Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

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

Irene Maffucci,[†,‡,¶] Damien Laage,[*,‡] Guillaume Stirnemann,[*,†] and Fabio Sterpone[*,†]

†*CNRS Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, PSL University, Sorbonne Paris Cité, 13 rue Pierre et Marie Curie, 75005, Paris, France*

‡*PASTEUR, Département de chimie, École Normale Supérieure, PSL University, Sorbonne Université, CNRS, 24 rue Lhomond, 75005 Paris, France*

¶*Sorbonne Universités, Université de Technologie de Compiègne,UMR CNRS 7025, Enzyme and Cell Engineering Laboratory, Rue Roger Couttolenc, CS 60319, 60203 Compiègne, France*

E-mail: damien.laage@ens.fr; stirnemann@ibpc.fr; fabio.sterpone@ibpc.fr

1. Simulation methodology

1.1. Structure preparation

The X-ray structures of *E. coli* DHFR (EcDHFR) in complex with folate and NADP+, of *T. maritima* DHFR (TmDHFR) in complex with methotrexate and NADPH were used as initial structures (PDB codes 1RX2[1] and 1D1G[2], respectively). The monomeric TmDHFR was obtained by removing one monomer from the TmDHFR crystallographic structure. In order to generate the apo states, all the ligands eventually present were manually removed. Conversely, for the Michaelis-Menten complex (MM complex) of EcDHFR the 7,8-dihydrofolate (FOL) and the NADP+ were replaced by the N5 protonated 7,8-dihydrofolate (FOL+) and NADPH, respectively. For the MM complex of TmDHFR, the methotrexate was mutated to FOL+ and its pterin ring moiety was flipped of 180deg, in order to correctly reproduce the FOL binding pose.[3] The protonation state of the residues of the obtained systems was set at physiological conditions (pH = 7, salinity= 0.10 M) and hydrogens were added with the Chimera software[4]. The parameters for NADPH and FOL+ were taken from the literature[5], protein atoms were described by the ff99SB Amber force field[6], and water molecules with the TIP3P[7] model. Each system was solvated with a cubic box of water molecules ensuring that all protein atoms were at least 10 Å from the box edges, and the negatively charged proteins were neutralized by adding an adequate number of Na+ ions.

1.2. Simulation parameters

Most simulations were performed with the NAMD 2.9 software[8]. PME algorithm (grid spacing = 1 Å) was used to handle long-range contributions of electrostatic interactions, while a cutoff of 9 Å was set for short-range interactions and real space contribution of electrostatic interactions. All bonds involving hydrogens were constrained.

1.3. Initial equilibration

After minimization, the systems were equilibrated under ambient conditions for 200 ns in the NPT ensemble using a Langevin thermostat (characteristic time 1 ps, T=300 K) and barostat (dumping time 50 fs, P = 1 atm) and an integration time of 2 fs.

1.4. REST2 simulations for conformational sampling and thermal stability

REST2 simulations were performed by using an in-house implementation in NAMD 2.9[9]. Within the REST2 scheme the replica evolve at a reference temperature β_{ref}, while the potential energy of the nth replica (E_n) is rescaled as:

$$E_n(\vec{X}) = \lambda_n E_{pp}(\vec{X}) + \sqrt{\lambda_n} E_{pw}(\vec{X}) + E_{ww}(\vec{X}) \tag{S1}$$

with $E_{pp}(\mathbf{X})$, $E_{pw}(\mathbf{X})$ and $E_{ww}(\mathbf{X})$ being the protein-protein, protein-solvent and solvent-solvent potential energies. Therefore, in each nth replica, the solvent evolves at the reference temperature, protein-solvent interactions at a temperature $\beta_n = \sqrt{(}\lambda_n)\beta_{ref}$, and protein-protein interactions at $\beta_n = \lambda_n \beta_{ref}$. For these latter, only the dihedral and the non-bonded degrees of freedom were rescaled, while protein bonds, angles and impropers are left unperturbed. This was ensured by rescaling the protein dihedral force constants and Lennard-Jones energies by λ_n and protein atomic charges by $\sqrt{(}\lambda_n))$[9,10]. Because protein-protein and protein-solvent interactions are scaled differently, we have shown that an effective temperature $\langle\beta'_n\rangle$ can be defined for each replica using the corresponding state principle and a mean field approximation[9]:

$$\langle\beta'_n\rangle = \beta_n \left(1 + \left(\sqrt{\frac{\beta_{ref}}{\beta_n}} - 1\right)\left\langle \frac{E_{pw}(\vec{X})}{E_{pw}(\vec{X}) + E_{pp}(\vec{X})} \right\rangle\right) \tag{S2}$$

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. 24 replica exchanging protein-protein corresponding temperatures of 289, 300, 311, 323, 335, 347, 360, 373, 387, 402, 417, 432, 448, 465, 482, 500, 519, 538, 558, 579, 600, 625, 634, 652 K were used and the replica were allowed to exchange every 10 ps (success rate 25%). The simulation protocol was similar to that used for the equilibration of the reactant state, except that atomic coordinates were output every 50 ps. Each simulation was run for 500 ns/replica, for a total of 12 μ. Overall, the replicas scanned an effective temperature window of Teff $\in$ [292 K, 500 K]. Only the last 250 ns of the simulations on EcDHFR and TmDHFR were considered for the analysis. In the case of TmDHFR the analyses were only performed on the rescaled monomer. In order to prevent the ligands from leaving the binding site, during the REST2 simulations of the MM complexes, harmonic restraints on three protein-FOL+ and three protein-NADPH distances were included. These distances were selected by evaluating the most stable hydrogen bonds and hydrophobic interactions between the protein and the ligands along the 200-ns classical equilibration of each complex (see Table S1). The distances between the donor atom and the acceptor atom for each hydrogen bond, or between two atoms involved in the hydrophobic interaction were computed and used to define the equilibrium distance for the harmonic restraint as the most occurring one in the relative distribution (Table S1). The harmonic force constant was set to 5 kcal $\cdot$ mol^{-1} $\cdot$ Å^{-2}.

1.5. Simulation convergence

Although REST2 simulations convergence is difficult to assess, the similar average of relevant metrics (RMSD and fraction of native contacts) as a function of the effective temperature showed little evolution over the 250 ns of the simulations (Figure S1).

Table S1: Atom selected for REST2 distqnce harmonic restraints and their equilibrium distances

EcDHFR			TmDHFR		
Atom1	Atom2	Eq. distance (Å)	Atom1	Atom2	Eq. distance (Å)
NPH-N7N	Ala7-O	3.0	NPH-O'A5	Ile46-N	2.7
NPH-O'N5	Gly97-N	3.0	NPH-O'N5	Lys103-N	3.0
NPH-OPA2	Thr46-OG1	2.7	NPH-OPA2	Thr47-OG1	3.4
FOL^+-O2	Arg57-NH2	2.8	FOL^+-C16	Phe31-CZ	4.0
FOL^+-N2	Asp27-OD1	2.8	FOL^+-O1	Arg58-NH1	2.7
FOL^+-N8	Ile5-O	3.0	FOL^+-N8	Val6-O	3.1

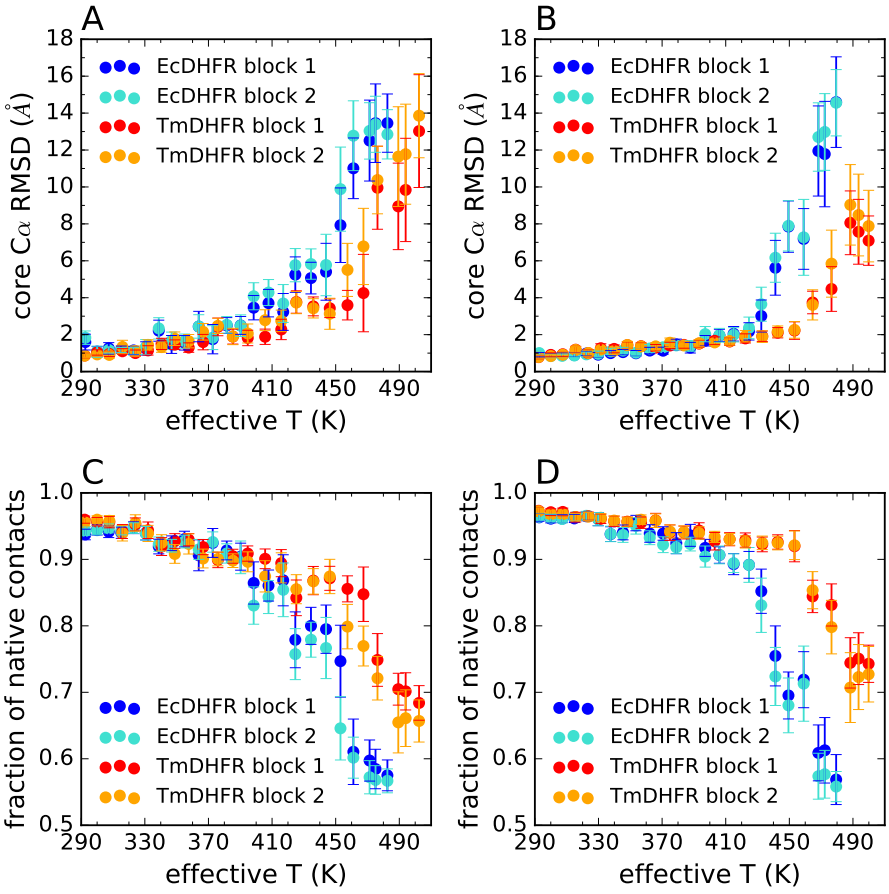

Figure S1: **Convergence assessment.** Analysis of the last 250 ns of apo state (A) and MM complex (B) REST2 replica in two blocks.

2. Structural analysis

2.1. 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.

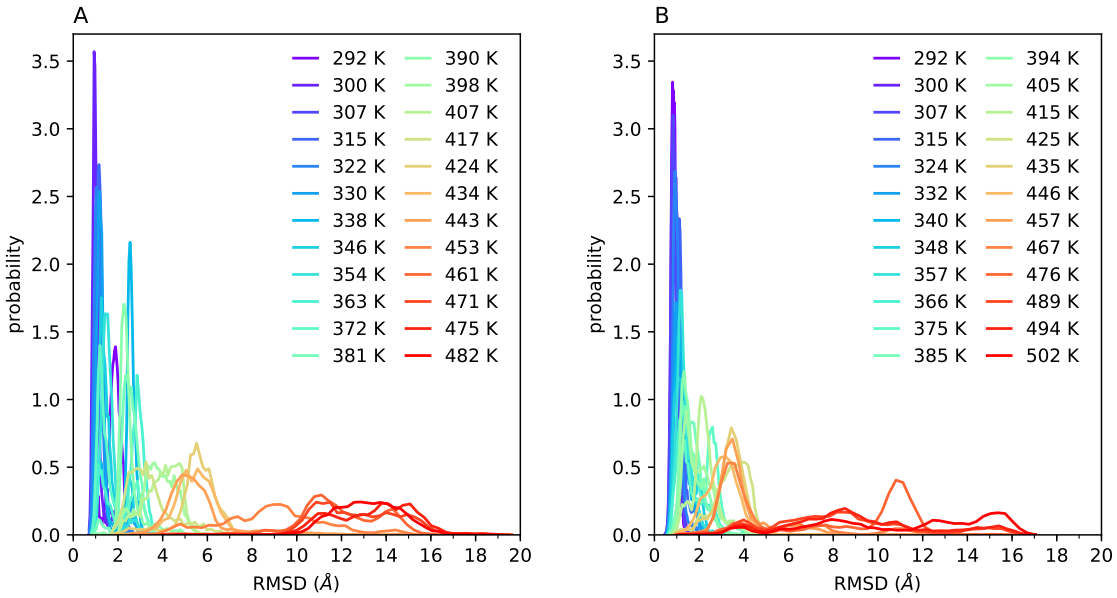

Figure S2: **Temperature effect on RMSD.** Distributions of the RMSD calculated for the rigid core of the proteins, EcDHFR (panel A) and TmDHFR (panel B), extracted from the REST2 trajectories.

The first is the root mean-square displacement (RMSD) computed on all non-loops C_α with respect to the average equilibrated structure. The folded fraction P(f) at each effective temperature was therefore defined as:

$$P(f) = \langle\langle(1-(RMSD_i - cutoff)^n)/(1-(RMSD_i - cutoff)^{2n})\rangle_w\rangle \tag{S3}$$

with RMSD_i being the RMSD of the ith frame. This way to express the folded fraction allows the smoothing of an otherwise discrete quantity. The RMSD cutoff between folded and

unfolded structures was set to 3.0 Å after the inspection of the per-replica RMSD distributions and the trajectories (Figure S2). Conversely, n was set to 30 in order to make the smoothing to occur in an interval of $\pm$ 0.5 Å around the cutoff (Figure S3). The inner average was calculated over a time window of 50 ns and the outer average was calculated along the trajectory over blocks of 50 ns each. In addition, 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 $\Delta T_m \simeq 31\,\mathrm{K}$ (Figure S4).

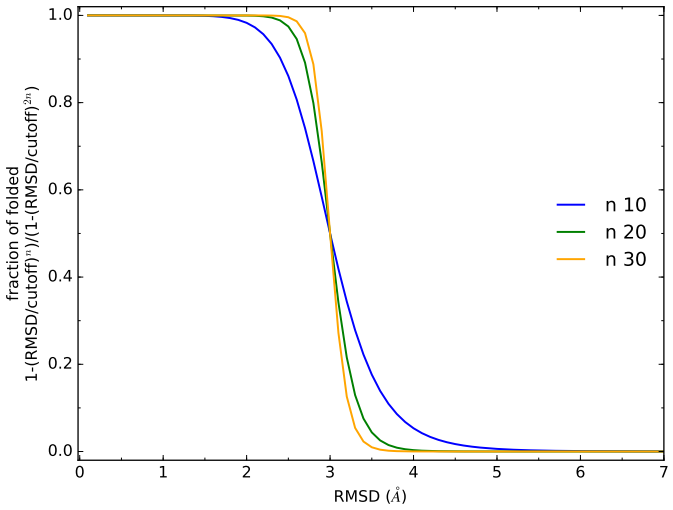

Figure S3: **Smoothing function.** Smoothing function for the definition of the fraction of folded conformations as a function of RMSD with a 3 Å cutoff and n = 10 (blue), 20 (green) and 30 (yellow). For the subsequent analysis, n was set to 30 to make the smoothing to occur in an interval of ± 0.5 Å around the cutoff.

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 Åin the equilibrated structure. The fraction of native contacts at time t (Q_t) is therefore:

$$Q_t = \frac{1}{N_{C\alpha}} \sum_{i=1}^{N_{C\alpha}} \frac{nc_i(t)}{nc_i'} \tag{S4}$$

with N_{C_α} being the number of C_α having nc_i^* native contacts in the reference state and $nc_i(t)$ is the number of native contacts at time t. We considered as folded the protein conformations possessing at least 85% of the native contacts. The stability curves obtained with this approach (Figure S5) are very similar to that obtained using a RMSD-based cutoff, suggesting that our results are not significantly by the choice of the metric used to define the folded protein fraction.

The thermodynamic fit of the stability curves was done by applying the Hawley's expression[11] for the free energy difference between the folded and the unfolded state, which is derived

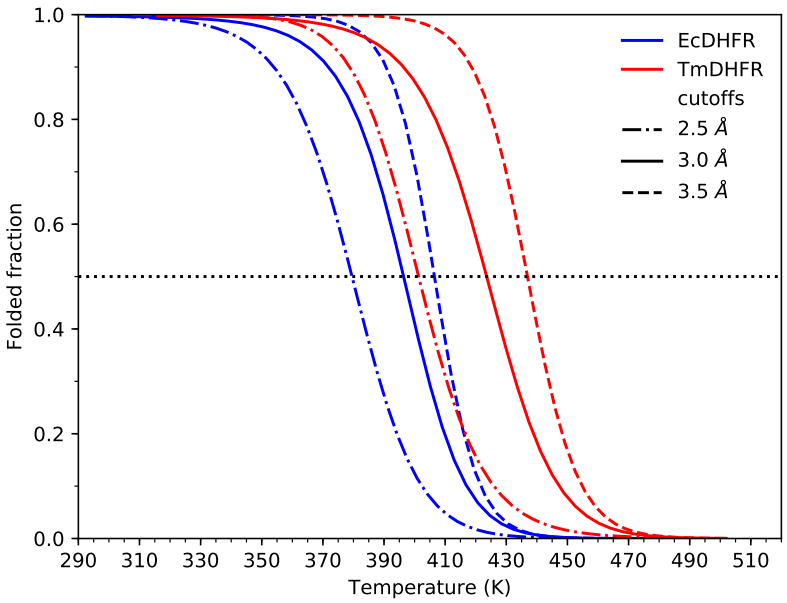

Figure S4: **Effect of RMSD cutoff on stability curves.** Stability curves obtained for EcDHFR (blue) and TmDHFR (red) using different RMSD cutoffs, namely 2.5 (dash-dotted), 3.0 (full line) and 3.5 Å(dashed).

from the folded fraction using Boltzmann statistics[11,12]:

$$\Delta G_u(T) = -\Delta C_p \left[T\left(ln\left(\frac{T}{T_m}\right) - 1 \right) + T_m \right] + \Delta H_u \left(1 - \frac{T}{T_m} \right) \tag{S5}$$

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. The fraction of folded protein $P(f)$ and $\Delta G(T)$ are in the following relationship:

$$P(f) = \frac{1}{1 + e^{-\frac{\Delta G_u}{k_B T}}} \tag{S6}$$

Finally, we have checked that the stability shift between EcDHFR and TmDHFR is observed both for the apo state (main text) and the MM complex (Figure S6). In the latter case, the presence of the ligands in the active site and the inclusion of distance harmonic restraints (Table

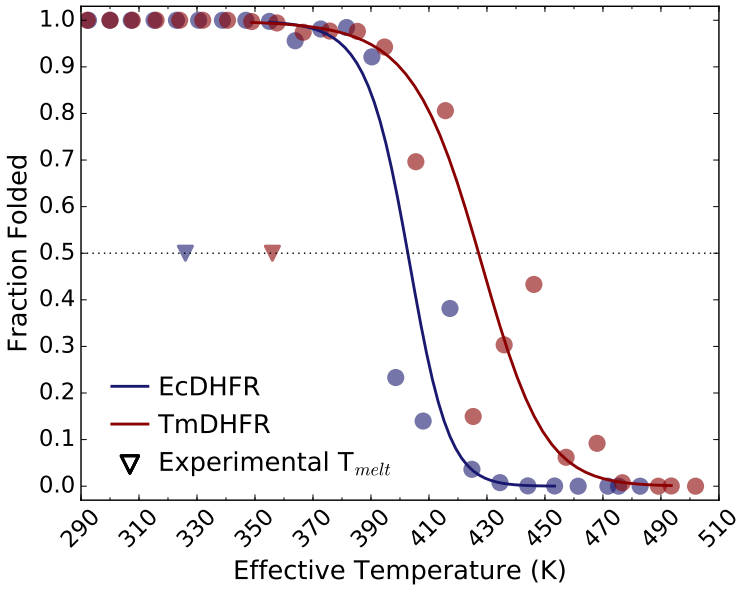

Figure S5: **Stability curves using native contacts as a metric.** Plot of the stability curves obtained by fitting data from REST2 trajectories and using different RMSD cut-off values to separate folded and unfolded configurations.

S1) slightly stabilize the protein. The melting temperatures are therefore shifted toward higher values, although the shift between EcDHFR and TmDHFR is conserved (26 K). Note that this effect has also been observed in the experiments, see e.g. ChemBioChem 2010, 11, 2010[13] on another DHFR homolog for which various ligands were observed to lead to a 3-15 K upshift of the protein stability.

Table S2: Thermodynamic data from our simulations.

DHFR	state	T_m K	ΔH_u kcal/mol	ΔC_p kcal/mol/K
Ec	apo	396	39.2	0.3
	MM	420	98.5	0.6
Tm dimer	apo	424	37.0	0.2
	MM	446	88.3	0.7
Tm mono	apo	403	45.3	0.2
	MM	422	45.7	-0.54

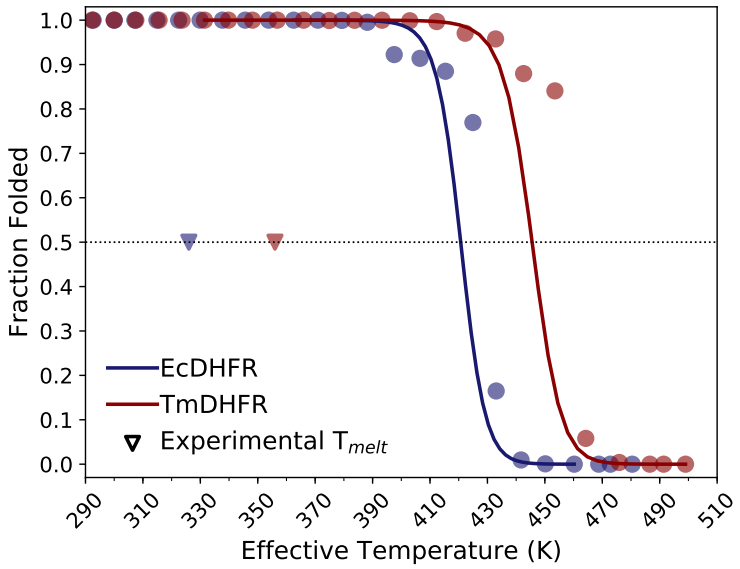

Figure S6: **Stability of MM complexes.** Fraction of folded protein as a function of the REST2 effective temperature for the MM complex of EcDHFR (blue circles) and TmDHFR (red circles) with the folded fraction defined by a smoothed RMSD cutoff. The melting temperature is defined as the temperature where half the systems are folded. Experimental melting temperatures for EcDHFR (326 K) and TmDHFR (356K) are indicated as blue and a red triangle, respectively. The calculated melting temperatures are 420 K and 446 K for EcDHFR and TmDHFR, respectively.

2.2. Melting of individual secondary structure elements

In order to identify the protein weak spots for thermal denaturation, we followed individually the unfolding of the protein structural elements as a function of temperature (Figure S7), by computing the RMSD with respect to the most populated protein cluster at 300 K. The respective melting temperatures have then been determined using a 2.5 Å cutoff, and are presented in Figure S8.

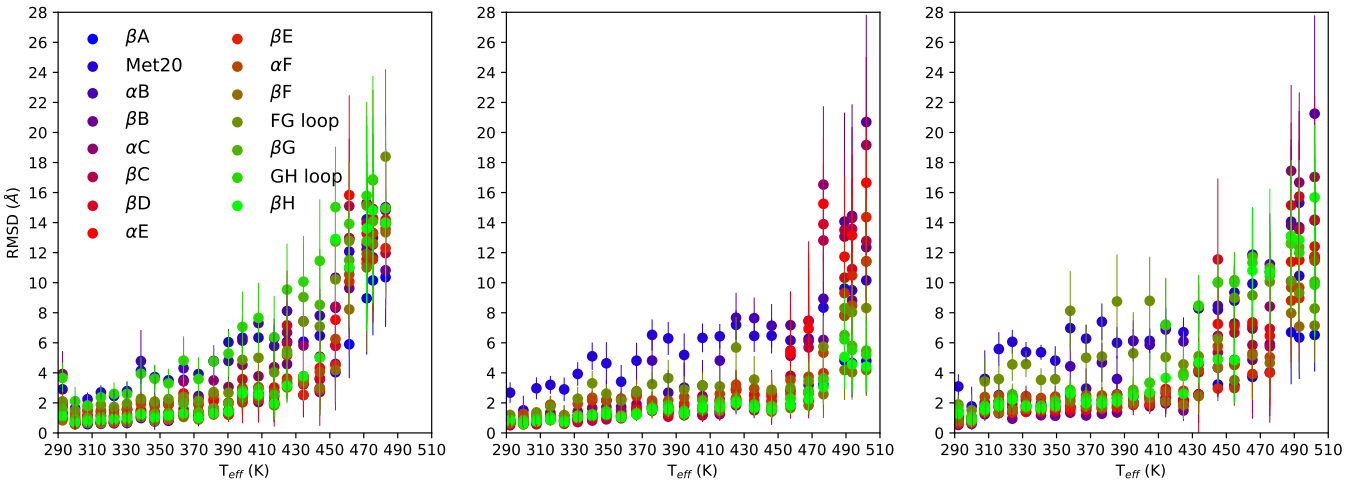

Figure S7: **Unfolding of secondary structure motifs.** RMSD as a function of the effective temperature for each secondary structure motif of EcDHFR (A) and TmDHFR (B) and monomeric TmDHFR (C) computed on the Cα with respect to the center of the most populated cluster at 300K.

2.3. Local and global flexibility

Local flexibility was assessed by computing the root means squared fluctuations (RMSF) of Cα on the last 250 ns of the REST2 simulations (Figure S9).

The global definition of flexibility was based on protein configurational entropy. The number of protein conformational states at each temperature is determined via a conformational clustering. This was performed on the last half of the apo state REST2 trajectories using the well-assessed leader algorithm[14]. The collective variable was set to be the protein Cα atoms

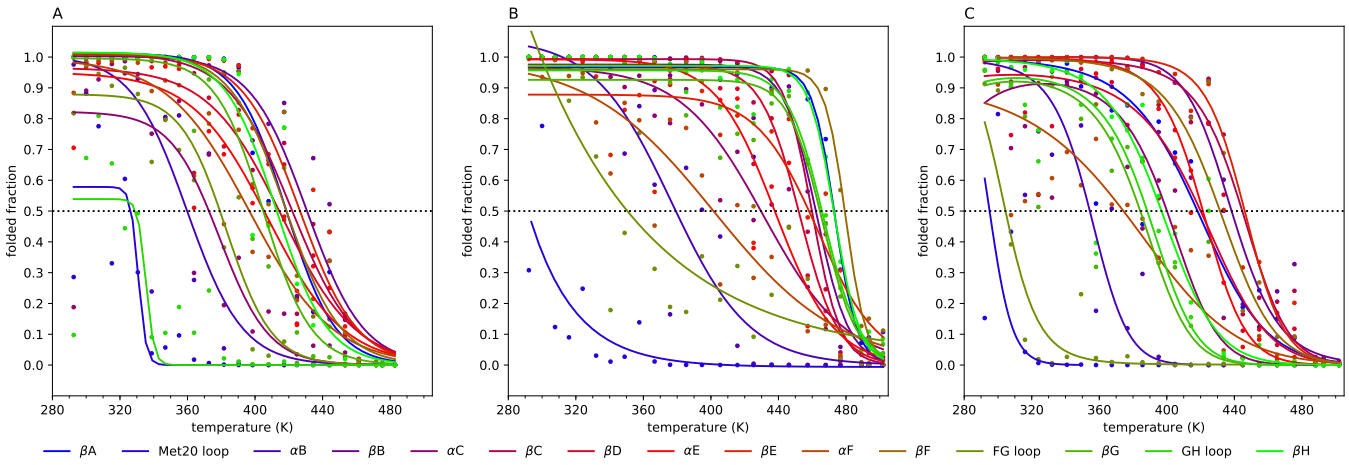

Figure S8: **Stability curves of secondary structure motifs.** Fraction of folded protein as a function of the REST2 effective temperature for each secondary structure motif of EcDHFR (A) and TmDHFR (B) and monomeric TmDHFR (C) with the folded fraction defined by a smoothed RMSD cutoff. The melting temperature is defined as the temperature where half the systems are folded.

RMSD between different frames of the trajectory after removing rigid body motions and using a cutoff of 2.5 Å to separate between different conformations.

2.4. Met20 loop conformations

We discriminated between Met20 loop open and close conformations by analyzing the distance between this loop and the αC helix (loop-helix distance). In particular, we looked at the distance between the EcDHFR Asn18 and His45 Cα atoms, which, after structural alignment, corresponds to the TmDHFR Val19 and Ile46 Cα atoms. If this distance is between 6 and 8 Å the Met20 loop is closed, while if it is longer than 15 Å the loop is open. Indeed, in the crystallographic structure of EcDHFR, where the loop is closed, the distance between Asn18 and His45 Cα atoms is 6.9 Å , while in the X-ray structure of TmDHFR, where the loop is open, the distance between Val19 and Ile46 Cα atoms is 18.4 Å . In addition, it has been experimentally observed that in the EcDHFR closed Met20 loop conformation Asn18 and His45 are in close contact with the possibility of having a H-bond between the Asn18 side chain

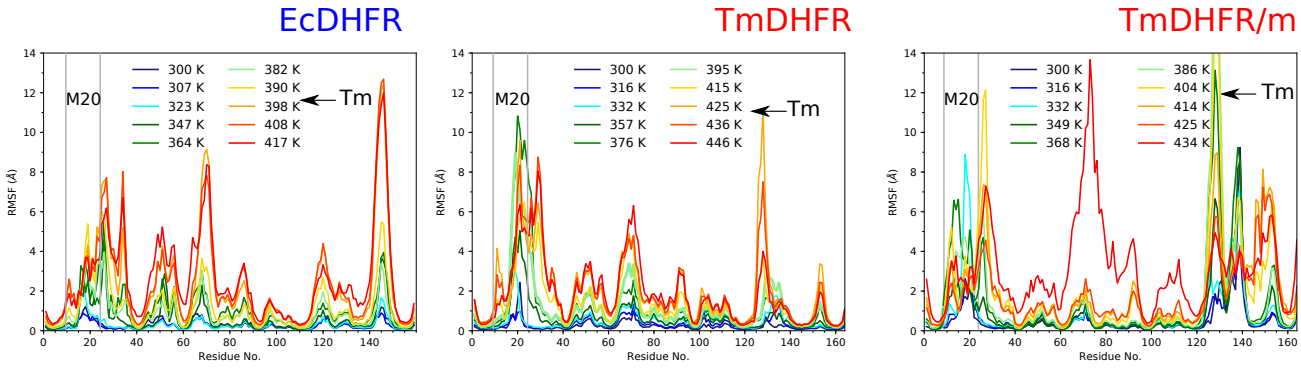

Figure S9: **Atomistic fluctuations.** Root mean square fluctuations computed on the C_α for the EcDHFR (left panel), TmDHFR dimer (mid panel), and TmDHFR monomer (right panel) at different effective temperatures.

nitrogen and the His45 backbone oxygen[1]. Conversely, in the Met20 loop open conformation it has been observed that the Val19 side chain creates hydrophobic interactions with Tyr125,[2] which belongs to the loop between the βF and βG strands and is more than 14 Å from His45 and Ile46 in EcDHFR and TmDHFR respectively.

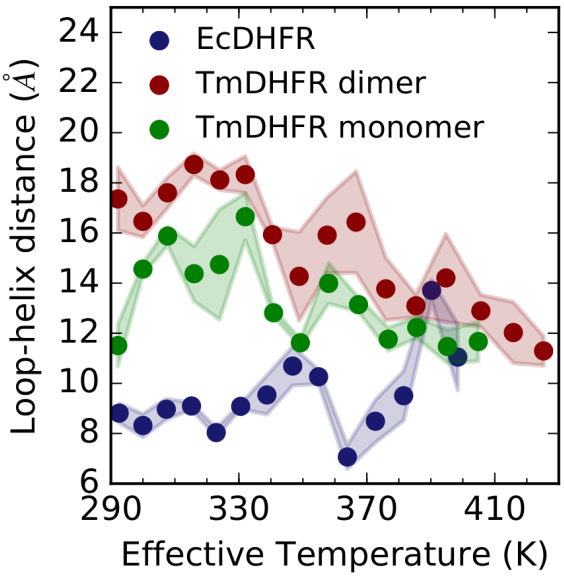

Figure S10: **Met20 loop conformations.** Met20 loop conformations as a function of the effective temperature for EcDHFR (blue), dimeric TmDHFR (red) and monomeric TmDHFR (green) in the apo state. The Met20 loop conformations are described by the Met20 loop - αC helix distance as above defined. Colored areas represent standard deviations of the mean.

2.5. CC distances

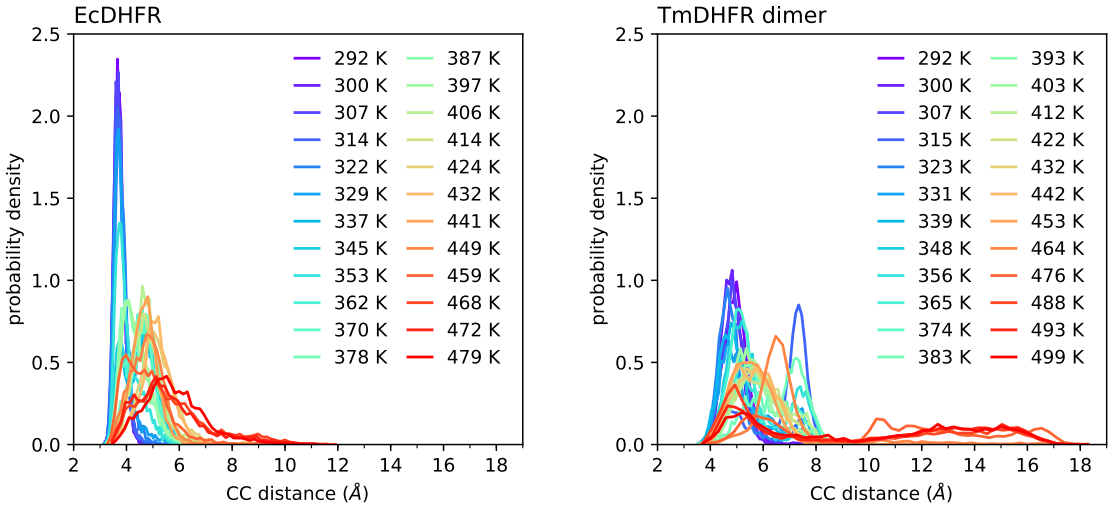

Figure S11: **CC distributions.** Probability distribution of the hydride donor-acceptor distance (CC) sampled in REST2 simulations for the MM state of EcDHFR (left) and TmDHFR dimer (right).

References

(1) Sawaya, M. R.; Kraut, J. Loop and Subdomain Movements in the Mechanism of Escherichia coli Dihydrofolate Reductase: Crystallographic Evidence. *Biochemistry* **1997**, *36*, 586–603.

(2) Dams, T.; Auerbach, G.; Bader, G.; Jacob, U.; Ploom, T.; Huber, R.; Jaenicke, R. The crystal structure of dihydrofolate reductase from Thermotoga maritima: molecular features of thermostability. *J. Mol. Biol.* **2000**, *297*, 659–672.

(3) Pang, J.; Pu, J.; Gao, J.; Truhlar, D. G.; Allemann, R. K. Hydride Transfer Reaction Catalyzed by Hyperthermophilic Dihydrofolate Reductase Is Dominated by Quantum Mechanical Tunneling and Is Promoted by Both Inter- and Intramonomeric Correlated Motions. *J. Am. Chem. Soc.* **2006**, *128*, 8015–8023.

(4) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(5) Liu, C. T.; Hanoian, P.; French, J. B.; Pringle, T. H.; Hammes-Schiffer, S.; Benkovic, S. J. Functional significance of evolving protein sequence in dihydrofolate reductase from bacteria to humans. *Proc. Natl. Acad. Sci.* **2013**, *110*, 10159–10164.

(6) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65*, 712–725.

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

(8) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(9) Stirnemann, G.; Sterpone, F. Recovering Protein Thermal Stability Using All-Atom Hamiltonian Replica-Exchange Simulations in Explicit Solvent. *J. Chem. Theory Comput.* **2015**, *11*, 5573–5577.

(10) Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431–9438.

(11) Hawley, S. A. Reversible Pressure-Temperature Denaturation of Chymotrypsinogen. *Biochemistry* **1971**, *10*, 2436–2442.

(12) Becktel, W. J.; Schellman, J. A. Protein stability curves. *Biopolymers* **1987**, *26*, 1859–1877.

(13) Evans, R. M.; Behiry, E. M.; Tey, L. H.; Guo, J.; Loveridge, E. J.; Allemann, R. K. Catalysis by dihydrofolate reductase from the psychropiezophile Moritella profunda. *ChemBioChem* **2010**, *11*, 2010–2017.

(14) Hartigan, J. A. *Clustering Algorithms*, 99th ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1975.