The role of oligomerization in the optimization of
cyclohexadienyl dehydratase conformational dynamics and
catalytic activity

Nicholas J. East\textsuperscript{1,2} | Ben E. Clifton\textsuperscript{3} | Colin J. Jackson\textsuperscript{1,2,4} | Joe A. Kaczmarski\textsuperscript{1,2}

\textsuperscript{1}ARC Centre of Excellence in Synthetic Biology, Australian National University, Canberra, Australia
\textsuperscript{2}Research School of Biology, Australian National University, Acton, Australian Capital Territory, Australia
\textsuperscript{3}Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology, Okinawa, Japan
\textsuperscript{4}ARC Centre of Excellence for Innovations in Peptide and Protein Science, Research School of Chemistry, Australian National University, Acton, Australian Capital Territory, Australia

Correspondence
Joe A. Kaczmarski, Research School of Biology, Australian National University. R.N Robertson Building, Building 46, Biology Place, Acton, Australian Capital Territory, Australia.
Email: joe.kaczmarski@anu.edu.au

Abstract
The emergence of oligomers is common during the evolution and diversification of protein families, yet the selective advantage of oligomerization is often cryptic or unclear. Oligomerization can involve the formation of isologous head-to-head interfaces (e.g., in symmetrical dimers) or heterologous head-to-tail interfaces (e.g., in cyclic complexes), the latter of which is less well studied and understood. In this work, we retrace the emergence of the trimeric form of cyclohexadienyl dehydratase from \textit{Pseudomonas aeruginosa} (PaCDT) by introducing residues that form the PaCDT trimer-interfaces into AncCDT-5 (a monomeric reconstructed ancestor of PaCDT). We find that single interface mutations can switch the oligomeric state of the variants and that trimerization corresponds with a reduction in the $K_M$ value of the enzyme from a promiscuous level to the physiologically relevant range. In addition, we find that removal of a C-terminal extension present in PaCDT leads to a variant with reduced catalytic activity, indicating that the C-terminal region has a role in tuning enzymatic activity. We show that these observations can be rationalized at the structural and dynamic levels, with trimerization and C-terminal extension leading to reduced sampling of non-catalytic conformational substates in molecular dynamics simulations. Overall, this work provides insight into how neutral sampling of distinct oligomeric states along an evolutionary trajectory can facilitate the evolution and optimization of enzyme function.

Keywords
cyclohexadienyl dehydratase, enzyme evolution, molecular dynamics simulations, oligomerization, protein evolution, size-exclusion chromatography, trimerization
1 | INTRODUCTION

A large proportion of proteins self-associate to form homo-oligomers comprising two or more identical subunits. Among other things, homo-oligomerization has been shown to play an important role in conferring gain of function, protection from degradation, cooperative binding properties, allosteric regulation of enzyme activity, and enhanced thermostability of proteins. Interestingly, while the allosteric regulation of enzyme activity, the PaCDT (1a) and enhanced thermostability of proteins. Shifts in oligomeric states have also been linked to enhanced thermostability during the laboratory-based evolution of proteins. Considering the prevalence and biological importance of homo-oligomers, understanding the evolutionary factors that drive the formation of new homo-oligomeric species and the effect that oligomerization has on protein properties are key goals in the study and design of protein complexes. In order to probe the drivers, mechanisms, and functional consequences of oligomerization, several studies have used evolution-based approaches to investigate how the initial emergence and/or diversification of oligomeric states affected protein properties during the natural or laboratory-based evolution of proteins. For example, ancestral sequence reconstruction (ASR) revealed that two historical substitutions conferred a switch between a dimeric and tetrameric form during the evolution of vertebrate hemoglobin, and that this shift in oligomeric state gave rise to beneficial cooperative binding of oxygen. Similarly, ASR revealed the role of quaternary structure plasticity during the evolution and diversification of Ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) function. Shifts in oligomeric states have also been linked to enhanced thermostability during the laboratory-directed evolution of an αE7 carboxylesterase and the tailoring of activity and structural stability amongst bacterial methionine S-adenosyltransferase (MAT) homologs. On the other hand, a dimerization event that occurred during the evolution of steroid hormone receptors provided no immediate selective advantage. While evolution-based studies such as these have provided valuable insight into the mechanisms and functional importance of historical dimerization events (leading to proteins with C2 or D2 symmetry, e.g.), less is known about the emergence of cyclic complexes such as homotrimers.

Cyclohexadienyl dehydratase from Pseudomonas aeruginosa (PaCDT) is a homotrimeric enzyme that evolved from a monomeric, non-catalytic ancestral solute-binding protein (Figure 1a). In our previous work, we observed a ~60-fold increase in catalytic efficiency between the monomeric AncCDT-5 (the most recent ancestor of PaCDT that we characterized) and modern PaCDT (a trimer), despite these two proteins sharing identical active-site residues. Interestingly, while the subunits of PaCDT and its ancestors share the same bilobed periplasmic binding protein-like II fold (comprising two subdomains linked by a flexible hinge region), we found that the monomeric ancestors of PaCDT, including AncCDT-5, mostly sampled open or wide-open states (in which their two subdomains were far apart), while trimeric PaCDT predominantly samples a more compact and catalytically relevant closed state (in which the two domains are closer together and the active site is pre-organized for catalysis). We speculated that the shift in open-closed sampling and the associated increase in dehydratase activity between AncCDT-5 and PaCDT may have been, at least in part, due to oligomerization of the protein restricting rigid-body motions of the individual chains.

In this work, we used crystal the structures of PaCDT to guide the design of a series of variants of AncCDT-5 that differed only at positions that form the trimer interfaces in PaCDT. By doing so, we (a) identified variants that differ by only one or two substitutions at the trimer interfaces but that have distinct oligomeric states, (b) showed that oligomerization coincides with a decrease in the $K_M$ of the enzyme, bringing it close to physiologically relevant concentrations of the substrate, and (c) showed that this is consistent with a shift in the conformational landscape toward more catalytically relevant states. Finally, we demonstrate that the removal of a C-terminal extension in PaCDT yields a variant with decreased activity, and, thus, we highlight the potential role this region plays in further refining the conformational sampling and activity of the modern enzyme.

2 | RESULTS

Designing trimer-interface variants of AncCDT-5. In our previous work, we hypothesized that the trimeric nature of PaCDT may have contributed to the >50-fold greater dehydratase catalytic efficiency, compared with its monomeric ancestors, including AncCDT-5. The PaCDT trimer is formed through interactions between the larger of the two subdomains in each of the three chains, as well as some residues in the flexible “hinge” region of each chain (Figure 1d–g). The three chains of the trimer are arranged around the point of three-fold symmetry (i.e., C3 symmetry) somewhat like a three-pronged pinwheel, with the larger subdomains of the three chains coming together near the “apex” of the trimer (Figure 1e). Since the smaller subdomains are not part of the trimer interface, this arrangement allows for some
FIGURE 1  Design of the trimer-interface variants of AncCDT-5. (A) A simplified version of the phylogenetic tree used in our previous work\textsuperscript{21} to explore the evolution of PaCDT. We previously characterized proteins in extant clades and ancestral nodes as being non-catalytic solute-binding proteins (yellow shading) or proteins with dehydratase activity (blue shading). Key ancestral nodes (AncCDT-1 to AncCDT-5) are labeled. Representative extant proteins are shown next to the clades in which they are found: Ws0279 (UniProt: Q7MAG0), Pu1068 (UniProt: Q4FLR5), Ea1174 (UniProt: K0ABP5), Ei3690 (UniProt: C5B978), and Mn4388 (UniProt: B8IBK5). (b) A zoom-in of the phylogenetic tree shown in (a), highlighting the extant CDT homologs and internal ancestral nodes between AncCDT-5 and PaCDT that were used to guide the design of the AncCDT-5 trimer-interface mutants. The number of sequences in each collapsed clade is shown in parentheses. Extant sequences: CDTmesorhi (UniProt: I2F416), CDTbradyrh (UniProt: A4YUK0), CDTchromob (UniProt: Q7P1H9), and CDTralston (UniProt: Q8XW22). (c) Table showing residues of AncCDT-5, PaCDT, and internal ancestral nodes at positions at the PaCDT trimer interfaces that differ between AncCDT-5 and PaCDT. The residues at the equivalent positions in an alternate reconstruction of AncCDT-5 (AncCDT-5 (WAG))\textsuperscript{21} are also shown. The residue numbering is based on that used for the crystal structure of AncCDT-5 (PDB 6WUP); PaCDT residue numbers (according to PDB 5HPQ) are shown in parentheses. (d–g) The crystal structure of PaCDT (PDB 5HPQ) and trimer interface residues. (d) A single chain from the crystal structure of PaCDT showing the small (gray) and large (teal) subdomains and the flexible hinge region (orange). The positions of residues that form the trimer interface are indicated by spheres. The active site region is shown as a black square. (e) The trimeric form of PaCDT, showing the three chains interacting mainly via the large domain of each chain. (f) Key trimer interface residues that differ between AncCDT-5 and PaCDT are shown as sticks (numbered according to PDB 5HPQ). (g) Key residue positions at the apex of the trimer in PaCDT.
rigid-body open-closed motions to occur in each of the chains. However, we reasoned that the packing of the chains in the trimer might sterically limit or restrict the sampling of wide-open conformations and lead to enhanced activity due to increased sampling of the catalytically competent closed state.\textsuperscript{23}

To determine if trimerization of the protein alone could have contributed to the increase in catalytic efficiency that we previously observed between AncCDT-5 and PaCDT (which differ at 88 residue positions),\textsuperscript{21,23} we aimed to identify variants of AncCDT-5 that had different oligomeric states but differed only by substitutions at the trimer-forming interfaces. Sequence and structural alignments of PaCDT (PDB 5HPQ) and AncCDT-5 (PDB 6WUP) revealed eight key positions spread across the trimer-forming interfaces that differed between AncCDT-5 and PaCDT (Figure 1c,f) as well as a single-residue deletion close to the interface (Gly216 is not found in PaCDT). Substitutions at the apex positions of the PaCDT trimer (including PaCDT residues Tyr101, Glu217, and Arg221; hereafter termed “YER”) were of particular interest (Figure 1g); we expected that these residues would likely be important for determining whether three chains could pack together to form the trimeric structure.

Having identified key residues involved in the PaCDT trimeric interfaces, we designed a series of variants that introduced these residues into AncCDT-5. Rather than generating all the possible combinations of the trimer-interface substitutions separating AncCDT-5 and PaCDT (which would have resulted in 256 variants to experimentally test), we used the principle of parsimony and the sequences of present-day descendants of AncCDT-5 to determine the order in which these substitutions likely accumulated during the historical evolution of PaCDT; we then introduced these substitutions into AncCDT-5 in a step-wise manner that reflected the historical accumulation of mutations. For example, since all extant descendants of AncCDT-5 had Leu94 and Phe206, the most parsimonious explanation is that mutations P94L and H206F were introduced earlier in the evolutionary trajectory leading to PaCDT. Therefore, “early” trimer interface variants in our series contained these two residues.

ASR further guided this approach. For example, the inferred maximum-likelihood ancestral sequences at nodes between AncCDT-5 and PaCDT (“A5a” and “A5b,” Figure 1b,c) were also consistent with H206F and P94L occurring earlier than the other trimer-interface substitutions. Ancestral sequences A5a and A5b also contained Phe101 and Val218 at trimer interface positions; these residues are not found in AncCDT-5 or PaCDT but are present in some modern descendants of AncCDT-5, including CDT homologs from \textit{Bradyrhizobium} sp. ORS 278 (UniProt A4YUK0) and \textit{Ralstonia solanacearum} CMR15 (UniProt D8N7P8) (Table S1). Although we did not know if the ancestors or these extant proteins formed oligomers or not, we generated variants that introduced these two residues into A5.1 because these substitutions may have occurred along the historical mutational path linking AncCDT-5 and PaCDT.

Together, the evolutionary-guided approach resulted in a series of variants that reflected the historical accumulation of substitutions at the trimer interfaces (Figure 2). Plasmids (pET28a) containing sequences encoding N-terminal His-tagged PaCDT, AncCDT-5, and the trimer-interface AncCDT-5 variants were produced. All proteins are expressed in their soluble, folded state in \textit{Escherichia coli}.

\textbf{Oligomeric structures of AncCDT-5 trimer-interface variants.} The oligomeric states of purified AncCDT-5, PaCDT, and AncCDT-5 trimer-interface variants were determined by size exclusion chromatography coupled to multiple angle laser light scattering (SEC-MALLS). Consistent with crystal structures and as previously observed,\textsuperscript{21,23} AncCDT-5 eluted from the SEC column primarily as a monomer when injected onto the column at 8 mg/ml, while PaCDT had an elution volume consistent with its being a trimer (Figures 2, S1, and S2).

AncCDT-5 variants containing all PaCDT trimer-interface residues (A5+CDT\textsubscript{interface} and A5+CDT\textsubscript{interface}\Delta G216) eluted from the SEC column with the estimated mass of trimers in solution, even at concentrations as low as 0.1 mg/ml (Figures 2, S1, and S2). This confirmed that AncCDT-5 could be converted into a trimer by only making substitutions at the trimer interfaces and that the formation of the trimer did not require any of the 79 substitutions that occurred elsewhere on the protein during the evolution of PaCDT from AncCDT-5.

The stepwise introduction of PaCDT trimer-interface residues into AncCDT-5 in a switch from monomer to trimer, as shown by SEC-MALLS experiments (Figures 2 and S1). The “early” variants A5+P94L, A5+H206F, A5+P94L+H206F (hereafter “A5.1”) and A5.1+T108R were unambiguously monomers in solution even when injected onto the column at 8 mg/ml. “Later” variants, including A5.1+T108R+YER (two substitutions away from the trimeric A5+CDT\textsubscript{interface} and A5.1+T108R+YER+K98R (one substitution, Q230I, away from the trimeric A5+CDT\textsubscript{interface} variant) eluted earlier than the monomeric proteins, but later than the trimeric A5+CDT\textsubscript{interface}, consistent with the presence of dimers in these samples (Figure S1). While we acknowledge that delayed elution profiles could also be explained by an equilibrium between monomer and trimeric states, light scattering was most consistent with these variants being dimers in solution (Figure S1). Interestingly, a single
substitution, Q230I, disrupted the formation of these complexes when introduced into A5.1+T108R+YER, but the same substitution was tolerated when introduced into A5.1+T108R+YER+K98R, indicating an epistatic interaction between these residues.

**FIGURE 2** Oligomeric states of trimer-interface variants of AncCDT-5. (a) Table showing residues in trimer-interface variants and predicted molecular weight (MW). MW was calculated by comparing elution volumes of proteins (injected at 5–8 mg/ml) from the size exclusion column to elution volumes of protein standards. Variants are colored according to their oligomeric state under the SEC-MALLS conditions (pink = monomeric, yellow = intermediate dimer or oligomeric mixture, blue = trimer). Residues found in PaCDT are shaded dark gray. Residues not found in either AncCDT-5 or PaCDT are shaded light-gray. (b) Normalized refractive index chromatograms showing elution peaks of trimer-interface variants. Proteins were injected onto the SEC column at 5–8 mg/ml. Vertical lines aligned with elution peaks of AncCDT-5 (monomer) and PaCDTΔC (trimer) are shown for reference. (c) Schematic showing mutational pathways linking AncCDT-5 and PaCDT via the interface variants in this study (colored by predicted oligomeric state as in (a)). (d) From left to right, structures of the monomeric (AncCDT-5, PDB 5T0W), putative dimeric (Alpha-Fold2 model of A5.1+D101F+P218V), and trimeric (PaCDT, PDB 6BQE) forms of the related proteins.

Putative dimeric forms of AncCDT-5 variants. In addition to A5.1+T108R+YER and A5.1+T108R+YER+K98R, there was evidence that indicated the formation of dimers in the chromatograms of some of the other AncCDT-5 variants. Indeed, a small (yet distinct) dimeric...
peak was observed in the chromatogram of AncCDT-5; this dimer peak became more prominent in an alternate reconstruction of AncCDT-5, AncCDT-5(WAG), which clearly samples both monomeric and dimeric states (Figure S3). Similarly, while A5.1+P218V and A5.1+D101F remained mostly monomeric when injected at 8 mg/ml, the variants A5.1+D101F+P218V and A5.1+T108R+D101F+P218V eluted at an intermediate volume, with light scattering consistent with them being dimers (Figure 2 and S1). Since these variants only differ from monomeric AncCDT-5 at positions on the PaCDT trimer interfaces, it is likely that any dimeric forms are asymmetric dimers that share the same interface as found between chains in the PaCDT trimer. Indeed, an AlphaFold2 structure of A5.1+D101F+P218V shows that the most likely dimeric form (ptm score of 0.84) is an asymmetric dimer that resembles the trimeric form with one of the chains missing (Figure 2d). It is likely that such complexes would be unstable and may only be present at the high concentrations used in the SEC-MALLS experiments; indeed, while A5.1+D101F+P218V eluted as an oligomer when injected at 8 mg/ml, it eluted as a monomer when injected onto the SEC column at 0.1 mg/ml (Figure S2).

Overall, our results show that the equilibrium between distinct oligomeric states can readily be shifted with a small number of mutations at the trimer-forming interfaces of these proteins. Interestingly, there also seems to be significant quaternary structure plasticity amongst extant homologs of PaCDT. For example, Ws0279, a L-lysine binding protein (26% sequence identity with PaCDT and a descendant from AncCDT-1), is a trimer in solution and crystals (Figure S4), while Pu1068 (a non-dehydratase homolog of PaCDT from Candidatus Pelagibacter ubique that is a descendant of AncCDT-2 and shares 27% sequence identity with PaCDT) is monomeric.21 Similarly, Ea1174 (a descendant of AncCDT-3 that shares 32% sequence identity with PaCDT) is a monomer (Figure S5) despite having dehydratase activity comparable to PaCDT in complementation assays.21 Thus, in the same manner that the point mutants of AncCDT-5 easily perturb the oligomeric equilibrium, extant proteins throughout this family appear to regularly interchange between different oligomeric states throughout the phylogeny.

Formation of the trimer corresponds with an increase in substrate affinity. The dehydratase activity of key trimer-interface variants was assessed to determine whether their propensity to form higher-order oligomeric states (i.e., dimers or trimers) was associated with an increase in dehydratase activity. Similar to our previous work,23 we observed a > 50-fold increase in the catalytic efficiency between AncCDT-5 and PaCDT (Table 1, Figure S6). None of the variants exhibited significant changes in terms of their turnover rates (kcat), which were close to those of AncCDT-5. In contrast, we observe a significant reduction (~7-fold) in the Kcat of these variants along the trajectory, with trimeric A5+CDT interface exhibiting a Kcat of ~80 μM in comparison to monomeric AncCDT-5, which has a Kcat of ~570 μM. This is particularly important in terms of the selective benefit, as the physiological concentration of the substrates is likely to be in the low micromolar range (PaCDT has a Kcat of ~27 μM, for example, and the concentrations of molecules in this biosynthetic pathway are typically ~14–18 μM in gram-negative bacteria.24) While we cannot be certain of the oligomeric state of the proteins at the low concentrations used in the enzyme assays, the fact that A5+CDT interface (which clearly forms trimers when injected at 0.1 mg/ml) has a reduced Kcat compared with monomeric A5.1, while only differing by six mutations at the trimer interface, suggests that trimerization contributes to the shift in its catalytic properties.

### Table 1: Prephenate dehydratase activity of the variants

|                | Kcat (s−1) | Kcat/Km (M−1 s−1) |
|----------------|------------|--------------------|
| A5             | 9.1 ± 0.2  | (1.60 ± 0.38) × 10⁴ |
| A5.1           | 10.3 ± 0.59| (1.81 ± 0.24) × 10⁴ |
| A5.1+D101F     | 10.5 ± 0.47| (6.22 ± 0.88) × 10⁴ |
| A5.1+P218V     | 5.9 ± 0.53 | (4.58 ± 1.38) × 10⁴ |
| A5.1+D101F+P218V| 6.9 ± 0.91| (4.36 ± 0.43) × 10⁴ |
| A5.1+T108R     | 8.4 ± 0.43 | (4.75 ± 0.76) × 10⁴ |
| A5.1+T108R+YER | 4.8 ± 0.47 | (1.37 ± 0.35) × 10⁴ |
| A5.1+T108R+YER+K98R | 3.9 ± 0.11| (1.04 ± 0.07) × 10⁴ |
| A5.1+T108R+YER+Q230I | 12.5 ± 0.43| (6.17 ± 0.64) × 10⁴ |
| A5+CDTint      | 4.6 ± 0.42 | (5.75 ± 1.93) × 10⁴ |
| PaCDT          | 25.1 ± 1.60| (9.40 ± 27.6) × 10⁴ |

Note: Values are mean ± standard error of fit (n = 2–3).
We also tested if an increase in the propensity to form oligomers was associated with an increase in the thermostability of the proteins; we reasoned that the formation of the trimer may have enhanced the thermostability, thereby providing an additional adaptive advantage. However, there was no clear correlation between the thermostability of the proteins and their propensity to form higher oligomeric states (Figure S7 and Table S2). PaCDT had the lowest thermostability with a $T_m$ of below 70°C, while AncCDT-5 was the most stable ($T_m$ of ~77°C); the introduction of PaCDT trimer-interface substitutions to AncCDT-5 led to a decrease in the thermostability of these variants, approximating that of PaCDT. Together, our results indicate that oligomerization affects the catalytic activity ($K_M$ in particular), but not the thermostability, of these proteins.

The role of the C-terminal region in the activity of PaCDT. While trimerization could account for a notable decrease in $K_M$ between AncCDT-5 and PaCDT, it could not fully account for the improved kinetic parameters of PaCDT relative to AncCDT-5. In particular, the trimeric AncCDT-5 variants, including A5+CDT_interface, still have $k_{cat}$ values that are significantly lower than that of PaCDT. Since trimerization did not fully account for the shift in kinetic properties observed between AncCDT-5 and PaCDT, we turned our attention to other structural differences between AncCDT-5 and PaCDT.

PaCDT has a nine-amino acid extension at its C-terminus compared with AncCDT-5 (Figure 3a,b); we wondered if this may have a role in tuning the activity of PaCDT. Indeed, C-terminal extensions and additional domains have been linked to functional differences between other periplasmic binding proteins. To probe the influence of the C-terminal extension on PaCDT activity, we created a truncated variant of PaCDT in which the additional residues were removed (yielding variant PaCDTΔC). While PaCDTΔC retained a trimeric state (Figure S8), its $k_{cat}$ fell at the same rate as AncCDT-5 (Figure 3c,d). This is notable because the trimerization of AncCDT-5 had no effect on $k_{cat}$. Thus, the C-terminal extension alone can account for most of the difference in $k_{cat}$ between trimeric AncCDT-5 and trimeric PaCDT. In contrast, the $K_M$ of truncated PaCDT is significantly lower than the $K_M$ of AncCDT-5 (147 ± 48.8 µM vs. 570 ± 9.1 µM, Figure 3e). Importantly, even though PaCDTΔC and A5+CDT_interface differ at 79 positions remote from the trimer interface, the $K_M$ values for these proteins are similar, supporting the
concept that the decrease in $K_M$ between AncCDT-5 and PaCDT is primarily due to trimerization rather than substitutions elsewhere in the protein.

Analysis of the sequences of close homologs of PaCDT reveals significant variation in the C-terminal region (even amongst descendants of AncCDT-5, Figure S9), suggesting that the C-terminal extension found in PaCDT was a recent adaptation (i.e., it evolved along the branch between A5b and PaCDT). It seems that C-terminal extensions may have arisen multiple times during the diversification of this protein family. Future work could investigate whether differences in the C-terminal region of CDT homologs could account for differences in the catalytic activity of these extant proteins.

**Trimerization and the C-terminal extension affect the conformational sampling of PaCDT.** In our previous work, we found that PaCDT sampled the catalytically relevant closed conformation more often compared with the ancestral sequences (which predominantly sampled states in which the two subdomains were far apart). This provides a plausible explanation for the difference in kinetic parameters: AncCDT-5, which mostly samples a "wide-open" state that requires significant reorganization to attain a Michaelis complex has a much higher $K_M$ and lower catalytic turnover than the extant PaCDT, which samples compact states around the Michaelis state. We speculated that trimerization or the addition of the C-terminal extension may have altered the open–closed conformational sampling in PaCDT by limiting the range of motion of the smaller domains, and that this is what caused the shift in the catalytic parameters.

As such, we used molecular dynamics (MD) simulations to probe whether either trimerization or the C-terminal extension could have affected the open–closed conformational sampling of these enzymes to provide a molecular-level rationalization of the kinetic results. Our previous work on PaCDT and its ancestors demonstrated that MD simulations were consistent with experimental measurement of conformational sampling using double electron–electron resonance spectroscopy distance measurements using unnatural amino acid: lanthanide labels. Accordingly, here, we ran additional MD simulations of PaCDT both with and without the C-terminal extension and compared these to simulations of monomeric AncCDT-5. As a simple proxy for the extent of the "openness" of the protein and the pre-organization of active site residues, we monitored how the distance between the alpha carbons of the catalytic residue Glu184 on the small domain (due to different numbering conventions, this is equivalent to the catalytic residue Glu173 mentioned in our previous work) and Tyr33 on the large domain (equivalent to Tyr22 in previous work) changed during the course of the simulations (Figure 4a).

Simulations initiated from structures based on the crystal structure of monomeric AncCDT-5 were consistent with our previous work, showing a protein that regularly samples wide-open states (in which the small domain completely twists away from the large domain, resulting in $\sim$3 nm between the two key residues) (Figures 4b and S10). When these wide-open states of AncCDT-5 are overlaid on the trimeric structure of PaCDT, it is clear that in a trimeric arrangement it would not be possible for more than one of the three chains to simultaneously sample this conformation since the small domains of each chain would clash (Figure 4c). Consistent with this, simulations initiated from PaCDT in which the C-terminal extension had been removed, which experimentally retained a trimeric structure and displayed similar $k_{cat}$ but significantly lower $K_M$ values than AncCDT-5, showed reduced sampling of the wide-open, twisted state, with more regular sampling of semi-open ($\sim$1.5 nm) and compact conformations (Figure 4d). Importantly, the arrangement of chains in the trimer prevented the chains opening to the same extent as observed in the AncCDT-5 simulations and also prevented all three chains from opening simultaneously (i.e., at least one chain was always in a more closed, catalytically relevant conformation; $<1.5$ nm; Figure 4e and S10). Finally, simulations of the full-length PaCDT trimer (i.e., with the C-terminal extension) showed that it sampled a relatively narrow distribution of closed states for the majority of the simulations, with limited sampling of a semi-open state ($\sim$1.5 nm) that would allow for substrate-binding and release and no sampling of the wide-open state (i.e., Glu184–Tyr33 distances remained below 2 nm; Figure 4f,g), consistent with previous work. Visual inspection of the trajectories revealed frequent bridging interactions between the small and large domains that were mediated by the C-terminal region. Thus, the formation of a trimer and the subsequent addition of a C-terminal extension progressively reshape the conformational landscape of CDT to freeze out non-catalytic conformations.

### 3 DISCUSSION

While the majority of experimentally characterized proteins with the periplasmic binding protein-like 2-fold are monomeric (with a few notable examples of dimeric assemblies); in this work, we show that PaCDT homologs and ancestral states interconvert easily between a range of oligomeric structures including monomers, dimers, and trimers. Indeed, the oligomeric state of the AncCDT-5 variants can be readily shifted with as few as one or two substitutions at the trimer-
forming interfaces, suggesting that different oligomeric states could be accessed easily via neutral genetic drift. This is consistent with observations from other studies that have traced the evolution of oligomerization. For example, as few as two mutations were required to switch between dimeric and tetrameric states during the evolution of vertebrate hemoglobin. Additionally, quaternary structure plasticity has been attributed to functional diversification in other protein families, including RuBisCO and MAT.

While other studies have retraced historical dimerization events (which typically involve the evolution of a single surface), the work presented here shows that cyclic oligomers that involve heterologous interfaces can also be readily accessible via a small number of mutations when these are introduced into an appropriate genetic background. This is notable considering that for cyclic complexes to emerge, it requires the evolution and optimization of two distinct complementary interfaces. Not only this, but the resulting geometry must also be such that it can accommodate the packing of multiple chains around the point of symmetry. Indeed, these factors contribute to the rarity of trimers among natural proteins.

**FIGURE 4** Molecular dynamics simulations of AncCDT-5, PaCDTΔC, and PaCDT. (a) Monomeric AncCDT-5 sampled a range of conformations ranging from a more closed state (left) in which the catalytic residues are close together to a distorted “wide-open” state (right) in which the small domain is twisted relative to the large domain catalytic residues are far apart (>2.5 nm). During simulations, we monitored the distance between the alpha-carbons of Glu184 and Tyr33 (red spheres). The position of the C-terminus is highlighted in yellow. The orientation of the helix on which Glu184 is found is highlighted in orange, and the small and large subdomains are shaded gray and blue, respectively. (b) Histogram showing the distribution of Glu184–Tyr33 distances during replicate simulations of AncCDT-5 (the first 50 ns of each simulation was omitted from histogram data). (c) Snapshots corresponding to closed (left) and wide-open (right) states of AncCDT-5 (cartoon representation) are mapped onto the trimeric structure of PaCDT, showing that sampling of the wide-open state would be restricted in a trimeric arrangement even when the other two chains (yellow and green surfaces) are in a closed conformation. (d) Histogram showing the distribution of Glu184–Tyr33 distances during replicate simulations of PaCDTΔC. (e) In simulations of PaCDTΔC, the small domains did not interact when they were all in a more closed state (left), but steric clashes between the small subunits of neighboring open chains corresponded to reduced sampling of the wide-open states (right). (f) Histogram showing the distribution of Glu184–Tyr33 distances during replicate simulations of full-length PaCDT. (g) Full-length PaCDT sampled a much more compact state (left), with the additional C-terminal residues keeping catalytic residues near even in the most open of the MD snapshots (right).
determining the oligomeric state of the protein. As such, this work furthers our understanding of the emergence of cyclic protein complexes, which are important targets for both drug design and protein engineering.\textsuperscript{15,19}

The effect of oligomerization on enzyme activity is often somewhat cryptic, as it is often unclear how an interface remote from an active site can affect activity. Moreover, it is difficult to test with extant proteins because of epistasis and the ratchet-like nature of evolution\textsuperscript{12,33,34}; if an interface is disrupted, how can we be sure the effects are from the loss of the oligomeric structure and no other structural changes? Constructive biochemistry approaches such as ASR are useful in this context, as they allow us to see the benefits of oligomerization in the “forward-direction.” However, traditional ASR approaches can still be limited in that ancestral sequences from consecutive nodes in the phylogeny are still often separated by many substitutions that are spread across the protein. Here, by introducing only those historical substitutions that occurred at the trimer interface into AncCDT-5 (and keeping other residues constant), we could be more confident in attributing the decrease in $K_M$ between AncCDT-5 and PaCDT to trimerization without having to determine the effect of substitutions elsewhere in the protein. Our MD simulations provided a structural explanation for this, showing how the architecture of the trimer restricts the sampling of unproductive conformations.

It is well established that historical events (including oligomerization) shape future evolutionary trajectories, even if the initial event is neutral or near-neutral in terms of providing a selective advantage.\textsuperscript{35–38} Genetic variation within protein families can lead to (often cryptic) variations in protein properties that influence the evolvability of a protein and subsequent evolutionary trajectories. Such substitutions and properties can become entrenched by subsequent mutations that prevent reversion to the ancestral state.\textsuperscript{14} While the formation of the trimer provided an immediate enhancement in $K_M$ during the evolution of PaCDT, trimerization would have also determined which future evolutionary trajectories were accessible and thus shaped any subsequent optimization of the enzyme; this may have included modifications in the C-terminal region that we showed contributed to an increase in the catalytic turn-over of the enzyme. At the same time, it is conceivable that efficient dehydratase activity may have been achievable without trimerization via other mutational routes; indeed, the modern enzyme Ea1174 seems to be an efficient dehydratase despite being a monomer. Structural diversification amongst the CDT homologs would have also likely been driven by differences in the environments in which the host organisms lived; with differences in oligomerization and C-terminal extensions providing routes for the fine-tuning of physiologically relevant enzyme activities in each case.

4 | CONCLUSION

Considering the prevalence and biological importance of homo-oligomers, understanding the drivers and consequences of oligomerization is important to understand protein biology and enable better protein design. Here, we show that the oligomeric state of PaCDT ancestors and homologs can be perturbed through a minimal number of substitutions at the interface, and that this can have significant effects on activity by altering the conformational sampling of the proteins. This suggests that new oligomeric states (such as the trimeric form seen in PaCDT and Ws0279) may have initially emerged via the accumulation of relatively neutral mutations, and that subsequent structural changes then entrenched the higher-order oligomeric state.

5 | MATERIALS AND METHODS

5.1 | Design of interface mutants

The sequences of trimer-interface variants were based on the sequence of AncCDT-5, an inferred ancestral sequence from our previous work.\textsuperscript{21,23} Unless otherwise stated, the residue numbering used in this study is based on the numbering used in the crystal structure of AncCDT-5 (PDB 6WUP); as such, the numbering convention used here differs from the residue numbering convention used in our previous work.\textsuperscript{21,23}

The trimer-interface-forming residues were identified by analyzing crystal structures of the trimeric PaCDT (PDB IDs 5HPQ and 6BQE) and using the InterfaceResidues script in PyMOL (https://pymolwiki.org/index.php/InterfaceResidues). Sequence and structural alignments between AncCDT-5 and PaCDT were then used to identify positions that differed between AncCDT-5 and PaCDT at these trimer-forming interface positions; this guided the design of AncCDT-5 variants that differed only at these trimer-interface positions. In order to roughly reflect the historical order in which trimer-forming residues were introduced, we also considered the sequences of PaCDT homologs that were identified as decedents of AncCDT-5 in our previous work.\textsuperscript{21} Further, PAML\textsuperscript{39} was used to infer the maximum-likelihood amino acid sequences at internal nodes between AncCDT-5 and PaCDT in the LG-inferred phylogenetic tree described in our previous work.\textsuperscript{21} The presence or absence of gaps in the inferred ancestral proteins was
decided based on parsimony and manual inspection of the multiple-sequence alignment. In addition to a variant of AncCDT-5 that had all the interface positions mutated to their corresponding PaCDT residues (“A5+CĐTinterface”), we also generated a similar protein with Gly216 removed (“A5+CĐTinterfaceΔG216”). We also constructed AncCDT-5 variants that introduced other residues found in other PaCDT homologs at the positions identified at the apex of the PaCDT trimer, including Val218 and Phe101. Sequences were back-translated, and the resulting DNA fragments encoding these variants were codon-optimized for expression in *E. coli*, synthesized and cloned into the pET28a vector (between BamHI and XhoI restriction sites) by Twist Bioscience. Variants A5.1, A5.1+P218V, A5.1+T108R and A5.1+T108R+YER+Q230I were generated using mutagenesis primers (Table S4) using a standard Gibson Assembly protocol.40 pET28a vectors containing the genes encoding AncCDT-5, PaCDT (UniProt: Q01269; residues 26–268), and the C-terminal-truncated PaCDT (UniProt: Q01269; residues 26–259) were also obtained. The pET28a vector adds a pET28a-encoded N-terminal hexahistidine tag, thrombin protease cleavage sequence, and T7 tag to the variant sequences. While the resulting protein sequences used in this study are similar to those encoded in pDOTS7 vectors used in our previous work,21,23 they do have slightly different N-terminal tags (e.g., an additional T7 tag), and the pET28a-encoded genes also omit the additional C-terminal leucine residue found in the pDOTS7-based constructs as a consequence of the Golden-Gate assembly method that was used. For oligomeric characterization of Ws0279 (UniProt: Q7MAG0; residues 24–258) and Ea1174 (UniProt: K0ABP5; residues 31–268), pDOTS7 vectors encoding the proteins were obtained from our previous work.21 Similarly, AncCDT-5(WAG) was characterized by expressing it from a pDOTS7 vector obtained from our previous work.

6 | **EXPRESSION AND PURIFICATION OF PROTEIN VARIANTS**

6.1 | **Protein expression**

In most cases, variants were expressed in BL21(DE3) cells (NEB): Plasmids were incorporated via electroporation, followed by recovery in LB media for 1 hr and incubating at 37°C on LB agar plates supplemented with 50 μg/ml kanamycin. The following day, single colonies were used to inoculate 5 ml starter cultures (LB media +50 μg/ml kanamycin), which were incubated for 5–6 hr at 37°C with shaking (~180 rpm). These were then used to inoculate 1 L LB + 50 μg/ml kanamycin cultures in Thomson Ultra Yield Flasks. Cultures were grown at 37°C with shaking (180 rpm) until OD600 reached approximately 0.6 before being induced with 0.5 mM β-d-1-isopropylthiogalactopyranoside (IPTG). Following induction, cultures were grown overnight at 30°C with shaking at 180 rpm. Cell pellets were collected by centrifugation and stored at −20°C. His-tagged Ws0279, Ea1174, and AncCDT-5(WAG) were expressed in BL21 (DE3) cells (NEB): Transformed cells were grown in LB media supplemented with 100 mg/L ampicillin to OD600 ~ 0.7 at 37°C, induced with 1 mM IPTG, and incubated for another 20 hr at 37°C.

6.2 | **Purification by immobilized-metal affinity chromatography**

All variants were initially purified from cell lysate using a standard immobilized-metal affinity chromatography (IMAC) approach. Briefly, cell pellets were thawed and resuspended in binding buffer (50 mM NaH₂PO₄, 500 mM NaCl, 20 mM imidazole, pH 7.4) and lysed by sonication on ice (50% power, two rounds of 6 min on, 6 min off). Lyzed cells were fractionated by ultracentrifugation (9,000 × g for 40 min at 4°C). The supernatant was filtered through a 0.45 μm syringe filter and loaded onto a 5 ml HisTrap HP column (GE Healthcare) equilibrated in binding buffer. The column was washed with 30 ml of binding buffer followed by 25 ml of wash buffer (50 mM NaH₂PO₄, 500 mM NaCl, 40 mM imidazole, pH 7.4). The His-tagged protein was eluted in an elution buffer (50 mM NaH₂PO₄, 500 mM NaCl, 500 mM imidazole, pH 7.4). Fractions containing protein were pooled. Samples collected from IMAC were exchanged into SEC buffer (20 mM Na₂HPO₄, 150 mM NaCl, pH 7.4) using either a HiPrep 26/10 desalting column (GE Healthcare) or through multiple rounds of concentration and dilution using an Amicon Ultra-15 filter unit with a 10 kDa molecular weight cutoff. Protein purity was determined through sodium dodecyl sulfate–polyacrylamide gel electrophoresis. The concentrations of purified protein samples were determined by measuring A₂₆₀ using a NanoDrop One (Thermo Scientific) and molar extinction coefficients calculated from ProtParam (http://expasy.org/tools/protparam.html).

7 | **PROTEIN CHARACTERIZATION**

7.1 | **SEC and comparison with protein standards**

SEC was performed at 4°C on an ÄKTA Pure. Samples were injected onto either a HiLoad 16/600 Superdex
200 column (GE Healthcare) or a Superdex 200 10/300 GL size-exclusion column (GE Healthcare) and purified in SEC buffer. Protein standards (GE Healthcare Gel Filtration Calibration Kits) were used for comparison of elution volumes and estimation of theoretical molecular masses (Figures S11 and S12).

7.2 | SEC coupled with multiple angle light scattering (SEC-MALLS)

Samples (100 μl) of each protein were loaded at 5–8 mg ml⁻¹ onto a pre-equilibrated Superdex 200 10/300 GL size-exclusion column (GE Healthcare) attached to multi-angle light scattering (DAWN HELEOS 8; Wyatt Technologies) and refractive index detection (Optilab rEX; Wyatt Technologies) units. A flow rate of 0.5 ml min⁻¹ was used. The multi-angle detectors were normalized using monomeric bovine serum albumin (Sigma, A1900). A dn/dc value of 0.186 g⁻¹ was used for each sample. The data were processed using ASTRA 5.3.4 (Wyatt Technologies). Unless otherwise stated, data were collected from a single experiment for each variant (n = 1).

7.3 | Differential scanning fluorimetry–thermostability measurements

Differential scanning fluorimetry (DSF) was performed on a QuantStudio 3 Real-Time-PCR-System (Thermo Fisher). In a 96-well PCR plate, reaction mixtures containing approximately 5 μM protein in DSF buffer (50 mM Na₂HPO₄, 150 mM NaCl, pH 7.4) and 1× Protein Thermal Shift Dye (Thermo Fisher) were heated from 20 to 99°C while monitoring emission at 623 nm with excitation at 580 nm. Measurements were completed as technical triplicates, with melting temperatures (Tₘ) calculated and determined from the steepest part of the curve (i.e., peaks in the derivative plot) in the Protein Thermal Shift Software v1.4 (Thermo Fisher).

7.4 | Prephenate dehydratase activity assays

Sodium prephenate was prepared by mixing 40 mM barium chorismite (Sigma) in water, followed by the addition of an equimolar amount of Na₂SO₄. To this, an equal volume of 100 mM Na₂HPO₄ was added, causing the precipitation of barium sulfate, which was subsequently removed through centrifugation. The resulting sodium chorismite solution was heated at 70°C for 1 hr to generate sodium prephenate. The concentration of prephenate was determined through acid conversion of prephenate to phenylpyruvate using 0.5 M HCl over a 15-min time course, followed by the addition of NaOH and measurement of absorbance at 320 nm using the extinction coefficient for phenylpyruvate (17,500 M⁻¹ cm⁻¹).

PDT activity was determined by using a discontinuous colorimetric assay to monitor the production of phenylpyruvate. Enzyme solutions were prepared in an SEC buffer to between 100 and 700 nM. Sodium prephenate was diluted to 1.4 mM in 50 mM Na₂HPO₄, and a series of seven 1-in-2 dilutions was created. To initiate the assay, 100 μl of the enzyme solution (final concentration of 100 nM) was mixed with 600 μl of substrate at 25°C. Every ~30 s, the absorbance of a 100 μl aliquot of enzyme-substrate solution was measured at 320 nm following the addition of 100 μl of 2 M NaOH. No-enzyme controls were performed; there was negligible product formation during the time course of the assays (~3 min). The concentration of phenylpyruvate was calculated assuming a pathlength of 0.5 cm, using the previously reported extinction coefficient (17,500 M⁻¹ cm⁻¹), and taking into account dilution factors. Data were plotted and analyzed using GraphPad Prism (version 9) using the Michaelis–Menten model.

7.5 | Modeling using ColabFold

The ColabFold: AlphaFold2 using MMseqs Google Colab notebook⁴¹,⁴² was used to generate the model of dimeric α5.1 + D101F + P218V: default options were used (with the addition of an amber minimization step).

7.6 | MD simulations

MD simulations were performed using Gromacs 2021.2⁴³ and the Charmm36m force field (charmm36-fep2021.ff).⁴⁴ Several systems were prepared (see Table S3 for details). In each case, protein structures were obtained from the PDB and then submitted through the PDB REDO server.⁴⁵ Small molecules (including water and acetate/HEPES) were removed. Protein structures were further prepared using the Schrodinger Protein Preparation Tool (Schrodinger 2020–4), including the removal of unwanted chains, ionization of residues (pH 7.4), optimization of H-bonds, and energy minimization. In cases where residues not observed in the crystal structure were modeled, the Schrödinger modeling tools were used to model the C-terminal extensions followed by local minimization in Maestro.

Protein chain termini were capped in Gromacs. Each protein was solvated in a rhombic dodecahedron with
SPC water molecules, such that the minimal distance of the protein to the periodic boundary was 14 Å, and an appropriate number of ions were added to neutralize the system (see Table S3). Energy minimization was performed on each system using the steepest descent algorithm, followed by a 100 ps isothermal-isochoric ensemble (NVT; 300 K) simulation with harmonic position restraints on the protein’s heavy atoms. This was followed by a 100-ps NPT simulation with harmonic position restraints on the protein’s heavy atoms. Production MD simulation runs were maintained at 300 K using a V-rescale thermostat ($\tau_T = 1$ ps) and 1 bar using the C-rescale barostat ($\tau_P = 2.0$ ps, compressibility $= 4.5 \times 10^{-5}$ bar$^{-1}$). Hydrogen bonds were constrained using the LINCS algorithm. The cut-off for short-range electrostatics was 1.2 Å. For each system, at least four random-seed replicates were performed. All systems were run for 500 ns per replicate. For analysis, frames corresponding to every 200 ps (i.e., 0.2 ns) were extracted (using gmx trjconv). RMSD and distance measurements were calculated using Gromacs tools (gmx rms and gmx distance). Data were plotted using matplotlib.

**AUTHOR CONTRIBUTIONS**

Nicholas J. East: Conceptualization (equal); data curation (lead); formal analysis (equal); investigation (equal); methodology (equal); writing – original draft (equal).

Ben E. Clifton: Data curation (supporting); writing – review and editing (equal).

Colin J. Jackson: Conceptualization (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal).

Joe A. Kaczmarski: Conceptualization (equal); supervision (lead); writing – original draft (equal); writing – review and editing (lead); Data curation (equal).

**ACKNOWLEDGMENTS**

This research was conducted by the Australian Research Council Centre of Excellence in Synthetic Biology (project number CE200100092) and funded by the Australian Government. Colin J. Jackson also acknowledges support from the ARC Centre of Excellence for Innovations in Peptide and Protein Science (CIPPS). Ben E. Clifton was supported by a JSPS Postdoctoral Fellowship for Overseas Researchers and KAKENHI Grant-in-Aid for Scientific Research (20F20705). This research/project was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. Open access publishing facilitated by Australian National University, as part of the Wiley - Australian National University agreement via the Council of Australian University Librarians.

**CONFLICT OF INTEREST**

The authors declare no conflicts of interest.

**DATA AVAILABILITY STATEMENT**

The data that support the findings of this study are openly available in Zenodo at http://doi.org/10.5281/zenodo.6519576.

**ORCID**

Nicholas J. East https://orcid.org/0000-0002-4471-449X

Ben E. Clifton https://orcid.org/0000-0002-7469-7424

Colin J. Jackson https://orcid.org/0000-0001-6150-3822

Joe A. Kaczmarski https://orcid.org/0000-0001-8549-0122

**REFERENCES**

1. Levy ED, Teichmann SA. Structural, evolutionary, and assembly principles of protein oligomerization. Prog Mol Biol Transl Sci. 2013;117:25–51. https://doi.org/10.1016/b978-0-12-386931-9.00002-7.

2. Marsh JA, Rees HA, Ahnert SE, Teichmann SA. Structural and evolutionary versatility in protein complexes with uneven stoichiometry. Nat Commun. 2015;6(1):6394. https://doi.org/10.1038/ncomms7394.

3. Levy ED, Pereira-Leal JB, Chothia C, Teichmann SA. 3D complex: A structural classification of protein complexes. PLoS Comput Biol. 2005;2(11):e155. https://doi.org/10.1371/journal.pcbi.0020155.

4. Kühner S, van Noort V, Betts MJ, et al. Proteome organization in a genome-reduced bacterium. Science. 2009;326(5957):1235–1240. https://doi.org/10.1126/science.1176343.

5. Schachman HK. From allostery to mutagenesis: 20 years with aspartate transcarbamoylase. Biochem Soc Trans. 1987;15(4):772–775. https://doi.org/10.1042/bst0150772.

6. Liu AK, Pereira JH, Kehl AJ, et al. Structural plasticity enables evolution and innovation of RuBiC0 assemblies. Sci Adv. 2022;8(34):eadc9440. https://doi.org/10.1126/sciadv.abc04940.

7. Perica T, Marsh JA, Sousa FL, et al. The emergence of protein complexes: Quaternary structure, dynamics and allostery. Biochem Soc Trans. 2012;40(3):475–491. https://doi.org/10.1042/bst20120056.

8. Pillai AS, Chandler SA, Liu Y, et al. Origin of complexity in haemoglobin evolution. Nature. 2020;581(7809):480–485. https://doi.org/10.1038/s41586-020-2292-y.

9. Bergendahl LT, Therese Bergendahl L, Marsh JA. Functional determinants of protein assembly into homomorphic complexes. Sci Rep. 2017;7(1):4932. https://doi.org/10.1038/s41598-017-05084-8.

10. Fraser NJ, Liu J-WW, Mabbutt PD, et al. Evolution of protein quaternary structure in response to selective pressure for increased Thermostability. J Mol Biol. 2016;428(11):2359–2371. https://doi.org/10.1016/j.jmb.2016.03.014.

11. Ruggiero A, Dattelbaum JD, Staiano M, Berisio R, D’Auria S, Vitagliano L. A loose domain swapping organization confers a remarkable stability to the dimeric structure of the arginine binding protein from Thermotoga maritima. PLoS One. 2014;9(5):e96560. https://doi.org/10.1371/JOURNAL.PONE.0096560.

12. Schulz L, Sendker FL, Hochberg GKA. Non-adaptive complexity and biochemical function. Curr Opin Struct Biol. 2022;73:102339. https://doi.org/10.1016/j.sbi.2022.102339.
13. Muñoz-Gómez SA, Bilolikar G, Wideman JG, Geiler-Samotere K. Constructive neutral evolution 20 years later. J Mol Evol. 2021;89(3):172–182. https://doi.org/10.1007/s00239-021-09996-y.

14. Hochberg GKA, Liu Y, Marklund EG, Metzger BPH, Laganowsky A, Thornton JW. A hydrophobic ratchet entrenches molecular complexes. Nature. 2020;588(7838):503–508. https://doi.org/10.1038/s41586-020-3021-2.

15. Zhu J, Avakyan N, Kakkis A, et al. Protein assembly by design. Chem Rev. 2021;121:13701–13796. https://doi.org/10.1021/acs.chemrev.1c00308.

16. Ueda G, Antanasićević A, Fallas JA, et al. Tailored design of protein nanoparticle scaffolds for multivalent presentation of viral glycoprotein antigens. Elife. 2020;9:e57659. https://doi.org/10.7554/elife.57659.

17. Park J, Selvaraj B, McShan AC, et al. De novo design of a homo-trimeric amantadine-binding protein. eLife. 2019;8:e47839. https://doi.org/10.22201/pdb6na/pdb.

18. Boyken SE, Chen Z, Groves B, et al. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. Science. 2016;352(6286):680–687. https://doi.org/10.1126/science.aad8865.

19. Fallas JA, Ueda G, Sheffler W, et al. Computational design of self-assembling cyclic protein homo-oligomers. Nat Chem. 2017;9(4):353–360. https://doi.org/10.1038/nchem.2673.

20. Kleiner D, Shapiro Tuchman Z, Shmulevich F, et al. Evolution of homo-oligomerization of methionine S-adenosyltransferases is replete with structure-function constraints. Protein Sci. 2022;31(7):e4352. https://doi.org/10.1002/pro.4352.

21. Clifton BE, Kaczmarski JA, Carr PD, Gerth ML, Tokuriki N, Jackson CJ. Evolution of cyclohexadienyl dehydratase from an ancestral solute-binding protein. Nat Chem Biol. 2018;14(6):542–547. https://doi.org/10.1038/s41589-018-0043-2.

22. Zhao GS, Xia TH, Fischer RS, Jensen RA. Cyclohexadienyl dehydratase from Pseudomonas aeruginosa. Molecular cloning of the gene and characterization of the gene product. J Biol Chem. 1992;267(4):2487–2493.

23. Kaczmarski JA, Mahawaththa MC, Feintuch A, et al. Altered conformational sampling along an evolutionary trajectory changes the catalytic activity of an enzyme. Nat Commun. 2020;11(1):5945. https://doi.org/10.1038/s41467-020-19695-9.

24. Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol. 2009;5(8):593–599. https://doi.org/10.1038/nchembio.186.

25. Gouridis G, Muthahari YA, de Boer M, et al. Structural dynamics in the evolution of a bilobed protein scaffold. Proc Natl Acad Sci USA. 2021;118(49):e2026165118. https://doi.org/10.1073/pnas.2026165118.

26. Cheepers GH, Lycklama A, Nijeholt JA, Poolman B. An updated structural classification of substrate-binding proteins. FEBS Lett. 2016;590(23):4393–4401. https://doi.org/10.1002/1873-3468.12445.

27. Shi R, Proteau A, Wagner J, et al. Trapping open and closed forms of FitE: A group III periplasmic binding protein. Proteins. 2009;75(3):598–609. https://doi.org/10.1002/prot.22272.

28. Li L, Ghimire-Rijal S, Lucas SL, et al. Periplasmic binding protein dimer has a second allosteric event tied to ligand binding. Biochemistry. 2017;56(40):5328–5337. https://doi.org/10.1021/acs.biochem.7b00657.

29. Gonin S, Arnoux P, Pierru B, et al. Crystal structures of an extractoplasmic solute receptor from a TRAP transporter in its open and closed forms reveal a helix-swapped dimer requiring a cation for α-keto acid binding. BMC Struct Biol. 2007;7(1):11. https://doi.org/10.1186/1472-6807-7-11.

30. Lynch M. Evolutionary diversification of the multimeric states of proteins. Proc Natl Acad Sci. 2022;110(30):E2821–E2828. doi:10.1073/pnas.1310980110.

31. Bai Y, Yu Z-H, Liu S, et al. Novel anticancer agents based on targeting the trimer interface of the PRL phosphatase. Cancer Res. 2016;76(16):4805–4815. https://doi.org/10.1158/0008-5472.can-15-2323.

32. Wrobel AG, Benton DJ, Hussain S, et al. Antibody-mediated disruption of the SARS-CoV-2 spike glycoprotein. Nat Commun. 2020;11(1):5337. https://doi.org/10.1038/s41467-020-19146-5.

33. Ben-David M, Koskine M, Dubovetskyi A, et al. Enzyme evolution: An epistatic ratchet versus a smooth reversible transition. Mol Biol Evol. 2020;37(4):1133–1147. https://doi.org/10.1093/molbev/msaz298.

34. Bridgman JT, Orlund EA, Thornton JW. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. Nature. 2009;461(7263):515–519. https://doi.org/10.1038/nature08249.

35. Harms MJ, Thornton JW. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. Nature. 2014;512(7513):203–207. https://doi.org/10.1038/nature13410.

36. Shah P, McCandlish DM, Plotkin JB. Contingency and entrenchment in protein evolution under purifying selection. Proc Natl Acad Sci. 2015;112(25):E3226–E3235. https://doi.org/10.1073/pnas.1429331112.

37. Xie VC, Pu J, Metzger BP, Thornton JW, Dickinson BC. Contingency and chance erase necessity in the experimental evolution of ancestral proteins. Elife. 2021;10:e67336. https://doi.org/10.7554/elife.67336.

38. Baier F, Hong N, Yang G, et al. Cryptic genetic variation shapes the adaptive evolutionary potential of enzymes. Elife. 2019;8:e40789. https://doi.org/10.7554/elife.40789.

39. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

40. Gibson DG. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods. 2009;6(5):343–345. https://doi.org/10.1038/nmeth.1318.

41. Mirdita M, Schütze K, Moriwa Y, Heo L, Ovchinnikov S, Steinwegger M. ColabFold: Making protein folding accessible to all. Nat Methods. 2022;19(6):679–682. https://doi.org/10.1038/s41592-022-01488-1.

42. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein prediction with AlphaFold. Nature. 2021;596(7873):481–487. https://doi.org/10.1038/s41586-021-03919-2.

43. Abraham MJ, Murtola T, Schulz R, et al. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14(1):71–73. https://doi.org/10.1038/nmeth.4067.
45. Joosten RP, Long F, Murshudov GN, Perrakis A. The PDB_REDO server for macromolecular structure model optimization. IUCrJ. 2014;1(4):213–220. https://doi.org/10.1107/S2052252514009324.

SUPPORTING INFORMATION
Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: East NJ, Clifton BE, Jackson CJ, Kaczmarski JA. The role of oligomerization in the optimization of cyclohexadienyl dehydratase conformational dynamics and catalytic activity. Protein Science. 2022;31(12):e4510. https://doi.org/10.1002/pro.4510