational diversity, but this alone is not sufficient to produce simulation-ready systems. Cofactors, metal centers, associated ions, and structurally important waters must first be restored, followed by explicit-solvent relaxation and QC-based screening. In this sense, the workflow is transferable mainly as a preparation and QC template: the same logic can be adapted to other cofactor- or metal-dependent enzymes, provided that system-specific cofactor parameters, metal-center definitions, reference geometries, and chemistry-aware QC thresholds are available. A suitable workflow therefore consists of three stages: generation of diverse starting conformations, reconstruction of chemically complete holo models, and explicit-solvent relaxation followed by QC-based screening.

The results also argue against applying a uniform filtering strategy across all subunits. Different subunits showed different dominant sensitivities, and these were not captured equally well by any single metric. A single global RMSD threshold would be too restrictive for AorA, where large backbone differences can coexist with acceptable local chemistry, whereas a purely cofactor-centered filter would not capture the broader post-QC behavior observed for AorB. The most appropriate approach is therefore a multi-criterion QC framework that combines general whole-system indicators, such as energy and backbone geometry, with subunit-specific measures focused on the relevant cofactors or metal centers.

The ensemble analysis also has practical implications for representative selection. In AorA, where the retained ensemble is more strongly structured in principal-component space, a smaller number of representative structures may capture much of the retained conformational variation. In AorB, broader coverage is likely to be necessary because the post-QC ensemble remains more diffuse and less compressible. AorC lies between these two cases. This means that even after QC, the effort required to design a representative starting set may remain subunit-dependent.

4.5 Limitations

Several limitations of the present study should be noted. First, the analysis was carried out at the level of individually prepared AorA, AorB, and AorC subunits rather than for the complete quaternary AOR structure described experimentally [24]. Second, the study focused primarily on structure preparation, minimization, and post-minimization QC, rather than on production-scale MD simulations or mechanistic analysis of catalysis. Although selected representative retained structures were propagated for 51 ns and remained stable without evident disruption of the reconstructed cofactor or metal-center environments, these simulations were used as practical stability checks rather than as a statistically complete validation of the full retained ensembles. Third, the clustering analysis was used as a descriptive tool for assessing the geometric organization

of the retained ensembles and was not intended to provide a rigorous estimate of equilibrium populations or transition kinetics.

An additional limitation is that the QC framework was designed primarily as a conservative screening procedure. As a result, the accepted structures can be interpreted more directly than the rejected ones: the former satisfy the energetic and geometry-based criteria adopted here, whereas the latter represent a heterogeneous set of models that failed one or more filters but were not analyzed further in mechanistic detail. Within these boundaries, the results still show that AI-derived monomer conformations can be converted into chemically plausible holo starting structures for a cofactor-rich oxidoreductase system with nontrivial requirements for Fe–S reconstruction, tungsten-site preparation, and FAD placement.

5 Conclusion

This study presented a subunit-specific BioEmu-to-AMBER workflow for the preparation of AorA, AorB, and AorC from the cofactor-rich aldehyde oxidoreductase of *A. aromaticum*. The results show that successful preparation is better assessed in terms of post-reconstruction physical and chemical plausibility than by simple recovery of the experimental reference backbone. Across all three subunits, the retained structures remained diverse after minimization while preserving local cofactor- or metal-centered geometry that stayed close to the corresponding reference-holo controls.

The three subunits also differed clearly in their preparation behavior. AorC was the most straightforward system, AorA remained globally diverse but could be prepared successfully when the Fe–S region was handled explicitly, and AorB showed the broadest and least compact retained ensemble after QC. Taken together, these results indicate that AI-derived conformational diversity can be used productively in cofactor-rich enzyme modeling, provided that it is combined with explicit holo reconstruction and subunit-specific QC.

Future work should extend this analysis to broader MD validation and to the larger AOR assembly. The representative MD checks performed support the practical stability of selected retained models, but a full dynamical comparison across retained ensembles remains a separate downstream objective. More generally, the same workflow may be applicable to other cofactor- or metal-dependent enzymes for which successful simulation setup requires careful reconstruction of the local chemical environment, including cofactors, metal centers, associated ions, and structurally retained waters.

Acknowledgments. The authors acknowledge the partial financial support of the Polish National Science Centre (OPUS26 2023/51/B/ST4/01224) and the European Union's HORIZON-EIC-2023- PATHFINDEROPEN-01 under grant agreement no. 101129798, European Innovation Council Pathfinder project W-BioCat, as well as financial support of high-performance computing infrastructure PL-Grid (HPC Center: ACK Cyfronet AGH, grant PLG/2025/018449) and from the AGH University of Krakow under the Excellence Initiative Research University (IDUB) programme.

Disclosure of Interests. The authors declare no competing interests relevant to the content of this article.

A Technical Supplementary Methods

A.1 BioEmu Sampling Details

BioEmu sampling was run independently for the FASTA sequence of each AOR subunit using model `bioemu-v1.2` with a fixed base seed of 1770289950782386813. For each subunit, 100 conformations were requested. No steering options were applied, and the `filter_samples` argument was not overridden, so BioEmu's default physicality filtering remained active [13]. The resulting exported sets comprised 96 structures for AorA, 91 for AorB, and 100 for AorC. The overall workflow and subunit scope are summarized in Fig. 2 and Table 1.

A.2 AMBER Preparation Details

Protonation states were assigned with H++ at pH 8.0, after which histidine protonation states were mapped to the corresponding AMBER residue names where needed [2,9]. LEaP assembly used `ff03.r1` for the protein, `gaff2` for nonstandard organic cofactors, and TIP3P water [6]. AorA was solvated with a 12.0 Å buffer, whereas AorB and AorC used a 10.0 Å buffer. Sodium counterions were added as needed for charge neutralization [6–8]. Custom templates, point charges, and bonded parameters for Fe–S clusters were based on the parameterization reported by Carvalho and Swart [4]. W-bis-MPT/Mg templates, partial charges, and bonded parameters were used as supplied from in-house files generated by M. Szaleniec and co-workers using density functional theory (DFT) calculations, following a protocol related to that described in Vilela-Alves et al. [22]. These files were used as supplied for the present benchmark and were not re-parameterized within this study.

A.3 Detailed Minimization Protocol

Energy minimization was performed in explicit solvent with `pmemd.cuda` under periodic boundary conditions at constant volume (`ntb = 1`) using a 10.0 Å nonbonded cutoff. For AorA and AorB, an additional Fe-S-specific pre-relaxation step was applied before the common minimization protocol. In this step (`min0`), the Fe-S cluster, coordinating cysteine residues, and residues within ±2 sequence positions of those cysteines were left unrestrained, whereas the remaining non-ion solute atoms were restrained with a force constant of 100 kcal mol^{-1} $\mathrm{Å}^{-2}$. Solvent molecules and counterions were not restrained. The purpose of this step was to relax the rebuilt FeS coordination region while limiting larger-scale displacement of the surrounding solute. The `min0` step should therefore be interpreted as a local pre-relaxation step rather than as the first point of a monotonic restraint-release schedule. Its output was then subjected to the common restrained-solute

minimization sequence, in which the restraint weight was reduced from 500 to 250, 100, and 10 kcal mol^{-1} $Å^{-2}$ during `min1`–`min4`, followed by two unrestrained minimization stages, `min5` and `min6`. AorC was minimized with the same protocol but without `min0`, since it does not contain an FeS center.

Stages `min0`–`min5` used mixed steepest-descent/conjugate-gradient minimization (`ntmin=1`), whereas `min6` used steepest descent only (`ntmin=2`). The corresponding numbers of minimization cycles were 20,000, 50,000, 50,000, 20,000, 50,000, 100,000, and 300,000 for `min0`–`min6`, respectively. For the mixed stages, `ncyc` values were 10,000 (`min0`), 2,500 (`min1`), 2,500 (`min2`), 5,000 (`min3`), 10,000 (`min4`), and 5,000 (`min5`). All minimizations were performed with Amber/24.3 and AmberTools/24.8, and downstream analysis used NumPy 2.2.6, pandas 2.3.3, matplotlib 3.10.8, scikit-learn 1.7.2, MDAnalysis 2.9.0, and `hdbscan` 0.8.41.

A.4 Additional PCA/HDBSCAN Implementation Details

For HDBSCAN, the `min_samples` parameter was left at `None`, which makes it default to the same value as `min_cluster_size`. This setting was used to avoid introducing an additional tuning parameter beyond the cluster-size rule applied uniformly across subunits.

References

1. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al.: Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature **630**(8016), 493–500 (2024). https://doi.org/10.1038/s41586-024-07487-w
2. Anandakrishnan, R., Aguilar, B., Onufriev, A.V.: H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. Nucleic Acids Res. **40**(W1), W537–W541 (2012). https://doi.org/10.1093/nar/gks375
3. Arndt, F., Schmitt, G., Winiarska, A., Saft, M., Seubert, A., Kahnt, J., et al.: Characterization of an aldehyde oxidoreductase from the mesophilic bacterium aromatoleum aromaticum ebn1, a member of a new subfamily of tungsten-containing enzymes. Front. Microbiol. **10**(JAN) (2019). https://doi.org/10.3389/fmicb.2019.00071
4. Carvalho, A.T., Swart, M.: Electronic structure investigation and parametrization of biologically relevant iron–sulfur clusters. J. Chem. Inf. Model. **54**(2), 613–620 (2014). https://doi.org/10.1021/ci400718m
5. Casadevall, G., Duran, C., Osuna, S.: AlphaFold2 and deep learning for elucidating enzyme conformational flexibility and its application for design. JACS Au **3**(6), 1554–1562 (2023). https://doi.org/10.1021/jacsau.3c00188
6. Case, D.A., Aktulga, H.M., Belfon, K., Cerutti, D.S., Cisneros, G.A., Cruzeiro, V.W.D., et al.: Ambertools. J. Chem. Inf. Model. **63**(20), 6183–6191 (2023). https://doi.org/10.1021/acs.jcim.3c01153
7. Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., et al.: The amber biomolecular simulation programs. J. Comput. Chem. **26**(16), 1668 (2005). https://doi.org/10.1002/JCC.20290

8. Eastman, P., et al.: OpenMM 7: rapid development of high performance algorithms for molecular dynamics. PLoS Comput. Biol. **13**(7) (2017). https://doi.org/10.1371/journal.pcbi.1005659
9. Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S., Onufriev, A.: H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res. **33**(SUPPL. 2) (2005). https://doi.org/10.1093/nar/gki464
10. Gowers, R.J., Linke, M., Barnoud, J., Reddy, T.J.E., Melo, M.N., Seyler, S.L., et al.: MDAnalysis: a python package for the rapid analysis of molecular dynamics simulations. SciPy 2016, pp. 98–105 (2016). https://doi.org/10.25080/majora-629e541a-00e
11. Iglewicz, B., Hoaglin, D.C.: How to Detect and Handle Outliers, ASQC Basic References in Quality Control, vol. 16. ASQC Quality Press, Milwaukee, WI (1993)
12. Jänes, J., Beltrao, P.: Deep learning for protein structure prediction and design-progress and applications. Mol. Syst. Biol. **20**(3), 162–169 (2024). https://doi.org/10.1038/s44320-024-00016-x
13. Lewis, S., Hempel, T., Jiménez-Luna, J., Gastegger, M., Xie, Y., Foong, A.Y., et al.: Scalable emulation of protein equilibrium ensembles with generative deep learning. Science **389**(6761) (2025). https://doi.org/10.1126/SCIENCE.ADV9817
14. Li, P., Merz, K.M.: MCPB.py: a python based metal center parameter builder. J. Chem. Inf. Modeling **56**(4), 599–604 (2016). https://doi.org/10.1021/acs.jcim.5b00674
15. Li, P., Merz, K.M.: Metal ion modeling using classical mechanics. Chem. Rev. **117**(3), 1564–1686 (2017). https://doi.org/10.1021/ACS.CHEMREV.6B00440
16. McInnes, L., Healy, J., Astels, S.: HDBSCAN: hierarchical density based clustering. J. Open Source Softw. **2**(11), 205 (2017). https://doi.org/10.21105/joss.00205
17. Meller, A., Bhakat, S., Solieva, S., Bowman, G.R.: Accelerating cryptic pocket discovery using AlphaFold. J. Chem. Theory Comput. **19**(14), 4355–4363 (2023). https://doi.org/10.1021/acs.jctc.2c01189
18. Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., Beckstein, O.: MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. **32**(10), 2319–2327 (2011). https://doi.org/10.1002/jcc.21787
19. Moulavi, D., Jaskowiak, P.A., Campello, R.J., Zimek, A., Sander, J.: Density-based clustering validation. Proceedings **2**, 839–847 (2014). https://doi.org/10.1137/1.9781611973440.96
20. Rothery, R.A., Stein, B., Solomonson, M., Kirk, M.L., Weiner, J.H.: Pyranopterin conformation defines the function of molybdenum and tungsten enzymes. Proc. Natl. Acad. Sci. **109**(37), 14773–14778 (2012). https://doi.org/10.1073/pnas.1200671109/-/DCSupplemental
21. Stein, R.A., McHaourab, H.S.: SPEACH_AF: sampling protein ensembles and conformational heterogeneity with AlphaFold2. PLoS Comput. Biol. **18**(8), e1010483 (2022). https://doi.org/10.1371/JOURNAL.PCBI.1010483
22. Vilela-Alves, G., Manuel, R.R., Martins, G., Carpentier, P., Raczyńska, A., Szaleniec, M., et al.: Structural insights into CO2 transport pathways in a w-formate dehydrogenase: Structural basis for co2 reduction. Angewandte Chemie - International Edition, p. e26133 (2026). https://doi.org/10.1002/anie.202526133

23. Wagner, J.R., Churas, C.P., Liu, S., Swift, R.V., Chiu, M., Shao, C., et al.: Continuous evaluation of ligand protein predictions: a weekly community challenge for drug docking. Structure (London, England : 1993) **27**(8), 1326 (2019). https://doi.org/10.1016/j.str.2019.05.012
24. Winiarska, A., et al.: A bacterial tungsten-containing aldehyde oxidoreductase forms an enzymatic decorated protein nanowire. Sci. Adv. **9**(22), eadg6689 (2023). https://doi.org/10.1126/sciadv.adg6689