Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Differences in thermal structural changes and melting between mesophilic and thermophilic dihydrofolate reductase enzymes

Irene Maffucci,[a,b‡] Damien Laage,[*b] Guillaume Stirnemann,[*a] and Fabio Sterpone[*a]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

A key aspect of life's evolution on Earth is the adaptation of proteins to be stable and work in a very wide range of temperature conditions. A detailed understanding of the associated molecular mechanisms would also help to design enzymes optimized for biotechnological processes. Despite important advances, a comprehensive picture of how thermophilic enzymes succeed in functioning under extreme temperatures remains incomplete. Here, we examine the temperature dependence of stability and of flexibility in the mesophilic monomeric *Escherichia coli (Ec)* and thermophilic dimeric *Thermotoga maritima (Tm)* homologs of the paradigm dihydrofolate reductase (DHFR) enzyme. We use all-atom molecular dynamics simulations and a replica-exchange scheme that allows to enhance the conformational sampling while providing at the same time a detailed understanding of the enzymes' behavior at increasing temperatures. We show that this approach reproduces the stability shift between the two homologs, and provides a molecular description of the denaturation mechanism by identifying the sequence of secondary structure elements melting as temperature increases, which is not straightforwardly obtained in the experiments. By repeating our approach on the hypothetical TmDHFR monomer, we further determine the respective effects of sequence and oligomerization in the exceptional stability of TmDFHR. We show that the intuitive expectation that protein flexibility and thermal stability are correlated is not verified. Finally, our simulations reveal that significant conformational fluctuations already take place much below the melting temperature. While the difference between the TmDHFR and EcDHFR catalytic activities is often interpreted via a simplified two-state picture involving the open and closed conformations of the key M20 loop, our simulations suggest that the two homologs' markedly different activity temperature dependences are caused by changes in the ligand-cofactor distance distributions in response to these conformational changes.

1 Introduction

A longstanding goal in biocatalysis has been to gain a better molecular understanding of enzyme thermal adaptation. Avoiding protein denaturation and enhancing protein thermal stability is essential to design optimized enzymes capable of functioning in the harsh conditions required by many technological applications. Comparing homologous proteins from organisms which have naturally adapted to different temperature conditions provides a rich source of inspiration. While the enzymatic activity optimum occurs at ambient temperature in mesophilic organisms, it is dramatically shifted to higher temperatures in (hyper)thermophiles[1–3]. The increased structural stability of thermophiles at high temperatures has often been attributed to a reduced flexibility, and thus a greater rigidity, compared to mesophiles[4,5].

While this flexibility picture is intuitive and appealing, experimental and theoretical studies have suggested that the molecular mechanisms causing thermal adaptation can be far more complex[6–24]. In addition, several major questions regarding protein flexibility remain unclear. A first question pertains to the precise definition of molecular flexibility and its thermodynamic implications[17,18,23,25]: proteins are large molecular objects made of very

[a] *CNRS Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, PSL University, Sorbonne Paris Cité, 13 rue Pierre et Marie Curie, 75005, Paris, France*
[b] *PASTEUR, Département de chimie, École Normale Supérieure, PSL University, Sorbonne Université, CNRS, 24 rue Lhomond, 75005 Paris, France*
‡ *Université de technologie de Compiègne, UPJV, CNRS, Enzyme and Cell Engineering, Centre de recherche Royallieu - CS 60319 - 60203 Cedex*
* *E-mail: damien.laage@ens.psl.eu, stirnemann@ibpc.fr, sterpone@ibpc.fr*
† Electronic Supplementary Information (ESI) available: Supporting data on applied methods and additional figures on protein stability and flexibility. See DOI: 00.0000/00000000.

different subunits, including e.g. loop, helices and sheets, and local subunit flexibilities can be markedly different from the global structural flexibility which is expected to be more relevant for thermal denaturation; such heterogeneity is also important in the context of catalytic activity since active site structural features essential for catalysis could be affected differently among homologs and are not necessarily connected to the global protein flexibility. Experimentally, the global melting temperature is well established, but gaining precise information about secondary structure melting is much more challenging[26–28]. Numerical simulations are therefore a promising technique to obtain a molecular resolution for the melting process[29–31]. A second challenge is to identify the molecular origin of the reduced flexibility observed in thermophilic proteins: several factors, including for example changes in protein sequence, differences in subunit interactions, and oligomerization, have been suggested[2,14], but it is not clear whether all subunits have an enhanced rigidity in thermophilic proteins. Finally, a critical question is to determine the consequences for the enzymatic catalytic activity. A reduced conformational flexibility can hinder the conformational rearrangement dynamics necessary for catalysis, and this has been suggested to explain the lower activity of thermophilic enzymes at ambient conditions[32,33]. However, dynamical effects have been suggested to play a very limited role in enzyme catalysis[34,35], which suggests that differences in activity between enzyme homologs may be rather caused by changes in reaction free energy barriers.

Here we address these questions using state-of-the-art enhanced sampling all-atom molecular dynamics simulations of two mesophilic and thermophilic proteins. We selected the paradigm dihydrofolate reductase enzymes[36] from the *Escherichia coli* mesophile (EcDHFR) and from the *Thermotoga maritima* thermophile (TmDHFR) (Figure 1). While EcDHFR operates at ambient temperature and has a melting temperature[37] of 53°C, TmDHFR is the most stable known DHFR and melts[38] at ≈83°C.

These two homologous enzymes have already been very extensively studied, both experimentally and via simulations[32,38–56]. While most prior work has been directed at understanding their different catalytic activities, the structural stability difference between the two proteins has been comparatively less studied. Crystal structure determinations[44,46] revealed than in contrast with the EcDHFR monomer, TmDHFR forms a very stable homodimer. Since the similarity between the EcDHFR and TmDHFR sequences is only 25%, whether the increased stability of TmDHFR arises from its different sequence or from dimerization is a major and much-studied[32,39,53] question. Experimentally, the TmDHFR monomer cannot be isolated, thus precluding its melting temperature determination[32]. Experiments employing a single-point mutant whose monomer is stable when combined with a surfactant[32,53] suggested that dimerization plays an important role in TmDHFR thermal stability. However, mutations (even single ones), as well as changes in the solvent, are known to have potentially dramatic effects on protein stability. Molecular dynamics simulations of the hypothetical TmDHFR monomer have been used[51] to study its flexibility and unfolding pathway, but the melting temperature was not determined. Finally, many prior studies have established the critical importance of conformational rearrangements in the DHFR catalytic activity, and in particular the key role played by the closed Met20 loop conformation[36,42,50,57] in stabilizing the Michaelis-Menten complex with the folate substrate and facilitating the catalyzed hydride transfer reaction (Figure 1). In TmDHFR this loop is located next to the interface between monomers and is constrained in the open conformation. While dimerization could thus possibly be the cause of the lower TmDHFR activity, experiments on a monomeric TmDHFR variant showed that the open loop conformation does not result from dimerization, and that the monomer activity is as low as that of the native dimer[32]; recent calculations[39] also concluded that the low TmDHFR activity is mostly a result of its sequence and not of dimerization.

A key feature of our present study is the use of enhanced sampling simulations via a solute tempering replica-exchange technique[58], which has already been successfully applied ta address a variety of other biochemical questions[59–61]. With respect to prior simulation studies of mesophilic and thermophilic DHFR, this approach has several major advantages. First, in comparison with brute-force simulations, it provides a better sampling of the protein conformational equilibrium and makes it easier to escape from local free energy minima[58]. Second, it offers a description of the system over the full temperature range from ambient conditions to elevated temperature where unfolding occurs[22,25,62], and it is not restricted to some specific temperatures[39–41,51]: this therefore allows identifying e.g. at which temperature each subunit undergoes a structural change. Finally, in contrast with simulations[51,63,64] where a single unfolding event is observed in trajectories propagated at very high temperature, the progressive heating among the different replicas provides access to multiple local folding/unfolding events in a more reversible manner.

As described in the remainder of this article, our simulations show that the enhanced sampling methodology provides a quantitative description of the melting temperature shift between protein homologs in agreement with experimental data, and offers unprecedented insight into the melting temperatures of the different protein subunits. Our results show that neither local flexibility, as commonly measured via structure root mean square fluctuations (RMSF), nor conformational flexibility, assessed through conformational clustering, correlate with thermal stability. We further evidence that some subunits undergo large structural changes at temperatures much below the global melting temperature. Finally, a major result is that a critical feature that distinguishes the catalytic activities of enzyme homologs does not arise from their flexibility, but rather from their different temperature-induced conformational changes, and specifically their different fractions of reactive conformations.

Results

Flexibility

The increased thermal stability of thermophilic proteins has often been assigned to their enhanced rigidity[2,3,65,66]. We therefore first examine whether this popular picture applies to the present DHFR systems, and we compare the EcDHFR, TmDHFR dimer and hypothetical TmDHFR monomer flexibilities. As local and

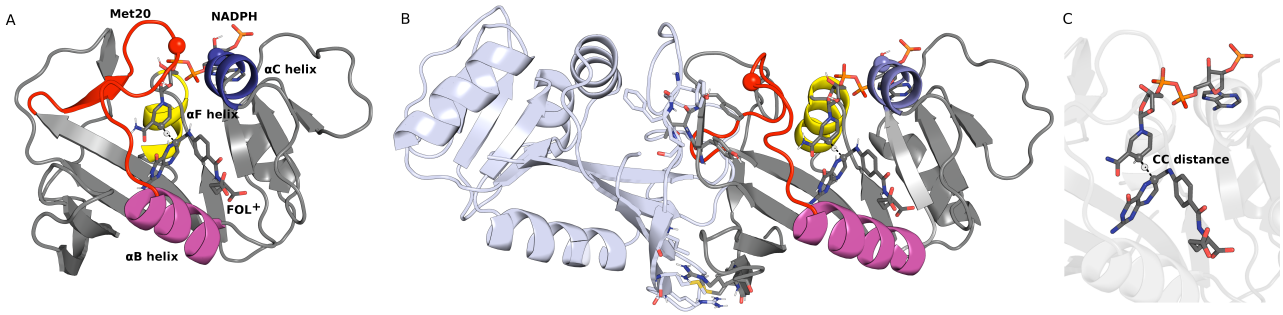

Fig. 1 Ec- and TmDHFR structures. Cartoon representations of the EcDHFR (A) and TmDHFR (B) protein 3D structures. Key structural elements are highlighted: Met20 loop (red), αB (magenta), αC (blue) and αF (yellow) helices. The Asn18/His45 (EcDHFR) and Val19/Ile46 (TmDHFR) pairs of residues used to determine the open and closed Met20 conformations based on the Met20 loop – αC helix distance are shown as plain balls. Substrate (FOL^+) and cofactor (NADPH) are reported as sticks. The TmDHFR interfacial residues are shown as sticks. (C) Magnified view of the active site FOL^+ and NADPH conformations with the CC donor–acceptor distance involved in the catalyzed hydride transfer.

global protein flexibilities could paint different images, we contrast the results obtained with the traditional RMSF calculation that provides a mostly local probe of flexibility with another approach reporting on global configurational entropy.

In order to enhance the sampling of protein conformations that would be too slow in brute-force molecular dynamics, our simulations employ solute-tempering Hamiltonian replica exchange REST2, which was already successfully applied to investigate protein melting in a broad range of conditions including dilute[22,62] and crowded solutions[67], together with powder systems[68]. Briefly, starting from the crystal structures and after minimization and equilibration in the apo state, each of the 24 replicas is propagated for 500 ns at the same physical temperature, but on a rescaled potential energy surface (see Methods and SI). Finally, we use a mean-field approximation and a corresponding state picture to translate the potential energy scaling factor into a temperature scale. The corresponding temperature which is effectively inferred from this approach is simply referred to as temperature T in the rest of this work.

We first analyze the enzyme local flexibility and characterize it using the backbone RMSF along the protein sequence (see Methods). Results for EcDHDR, TmDHFR monomer and dimer are shown in Figure 2 at ambient temperature ($T = 300$ K). Quite surprisingly, the TmDHFR dimer is overall not significantly more rigid than EcDHFR. While the TmDHFR protein core and most residues at the dimer interface exhibit smaller fluctuations than in EcDHFR, regions that are opposite to the interface are more mobile. Strikingly, this is also the case for the Met20 loop, despite its proximity to the dimer interface: while the interface constrains the Met20 loop global conformation (as examined in detail further), it does not prevent its fluctuations. The comparison between the TmDHFR monomer and dimer further reveals that the dimer RMSFs probably result from two counteracting effects: first, the sequence clearly favors fluctuations that are much more pronounced than in EcDHFR, especially in the Met20 loop region and its vicinity. On the other hand, dimerization leads to a rigidification of the regions adjacent to the interface, including the Met20 loop. The results are in agreement with recent simulations results published while this work was completed[39], and the same trend is observed when temperature is increased (see Figure 3B and Supplementary Information).

However, RMSF analyses have important limitations in the study of thermal stability. RMSF results probe the amplitude of backbone fluctuations around the average structure, and in typical brute-force simulations where the conformational space exploration is very limited, they therefore report mostly on local fluctuations around a single conformation. In enhanced sampling where conformational distribution is stochastically generated, RMSF values include contributions from both local fluctuations and conformational rearrangements but these respective contributions cannot be separated. Since global conformational flexibility is much more relevant than local fluctuations to understand thermal stability, we now examine another flexibility definition specifically based on protein global configurational entropy, and already applied to compare other mesophilic and thermophilic homologues[20,22,23]. The number of protein conformational states at each temperature is determined via a conformational clustering analysis performed on the last half of the apo protein REST2 trajectories. Protein conformations are compared by computing the RMSD of their C_α atoms once rigid body motions are removed, and the conformational states are identified with the well-established leader algorithm[69] using a 2.5 Å RMSD cutoff[23].

The numbers of conformational states are reported in Figure 3A for the three systems, and can be contrasted with the RMSF averaged along the protein sequence (Figure 3B and SI Fig. S9) at increasing temperatures. The global flexibility analysis yields results that confirm some of the observations made on the RMSF: for example, the conformational flexibility of TmDHFR monomer is greatly enhanced compared to the other two systems. Differences between EcDHFR and TmDHFR dimer are not obvious from the RMSF temperature dependence, because of compensations between more flexible and more rigid protein regions as can be seen in the 300 K RMSF results along the sequence (Figure 2). However, Figure 3A clearly demonstrates that EcDHFR visits a larger number of conformations than the TmDHFR dimer along the entire temperature range. A sudden increase in the EcDHFR conformational entropy is also evidenced above 330–340 K, revealing the onset of large conformational changes, in contrast with the TmDHFR monomer and dimer where this transition oc-

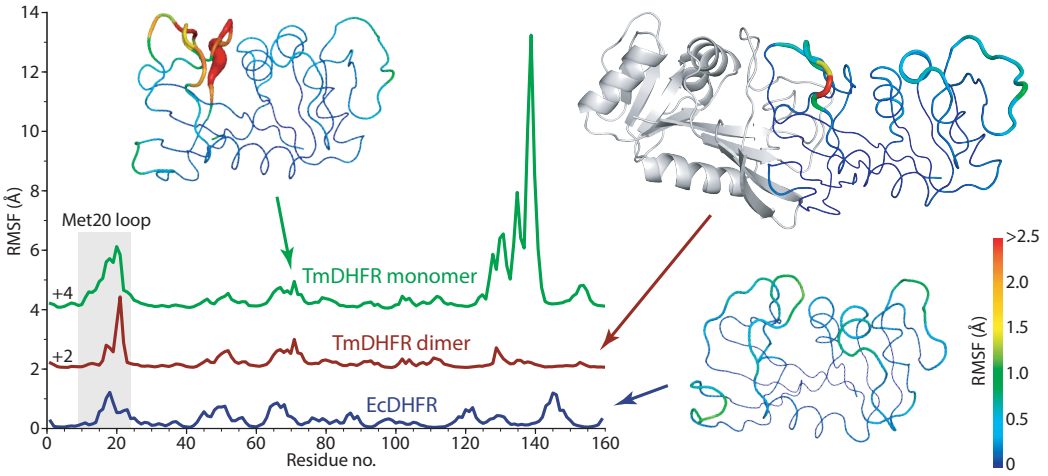

Fig. 2 Local flexibility Atomistic root-mean square fluctuations at 300 K along the total protein sequence for EcDHFR (blue), TmDHFR dimer (red) and monomer (green). For better visibility, the curves for TmDHFR are shown with an offset of 2 and 4 Å, respectively. The corresponding fluctuations are shown for each enzyme structure, with a color scale indicated on the right hand side. The thickness of the ribbons are proportional to the RMSF for values > 2.5 Å.

curs at higher temperatures.

Our results therefore suggest that first, the TmDHFR sequence tends to favor more local fluctuations and a larger number of accessible enzyme conformations on a broad temperature range. Second, the number of accessible conformations is noticeably larger for EcDHFR as compared to TmDHFR, especially above 330 K. Following the idea that increased rigidity leads to a greater resistance to temperature, it has often been argued that the dimerization of TmDHFR is responsible for its increased thermal stability[44,47,53]. While it is not apparent from the RMSF data, the conformational entropy perspective does confirm the TmDHFR dimer greater rigidity.

Protein melting temperature

We now turn to the determination of the homologs' melting temperatures. For each system, we analyze the distribution of conformations with increasing temperature, and we construct the protein stability curve which provides the fraction of folded proteins at each temperature. To do so, we first measure the C_α RMSD with respect to a reference ambient temperature equilibrium conformation for all replicas (Figure 4). Thermal denaturation results in the sampling of wider RMSD distributions around larger average values as temperature increases. Noticeably, EcDHFR (Figure 4A) starts sampling large deviations from the reference folded structure at lower temperatures than TmDHFR, both in its dimeric (Figure 4B) and monomeric (Figure 4C) forms, thus providing a first glimpse about their distinct thermal stabilities. These observations are further quantified by determining the fraction of folded conformations based on a C_α RMSD cutoff of 3 Å. We pause to note that the latter value was shown in prior work to properly separate folded and unfolded conformations[22,68]. In this calculation, loops are excluded, and for the TmDHFR dimer, only the thermally excited monomer is considered (see further details in the Methods section).

Figure 5 shows the simulated melting curves for EcDHFR, TmDHFR monomer and dimer, together with their melting temperatures T_m. While the simulated T_m exhibit a systematic shift with respect to the experimental values, the shift between the mesophilic EcDHFR and thermophilic TmDHFR dimer T_m is remarkably well captured by the simulations: $\Delta T_m^{sim} = 28$ K vs $\Delta T_m^{exp} = 30$ K[37,38].

The systematic shift between experimental and simulated absolute T_m values is not surprising, since it has already been reported for other systems[22,68] and is a combination of several factors like like force field and sampling limitation citeStirnemann2015. Concerning The most important one is that molecular force fields are traditionally calibrated on ambient-temperature data or calculations and are known to over-stabilize folded structures over unfolded configurations[70]. Even recent modifications do not allow to reproduce correctly the melting for many proteins[31]. For instance, temperature replica-exchange on much smaller polypeptides (a technique not applicable to our larger proteins), have shown that simulated melting temperatures are often overestimated, sometimes by more than 100 K[63,71–74].

We performed several checks of our results' robustness. First, we verified the simulation convergence and confirmed that the obtained melting curves do not significantly evolve during the last half of the 500 ns trajectories (SI Fig. S2), suggesting that simulations are converged on this accessible timescale. Second, we verified that the chosen RMSD cutoff value does not have a large impact on the resulting shift in thermal stability: for example, with a 3.5 Å cut-off we obtain $\Delta T_m \simeq 31$ K (see SI Fig. S4). Third, the choice of another collective variable to describe folding, i.e. the number of native contacts, leads to results in quantitative agreement with these found using the RMSD cutoff (see Methods and SI Figure S5). Finally, we have checked that the stability shift between EcDHFR and TmDHFR is observed for both the holo and apo states, respectively with and without the DHFR substrate and cofactor (SI Figure S6). In the former case, the presence of ligands in the active site slightly stabilizes the protein. The melting temperatures are therefore shifted toward

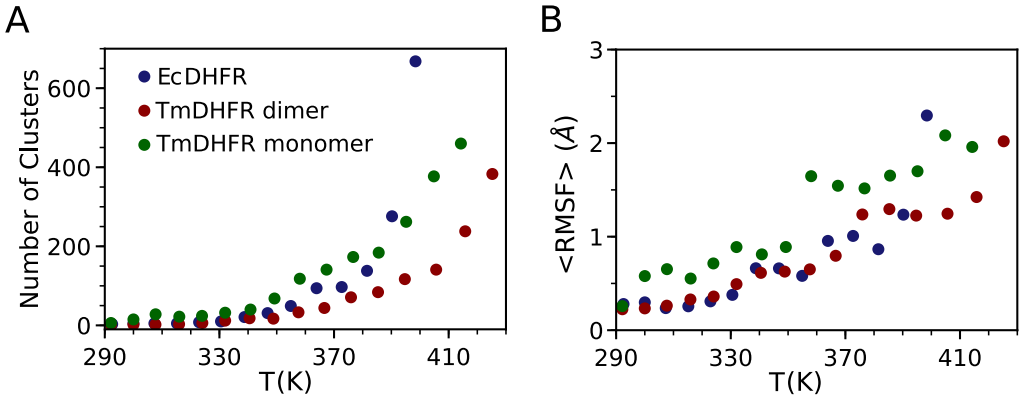

Fig. 3 Temperature-dependence of enzyme conformational and local flexibilities (A) Number of conformational states obtained by cluster analysis and (B) atomistic root-mean squared fluctuations (RMSF) averaged over the total protein sequence for EcDHFR (blue), Tm dimer (red) and monomer (green), for replicas whose temperature lies below the calculated melting temperature of each system.

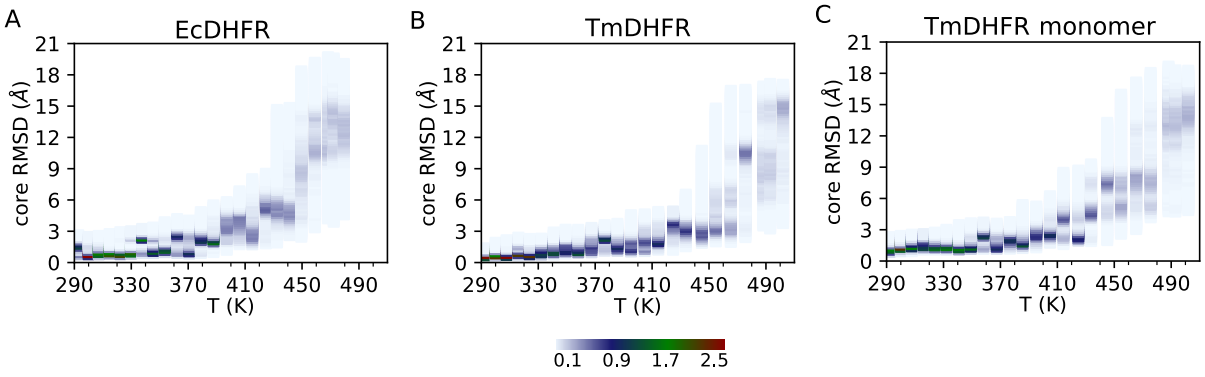

Fig. 4 Thermal stability: RMSD distributions Evolution of RMSD distributions with temperature for EcDHFR (A), TmDHFR dimer (B), and TmDHFR monomeric form (C).

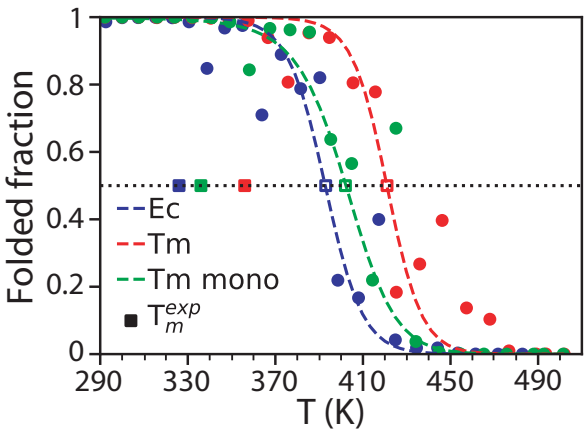

Fig. 5 Thermal stability: melting curves Fraction of folded protein conformations (colored circles) for EcDHFR (blue), TmDHFR dimer (red) and TmDHFR monomer (green), defined with a smoothed RMSD cutoff applied to the REST2 simulations in the apo state (see Supplementary Information), together with the corresponding fitted stability curves (dashes) and experimental (full squares) and resulting simulated (open squares) melting temperatures T_m.

higher values, although the shift between EcDHFR and TmDHFR is conserved (26 K, Figure S6). Note that this effect has also been observed in the experiments[75], on another DHFR homolog for which different ligands were measured to lead to a 3–15 K protein stability upshift.

The very good agreement between our simulation results and the experimental T_m shift between EcDHFR and TmDHFR dimer suggests that our methodology should also provide a good prediction of the T_m shift for the TmDHFR monomer.

Experimentally, this T_m value is not accessible because the TmDHFR monomer is not stable due to the apolar nature of the dimeric interface. Measurements for a monomeric TmDHFR mutant stabilized by surfactants[53] are used to deduce for this mutant the melting temperature. For the monomeric mutant in its apo state, T_m would be 3 K lower than that of EcDHFR in the absence of surfactants and 8 K higher in the presence of the surfactants[53]. While the authors noted that this mutant should be predominantly monomeric in both these conditions, it remains unclear how their results would compare to that of a hypothetical monomeric TmDHFR, in the absence of any mutation and surfactant effects. Our simulations thus provide the first determination of the TmDHFR monomer T_m melting temperature, and suggest that it is 7 K higher than that of EcDHFR and remains 21 K below that of the TmDHFR dimer. Our simulations therefore confirm that a major contribution to the exceptional TmDHFR thermal stability arises from its dimeric nature. However, they also illustrate that a reduced protein flexibility (characterized here with both local and global scales) is not necessarily associated with enhanced thermal stability. The TmDHFR monomer is much more flexible than the other two systems, but its melting temperature is intermediate. Such discrepancy between melting temperature and local flexibility was previously reported for the comparison between EcDHFR and the (monomeric) thermophilic BsDHFR[76,77], and more generally, a number of recent investigations[18,23] have questioned the relationship between flexibility and thermal stability.

Subunit melting temperatures and unfolding pathway

Another advantage of enhanced sampling molecular dynamics with respect to simulations run at some specific temperature[39–41,51] is that, in addition to the global melting temperature, they also provide valuable insight on the melting temperatures of individual protein subunits that are very challenging to access experimentally. Identifying which protein parts lose their native structure first when the temperature is increased offers a molecular picture of the unfolding pathway.

Following the same approach as described in the previous section for the global melting temperature, we determined the RMSD of each protein secondary structure (see SI Fig. S7) and repeated the melting curve calculations to determine their T_m values (SI Fig. S8). The results are presented in Figure 6, with the individual T_m values for EcDHFR, TmDHFR monomer and dimer (panels A-C). The T_m shifts between EcDHFR and TmDHFR monomer and between TmDHFR monomer and dimer are shown in Figure 7. The molecular resolution provided by these results allows identifying the weak spots where unfolding initiates, and the comparison between the three proteins offers a detailed analysis of the sequence and dimerization impacts on structural stability. We now discuss these different aspects.

We first address the unfolding mechanism. For all three systems, loops are the most flexible regions and undergo large deviations from their native conformation already 100 K below the global melting temperature. All αhelices exhibit lower melting temperatures than βsheets, in agreement with their well-established general stability difference. In all three systems, the αB helix unfolds first (see nomenclature in Figure 1, more than 40 K below the global melting temperature. Turning now to other helices, in EcDHFR the αC helix undergoes a significant structural change $\approx$ 20 K below T_m, while in TmDHFR the αF helix – very close to the Met20 loop and to αC, next to the active site (Figure 1) – unfolds in the same temperature range. These results provide a direct molecular interpretation of several experimental studies, where the presence of a significant population of non-native protein conformations for EcDHFR was readily seen at temperatures well-below melting[37,78,79]. The obtained results are also in agreement with a NMR study on another DHFR mutant[45], which showed that αB/C helices and the Met20 loop exhibit very low melting enthalpies as compared to other protein regions. We note that the unfolding pathway suggested from the different secondary structure T_m values is quite different from prior high temperature nonequilibrium unfolding simulations[51], which shows the importance of extensive sampling of the protein conformational equilibrium.

To understand the origin of TmDHFR dimer's exceptional thermal stability and the respective roles played by primary and quaternary structures, analyzing the secondary structure elements T_m considerably extends the first elements offered by the global T_m and provides a much more detailed investigation of the sequence and dimerization effects. The sequence change effects are as-

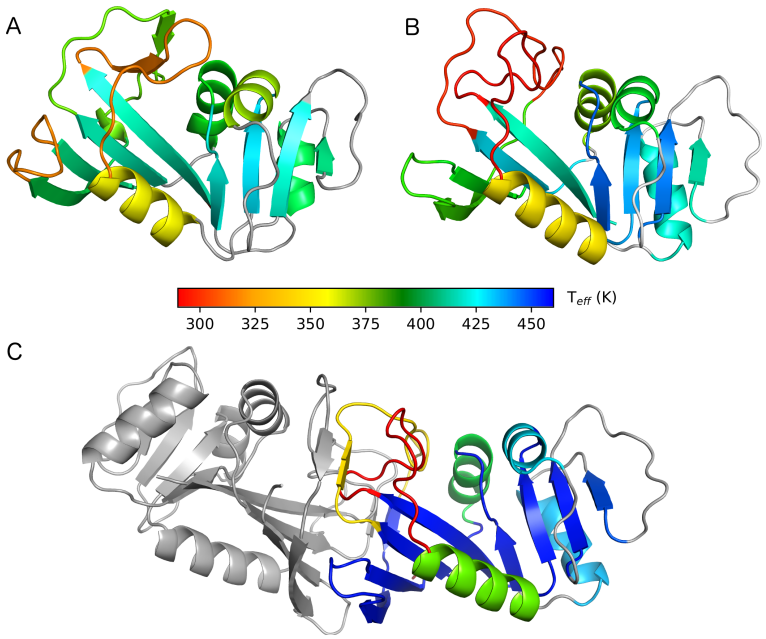

Fig. 6 Local stability: melting of secondary structure elements. Enzyme structures where each secondary structure element is colored according to its melting temperature, estimated individually for each of them using a RMSD cutoff of 2.5 Å (see Supplementary Information), for (A) EcDHFR, (B) an isolated monomer of TmDHFR, and (C) the TmDHFR dimer. Melting temperatures haven't been computed for the regions in grey.

sessed via the subunit T_m shifts between EcDHFR and monomeric TmDHFR, while the dimerization impact is measured by the T_m shifts between monomeric and dimeric TmDHFR. Our results in Figure 7 reveal that the different amino-acid sequence in TmDHFR with respect to EcDHFR has contrasted effects on the secondary structure stabilities: while the fragile αB helix is more stable in TmDHFR monomer, other protein parts including e.g. the αC helix and some β sheets are actually less stable in TmDHFR than in EcDHFR. We now turn to the effect of dimerization. Figure 7 reveals that dimerization strongly enhances the stability of all α helix and β sheet subunits. In addition, a clear gradient is observed away from the interface between monomers: the β sheet next to the dimerization interface is most affected and its melting temperature increases by more than 60 K; other subunits further from the interface are stabilized to a more moderate extent. Our results thus show that dimerization is essential for the thermal stability of TmDHFR, in agreement with prior suggestions[53] but also reveal that individual subunits are not stabilized to the same extent. This picture provides the first direct probe of sequence and dimerization effects for TmDHFR.

Consequences for catalytic activity

We now discuss the possible impacts of the identified structural features on these enzymes' catalytic activity. The chemical reaction catalyzed by DHFR is a hydride transfer leading to the reduction of 7,8-dihydrofolate (FOL) into 5,6,7,8-tetrahydrofolate (THF) that uses NADPH as a cofactor (Fig. 1C). Experimental measurements have shown that, at its optimum working temperature, the hydride transfer in EcDHFR is about 2 orders of magnitude faster than in the native (dimeric) TmDHFR[38,43,80], and that the monomeric TmDHFR mutant is an even poorer catalyst than the dimer[32,53]. We will now successively discuss two possible origins for these significant differences: first, the different conformational flexibilities in these systems, and second, the temperature impact in a protein region that is key to catalysis.

A decreased conformational flexibility can hinder the conformational rearrangement dynamics necessary for catalysis, and this has been suggested to explain the reduced activity of thermophilic enzymes at ambient conditions[1,81]. However, as temperature increases, the enzyme becomes more mobile, and the corresponding state picture suggested that thermophilic enzymes can recover activities similar to their mesophilic counterpart once they become flexible enough[1,4]. However, this idea has been strongly questioned by a number of experimental and theoretical investigations, and for the three DHFR systems studied here, it fails to account for the changes in activity. First, the TmDHFR monomer is more flexible than both the dimer and EcDHFR, but is the least catalytically active of the three[32]. The corresponding state picture does not hold either for the TmDHFR dimer as compared to EcDHFR. Our results have shown that the dimer is indeed more rigid than EcDHFR along the entire temperature range (Figure 3), but when examined along a reduced temperature scale T/T_m(Figure 8), both homologs exhibit very similar conformational flexibilities. On a wide temperature range below their respective melting temperature, TmDHFR samples a slightly larger number of conformational substates than EcDHFR, and a flexibility-based argument cannot explain why the TmDHFR optimum activity is two orders of magnitude lower than that of EcDHFR[38,53]. The global conformational rigidity enhancement from Ec to TmDHFR is thus directly connected to the increased thermal stability of TmDHFR, but it cannot be the main cause of the latter's much lower activity, as already suggested by experimental studies[32], as well as very recent simulations[39]. (We note

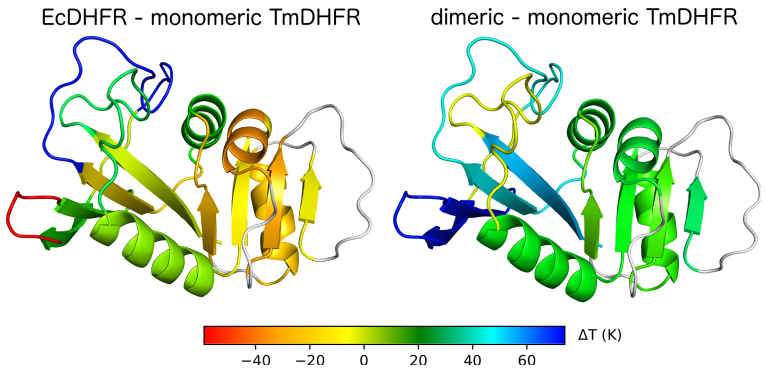

Fig. 7 Local stability: direct comparison between the three systems. Data identical to that of Figure 6, except that the *difference* in melting temperatures of each elements relative to the TmDHFR monomer is shown for EcDHFR (left) and the TmDHFR dimer (right, the unperturbed monomer is not shown). Melting temperatures haven't been computed for the regions in grey.

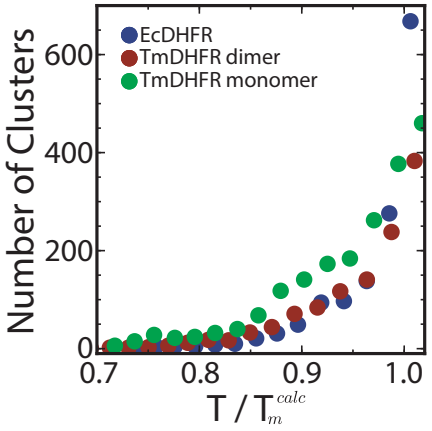

Fig. 8 Conformational flexibility Number of conformational states determined by cluster analysis along a reduced temperature obtained by rescaling T by the calculated melting temperature for EcDHFR (blue), TmDHFR dimer (red) and monomer (green).

that, not surprisingly, in *T. maritima*, the product of DHFR activity (tetrahydrofolate) is mainly supplied by another enzyme, the flavin-dependent thymidylate synthase[82].)

We now focus on the conformational changes occurring in EcDHFR and TmDHFR below their melting temperature and relevant for catalysis. We show that a key difference between the two homologs lies in the different fractions of protein–substrate conformations where the catalyzed hydride transfer reaction can take place with a low barrier.

Several prior studies (see, e.g, refs 36,48,57,83,84) stressed the important role played by the Met20 loop during the chemical step. In particular, closed loop conformations contribute to the stabilization of the Michaelis-Menten complex with the folate substrate and facilitate the catalyzed hydride transfer reaction (Figure 1). In EcDHFR, crystal structures indicate that this loop is closed under ambient conditions[36]. In contrast, in TmDHFR, the loop is located next to the interface between the two monomers[44] and lies in an open conformation. The lowered catalytic activity of TmDHFR has been attributed to the inability of this loop to close during the catalytic turnover[47]. Further studies on a monomeric TmDHFR variant showed that the poor activity does not result from dimerization[32], suggesting that the Met20 loop also lies in a weakly reactive conformation in the monomer as well. Eventually, recent calculations on the TmDHFR dimer and its monomer (which cannot be isolated in the experiments without mutations and addition of surfactant[53]) concluded that the low TmDHFR activity is mostly a result of its sequence and not of dimerization[39]. Because it is not directly accessible in the experiments, the detailed Met20 loop conformation in the monomer have not been fully characterized. We now discuss the results of our simulations for the holo states of EcDHFR, TmDHFR monomer and TmDHFR dimer on a wide temperature range. The simulation protocol is identical to that presented before for the apo state (see Methods and Supplementary Information), except that we have explicitly included the two reacting substrates to be able to demonstrate how the molecular information gained on the temperature-dependence of this loop can provide an interpretation to the temperature dependence of catalytic rate.

In our simulations, we monitor the Met20 loop opening and closing via the Met20 – αC helix distance (d_{M20}) as shown in Figure 1. The results for the EcDHFR, dimeric and monomeric TmDHFRs are reported in Figure 9A. d_{M20} takes values characteristic of a closed conformation in the EcDHFR holo state at ambient temperatures (typically between 6 and 8 Å[36]), but exhibits much larger values in the TmDHFR dimer, which correspond to an open conformation[44] (above 8 Å), as expected from the previous crystallographic studies[36,44]. The different Met20 loop configurations in Ec and TmDHFR could be due to TmDHFR's dimeric structure[44,46,51,85] or to the difference in sequences, as suggested by experiments on a TmDHFR monomeric mutant[32]. We therefore repeated our calculations on the TmDHFR monomer and found that the Met20 loop remains open over the entire temperature range (albeit taking smaller values than in the dimer), thus confirming that the open conformation is not a consequence of dimerization, but rather results from the different sequences. This is also in agreement with very recent calculations done with a different computational approach published while this work was completed[39].

As temperatures increases, the Met20 loop always remains open in both TmDHFR enzymes (Figure 9A), and interestingly,

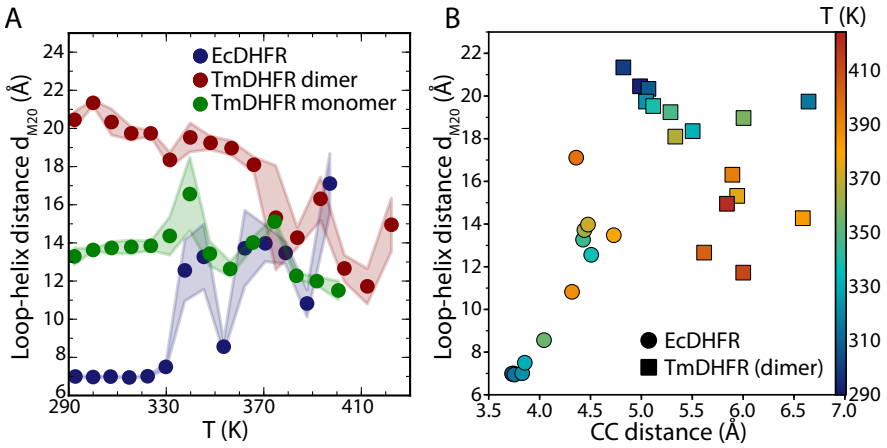

Fig. 9 Met20 Loop conformation and impact on CC distance. (A) Met20 loop conformations (defined as the as the Met20 loop – αB helix distance) as a function of temperature for EcDHFR (blue), dimeric TmDHFR (red) and monomeric TmDHFR in their holo states and for temperatures below their respective meltings. Colored areas represent standard deviations of the mean (Fig. S10 shows that apo state results are similar) (B) Correlation between average Met20 loop - αC helix distance and average CC distance for EcDHFR (spheres) and TmDHFR dimer (squares) at increasing temperatures (from blue to red).

the monomer d_{M20} remains almost unchanged, while the dimer d_{M20} decreases toward the monomer value. This is probably because the loop is less frustrated by the interface as temperature increases, and can adopt conformations that resemble more and more that in the monomer. By contrast, a clear transition from closed to open conformations occurs in EcDHFR around 330 K. We note that this transition is located well below the calculated melting temperature of EcDHFR (396 K). We now make the connections between the Met20 loop conformations and the hydride donor substrate – acceptor substrate distance (noted CC in the following). Smaller separations facilitate the hydride transfer[49] and are thus associated with lower reaction free-energy barriers.

Figure 9B shows that there is a clear correlation between d_{M20} and the average CC distance: short CC distances are only accessible in EcDHFR, for closed conformations of the Met20 loop. The opening of the loop at higher temperatures results in the exploration of longer distances. For TmDHDR, only open conformations are sampled at all investigated temperatures. For this system, the CC distance is uncorrelated to d_{M20}, and overall, longer average CC distances are explored as temperature increases.

While these average d_{M20} and CC distance values already provide significant insights, the impact on catalysis is better evidenced by considering the CC distance distributions (Figure 10A and B). For EcDHFR (Figure 10A), the loop remains closed up to 330 K, which results in a narrow CC distribution centered on ~ 3.7 Å. As the loop opens above this temperature, this peak strongly decreases (we note that a small residual population of closed conformations remains up to 390 K), much larger CC distances are sampled, and the distribution becomes broader. This increasing width is likely to be correlated with the increased conformational flexibility discussed earlier. For TmDHFR (Figure 10B), the absence of closed loop conformations at any temperature results in the absence of a peak at 3.7 Å. Instead, the enzyme samples large distances at all temperatures, and the width of the distribution increases with temperature, once again in agreement with the increasing number of accessible enzyme conformations.

As a consequence, short CC distances, which are necessary for the enzyme catalytic efficiency, are very rarely accessed in TmDHFR, which is likely to be correlated to its much lower catalytic activity. In contrast, they can be sampled for EcDHFR up to a point where the temperature increase leads to a loop opening. While a proper connection between the CC distance and the reaction free-energy barriers would involve an explicit treatment of the chemical reaction and its associated electronic structure rearrangements, together with an explicit determination of the CC distributions in the transition state geometries, our current results can already be analyzed in a more quantitative framework. At a given temperature, the catalytic efficiency of the enzymes will likely be dominated by the fraction of conformations with short CC distances, with the population with larger CC distances only contributing marginally to the chemical rate. (At ambient temperature, an increase of ~ 1.2 Å between EcDHFR and TmDHFR, in the average CC distance leads to a 4.2 kcal/mol difference in the free-energy barrier[43].) We estimated, at each temperature, the fraction of conformations P that exhibit CC distances shorter than the value corresponding to the maximum of the distributions at ambient temperature (located at $\mathrm{CC}^{max}_{300K} = 3.6$ Å for EcDHFR and $\mathrm{CC}^{max}_{300K} = 4.8$ Å for TmDHDR). The position of this cut-off is somehow arbitrary, but we checked that this choice did not significantly affect our following conclusions. We then define the average free-energy cost $\Delta G_{\mathrm{contraction}}(T)$ to reach such values as

$$\Delta G_{\mathrm{contraction}}(T) = -k_B T \ln P(\mathrm{CC} < \mathrm{CC}^{max}_{300\ \mathrm{K}}; T), \qquad (1)$$

where k_B is the Boltzmann constant. We report in Figure 11 the obtained $\Delta G_{\mathrm{contraction}}(T)$ values for both systems relative to their value at ambient temperature. This thus allows to estimate the expected increase in the reaction free-energy barrier due to the increased cost for the contraction of the CC distance. For EcDHFR, the temperature-dependence of $\Delta G_{\mathrm{contraction}}$

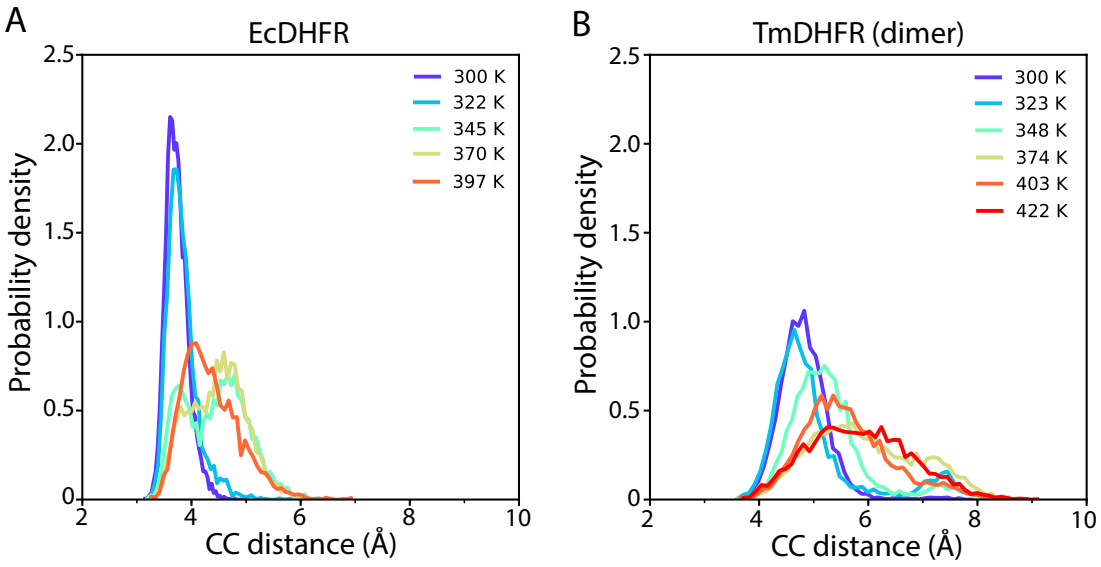

Fig. 10 CC distributions. Probability distribution of the hydride donor-acceptor distance (CC) sampled in REST2 simulations for a selection of replicas located every 20–30 K at temperatures below melting for the holo EcDHFR (A) and TmDHFR dimer (B) (see Fig. S11 for other temperature replica results).

is more pronounced because of the loop opening which results in a small fraction of reactive systems with short CC distances (Figure 10A). For TmDHFR, the increase is more moderate because no such transition occurs, and it only corresponds to the progressive broadening of the CC distributions and the exploration of larger (and increasingly poorly reactive) CC distances. These results are in agreement with experimental estimations of the reaction entropy for both systems that also suggest a more pronounced temperature-dependence of the reaction free-energy barriers in EcDHFR[43] ($\Delta S^{\ddagger}_{exp} = -26.1$ cal/mol/K for EcDHFR and $\Delta S^{\ddagger}_{exp} = -21.3$ cal/mol/K for TmDHFR). While our simple model underestimates $\Delta S^{\ddagger}$, the relative difference between both systems is qualitatively reproduced with $\Delta\Delta S^{\ddagger}_{sim} = 8.6$ cal/mol/K. Further extensions of this model will include an explicit treatment of electronic structure rearrangements in order to provide a more complete picture including intrinsic temperature effects on the barrier for a given conformation.

Finally, while the difference between the TmDHFR and EcDHFR catalytic activities is often interpreted via a simplified two-state picture involving the key M20 loop open and closed conformations, our results stress that the ligand-cofactor distance distributions also play a key role in the temperature-induced changes in the catalytic barrier. While at ambient temperature the different catalytic activities of EcDHFR and TmDHFR have been interpreted as arising from the different active site polarities caused by the closed and open loop conformations[86], our simulations show that the two DHFR homologs' markedly different activity temperature dependences are mostly caused by changes in the ligand-cofactor distance distributions.

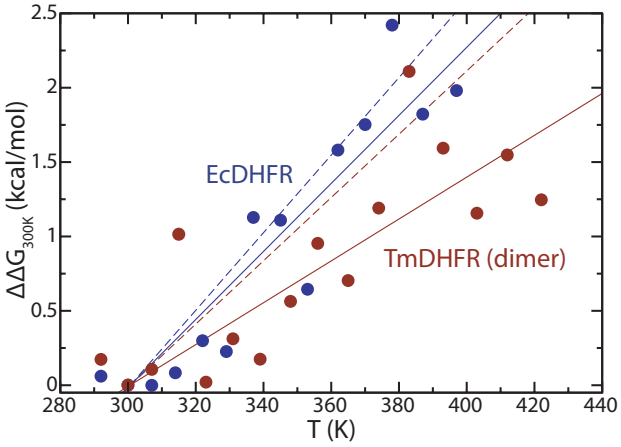

Fig. 11 Free-energy cost to contract the CC distance. Free-energy cost relative to its value at 300 K ($\Delta\Delta G = \Delta G_{\text{contraction}}(T) - \Delta G_{\text{contraction}}(300\ \text{K})$) to contract the CC distance (Equation 1), as a function of temperature for EcDHFR (blue dots) and TmDHFR (red dots). A linear fit (such that $\Delta\Delta G = 0$ at 300 K) of the simulation data is shown as a plain line with a similar color code, while dashed lines indicate the temperature-dependence of the free-energy barrier with respect to its value at 300 K that can be predicted from experimental reaction entropy estimations[43].

Conclusions

In this work we performed state-of-the-art enhanced sampling all-atom molecular dynamics simulations of two mesophilic and

thermophilic DHFR enzymes, with a special focus on the consequences of dimerization in the thermophilic homolog. Our approach allows a detailed analysis of the relationship between enzyme flexibilities (of which we used two definitions at different lengthscales) and thermal stability. As already observed for other systems[9,18,22,23], we find that an increased melting temperature is not necessarily associated with a more rigid protein matrix. The origin of the increased thermal stability of TmDHFR has often been attributed to its dimeric (and thus less flexible) structure. We find that this dimer is not significantly more rigid than the monomeric EcDHFR, which melts at lower temperatures. By comparing this dimer to its hypothetical monomeric state, we confirm that dimerization greatly contributes to the increased thermal stability of the dimer. We show that the melting temperature of the TmDHFR monomer, which cannot be experimentally isolated, is intermediate between that of EcDHFR and the TmDHFR dimer, whose stability shift is in very good agreement with experimental data. Only one third of the increased thermal resistance of TmDHFR over EcDHFR can be attributed to the difference in sequences, while the remaining 2/3 are due to dimerization. But strikingly, the TmDHFR monomer is much more flexible than both the dimer and EcDHFR. Therefore, the observed trend in flexibility does not correlate with that observed in melting temperatures.

The ability of our approach to achieve an extended sampling of the protein conformational landscape allowed to individuate the critical weakest points of the three systems during thermal denaturation and single out the role of the dimeric interface on the TmDHFR stability. Dimerization induces an increased stability in all secondary structure motifs of the monomer, with the strongest effects closer to the interface. The ability to individuate the stability hierarchy of structural elements in a protein fold could be applied to other systems and might provide important information for the design of thermally stable and active mutants.

We finally discuss the consequences of these observations on the catalytic activity of the two homologs. Experimentally, TmDHFR is much less efficient than EcDHFR. First, we rule out an explanation of these differences based on distinct enzyme flexibilities. Strikingly, we show that the TmDHFR monomer exhibits large conformational flexibility at all investigated temperatures, but still lacks the mesophilic EcDHFR catalytic efficiency. Moreover, conformational fluctuations are very similar in the TmDHFR dimer and EcDHFR when examined along a reduced temperature scale when approaching melting. The stability/activity trade-off[1,4,81] clearly does not hold for this system, and the modulation of its reactivity instead stems from key conformational differences and their temperature-dependence between both homologs. By contrast with the thermal stability, which mainly results from dimerization, the lower catalytic activity of TmDHFR as compared to EcDHFR is more likely to originate from its different sequence, as already suggested[53].

Our simulations reveal that at all investigated temperatures, the Met20 loop, which is crucial for catalysis, stays in an open (and poorly reactive) conformation in both TmDHFR dimer and its monomeric variant. This configuration results in large donor – acceptor distances between substrate molecules, which are very likely to lead to higher free-energy barriers. For EcDHFR, this loop is closed at ambient temperature, favoring much shorter donor – acceptor CC distances, and thus lower barriers. However, as temperature increases, a clear loop opening occurs and EcDHFR starts to explore less reactive configurations well below its melting temperature. For both homologs, we estimated the expected increase in the reaction free-energy barrier with temperature due to the increased cost for the CC distance contraction, and we showed that the effect of temperature on the free-energy barrier is more pronounced for EcDHFR, as measured experimentally. Interestingly, experimental measurements[38,43,80] have shown that the activation energy, i.e. the rate constant temperature dependence, is much smaller in Ec than in TmDHFR, while their reaction free-energy barriers differ by a few kcal/mol only. In a future work involving an explicit treatment of electronic structure rearrangements, we will focus on the implications of our current findings, in particular regarding how temperature-induced changes in the Met20 loop conformations affect the temperature-dependence of the catalytic rate.

Methods

A complete description of the protocols followed for systems (EcDHFR, TmDHFR et monomeric TmDHFR in both the apo and holo states) preparation, simulations and analyses performed are reported in the Supplementary Information. Here we report a brief summary of the employed methods.

Conformational sampling

A solute tempering Hamiltonian replica exchange (REST2)[58,64] was used for the conformational sampling of EcDHFR, TmDHFR and monomeric TmDHFR. Within this simulation scheme, we propagated each of the 24 replica for 500 ns (for a total of 12μs) at the same physical temperature, but at a rescaled protein potential energy. Successively, an effective temperature can be defined for each replica using the corresponding state principle and a mean field approximation[62].

For TmDHFR, only one of the two monomers was subjected to the rescaling scheme and the other one treated as solvent molecules. In the case of the MM complexes, the ligands were also treated as solvent to avoid unrealistic molecular geometries arising from potential energy rescaling. Only the last 250 ns of the simulations on EcDHFR and TmDHFR were considered for the analysis, after a careful assessment of convergence. In the case of TmDHFR the analyses were only performed on the rescaled monomer.

Fraction of folded proteins and stability curves

The fractions of folded protein were computed by analyzing the REST2 simulations on the apo states averaged over 50-ns trajectory portions. In order to define the fraction of protein in the folded state two different collective variables were used. The first is the root mean-square displacement (RMSD) computed on all non-loops C_α with respect to the average equilibrated structure and using a smoothed RMSD cutoff between folded and unfolded structures of 3.0 Å.

The second metric for the definition of the folded fraction is

based on the number of native contacts nc_i^* for a given C_α, which is the number of C_α atoms separated from it by at least 4 residues and located within a distance of 8.0 Å in the equilibrated structure.

Melting temperatures were obtained by performing a thermodynamic fit of the stability curves by applying the Hawley's expression[87] for the free energy difference between the folded and the unfolded state, which is derived from the folded fraction using Boltzmann statistics[87,88]:

$$\Delta G_u(T) = -\Delta C_p \left[T\left(ln\left(\frac{T}{T_m} \right) - 1 \right) + T_m \right] + \Delta H_u \left(1 - \frac{T}{T_m} \right) \quad (2)$$

with T_m being the melting temperature, ΔC_p the change in heat capacity when going from the folded to the unfolded state (constant), and ΔH_u the unfolding enthalpy, see SI Table S2. The fraction of folded protein $P(f)$ and $\Delta G(T)$ are in the following relationship:

$$P(f) = \frac{1}{1 + e^{-\frac{\Delta G_u}{k_B T}}} \quad (3)$$

The same approach was used to compute the melting temperatures of the different secondary structure motifs of the three systems, though only the Cα RMSD was used as a metric with a smoothed cutoff of 2.5 Å.

Protein flexibility

Local protein flexibility was assessed by computing the Cα RMSF on the last 250 ns of the simulation. Conversely, the global protein flexibility was investigated on the bases of the concept of protein configurational entropy. The number of conformational substates is determined via a cluster analysis performed on the above-described REST2 trajectories, with a Leader algorithm[69] and an Cα RMSD-based cutoff of 2.5 Å. For TmDHFR, only the rescaled monomer is considered.

Met20 loop conformations

We discriminate between the Met20 loop open and closed conformations by analyzing the distance between this loop and the C helix (loop-helix distance). In particular, we considered the distance between the EcDHFR Asn18 and His45 Cα atoms, which, after structural alignment, corresponds to the TmDHFR Val19 and Ile46 Cα atoms. The loop is considered closed if this distance is between 6 and 8 Å.

Conflicts of interest

The authors declare no competing interests.

Acknowledgements

The research leading to these results has received funding from PSL University (ANR-10-IDEX-0001-02 PSL*) through a collaborative chemistry research program, from the European Research Council (FP7/2007-2013, grant agreements 258748 to F.S. and 279977 to D.L.), and from the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01). Part of this work was performed using HPC resources from LBT, ENS, GENCI [CINES and TGCC] (Grants x201776818 and x2017077156) and PRACE allocation (Pra13-3298, 13th call).

Notes and references

1 G. N. Somero, *Annu. Rev. Ecol. Syst.*, 1978, **9**, 1–29.
2 C. Vieille and G. J. Zeikus, *Microbiol. Mol. Biol. Rev.*, 2001, **65**, 1–43.
3 S. Kumar and R. Nussinov, *Cell. Mol. Life Sci.*, 2001, **58**, 1216–1233.
4 G. N. Somero, *Annu. Rev. Physiol.*, 1995, **57**, 43–68.
5 F. H. Arnold, P. L. Wintrode, K. Miyazaki and A. Gershenson, *Trends Biochem. Sci.*, 2001, **26**, 100–106.
6 R. Jaenicke and G. Böhm, *Curr. Opin. Struct. Biol.*, 1998, **8**, 738–748.
7 M. Wolf-Watz, V. Thai, K. Henzler-Wildman, G. Hadjipavlou, E. Z. Eisenmesser and D. Kern, *Nat. Struct. Mol. Biol.*, 2004, **11**, 945–949.
8 P. Zavodszky, J. Kardos, A. Svingor and G. A. Petsko, *Proc. Natl. Acad. Sci. U.S.A*, 1998, **95**, 7406–7411.
9 G. Hernandez, F. E. Jenney, M. W. W. Adams and D. M. LeMaster, *Proc. Natl. Acad. Sci. U.S.A*, 2000, **97**, 3166–3170.
10 M. Tehei, D. Madern, B. Franzetti and G. Zaccai, *J. Biol. Chem.*, 2005, **280**, 40974–40979.
11 J. Fitter and J. Heberle, *Biophys. J*, 2000, **79**, 1629–1636.
12 R. Merkl and R. Sterner, *Perspect. Sci.*, 2016, **9**, 17–23.
13 V. Nguyen, C. Wilson, M. Hoemberger, J. B. Stiller, R. V. Agafonov, S. Kutter, J. English, D. L. Theobald and D. Kern, *Science*, 2017, **355**, 289–294.
14 F. Sterpone and S. Melchionna, *Chem. Soc. Rev.*, 2012, **41**, 1665–1676.
15 T. Lazaridis, I. Lee and M. Karplus, *Protein Sci.*, 1997, **6**, 2589–2605.
16 B. N. Dominy, H. Minoux and C. L. Brooks, 3rd, *Proteins*, 2004, **57**, 128–41.
17 L. Sawle and K. Ghosh, *Biophys. J.*, 2011, **101**, 217–227.
18 L. S. J. Huihui and K. Ghosh, *J. Chem. Theory Comput.*, 2017, **13**, 5065–6075.
19 M. Katava, M. Maccarini, G. Villain, A. Paciaroni, M. Sztucki, O. Ivanova, D. Madern and F. Sterpone, *Sci. Reports.*, 2016, **7**, 41092.
20 M. Katava, M. Kalimeri, G. Stirnemann and F. Sterpone, *J. Phys. Chem. B*, 2016, **120**, 2721–2730.
21 T. Zeiske, K. A. Stafford and A. G. Palmer, *J. Chem. Theory Comput.*, 2016, **12**, 2489–2492.
22 G. Stirnemann and F. Sterpone, *J. Phys. Chem. Lett.*, 2017, **8**, 5884–5890.
23 M. Kalimeri, O. Rahaman, S. Melchionna and F. Sterpone, *J. Phys. Chem. B*, 2013, **117**, 13775–13785.
24 M. Katava, M. Marchi, D. Madern, M. Sztucki, M. Maccarini and F. Sterpone, *J. Phys. Chem. B*, 2020, **124**, 1001–1008.
25 S. Timr, D. Madern and F. Sterpone, in *Protein thermal stability*, Elsevier, 2020.
26 B. E. Bowler, *Curr. Opin. Struct. Biol.*, 2012, **22**, 4–13.
27 S. S. Sarkar, J. B. Udgaonkar and G. Krishnamoorthy, *Biophys.*

J., 2013, **105**, 2392–402.
28 S. Sen Mojumdar, Z. N. Scholl, D. R. Dee, L. Rouleau, U. Anand, C. Garen and M. T. Woodside, *Nature Commun.*, 2017, **8**, 1881.
29 V. Daggett, *Acc. Chem. Res.*, 2002, **35**, 422–429.
30 K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana and D. E. Shaw, *J. Am. Chem. Soc.*, 2012, **134**, 3787–3791.
31 S. P. Paul Robustelli and D. E. Shaw, *Proc. Natl. Acad. Sci. U.S.A*, 2018, **115**, E4758–E4766.
32 J. Guo, E. J. Loveridge, L. Y. P. Luk and R. K. Allemann, *Biochemistry*, 2013, **52**, 3881–3887.
33 M. Elias, G. Wieczorek, S. Rosenne and D. S. Tawfik, *Trends Biochem. Sci.*, 2014, **39**, 1–7.
34 S. C. L. Kamerlin and A. Warshel, *Proteins*, 2010, **78**, 1339–75.
35 I. Tuñón, D. Laage and J. T. Hynes, *Arch. Biochem. Biophys.*, 2015, **582**, 42–55.
36 J. R. Schnell, H. J. Dyson and P. E. Wright, *Annu. Rev. Biophys. Biomol. Struct.*, 2004, **33**, 119–140.
37 E. Ohmae, T. Kurumiya, S. Makino and K. Gekko, *J. Biochem.*, 1996, **120**, 946–953.
38 G. Maglia, M. H. Javed and R. K. Allemann, *Biochem. J.*, 2003, **374**, 529–535.
39 J. J. Ruiz-Pernía, I. Tuñón, V. Moliner and R. K. Allemann, *ACS Cat.*, 2019, **9**, 5902–5911.
40 Y. Y. Sham, B. Ma, C.-J. Tsai and R. Nussinov, *Proteins Struct. Funct. Genet.*, 2002, **46**, 308–320.
41 J. Tian, J. C. Woodard, A. Whitney and E. I. Shakhnovich, *PLOS Comput. Biol.*, 2015, **11**, e1004207.
42 A. R. Mhashal, A. Vardi-Kilshtain, A. Kohen and D. T. Major, *J. Biol. Chem.*, 2017, **292**, 14229–14239.
43 E. J. Loveridge, L.-H. Tey, E. M. Behiry, W. M. Dawson, R. M. Evans, S. B.-M. Whittaker, U. L. Günther, C. Williams, M. P. Crump and R. K. Allemann, *J. Am. Chem. Soc.*, 2011, **133**, 20561–20570.
44 T. Dams, G. Auerbach, G. Bader, U. Jacob, T. Ploom, R. Huber and R. Jaenicke, *J. Mol. Biol.*, 2000, **297**, 659–672.
45 O. A. Oyeyemi, K. M. Sours, T. Lee, K. A. Resing, N. G. Ahn and J. P. Klinman, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 10074–10079.
46 T. Dams and R. Jaenicke, *Biochemistry*, 1999, **38**, 9169–78.
47 R. K. Allemann, E. J. Loveridge and L. Y. P. Luk, *Protein Motions, Dynamic Effects and Thermal Stability in Dihydrofolate Reductase from the Hyperthermophile Thermotoga maritima*, Springer International Publishing, Cham, 2015, pp. 99–113.
48 J. L. Radkiewicz and C. L. Brooks, *J. Am. Chem. Soc.*, 2000, **122**, 225–231.
49 M. Roca, H. Liu, B. Messer and A. Warshel, *Biochemistry*, 2007, **46**, 15076–15088.
50 K. Arora and C. L. Brooks, *Top. Curr. Chem.*, 2013, **337**, 165–187.
51 J. Pang and R. K. Allemann, *Phys. Chem. Chem. Phys.*, 2007, **9**, 711–718.
52 J. Pang, J. Pu, J. Gao, D. G. Truhlar and R. K. Allemann, *J. Am. Chem. Soc.*, 2006, **128**, 8015–8023.
53 E. J. Loveridge, R. J. Rodriguez, R. S. Swanwick and R. K. Allemann, *Biochemistry*, 2009, **48**, 5922–5933.
54 A. Kohen, *F1000Research*, 2015, **4**, 1464.
55 E. J. Loveridge, E. M. Behiry, R. S. Swanwick and R. K. Allemann, *J. Am. Chem. Soc.*, 2009, **131**, 6926–6927.
56 L. Y. P. Luk, E. J. Loveridge and R. K. Allemann, *J. Am. Chem. Soc.*, 2014, **136**, 6862–6865.
57 M. R. Sawaya and J. Kraut, *Biochemistry*, 1997, **36**, 586–603.
58 L. Wang, R. A. Friesner and B. J. Berne, *J. Phys. Chem. B*, 2011, **115**, 9431–9438.
59 A. Niitsu, S. Re, H. Oshima, M. Kamiya and Y. Sugita, *J. Chem. Inf. Model.*, 2019, **59**, 3879–3888.
60 P. Kührová, R. B. Best, S. Bottaro, G. Bussi, J. Šponer, M. Otyepka and P. Banáš, *J. Chem. Theory Comput.*, 2016, **12**, 4534–48.
61 L. Wang, B. J. Berne and R. A. Friesner, *Proc. Natl. Acad. Sci. U.S.A*, 2012, **109**, 1937–42.
62 G. Stirnemann and F. Sterpone, *J. Chem. Theory Comput.*, 2015, **11**, 5573–5577.
63 A. G. Rocco, L. Mollica, P. Ricchiuto, A. M. Baptista, E. Gianazza and I. Eberini, *Biophys. J.*, 2008, **94**, 2241–51.
64 G. Stirnemann, S.-G. Kang, R. Zhou and B. J. Berne, *Proc. Natl. Acad. Sci. U.S.A*, 2014, **111**, 3413–8.
65 R. Jaenicke and P. Závodszky, *FEBS Lett.*, 1990, **268**, 344–349.
66 A. Karshikoff, L. Nilsson and R. Ladenstein, *FEBS Journal*, 2015, **282**, 3899–3917.
67 D. Gnutt, S. Timr, J. Ahlers, B. König, E. Manderfeld, M. Heyden, F. Sterpone and S. Ebbinghaus, *J. Am. Chem. Soc.*, 2019, **141**, 4660–4669.
68 M. Katava, G. Stirnemann, M. Zanatta, S. Capaccioli, M. Pachetti, K. L. Ngai, F. Sterpone and A. Paciaroni, *Proc. Natl. Acad. Sci.*, 2017, **114**, 9361–9366.
69 J. A. Hartigan, *Clustering Algorithms*, John Wiley & Sons, Inc., New York, NY, USA, 99th edn, 1975.
70 S. Piana, J. L. Klepeis and D. E. Shaw, *Curr. Opin. Struct. Biol.*, 2014, **24**, 98–105.
71 P. Kührová, A. De Simone, M. Otyepka and R. B. Best, *Biophys J*, 2012, **102**, 1897–906.
72 D. Paschek, S. Hempel and A. E. García, *Proc. Natl. Acad. Sci. U.S.A*, 2008, **105**, 17754–9.
73 C. Yang, S. Jang and Y. Pak, *Nature Commun.*, 2014, **5**, 5773.
74 B. Adamczak, M. Kogut and J. Czub, *Phys. Chem. Chem. Phys.*, 2018, **20**, 11174–11182.
75 R. M. Evans, E. M. Behiry, L. H. Tey, J. Guo, E. J. Loveridge and R. K. Allemann, *ChemBioChem*, 2010, **11**, 2010–2017.
76 L. Meinhold, D. Clement, M. Tehei, R. Daniel, J. L. Finney and J. C. Smith, *Biophys J*, 2008, **94**, 4812–8.
77 J. Guo, L. Y. P. Luk, E. J. Loveridge and R. K. Allemann, *Biochemistry*, 2014, **53**, 2855–63.
78 A. C. Clark and C. Frieden, *J. Mol. Biol.*, 1999, **285**, 1765–76.
79 R. M. Ionescu, V. F. Smith, J. C. O'Neill, Jr and C. R. Matthews,

Biochemistry, 2000, **39**, 9540–50.
80 R. S. Sikorski, L. Wang, K. A. Markham, P. T. R. Rajagopalan, S. J. Benkovic and A. Kohen, *J. Am. Chem. Soc.*, 2004, **126**, 4778–4779.
81 G. Feller, *J. Phys.: Condens Matter*, 2010, **22**, 323101.
82 K. Karunaratne, N. Luedtke, D. M. Quinn and A. Kohen, *Arch. Biochem. Biophys.*, 2017, **632**, 11–19.
83 P. K. Agarwal, S. R. Billeter, P. T. R. Rajagopalan, S. J. Benkovic and S. Hammes-Schiffer, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 2794–2799.
84 S. J. Benkovic and S. Hammes-Schiffer, *Science*, 2003, **301**, 1196–1202.
85 T. Dams, G. Böhm, G. Auerbach, G. Bader, H. Schurig and R. Jaenicke, *Biol. Chem.*, 1998, **379**, 367–71.
86 L. Y. P. Luk, J. J. Ruiz-Pernia, A. S. Adesina, E. J. Loveridge, I. Tuñón, V. Moliner and R. K. Allemann, *Angew Chem Int Ed Engl*, 2015, **54**, 9016–9020.
87 S. A. Hawley, *Biochemistry*, 1971, **10**, 2436–2442.
88 W. J. Becktel and J. A. Schellman, *Biopolymers*, 1987, **26**, 1859–1877.