Calculation of Protein Folding Thermodynamics using Molecular Dynamics Simulations

Juan José Galano-Frutos[1, 2], Francho Nerín-Fonz[1, §], Javier Sancho[1, 2, 3]*

[1]Department of Biochemistry, Molecular and Cell Biology. Faculty of Science, University of Zaragoza, 50009 Zaragoza, Spain

[2]Biocomputation and Complex Systems Physics Institute (BIFI). Joint Units BIFI-IQFR (CSIC) and GBs-CSIC, University of Zaragoza, 50018 Zaragoza, Spain

[3]Aragon Health Research Institute (IIS Aragón), 50009 Zaragoza, Spain

[§]Current Address: Biomedical Research Foundation, Academy of Athens, 11527 Athens, Greece

***Corresponding author:** Javier Sancho, **e-mail:** jsancho@unizar.es

Abstract

Despite impressive advances by AlphaFold2 in the field of computational biology, the protein folding problem remains an enigma to be solved. The continuous development of algorithms and methods to explore longer simulation timescales of biological systems, as well as the enhanced accuracy of potential functions (force fields and solvent models) have not yet led to significant progress in the calculation of the thermodynamics quantities associated to protein folding from first principles. Progress in this direction can help boost related fields such as protein engineering, drug design, or genetic interpretation, but the task seems not to have been addressed by the scientific community. Following an initial explorative study, we extend here the application of a Molecular Dynamics-based approach −with the most accurate force field/water model combination previously found (Charmm22-CMAP/Tip3p)− to computing the folding energetics of a set of two-state and three-state proteins that do or do not carry a bound cofactor. The proteins successfully computed are representative of the main protein structural classes, their sequences range from 84 to 169 residues, and their isoelectric points from 4.0 to 8.9. The devised approach enables accurate calculation of two essential magnitudes governing the stability of proteins −the changes in enthalpy and in heat capacity associated to protein unfolding−, which are used to obtain accurate values of the change in Gibbs free-energy, also known as the protein conformational stability. The method proves to be also suitable to obtain changes in stability due to changes in solution pH, or stability differences between a wild-type protein and a variant. The approach addresses the calculation by difference, a shortcut that avoids having to simulate the protein folding time, which is very often unfeasible computationally.

Introduction

Proteins are very versatile biological molecules[1] and thermodynamics can greatly help to understand how they fold and perform useful tasks[2,3]. Molecular Dynamics (MD) simulation has become a powerful tool to study protein folding and other related processes[4–12]. However, despite great efforts in developing algorithms and methods to enable longer and better sampled simulations and in improving the accuracy of force fields and water models, significant challenges remain[13]. On one hand, simulating the protein folding time (from microseconds up to tens of seconds or beyond) in explicit solvent remains inaccessible, except for small fast-folding proteins[5–7,9]. On the other, work on improving the accuracy of MD force fields seems to have focused on reproducing structural, dynamic and mechanistic aspects of protein behavior[14–18], and has paid less attention to try to reproduce protein potential energy. One reason for this is the difficulty in obtaining accurate structural models of unfolded ensembles, which has prevented comprehensive studies of this side of the problem, making fine-tuning of the force field parameters difficult. The experimental limitations inherent in quantifying individual atomic interactions, and the massive cancellation of interactions that takes place in a protein folding reaction[19] adds to the complexity of the goal[2]. All of the above has perhaps frustrated the interest of scientists in the use of MD simulations to quantitatively study protein thermodynamics, hindering progress in many applied fields, such as protein design[20], drug design[21], genetic interpretation[22], protein engineering[23] or cell engineering[24].

Recently, we addressed this issue by carrying out accurate, quantitative calculations of conformational stability on two two-state model proteins (barnase and nuclease) through an all-atom MD-based approach[25]. The approach circumvents the simulation of the whole folding/unfolding time and it is based on separately simulating the two main conformations: the folded and unfolded states. The folded state is modeled starting from an experimentally determined structure that is conveniently solvated and sampled conformationally. The unfolded state is modeled and sampled from an ensemble of completely unfolded conformations generated by the ProtSA server[26] that are similarly solvated. From the simulations, the enthalpy change of unfolding (ΔH_{unf}) is calculated by difference (the unfolded minus the folded state enthalpy averages), while the heat capacity change at constant pressure (ΔCp_{unf}) is obtained from the temperature dependence of the calculated enthalpy change. As a final step, the calculated thermodynamic quantities (ΔH_{unf} and ΔCp_{unf}) are combined with the experimentally determined mid denaturation temperature (T_m) to calculate the conformational stability of the protein (ΔG_{unf}) as a function of temperature by means of the Gibbs-Helmholtz equation[27]. One goal of the approach was testing the ability of classical force fields, e.g. Charmm22-CMAP[18] and AmberSB99-ILDN[17] (or the more recently released Amber force field, AmberSB99-disp[16], tuned to be used both with folded and disordered proteins) to yield accurate folding energetics by difference using systems solvated with explicit water. Therefore, the indicated force fields were combined with seven explicit water models, Tip3p[28], Tip4p[28], Tip4p-d[29], Tip4-d-mod[16], Tip5p[30], Spc[31], and Spc/E[32].

Results obtained by setting short MD simulations (2-ns productive trajectories per replica) and the combinations of either Charmm22-CMAP or AmberSB99-ILDN with Tip3p allowed, for the two proteins indicated, to finely capture the energy balance between the numerous interactions established between protein and solvent atoms in both the native state and the unfolded ensemble[33,34].

In this work, we generalize the described methodology using the most accurate combination of force field and water model found[25] and a larger conformational sampling, and show the good accuracy obtained on a set of two-state, three-state, apo, holo, wild-type (WT) or mutated proteins relative to their corresponding experimentally determined thermodynamics. In addition to barnase[35,36] and nuclease[37,38] energetics (here calculated anew with higher precision[25]), we present the calculation for additional two-state proteins: Barley chymotrypsin inhibitor 2 (CI2, truncated variant[39,40]) and Phage T4 lysozyme[41] (the WT and pseudo-WT variants), a three-state protein (apoflavodoxin from *Anabaena* PCC 7119[42,43–45]) for which the energetics involved in the two unfolding transitions ($\Delta G_{unf(F\text{-}to\text{-}I)}$, and $\Delta G_{unf(I\text{-}to\text{-}U)}$) are obtained, and a holoprotein (flavodoxin from *Anabaena* PCC 7119) containing the FMN cofactor. Furthermore, we evaluate the capability and limits of the approach to capture small stability changes or small differences between similar systems, e.g. those associated to mutation ($\Delta\Delta H_{mut\text{-}nat}$ and $\Delta\Delta G_{mut\text{-}nat}$), changes in pH ($\Delta\Delta H_{pH1\text{-}pH2}$ and $\Delta\Delta G_{pH1\text{-}pH2}$) or individual steps within a multi-state unfolding ($\Delta H_{unf(F\text{-}to\text{-}I)}$, $\Delta Cp_{unf(F\text{-}to\text{-}I)}$, $\Delta G_{unf(F\text{-}to\text{-}I)}$ or $\Delta H_{unf(I\text{-}to\text{-}U)}$, $\Delta Cp_{unf(I\text{-}to\text{-}U)}$ and $\Delta G_{unf(I\text{-}to\text{-}U)}$). The method requires a reliable structural model for the folded conformation, which sometime is not available. Hopefully, advances in high resolution protein models[46,47] will allow the application of the protocol to the entire proteome.

Target Proteins and Case Studies

<u>Barnase from *B. amyloliquefaciens* and nuclease from *S. aureus*.</u> The 110- and 149-residue proteins barnase[14,48–57] and nuclease[58–61] (C-terminal fragment), respectively, are well characterized two-state proteins, as summarized in Galano-Frutos et.al.[25], where their folding energetics were first calculated. Here, the two-state unfolding energetics of WT barnase and nuclease are again determined using the present computational approach and the reported effect of pH on nuclease stability is also analyzed (see **Table 1**).

<u>CI2 from barley seeds.</u> CI2[39,40] is a small, 84-residue, globular serine proteinase inhibitory protein extensively studied and reported to fold in a two-state manner under thermal conditions[62–64]. Its 19-residue N-terminal tail is completely unstructured[65,66]. We focus here on a truncated form of CI2 lacking the unstructured N-terminal tail because the structure of the full length protein is not available and because it has been shown that the tail does not contribute to the protein stability[63,64]. Thermodynamic data for the WT CI2 truncated form and for a broad set of point mutants analyzed under different solvent conditions (varying in pH and ionic strength) are available[62–64] (**Table 1**). The CI2 variant Ile76Ala which, relative to WT in identical solvent, shows a significantly lower unfolding

enthalpy change and a large destabilization[64] (**Table 1**) has been selected to evaluate the feasibility of the approach to calculate the effect of single amino acid replacements on protein stability. On the other hand, the significantly different thermodynamics quantities reported for WT CI2[64] under pH 3.0 and 6.3 (see **Table 1**), invited us to also evaluate the sensibility of the method to solvent effects.

Phage T4 lysozyme. T4 endolysin (lysozyme)[41] is a two-domain, 164-residue globular protein that has also been subject of extensive study, and widely used to investigate the role of hydrophobic interactions in protein structural stabilization[67–69]. The thermal stability of the WT protein and many variants thereof have been measured[70–73]. Over 500 X-ray structures of T4 lysozyme –including those of an engineered pseudo-lysozyme (see below) and many variants thereof– have been obtained under a variety of experimental conditions (buffer, pH, ionic strength)[74]. WT lysozyme carries two cysteine residues at positions 54 and 97. To ease experimental work on the protein, a Cys54Thr/Cys97Ala variant (termed pseudo-WT lysozyme) has often been studied. WT and pseudo-WT lysozymes[75] slightly differ in structure and thermodynamics[70–73,76] (**Table 1**). For the sake of testing the method, the energetics of these two lysozyme variants are calculated. Besides, the energetics of the non-pseudo lysozyme variant, Ile3Glu[73], is addressed as a further attempt to capture the effect of single amino acid replacements, and the pseudo-WT lysozyme[76] is simulated in different solvent conditions (different pH values) to assess, as with nuclease and CI2, whether the method can capture pH-related effects on protein stability (**Table 1**).

Anabaena flavodoxin (Fld). Fld[42,77] is a 169-residue protein that carries electrons from photosystem I to ferredoxin-NADP$^+$ reductase[78,79] in the cyanobacteria *Anabaena* PCC 7119. Fld capability to transfer electrons is conferred by the presence of a molecule of non-covalently bound FMN cofactor. Reversible removal of the cofactor from the holoprotein (holoFld) leads to the apo form (apoFld). Fld has been widely studied to investigate protein/cofactor interactions[80,81], as well as non-native protein conformations[44,82–85]. While apoFld thermal unfolding equilibrium is three-state[43–45], binding of FMN greatly stabilizes the complex so that holoFld unfolds following a two-state mechanism[82,86]. A detailed picture of Fld folding and binding thermodynamics is available[43–45,81,82,84–87]. The reasonably high enthalpy and heat capacity changes (**Table 1**) of the two apoFld unfolding transitions, folded to intermediate (F-to-I) and intermediate to unfolded (I-to-U), together with the availability of a representative structure of its intermediate conformation[84] made us select this protein to test the simulation approach for the calculation of unfolding energetics in three-state proteins.

Results

Energetics of two-state proteins: barnase, nuclease, CI2 and lysozyme unfolding

The thermal unfolding of barnase, nuclease, CI2 and lysozyme at equilibrium has been described to follow a two-state mechanism. Accordingly, we have calculated their unfolding energetics using the general workflow (see **Methods** and **Figure 1**), where the numbers of simulated replicas of the folded state and of simulated structures in the unfolded ensemble has been increased relative to the initial formulation of the method[25].

Barnase is simulated (**Figure 2a**) at pH ~4.1 (see **Table 1** and **SI Table 2**) in solvating conditions similar to those used in the experimental measurements reported and also to those previously used to model the protein[25]. In the previous modelling, a reasonable agreement was found between experimental and calculated data. Here, the calculated values of $\Delta H_{unf}(T_m)$, ΔCp_{unf} and ΔG_{unf} at 298.15 K (ΔG^0_{unf}) obtained with the extended sampling (110.4 ± 3.1 kcal/mol, 1.0 ± 0.1 kcal/mol·K and 7.5 ± 1.2 kcal/mol, respectively, **Table 2**) agree very well with the averaged experimental energetics (118.7 ± 4.9 kcal/mol, 1.4 ± 0.1 kcal/mol·K and 7.8 ± 0.4 kcal/mol, **Table 1**). Due to this fine agreement, the experimental and calculated temperature dependencies of ΔG_{unf} (stability curves), excess heat capacity (thermograms), and state fractions (**Figure 2b-d)** nearly coincide. The agreement between calculated and experimental magnitudes is better than that obtained in the previous calculation[25] (92.3 ± 5.7 kcal/mol, 0.9 ± 0.1 kcal/mol·K and 6.5 ± 0.8 kcal/mol, respectively) that was performed using the same conditions and workflow but with a smaller sampling.

Nuclease unfolding thermodynamic data are available over a range of pH (from 3 to 8.5) and solvating conditions[59,60]. WT nuclease is simulated here (**Extended Data Figure 1a**) at three pH values: 7.0, 5.0 and 4.1 (see solvating conditions and protonation states in **Table 1** and **SI Table 2**). At pH 7.0, the calculated $\Delta H_{unf}(T_m)$ (75.1 ± 4.5 kcal/mol), ΔCp_{unf} (1.7 ± 0.3 kcal/mol·K) and ΔG_{unf} (4.8 ± 1.7 kcal/mol) at 293.15 K (the temperature at which experimental data are mostly reported for this protein, **Table 2**) match very well the averaged experimental values (82.1 ± 4.7 kcal/mol, 2.3 ± 0.3 kcal/mol·K and 4.3 ± 0.3 kcal/mol, respectively, **Table 1**). This excellent agreement is reflected, as seen above for barnase, in the fine correspondence between the experimental and calculated Gibbs free-energy difference, excess heat capacity, and molar fractions temperature dependencies (**Extended Data Figure 1b-d**). The second solvating condition simulated for nuclease reproduces a protonation scheme previously used[25], corresponding to pH 5.0. Under this condition, our simulations yield calculated energetics ($\Delta H_{unf}(T_m)$ = 71.0 ± 4.5 kcal/mol, ΔCp_{unf} = 1.5 ± 0.4 kcal/mol·K and $\Delta G^{293.15K}_{unf}$ = 4.4 ± 2.8 kcal/mol, **Table 2**) that matches fairly well the mean experimental values (73.1 ± 0.1 kcal/mol, 2.3 ± 0.1 kcal/mol·K and 3.5 ± 0.1 kcal/mol, respectively, **Table 1** and **Extended Data Figure 1e-f**). The application here of a more exhaustive sampling reproduces similarly well the experimental results of nuclease compared to the implementation carried out with a smaller

sampling[25]: calculated $\Delta H_{unf}(T_m)$ = 76.0 ± 8.1 kcal/mol, ΔCp_{unf} = 1.8 ± 0.1 kcal/mol·K and $\Delta G^{293.15K}_{unf}$ = 4.6 ± 1.4 kcal/mol, respectively[25]. Thus, nuclease stability is accurately calculated in the 5.0-7.0 pH range. However, at lower pH, pH 4.1, the method overestimates $\Delta H_{unf}(T_m)$ and ΔCp_{unf}, which leads to less accurate calculated stability (4.8 ± 2.2 kcal/mol, **Table 2** versus a mean experimental value of 2.9 ± 0.3 kcal/mol, **Table 1** and **Extended Data Figure 1g-h**).

WT CI2 is simulated (**Extended Data Figure 2a**) at two pH values for which reliable experimental data are available (**Table 1** and **SI Table 2**). At pH 3.0, the calculated $\Delta H_{unf}(T_m)$ and ΔCp_{unf} values (46.1 ± 1.9 kcal/mol and 0.4 ± 0.03 kcal/mol·K, respectively) are a bit lower than the corresponding experimental ones (61.0 ± 2.3 kcal/mol and 0.72 kcal/mol·K). Notwithstanding, the calculated ΔG^0_{unf} at this pH (4.3 ± 0.4 kcal/mol) virtually agrees within error with the experimental stability reported at 298.15 K (5.4 ± 0.7 kcal/mol). At pH 6.3, CI2 is more stable. At this pH, the experimental $\Delta H_{unf}(T_m)$ and ΔCp_{unf} values (78.4 ± 0.7 kcal/mol and 0.8 ± 0.1 kcal/mol·K, respectively) lead to a higher conformational stability (ΔG^0_{unf} = 7.2 ± 0.4 kcal/mol). The increases in $\Delta H_{unf}(T_m)$ and ΔCp_{unf} values at pH 6.3 relative to pH 3.0 are captured by our simulations (calculated values at pH 6.3: $\Delta H_{unf}(T_m)$ = 57.1 ± 0.5 kcal/mol and ΔCp_{unf} = 0.5 ± 0.07 kcal/mol·K, respectively), as well as the increase in conformational stability (calculated ΔG^0_{unf} = 6.9 ± 0.6 kcal/mol, in agreement with the experimental value of 7.2 ± 0.4 kcal/mol). To assess the capability of the simulation approach to detect changes in stability associated to point mutations, we compute the energetics of the CI2 variant Ile76Ala at pH 3.0 and compared it to that of WT CI2 at the same pH. Substitution of the bulky WT isoleucine residue by alanine creates a cavity that severely destabilizes the folded structure of the mutant. The reduced stability of Ile76Ala CI2 compared to WT is evidenced in its experimental unfolding energetics ($\Delta H_{unf}(T_m)$ = 30.2 kcal/mol, ΔCp_{unf} = 0.7 kcal/mol·K and ΔG^0_{unf} = 1.1 ± 0.3 kcal/mol, **Table 1**), which is accurately obtained from our simulations: calculated $\Delta H_{unf}(T_m)$ = 27.7 ± 1.7 kcal/mol, ΔCp_{unf} = 0.5 ± 0.01 kcal/mol·K and ΔG^0_{unf} = 1.0 ± 0.2 kcal/mol (**Table 2**). Thus, the simulation workflow allows to capture the experimental observations that: 1) WT CI2 is stabilized by raising the pH from 3.0 to 6.3 (experimental $\Delta\Delta G_{unf(pH3\rightarrow pH6.3)}$ = +1.8 ± 1.1 kcal/mol; calculated value = +2.5 ± 1.2 kcal/mol) and 2) WT CI2 is severely destabilized by replacing Ile76 by Ala (experimental $\Delta\Delta G^0_{unf(WT\rightarrow I76A)}$ = −4.3 ± 1.0 kcal/mol; calculated value = −3.3 ± 0.6 kcal/mol). Experimental and calculated stability curves, thermograms, and state fractions of WT (pH 3.0), WT (pH 6.3) and Ile76Ala CI2 mutant (pH 3.0) are compared in **Extended Data Figure 2b-h**. A good agreement between calculated and experimental data can be observed that is particularly remarkable for the Ile76Ala CI2 variant (**Extended Data Figure 2g-h**).

Lysozyme is simulated (**Extended Data Figure 3a**) at pH 2.4 (WT and Ile3Glu mutant) and at pH 3.0 and 3.7 (pseudo-WT, **Extended Data Figure 4a**). The experimental ΔCp_{unf} is calculated accurately for the pseudo-WT but underestimated for the WT. For the four simulated lysozyme variants or pH condition (**Table 1**), the calculated ΔH_{unf} values (**Table 2**) clearly overestimate the

corresponding experimental ones (**Table 1**). As a consequence, the stabilities calculated at 298.15 K also overestimate the experimental values, and the stability versus temperature dependencies (**Extended Data Figures 3b-f** and **4b-f**) do not match the calculated ones. Thus, the actual lysozyme stabilities are not correctly calculated by this MD-based approach (possible reasons are indicated in the **Discussion** section). Still, both the lower stability of the Ile3Glu mutant relative to WT at pH 2.4 ($\Delta\Delta G^0_{unf(WT\rightarrow Ile3Glu)} = -1.0$ kcal/mol), and the higher stability of pseudo-WT at pH 3.7 compared to pH 3.0 ($\Delta\Delta G_{unf(pH3.0\rightarrow pH3.7)} = +3.3$ kcal/mol) are qualitatively captured (−2.8 and +6.5 kcal/mol, respectively).

Energetics of a three-state protein: apoFld unfolding

ApoFld thermal unfolding is three-state, with a well-defined intermediate accumulating at equilibrium with the folded and unfolded conformations. The unfolding enthalpy changes of the sequential partial unfolding equilibria (F-to-I and I-to-U) are separately calculated using the general workflow (**Figure 1**) from the structures or ensembles representing the states involved in each transition (**Figure 3a**). The calculated enthalpy changes of the two unfolding transitions, $\Delta H_{unf(F\text{-}to\text{-}I)} = 35.6 \pm 6.0$ and $\Delta H_{unf(I\text{-}to\text{-}U)} = 48.1 \pm 4.1$ kcal/mol (**Table 2**), are in excellent agreement with the corresponding experimental enthalpies of 32.0 ± 1.1 and 55.6 ± 2.0 kcal/mol (**Table 1**). The heat capacity changes calculated for each partial unfolding step, $\Delta Cp_{unf(F\text{-}to\text{-}I)} = 1.5 \pm 0.1$ and $\Delta Cp_{unf(I\text{-}to\text{-}U)} = 1.0 \pm 0.0$ kcal/mol·K, respectively (2.5 ± 0.1 kcal/mol·K for the global transition, **Table 2**) are also in fair agreement with the experimental values of 1.35 ± 0.3 and 1.55 ± 0.3 kcal/K·mol, respectively (2.9 ± 0.6 kcal/mol·K for the global transition, **Table 1**). From these calculated data and the experimental T_ms of the two transitions (**Table 1**), the Gibbs free-energy changes of the individual and global unfolding transitions are calculated at 298.15 K, with the Gibbs-Helmholtz equation[27]. A fine correspondence between the calculated values, $\Delta G^0_{unf(F\text{-}to\text{-}I)} = 1.3 \pm 1.7$ kcal/mol, $\Delta G^0_{unf(I\text{-}to\text{-}U)} = 3.0 \pm 0.9$ kcal/mol and $\Delta G^0_{unf(F\text{-}to\text{-}U)} = 4.3 \pm 2.6$ (**Table 2**), and the corresponding experimental ones, 1.1 ± 1.4, 2.9 ± 1.3 and 4.0 ± 2.7 kcal/mol, is observed. This outstanding correspondence between the calculated and the experimentally determined apoFld thermal unfolding thermodynamics can be also visually assessed from the stability curves[27], thermograms, and folded/intermediate/unfolded state-fractions (**Figure 3b-d**).

For the sake of comparison, an otherwise identical calculation of apoFld thermal unfolding thermodynamics has been carried out using the Amber99SB-ILDN force field instead of Charmm22-CMAP. Although accurate heat capacity changes (ΔCp_{unf}) are calculated for the two equilibria with Amber99SB-ILDN (1.4 ± 0.1 and 1.1 ± 0.1 kcal/K·mol, respectively, **Table 2**), the calculated enthalpy changes (**Table 2**) do not agree well with the experimental values (**Table 1**), which results in less accurate calculations of the sequential Gibbs free-energy changes (**Table 2**), compared to those obtained with Charmm22-CMAP. A better agreement of Charmm22-CMAP thermodynamics calculations with experimental values was already reported for barnase and nuclease[25].

Energetics of a holoprotein: holoFld unfolding

The calculation of the thermal unfolding energetics of a holoprotein (a protein carrying a non-covalently bound cofactor) is performed as described in **Methods** and depicted in **Figure 4a**. The results obtained for the three different FMN parameterizations tested are presented in **Table 2** and **Figure 4b**. The ΔH_{unf} calculated for holoFld with any of the three setups (ranging from 103.0 ± 6.5 to 114.2 ± 7.8 kcal/mol) are in fair agreement with the experimental value reported by Lamazares and coworkers[82] from DSC measurements (101.9 ± 0.6 kcal/mol, **Table 1**).

The holoFld ΔCp_{unf} has not been reported, but an estimation can be done by adding the reported value for FMN dissociation ($\Delta Cp_{diss} = -\Delta Cp_{bind}$ = 0.6 ± 0.0 kcal/mol·K) to the apoFld ΔCp_{unf} (2.9 ± 0.6 kcal/mol·K, **Table 1**). Thus, holoFld ΔCp_{unf} is estimated to be 3.5 ± 0.6 kcal/mol·K. Our calculated holoFld ΔCp_{unf} values are reported in **Table 2** for the three FMN parameterizations above referred and depicted as the slope of fitting lines in **Figure 4b**. The values obtained with either FMN Par.-1 or FMN Par.-2 (3.0 ± 0.2 and 2.9 ± 0.6 kcal/mol·K, respectively) agree within experimental error, and that obtained with FMN Par.-3 (2.6 ± 0.1 kcal/mol·K) while lower, is still above the value previously calculated for apoFld (2.5 ± 0.1 kcal/mol·K, **Table 2**), in agreement with the observed positive value of ΔCp for FMN dissociation.

The stability of holoFld at 298.15 K is obtained through **SI Eq. 5** (see the derivation in **SI Methods**). To the apoprotein Gibbs free-energy, **SI Eq. 5** applies a correction due to the ligand concentration, incorporating the van't Hoff approximation[88] to account for the temperature dependence of the binding constant, $K_b(T)$. Thus, **Eq. S5** is not based on the thermodynamics derived from the holoFld simulations (ΔH^{holo}_{unf}, ΔCp^{holo}_{unf}) but on those of the apoprotein (ΔH^{apo}_{unf}, ΔCp^{apo}_{un}) plus the cofactor energetics. Using **SI Eq. 5,** the $\Delta G^{298.15K}_{unf}$ value calculated (17.3 ± 2.6 kcal/mol, **Table 2**) is in close agreement with the experimental value (17.1 ± 2.7 kcal/mol, **Table 1**) similarly obtained with **SI Eq. 5** using experimental ΔH^{apo}_{unf} and ΔCp^{apo}_{unf} data. Importantly, the calculated $\Delta G^{298.15K}_{unf}$ also matches, within error, the experimental stability of holoFld directly obtained from thermal unfolding curves (19.0 ± 0.9 kcal/mol)[86].

The van't Hoff approximation to model the temperature dependence of binding constant[88] –therefore the correction to stability due to the ligand concentration– should work fine as long as the conformation of the protein binding site does not change significantly. However, as this will not be the case at temperatures where the apoprotein begins to unfold, we consider the ΔG_{unf} temperature profile of the holoprotein (**Figure 4c**) not to be reliable beyond the first mid denaturation temperature (T_m1) of the apoprotein (316.2 K in the case of apoFld).

Discussion

Overall accuracy of calculated energetics

The devised MD-based workflow allows the calculation of ΔH_{unf}, ΔCp_{unf}, and ΔG_{unf}, the main thermodynamic magnitudes governing the stability of proteins. The overall accuracy of the method can be assessed from lineal plots of calculated versus experimentally determined values of each of those magnitudes.

The primary outcome is the unfolding enthalpy change of the proteins investigated. With the exception of lysozyme (simulated in four conditions) and nuclease (when simulated at low pH, pH 4.1), which are clear outliers, the linear plot (**Figure 5a**) can be fitted to a straight line with ordinate close to zero, slope close to unity (0.95) and a correlation of R^2 = 0.93. The fitting includes the data from ten simulated systems (barnase, nuclease at two pH values, two partial unfolding equilibria as well as the whole transition of three-state apoFld, CI2 at two pH values plus one mutant, and holoFld) spanning a range of ΔH_{unf} values from 30 to 120 kcal/mol. It is thus clear that ΔH_{unf} can be accurately calculated using the approach.

The second outcome is the unfolding heat capacity change (**Figure 5b**), which is also captured for those 10 protein systems well fitted in **Figure 5a**. The four lysozyme systems simulated (WT at two pH values, pseudo-WT and a variant of pseudo-WT), as well as nuclease at pH 4.1, fit worse than the other 10 systems. Albeit their calculated ΔCp_{unf} values do not differ too much from their experimental ones, they are also treated as outliers, for consistency. The linear fit with data from the 10 simulated systems also yields a straight line with ordinate close to zero, slope close to unity and a correlation of R^2 = 0.94, indicating that the change in heat capacity of unfolding can be also calculated in an accurate manner using the approach. The range of ΔCp_{unf} values spanned goes from 0,6 to 3.5 kcal/mol·K.

The third outcome is the change in Gibbs free-energy upon unfolding (ΔG_{unf}), i.e. the conformational stability of the protein. To derive it, our workflow combines the calculated enthalpy and heat capacity changes with experimental values of melting temperatures, using the Gibbs-Helmholtz equation (**Eq. 1**) for apoproteins, or an analogous equation for holoproteins (**SI Eq. 5**). As expected, in the linear plot of calculated versus experimentally determined stabilities **(Figure 5c)** lysozyme yields outliers as the high enthalpy changes calculated for this protein systems are carried over in the calculation of the stability. Although, nuclease at pH 4.1 is not a clear outlier in the stability representation, it is kept as such for consistency. The fitting of the calculated and experimental values for the other 10 systems simulated gives rise again to a straight line with close to zero intercept, close to unity slope and a high correlation of R^2 = 0.99. It seems thus that conformational stability is also accurately calculated from first principles using the described

simulation workflow. The range of Gibbs free-energies spanned in the plot goes from 1 to 17 kcal/mol.

Applicability and limitations of the approach

The present MD-based workflow accurately calculates the protein changes in enthalpy, heat capacity and Gibbs free-energy upon unfolding and can also be used to compare the stability of a protein under different pH values, or to compare the stability of a wild-type protein with mutants thereof. According to our literature search, no similar approach for the calculation of protein folding energetics has been described, which precludes direct comparison with other methods. The systems successfully calculated here contain representatives of the main protein classes (mainly-alpha, mainly-beta, and alpha-beta)[89], with sequences ranging from 84 to 169 residues, and isoelectric points from 4.0 to 8.9. They include proteins that undergo two- or three-state thermal unfolding, as well as proteins that do or do not carry a tightly bound cofactor. As a whole, these proteins offer a fair representation of natively folded proteins for which the unfolding process leads to fully unfolded conformations. As detailed thermodynamic studies on much larger proteins scarce, the approach has not been tested on large proteins. We foresee no reasons why the energetics of larger proteins cannot be calculated with similar accuracy using sufficient sampling, provided they exhibit fully unfolded conformations after heating. This is a requisite of our protocol, necessary to be able to build realistic models of the unfolded ensemble using ProtSA.[26]

However, for one of the proteins simulated, lysozyme, the calculations have consistently led to overestimated ΔH_{unf} values, which has translated in overestimated stability. In principle, the method could have failed due to insufficient quality of the models used to represent the lysozyme folded and unfolded conformations. This is unlikely as the folded structures have been solved in a highly experience lab[90] and they get good marks when subjected to quality control with the MolProbity server[91] (not shown). Besides, the model of the unfolded ensemble generated by ProtSA[26] would be wrong if the lysozyme unfolded state were compact, but we have found no reports pointing to that. A different reason for the inaccurate lysozyme calculation may relate to small inaccuracies in force field parameters. Although the same force field has been used in lysozyme and in the successful calculations, it should be noticed that force field parameters are globally optimized and individual optimal performance of each of them cannot be taken for granted. In this respect, of all the systems simulated here, lysozyme stands out as the one containing the highest net (positive) charge (**Table S2**), only paralleled by the high net (positive) charge of nuclease in the single simulation condition (pH 4.1) where inaccurate results have also been obtained for this protein. It is possible that the discrepancy between calculated and experimental lysozyme unfolding magnitudes is related to insufficient tuning of coulombic treatment by the Charmm22-CMAP force field[18], at least for lysine and arginine protonated side chains. In addition, some uncertainty in the protonation state of lysozyme

carboxyl groups at the acidic pH of the simulations could also contribute to inaccuracy. Whatever the reason, the poorer performance of the method on lysozyme suggests it should be used with caution when highly positively charged proteins are simulated at acidic pH values. On the other hand, as proteins are rarely studied experimentally in basic pH conditions, we have not tested the performance of the method in alkaline solutions.

The sampling requirements of the protocol here presented are affordable with ordinary CPU- or GPU-clusters. The biggest, more challenging system here computed, 168-residue three-state apoFld (2.2 µs simulation: ~3.0 ns per 40 folded, 100 intermediate and 100 unfolded replicas, each at three temperatures), takes 2-3 weeks of essentially unattended calculation on 32 cores (32 OpenMP threads/1 MPI task) of an Intel Xeon E5-2680v3-based cluster (170 CPUs arranged in 85 nodes, https://cesar.unizar.es/hpc/).

Although the described approach is based on a specific force field and water model, it suggests that current force fields are already close to capturing the complexity of protein folding energetics. We hope that our results will encourage further improvement of force fields and water models. Toward that goal, the described methodology constitutes an effective and efficient way to assess their ability to replicate the changes in energy that govern protein equilibria.

Methods

General MD-based workflow for calculation of unfolding energetics (ΔH_{unf}, ΔCp_{unf} and ΔG_{unf}) in apoproteins

The presented workflow (**Figure 1**) is based on that described in Galano-Frutos et.al.[25]. The current version relies on a higher sampling of the folded and unfolded states. Briefly, X-ray crystal structures with the highest resolution and sequence coverage are retrieved from the RCSB Protein Data Bank (https://www.rcsb.org/[92,93], see PDB codes below) and taken as the starting structures for modeling the native (folded) state. When needed, the initial crystal structure is used to model the amino acid replacement leading to the mutant simulated (e.g. the CI2 Ile76Ala and lysozyme Ile3Glu variants).

Forty replicas of the folded structure are simulated, each consisting of a single protein molecule solvated with water molecules in a specified simulation box additionally containing, when required, ions (Na^+ and/or Cl^-). On the other hand, large unfolded ensembles (~2000 structures) are generated from the protein sequence using the ProtSA server[26] and a random sample of 100 unfolded structures is extracted to model the unfolded state (see **Figure 1** and panel **a** in **Figures 2-4** and **Extended Data Figures 1-4**). To avoid using too large simulation boxes, which may increase the simulation time as well as add noise to the results, the most extended conformations (~10 %) are previously identified and removed as described[25] (**SI Figure 1**). The selected 100 unfolded conformations are simulated in

boxes containing one unfolded molecule and exactly the same number of water molecules, ions and cofactors –when it is the case– as in the corresponding boxes used to simulate the folded conformations of the same protein. For three-state proteins, in addition to the overall enthalpy change, those of the individual steps (F-to-I and I-to-U) can be obtained if the absolute enthalpy of an additional box containing one molecule of protein in the intermediate conformation and the same number of water and ion entities is calculated (see **Figure 1** and **Figure 3a**). To model the intermediate conformation, a suitable structural model is needed. In the three-state apoFld, a 20-model NMR ensemble previously described[84] is used. In this case, five replicas are simulated for each of the 20 structures, totaling 100 replicas, the same number of unfolded conformations modeled (**Figure 3a**).

For each replica, a short 2-ns productive trajectory (see **SI Table 1**) is run and the individual time-averaged enthalpy (H^i_F, H^i_U or H^i_I) retrieved. The individual enthalpies of replicas of the same conformational state (i.e. folded, unfolded or intermediate) are then ensemble-averaged to obtain the corresponding enthalpy to each folding state ($\langle H_F \rangle$, $\langle H_U \rangle$ or $\langle H_I \rangle$). Subsequently, the unfolding enthalpy change, ΔH_{unf}, is calculated by difference, i.e. by subtracting the calculated ensemble-averaged enthalpy obtained from simulations of the folded state from the ensemble-averaged enthalpy obtained from simulations of the unfolded state: $\Delta H_{unf} = \langle H_U \rangle - \langle H_F \rangle$. For three-state proteins, enthalpy changes corresponding to the first unfolding transition (F-to-I) and the second one (I-to-U) are calculated likewise: $\Delta H_{unf(F\text{-}to\text{-}I)} = \langle H_I \rangle - \langle H_F \rangle$ and $\Delta H_{unf(I\text{-}to\text{-}U)} = \langle H_U \rangle - \langle H_I \rangle$ (**Figure 1**).

The calculation of the heat capacity change upon unfolding (ΔCp_{unf}) relies on the linear dependency of ΔH_{unf} with temperature. For each protein, three not-distant temperatures are selected so that the temperature range covered contains the experimental T_m of the simulated protein. The three calculated ΔH_{unf} values are represented as a function of simulation temperature and the ΔCp_{unf} is calculated as the slope of a linear fit. For three-state proteins (e.g. apoFld), $\Delta Cp_{unf(F\text{-}to\text{-}I)}$ and $\Delta Cp_{unf(I\text{-}to\text{-}U)}$ are obtained as the temperature dependence of the calculated enthalpy changes of each unfolding transition.

The calculation of the protein stability curves (ΔG_{unf} as a function of temperature) is done through the Gibbs-Helmholtz equation[27] (**Eq. 1**):

$$\Delta G(T) = \Delta H_{Tm} \times (1 - T/T_m) - \Delta Cp \times [T_m - T + T \times \ln(T/T_m)] \quad \textbf{(1)}$$

introducing the calculated ΔH_{unf} and ΔCp_{unf} values, and the reported experimental T_m.

MD-based specific workflow for calculation of unfolding energetics (ΔH_{unf}, ΔCp_{unf} and ΔG_{unf}) in holoproteins

In the case of holoproteins (non-covalent complexes of apoprotein and cofactor; e.g. holoFld), the ensemble-averaged enthalpy of the folded (bound) state, $\langle H_{holo(F)} \rangle$, is obtained from simulations (40 replicas) each consisting of one molecule of holoFld solvated with water molecules and ions, as

needed (**Figure 1** and **Figure 4a**). Similarly, the energetics of the unfolded (unbound) state is modeled from simulations (100 replicates) in which one unfolded protein molecule generated with ProtSA and one cofactor molecule (placed at a minimum distance of 3 nm from the protein) are put together in a box, where they are solvated in the same way (**Figure 4a**). The ensemble-averaged enthalpy of such boxes, $\langle H_{apo(U)+cofactor} \rangle$, is obtained following the averaging scheme of the general workflow. Then, the unfolding enthalpy change is calculated as: $\Delta H_{unf} = \langle H_{apo(U)+cofactor} \rangle - \langle H_{holo(F)} \rangle$.

As required for this enthalpy change calculation by difference, the number of water molecules and ions in the box containing unfolded protein and cofactor must equate those in the box containing folded holoprotein (**SI Table 2**). The simulations are also performed at three different temperatures so that the unfolding ΔCp_{unf} can be obtained as the slope of a ΔH_{unf} versus temperature plot (**Figure 4a-b**).

Similarly to the case of apoproteins, the ΔH_{unf} and ΔCp_{unf} values calculated for the holoprotein can be combined with the experimental T_m to obtain the protein stability curves (ΔG_{unf} as a function of temperature). However, as the conformational stability of holoproteins is cofactor concentration dependent, a modified Gibbs-Helmholtz equation that takes into account the binding energetics (**SI Eq. 5**, see details in **SI Methods**) has been used to calculate the conformational stability as a function of temperature and concentration of free cofactor.

Solvation conditions and MD simulation general details

Solvation conditions on the simulated proteins (i.e. protonation states and the number of ions added) were selected in each case to reproduce the experimental pH and ionic strength (IS) under which the experimental thermodynamics measurements were performed (see detailed information in **SI Methods** and **SI Table 2)**. Box dimensions were adopted from the diameter of the most elongated structure in the unfolded ensemble sampled for the protein in question plus a minimum distance of 1 nm from protein atoms to the simulation box edges. The MD simulation setup is similar to that described in Galano-Frutos et.al.[25] (details are also given in **SI Table 1**). All the systems are simulated with the force field Charmm22 with CMAP correction (version 2.0)[18] and the explicit water model Tip3p[28]: the most accurate force field/water model combination reported in previous work[25]. The Amber99SB-ILDN[17] force field is again tested combined with Tip3p[28] by modeling the apoFld unfolding thermodynamics. MD simulations are run and analyzed with Gromacs 2020 package[94]. The usage of short 2-ns productive trajectories in the workflow[25] avoids the known issue of structure overcompaction in long simulations[16,25] for force fields like Charmm22-CMAP[18] and Amber99SB-ILDN[17]. In addition, the simulations performed here have been tested for overcompaction through analysis of the radius of gyration (Rg) evolution along the trajectories (**SI Table 3**). Results of this analysis confirm that no significant protein compaction takes place over the trajectories of the proteins simulated (**SI Table 3)**. The mutant variants here tested (of CI2 and lysozyme) have been modeled by

replacing the wild-type residue by the new one using the mutator tool of Chimera (v.1.15)[95], as no solved structures were available. No clashes were observed in the final mutant structures of lowest energy obtained after accommodating the new residues, which were taken as the starting structures in simulations of their folded states. In the case of the apoFld intermediate state, the representative model used (see below) was mutated back to its wild type sequence (Chimera v.1.15)[95] in order to keep the same amino acids sequence as that of the other structural models used in simulations of apo and holoFld. No clashes were observed after this replacement either. Crystal waters and any other non-protein molecule were removed from the PDB structural models chosen (see below). Computational core facilities used in the current implementation consisted of a CPU-based cluster (170 Intel Xeon E5-2680v3 CPUs arranged in 85 nodes, https://cesar.unizar.es/hpc/), which was exploited through an Open MPI parallel framework (most simulations were launched on 32 cores: 32 OpenMP threads/1 MPI task).

Structure models (PDBs) and coverage

The starting structures used to simulate the folded state of the proteins analyzed are those with the highest resolution available in the RCSB Protein Data Bank[92,93] at the time of writing this manuscript. Namely: 1A2P (1.5 Å resolution)[96] for barnase, 2SNS (1.5 Å)[97] for nuclease (C-ter fragment), 2CI2 (2.0 Å)[98] for CI2 (truncated form), 6LZM (1.8 Å)[75] for lysozyme, 1L63 (1.75 Å)[99] for pseudo-lysozyme, 1FTG (2.0 Å)[100] for apoFld and 1FLV (2.0 Å)[101] for holoFld. On the other hand, the thermal unfolding intermediate state of apoFld is represented by 2KQU[84], a 20-model NMR ensemble of the Phe99Asn mutant previously shown to be a reliable representation of this state[83,84,102] According to the reference sequences in UniProt[103] the structural coverage of the solved sequences is 3-110 (barnase), 83-231 (nuclease C-terminal fragment), 20-84 (WT CI2 and Ile76Ala mutant), 1-162 (WT lysozyme and Ile3Glu mutant), 1-162 (pseudo-WT lysozyme), 3-170 (apo and holoFld). .

FMN parameterization

Three different parameterizations of the FMN molecule (charge –2) have been tested. Namely, 'Par.-1' was done *ad hoc* assisted by the AmberTools20 package[104] and the Gaussian 09 program[105]; 'Par.-2' was that reported by Schulten et.al.[106]; and 'Par.-3' was obtained through the SwissParam server[107]. FMN coordinates were extracted from the crystal structure of holoFld (PDB ID: 1FLV)[101]. For *ad-hoc* 'Par.-1', partial atomic charges were modeled with Gaussian 09 (HF/6-31G*), which were then fitted through the RESP method[108,109] (with Antechamber)[104,110], and finally, parameters obtained from the General Amber Force Field (GAFF[111], Antechamber[104,110]). FMN coordinates were uploaded to SwissParam[107] ('Par.-3') in mol2 format after adding hydrogen atoms. Except for van der Waals parameters, which are taken from the closest atom type in Charmm2, parameters and charges with this server are derived/taken from Merck Molecular Force Field (MMFF)[107].

Increased sampling for higher precision

As previously shown[25], individual enthalpies (H^i_F, H^i_U or H^i_I) of the simulated systems (i.e. boxes containing one protein molecule, several ions and thousands of water molecules) can mount to 10^5 kcal/mol (negative values) or even higher (see **SI Table 4**). These big figures owe to the large number of water molecules present in the large simulation boxes required to solvate the unfolded conformations. In general, the larger the protein, the larger the negative enthalpy of the simulated box. Therefore, calculation of unfolding thermodynamics by difference requires a high precision (a low standard error in the calculation) to be able to assess the accuracy of the approach (the difference between experimental and *in silico* results). Since the enthalpy change of a partial thermal unfolding step of a protein (e.g. the apoFld F-to-I or I-to-U transitions) can be significantly lower than the enthalpy changes previously modeled for barnase and nuclease[25] (see **Table 1**), a higher precision (standard error ≤ 10 kcal/mol) than that achieved in previous work[25] is here guaranteed *a priori* by running a higher number of replicas (extended sampling, **Table 2**). For each system the minimum sample size necessary to meet such precision was estimated as reported (see the **Supplementary Information** of Galano-Frutos et al.[25]).

References

1. Berg, J. M., Tymoczko, J. L. & Stryer, L. *Protein Structure and Function*. (W. H. Freeman, 2002).

2. Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *Science* vol. 338 1042–1046 (2012).

3. Goldenzweig, A. & Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **87**, 105–129 (2018).

4. Best, R. B. Atomistic molecular simulations of protein folding. *Curr. Opin. Struct. Biol.* **22**, 52–61 (2012).

5. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science (80-.).* **334**, 517–520 (2011).

6. Lei, H., Wu, C., Liu, H. & Duan, Y. Folding free-energy landscape of villin headpiece subdomain from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 4925–4930 (2007).

7. Miao, Y., Feixas, F., Eun, C. & McCammon, J. A. Accelerated molecular dynamics simulations of protein folding. *J. Comput. Chem.* **36**, 1536–1549 (2015).

8. Daggett, V. & Levitt, M. Protein unfolding pathways explored through molecular dynamics

simulations. *J. Mol. Biol.* **232**, 600–619 (1993).

9. Jiang, F. & Wu, Y. D. Folding of fourteen small proteins with a residue-specific force field and replica-exchange molecular dynamics. *J. Am. Chem. Soc.* **136**, 9536–9539 (2014).

10. Gsponer, J. & Caflisch, A. Molecular dynamics simulations of protein folding from the transition state. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6719–6724 (2002).

11. Sedov, I. A. & Magsumov, T. I. Molecular dynamics study of unfolding of lysozyme in water and its mixtures with dimethyl sulfoxide. *J. Mol. Graph. Model.* **76**, 466–474 (2017).

12. Lindorff-Larsen, K., Trbovic, N., Maragakis, P., Piana, S. & Shaw, D. E. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* **134**, 3787–3791 (2012).

13. Freddolino, P. L., Harrison, C. B., Liu, Y. & Schulten, K. Challenges in protein-folding simulations. *Nat. Phys. 2010 610* **6**, 751–758 (2010).

14. Vuillleumier, S. & Fersht, A. R. Insertion in barnase of a loop sequence from ribonuclease T1: Investigating sequence and structure alignments by protein engineering. *Eur. J. Biochem.* **221**, 1003–1012 (1994).

15. Lindorff-Larsen, K. *et al.* Systematic validation of protein force fields against experimental data. *PLoS One* **7**, (2012).

16. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E4758–E4766 (2018).

17. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma.* **78**, 1950–1958 (2010).

18. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation. *J. Comput. Chem.* **25**, 1400–1415 (2004).

19. Ben-Naim, A. The Rise and Fall of the Hydrophobic Effect in Protein Folding and Protein-Protein Association, and Molecular Recognition. *Open J. Biophys.* **01**, 1–7 (2011).

20. Baker, D. What has de novo protein design taught us about protein folding and biophysics? *Protein Sci.* **28**, 678 (2019).

21. Bender, B. J. *et al.* A practical guide to large-scale docking. *Nat. Protoc.* **16**, 4799 (2021).

22. Galano-Frutos, J. J., Garciá-Cebollada, H. & Sancho, J. Molecular dynamics simulations for genetic interpretation in protein coding regions: where we are, where to go and when. *Brief.*

Bioinform. **22**, 3–19 (2021).

23. Castro, K. M., Scheck, A., Xiao, S. & Correia, B. E. Computational design of vaccine immunogens. *Curr. Opin. Biotechnol.* **78**, 102821 (2022).

24. Zielinski, D. C., Patel, A. & Palsson, B. O. The Expanding Computational Toolbox for Engineering Microbial Phenotypes at the Genome Scale. *Microorganisms* **8**, 1–18 (2020).

25. Galano-Frutos, J. J. & Sancho, J. Accurate Calculation of Barnase and SNase Folding Energetics Using Short Molecular Dynamics Simulations and an Atomistic Model of the Unfolded Ensemble: Evaluation of Force Fields and Water Models. *J. Chem. Inf. Model.* **59**, 4350–4360 (2019).

26. Estrada, J., Bernadó, P., Blackledge, M. & Sancho, J. ProtSA: A web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. *BMC Bioinformatics* **10**, 104 (2009).

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

28. 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.* **79**, 926–935 (1983).

29. Piana, S., Donchev, A. G., Robustelli, P. & Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **119**, 5113–5123 (2015).

30. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **112**, 8910–8922 (2000).

31. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F. & Hermans, J. Interaction Models for Water in Relation to Protein Hydration. in 331–342 (Springer, Dordrecht, 1981). doi:10.1007/978-94-015-7658-1_21.

32. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).

33. Deller, M. C., Kong, L. & Rupp, B. Protein stability: A crystallographer's perspective. *Acta Crystallographica Section:F Structural Biology Communications* vol. 72 72–95 (2016).

34. Sancho, J. The stability of 2-state, 3-state and more-state proteins from simple spectroscopic techniques... plus the structure of the equilibrium intermediates at the same time. *Archives of Biochemistry and Biophysics* vol. 531 4–13 (2013).

35. Paddon, C. J. & Hartley, R. W. Cloning, sequencing and transcription of an inactivated copy of Bacillus amyloliquefaciens extracellular ribonuclease (barnase). *Gene* **40**, 231–239 (1985).

36. Hartley, R. W. & Barker, E. A. Amino-acid sequence of extracellular ribonuclease (barnase) of bacillus amyloliquefaciens. *Nat. New Biol.* **235**, 15–16 (1972).

37. Davis, A., Moore, I. B., Parker, D. S. & Taniuchi, H. Nuclease B. A possible precursor of nuclease A, an extracellular nuclease of Staphylococcus aureus. *J. Biol. Chem.* **252**, 6544–6553 (1977).

38. Cone, J. L., Cusumano, C. L., Taniuchi, H. & Anfinsen, C. B. Staphylococcal nuclease (Foggi strain). II. The amino acid sequence. *J. Biol. Chem.* **246**, 3103–3110 (1971).

39. Williamson, M. S., Forde, J., Buxton, B. & Kreis, M. Nucleotide sequence of barley chymotrypsin inhibitor-2 (CI-2) and its expression in normal and high-lysine barley. *Eur. J. Biochem.* **165**, 99–106 (1987).

40. Svendsen, I., Martin, B. & Jonassen, I. Characteristics of Hiproly barley III. Amino acid sequences of two lysine-rich proteins. *Carlsberg Res. Commun.* **45**, 79–85 (1980).

41. Tsugita, A. & Inouye, M. Purification of bacteriophage T4 lysozyme. *J. Biol. Chem.* **243**, 391–397 (1968).

42. Fillat, M. F., Edmondson, D. E. & Gomez-Moreno, C. Structural and chemical properties of a flavodoxin from Anabaena PCC 7119. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* **1040**, 301–307 (1990).

43. Irún, M. P., Maldonado, S. & Sancho, J. Stabilization of apoflavodoxin by replacing hydrogen-bonded charged Asp or Glu residues by the neutral isosteric Asn or Gln. *Protein Eng.* **14**, 173–181 (2001).

44. Irún, M. P., Garcia-Mira, M. M., Sanchez-Ruiz, J. M. & Sancho, J. Native hydrogen bonds in a molten globule: the apoflavodoxin thermal intermediate. *J. Mol. Biol.* **306**, 877–888 (2001).

45. Lamazares, E., Clemente, I., Bueno, M., Velázquez-Campoy, A. & Sancho, J. Rational stabilization of complex proteins: A divide and combine approach. *Sci. Rep.* **5**, 1–11 (2015).

46. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nat. 2021 5967873* **596**, 583–589 (2021).

47. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* (2021) doi:10.1093/NAR/GKAB1061.

48. Serrano, L., Matouschek, A. & Fersht, A. R. The folding of an enzyme. VI. The folding pathway of barnase: Comparison with theoretical models. *J. Mol. Biol.* **224**, 847–859 (1992).

49. Matouschek, A., Serrano, L., Meiering, E. M., Bycroft, M. & Fersht, A. R. The folding of an

enzyme. V. solH2H exchange-nuclear magnetic resonance studies on the folding pathway of barnase: Complementarity to and agreement with protein engineering studies. *J. Mol. Biol.* **224**, 837–845 (1992).

50. Oliveberg, M., Vuilleumier, S. & Fersht, A. R. Thermodynamic Study of the Acid Denaturation of Barnase and Its Dependence on Ionic Strength: Evidence for Residual Electrostatic Interactions in the Acid/ Thermally Denatured State. *Biochemistry* **33**, 8826–8832 (1994).

51. Matouschek, A., Serrano, L. & Fersht, A. R. The folding of an enzyme. IV. Structure of an intermediate in the refolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 819–835 (1992).

52. Serrano, L., Matouschek, A. & Fersht, A. R. The folding of an enzyme. III. Structure of the transition state for unfolding of barnase analysed by a protein engineering procedure. *J. Mol. Biol.* **224**, 805–818 (1992).

53. Serrano, L., Kellis, J. T., Cann, P., Matouschek, A. & Fersht, A. R. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* **224**, 783–804 (1992).

54. Griko, Y. V., Makhatadze, G. I., Privalov, P. L. & Hartley, R. W. Thermodynamics of barnase unfolding. *Protein Sci.* **3**, 669–676 (1994).

55. Martínez, J. C. *et al.* A Calorimetric Study of the Thermal Stability of Barstar and its Interaction with Barnase. *Biochemistry* **34**, 5224–5233 (1995).

56. Makarov, A. A. *et al.* Comparative study of thermostability and structure of close homologues - bamase and binase. *J. Biomol. Struct. Dyn.* **10**, 1047 (1993).

57. Matouschek, A., Matthews, J. M., Johnson, C. M. & Fersht, A. R. Extrapolation to water of kinetic and equilibrium data for the unfolding of barnase in urea solutions. *Protein Eng. Des. Sel.* **7**, 1089–1095 (1994).

58. Shortle, D. & Meeker, A. K. Mutant forms of staphylococcal nuclease with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins Struct. Funct. Bioinforma.* **1**, 81–89 (1986).

59. Shortle, D., Meeker, A. K. & Freire, E. Stability Mutants of Staphylococcal Nuclease: Large Compensating Enthalpy-Entropy Changes for the Reversible Denaturation Reaction. *Biochemistry* **27**, 4761–4768 (1988).

60. Carra, J. H., Anderson, E. A. & Privalov, P. L. Thermodynamics of staphylococcal nuclease denaturation. I. The acid-denatured state. *Protein Sci.* **3**, 944–951 (1994).

61. Eftink, M. R., Ghiron, C. A., Kautz, R. A., Fox, R. O. & Fox, R. O. Fluorescence and Conformational Stability Studies of Staphylococcus Nuclease and Its Mutants, Including the Less Stable Nuclease-Concanavalin a Hybrids. *Biochemistry* **30**, 1193–1199 (1991).

62. Tan, Y. J., Oliveberg, M., Davis, B. & Fersht, A. R. Perturbed pKA-values in the denatured states of proteins. *J. Mol. Biol.* **254**, 980–992 (1995).

63. Jackson, S. E. & Fersht, A. R. Folding of Chymotrypsin Inhibitor 2.1. Evidence for a Two-State Transition. *Biochemistry* **30**, 10428–10435 (1991).

64. Jackson, S. E., Moracci, M., elMasry, N., Johnson, C. M. & Fersht, A. R. Effect of Cavity-Creating Mutations in the Hydrophobic Core of Chymotrypsin Inhibitor 2. *Biochemistry* **32**, 11259–11269 (1993).

65. Kjær, M. & Poulsen, F. M. Secondary structure of barley serine proteinase inhibitor 2 determined by proton nuclear magnetic resonance spectroscopy. *Carlsberg Res. Commun.* **52**, 355–362 (1987).

66. Kjær, M. *et al.* Sequence specific assignment of the proton nuclear magnetic resonance spectrum of barley serine proteinase inhibitor 2. *Carlsberg Res. Commun.* **52**, 327–354 (1987).

67. Matthews, B. W. Genetic and Structural Analysis of the Protein Stability Problem. *Biochemistry* **26**, 6885–6888 (1987).

68. Hawkes, R., Grutter, M. G. & Schellman, J. Thermodynamic stability and point mutations of bacteriophage T4 lysozyme. *J. Mol. Biol.* **175**, 195–212 (1984).

69. Klemm, J. D., Wozniak, J. A., Alber, T. & Goldenberg, D. P. Correlation between Mutational Destabilization of Phage T4 Lysozyme and Increased Unfolding Rates. *Biochemistry* **30**, 589–594 (1991).

70. Kitamura, S. & Sturtevant, J. M. A Scanning Calorimetric Study of the Thermal Denaturation of the Lysozyme of Phage T4 and the Arg 96 →His Mutant Form Thereof. *Biochemistry* **28**, 3788–3792 (1989).

71. Matsumura, M., Becktel, W. J. & Matthews, B. W. Hydrophobic stabilization in T4 lysozyme determined directly by multiple substitutions of Ile 3. *Nature* **334**, 406–410 (1988).

72. Connelly, P. *et al.* A Differential Scanning Calorimetric Study of the Thermal Unfolding of Seven Mutant Forms of Phage T4 Lysozyme. *Biochemistry* **30**, 1887–1891 (1991).

73. Ladbury, J. E., Sturtevant, J. M. & Hu, C. Q. A Differential Scanning Calorimetric Study of the Thermal Unfolding of Mutant Forms of Phage T4 Lysozyme. *Biochemistry* **31**, 10699–10702 (1992).

74. Burley, S. K. *et al.* RCSB Protein Data Bank: Biological macromolecular structures enabling

research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.* **47**, D464–D474 (2019).

75. Bell, J. A. *et al.* Comparison of the crystal structure of bacteriophage T4 lysozyme at low, medium, and high ionic strengths. *Proteins Struct. Funct. Bioinforma.* **10**, 10–21 (1991).

76. Carra, J. H., Murphy, E. C. & Privalov, P. L. Thermodynamic effects of mutations on the denaturation of T4 lysozyme. *Biophys. J.* **71**, 1994–2001 (1996).

77. Sancho, J. Flavodoxins: sequence, folding, binding, function and beyond. *Cell. Mol. Life Sci. C. 2006 637* **63**, 855–864 (2006).

78. Goñi, G. *et al.* Flavodoxin: A compromise between efficiency and versatility in the electron transfer from Photosystem I to Ferredoxin-NADP+ reductase. *Biochim. Biophys. Acta - Bioenerg.* **1787**, 144–154 (2009).

79. Casaus, J. L. *et al.* Anabaena sp. PCC 7119 flavodoxin as electron carrier from photosystem I to ferredoxin-NADP + reductase. Role of Trp 57 and Tyr 94. *J. Biol. Chem.* **277**, 22338–22344 (2002).

80. Lostao, A. *et al.* How FMN binds to Anabaena apoflavodoxin: A hydrophobic encounter at an open binding site. *J. Biol. Chem.* **278**, 24053–24061 (2003).

81. Lostao, A. *et al.* Dissecting the energetics of the apoflavodoxin-FMN complex. *J. Biol. Chem.* **275**, 9518–9526 (2000).

82. Lamazares, E. *et al.* Direct examination of the relevance for folding, binding and electron transfer of a conserved protein folding intermediate. *Phys. Chem. Chem. Phys.* **19**, 19021–19031 (2017).

83. García-Fandiño, R., Bernadó, P., Ayuso-Tejedor, S., Sancho, J. & Orozco, M. Defining the Nature of Thermal Intermediate in 3 State Folding Proteins: Apoflavodoxin, a Study Case. *PLoS Comput. Biol.* **8**, e1002647 (2012).

84. Ayuso-Tejedor, S. *et al.* Design and Structure of an Equilibrium Protein Folding Intermediate: A Hint into Dynamical Regions of Proteins. *J. Mol. Biol.* **400**, 922–934 (2010).

85. Campos, L. A., Bueno, M., Lopez-Llano, J., Jiménez, M. Á. & Sancho, J. Structure of stable protein folding intermediates by equilibrium φ-analysis: The apoflavodoxin thermal intermediate. *J. Mol. Biol.* **344**, 239–255 (2004).

86. Campos, L. A. & Sancho, J. Native-specific stabilization of flavodoxin by the FMN cofactor: Structural and thermodynamical explanation. *Proteins Struct. Funct. Genet.* **63**, 581–594 (2006).

87. Lostao, A., Gómez-Moreno, C., Mayhew, S. G. & Sancho, J. Differential stabilization of the

three FMN redox forms by tyrosine 94 and tryptophan 57 in flavodoxin from Anabaena and its influence on the redox potentials. *Biochemistry* **36**, 14334–14344 (1997).

88. Fukada, H., Sturtevant, J. M. & Quiocho, F. A. Thermodynamics of the binding of L-arabinose and of D-galactose to the L-arabinose-binding protein of Escherichia coli. *J. Biol. Chem.* **258**, 13193–13198 (1983).

89. Sillitoe, I. *et al.* CATH: increased structural coverage of functional space. *Nucleic Acids Res.* **49**, D266 (2021).

90. Baase, W. A., Liu, L., Tronrud, D. E. & Matthews, B. W. Lessons from the lysozyme of phage T4. *Protein Sci.* **19**, 631 (2010).

91. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293 (2018).

92. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

93. Burley, S. K. *et al.* RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **49**, D437–D451 (2021).

94. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* vol. 26 1701–1718 (2005).

95. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

96. Martin, C., Richard, V., Salem, M., Hartley, R. & Mauguen, Y. Refinement and structural analysis of barnase at 1.5 Å resolution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55**, 386–398 (1999).

97. Legg, M.J., Cotton, F.A., Hazen Jr., E. E. *RCSB PDB - 2SNS: Staphylococcal Nuclease. Proposed Mechanism of Action Based on Structure of Enzyme-Thymidine 3(Prime), 5(Prime)-Biphosphate-Calcium Ion Complex at 1.5-Angstroms Resolution*. https://www.rcsb.org/structure/2SNS (1982) doi:10.2210/pdb2SNS/pdb.

98. McPhalen, C. A. & James, M. N. Crystal and Molecular Structure of the Serine Proteinase Inhibitor CI-2 from Barley Seeds. *Biochemistry* **26**, 261–269 (1987).

99. Nicholson, H., Anderson, D. E., Dao-pin, S. & Matthews, B. W. Analysis of the Interaction between Charged Side Chains and the α-Helix Dipole Using Designed Thermostable Mutants of Phage T4 Lysozyme. *Biochemistry* **30**, 9816–9828 (1991).

100. Genzor, C. G., Perales-Alcón, A., Sancho, J. & Romero, A. Closure of a tyrosine/tryptophan aromatic gate leads to a compact fold in apo flavodoxin. *Nat. Struct. Mol. Biol.* **3**, 329–332 (1996).

101. Rao, S. T. *et al.* Structure of the oxidized long-chain flavodoxin from anabaena 7120 at 2 å resolution. *Protein Sci.* **1**, 1413–1427 (1992).

102. Ayuso-Tejedor, S., García-Fandiño, R., Orozco, M., Sancho, J. & Bernadó, P. Structural analysis of an equilibrium folding intermediate in the apoflavodoxin native ensemble by small-angle X-ray scattering. *J. Mol. Biol.* **406**, 604–619 (2011).

103. Bateman, A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).

104. Case, D. A., et al. AMBER 2020. (2018).

105. Frisch M. J. et al. Gaussian 09, Revision A.02. (2016).

106. Freddolino, P. L., Gardner, K. H. & Schulten, K. Signaling mechanisms of LOV domains: New insights from molecular dynamics studies. *Photochem. Photobiol. Sci.* **12**, 1158–1170 (2013).

107. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32**, 2359–2368 (2011).

108. Cieplak, P., Cornell, W. D., Bayly, C. & Kollman, P. A. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* **16**, 1357–1377 (1995).

109. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).

110. Wang, J., Wang, W., Kollman, P. A. & Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006).

111. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

112. Bueno, M., Campos, L. A., Estrada, J. & Sancho, J. Energetics of aliphatic deletions in protein cores. *Protein Sci.* **15**, 1858–1872 (2006).

113. Budavari, S. *The Merck index : an encyclopedia of chemicals, drugs, and biologicals*. (Merck and Co., Inc., 1996).

Tables

Table 1. Experimental thermal unfolding data

Protein	Method	Ref.	pH & IS (mM)[a]	T_m(s) (K)[b,c]	ΔH_{unf} (kcal/mol)[c,d]	ΔCp_{unf} (kcal/mol·K)[c]	ΔG^0_{unf} (kcal/mol)[c,e]
barnase	DSC	56	3.8 & 0.4[f]	321.0 ± 0.1	107.0 ± 7.0	0.9 ± 0.1	6.9 ± 0.7
		54	4.0 & 0.2[g]	321.7[g]	106.4[g]	1.4[g]	6.6
		55	4.0 & 8[h]	323.5[g]	125.0 ± 6.3	1.6 ± 0.2	8.2 ± 0.9
		14	4.4 & 6[i]	324.5[g]	126.7[g]	1.6[g]	8.5[g]
		57	4.4 & 6[i]	324.5 ± 0.0	128.6 ± 1.5	1.6 ± 0.2	8.7 ± 0.1
	(Ave ± SE)		**~4.1 & 4.0**	**323.0 ± 0.7**	**118.7 ± 4.9**	**1.4 ± 0.1**	**7.8 ± 0.4**
nuclease[j]	DSC	60	7.0 & 141[k]	327.3 ± 0.5	77.4 ± 7.7	2.2 ± 0.5	4.0 ± 3.0
			5.0 & 141[l]	323.8 ± 0.5	73.1 ± 7.3	2.2 ± 0.5	3.6 ± 2.7
			4.1 & 0.4[m]	316.4 ± 0.5	49.2 ± 4.9	1.2 ± 0.5	2.6 ± 1.5
		59	7.0 & 151[n]	326.5 ± 0.2	86.8 ± 1.4	2.4 ± 0.4	4.6 ± 1.4
			4.9 & 112[o]	321.5 ± 0.3	73.0 ± 6.6	2.4 ± 0.4	3.4 ± 1.9
			4.1 & 103[o]	314.4 ± 0.3	59.0 ± 5.3	1.2 ± 0.4	3.1 ± 1.0
	(Ave ± SE)		**pH 7.0[p]**	**326.9 ± 0.4**	**82.1 ± 4.7**	**2.3 ± 0.1**	**4.3 ± 0.3**
			pH 5.0[p]	**322.7 ± 1.1**	**73.1 ± 0.1**	**2.3 ± 0.1**	**3.5 ± 0.1**
			pH 4.1[p]	**315.4 ± 1.0**	**54.1 ± 4.9**	**1.2 ± 0.0**	**2.9 ± 0.3**
WT CI2[q]	DSC	64	3.0 & 1.9[f]	337.2 ± 0.2	61.0 ± 2.3	0.7[g]	5.4 ± 0.7
	Spectroscopy	62,63	6.3 & 33[r]	358.7[g,s]	78.4 ± 1.4[t]	0.8 ± 0.1[t]	7.2 ± 0.4[u]
Ile76Ala CI2[q]	DSC	64	3.0 & 1.9[f]	311.5[g]	30.2[g]	0.7[g]	1.1 ± 0.3
WT lysozyme	DSC	70	2.4 & 34[v]	317.4 ± 0.6[w]	109.0 ± 4.2[w]	2.3 ± 0.2[w]	5.3 ± 0.8
		73	2.4 & 34[v]	317.8 ± 0.2[x]	104.3 ± 2.2[x]	2.5 ± 0.0[x]	4.9 ± 0.6
	(Ave ± SE)		**2.4 & 34**	**317.6 ± 0.2**	**106.6 ± 2.3**	**2.4 ± 0.1**	**5.1 ± 0.6**
Ile3Glu lysozyme	DSC	73	2.4 & 34[v]	312.9 ± 0.3[x]	105.1 ± 3.0[x]	2.5 ± 0.0[x]	4.1 ± 0.8
pseudo-WT lysozyme[y]	DSC	76	3.0 & 3.8[z]	327.1 ± 0.5	127.4 ± 12.7	1.6 ± 0.2[§]	9.3 ± 1.5
		76	3.7 & 0.9[z]	337.6 ± 0.5	138.9 ± 13.9	1.6 ± 0.2[§]	12.5 ± 1.1
apoFld	Spectroscopy	44	7.0 & 17[‡]	316.5 ± 0.3 329.6 ± 0.1	28.1 ± 0.6 54.9 ± 1.2	1.5 ± 0.3 1.6 ± 0.5	0.8 ± 0.7 2.8 ± 1.1
	Spectroscopy	45	7.0 & 17[‡]	316.0 ± 0.3 328.3 ± 0.4	34.7 ± 2.0 55.6 ± 2.0		
	Spectroscopy	85	7.0 & 17[‡]	317.3 ± 0.2 329.0 ± 0.1	33.9 ± 0.1 52.7 ± 0.6		
	Spectroscopy	112	7.0 & 17[‡]	317.3 ± 0.2 329.0 ± 0.1	34.1 ± 0.1 52.9 ± 0.6		
	Spectroscopy	84	7.0 & 17[‡]	317.2 ± 0.3 328.9 ± 0.1	32.2 ± 0.1 51.0 ± 1.1		
	DSC	84	7.0 & 17[‡]	317.1 ± 0.6 329.4 ± 0.1	32.7 ± 0.4 66.0 ± 0.2	1.2 ± 0.1 1.5 ± 0.1	1.3 ± 0.5 4.0 ± 0.3

	DSC	44	7.0 & 17‡	311.8 ± 2.5 326.5 ± 1.2	28.0 ± 3.7 56.0 ± 2.5	≤1.5 ± <0.3 ≥1.0 ± <0.3	≥0.8 ± 4.2 ≤3.6 ± 1.9
	(Ave ± SE)		**7.0 & 17**	**316.2 ± 0.8** **328.7 ± 0.4**	**32.0 ± 1.1** **55.6 ± 2.0**	**1.4 ± 0.3** **1.6 ± 0.3**	**1.1 ± 1.4** **2.9 ± 1.3**
holoFld	DSC	82	7.0 & 17‡	340.7 ± 0.1	101.9 ± 0.6	3.5 ± 0.6ϕ	17.1 ± 2.7‡

[a] Experimental pH and ionic strength (IS) conditions. IS reported or calculated according to buffer, concentration, and pH reported.
[b] Mid-denaturation temperature (T_m) reported or calculated from a reported empirical equation.
[c] For three-state apoFld, two values are shown per row. The first one corresponds to the Native-to-Intermediate transition, and the second one to the Intermediate-to-Unfolded transition.
[d] Enthalpy change upon thermal unfolding (ΔH_{unf}) either reported or calculated from a given empirical equation at T_m.
[e] Standard (298.15 K) conformational stability (ΔG^0_{unf}) obtained from the Gibbs-Helmholtz equation[27] (**Eq. 1**), except otherwise noted. When more than one experimental data are reported, the ΔG^0_{unf} values shown in the 'Ave±SE' row are the average among those values (Ave), and the standard error is obtained by dividing the standard deviation (SD) between the square root of the number of data (SD/√n) (it is not the value calculated through the Gibbs-Helmholtz equation and its propagated associate error). For nuclease, values are calculated at 293.15 K, as experimental data appear reported at that temperature.
[f] 10mM glycine hydrochloride. IS calculated from the Henderson-Haselbach equation and the Glycine pKa values of 2.37 and 9.78[113].
[g] No error reported.
[h] 50 mM sodium acetate.
[i] 20 mM sodium acetate.
[j] The modeled nuclease is the 149-residue C-ter fragment of the protein.
[k] 20 mM sodium phosphate, 100 mM NaCl, 1 mM EDTA.
[l] 20 mM sodium acetate, 100 mM NaCl, 1 mM EDTA.
[m] 20 mM glycine hydrochloride. The influence of salt concentration (between 0-800 mM) on measurements seems negligible (see Fig. 1c of the reference paper)[60].
[n] 25 mM sodium phosphate, 100 mM NaCl.
[o] 20 mM sodium acetate, 100 mM NaCl.
[p] As measurements of nuclease unfolding thermodynamics are independent of IS[60] and this parameter largely varied in the experiments reported, the buffer IS is not taken into account in the modeling of this protein.
[q] Truncated wild type CI2 and Ile76Ala variant lacking the first 19 amino acid residues.
[r] 50 mM MES, as reported by Jackson et. al.[63]
[s] T_m reported by Tan et. al.[62]
[t] Obtained by extrapolating at T_m after doing a ΔH_{unf} *vs.* T_m fitting with reported data[63], the slope being ΔCp_{unf}.
[u] Value extrapolated to [GdnHCl] = 0 M from thermal denaturation data[63].
[v] 20 mM potassium phosphate, 25 mM KCl, 0.5 mM dithiothreitol.
[w] Values obtained from the reported empirical equations $T_m = 9.63 + 14.41\times pH$ and $\Delta H_{unf} = 5.97 + 2.33\times T$. ΔCp_{unf} is the slope of this fitting equation.
[x] Values obtained from the reported empirical equations $T_m = 9.13 + 14.81 \times pH$ and $\Delta H_{unf} = -10.51\ (\pm 0.83) + 2.57\ (\pm 0.02) \times T$ for the wild-type protein, $T_m = -0.62\ (\pm 0.13) + 16.84\ (\pm 0.05) \times pH$ and $\Delta H_{unf} = 5.22\ (\pm 1.14) + 2.51\ (\pm 0.03) \times T$ (T in Celsius degrees) for the Ile3Glu variant. ΔCp_{unf} is the slope of the ΔH_{unf} *vs.* T fitting line.
[y] Lysozyme variant where residues 54 and 97 appear replaced by a threonine and an alanine, respectively.
[z] 20 mM glycine hydrochloride. IS calculated from the Henderson-Haselbach equation and the glycine pKa values of 2.37 and 9.78[113].
[§] Value obtained from the ΔH_{unf} *vs.* T linear fitting plot in Fig. 6a of the reference paper[76].
[‡] 50 mM MOPS at 298.15 K.
[‡] Standard Gibbs free-energy of unfolding ([FMN] = 1M) obtained from **Eq. S5** (includes the correction of temperature and ligand concentration, see the **SI Methods**). For the calculation of this stability, the average ($K_b = 3.61(\pm 1.4) \times 10^9$ M) of binding constants reported for FMN[81,82,86], as well as the enthalpy ($\Delta H_{bind} = -11.0 \pm 0.2$ kcal/mol) and heat capacity changes ($\Delta Cp_{bind} = -0.6 \pm 0.02$ kcal/mol·K) upon binding[81] were used. As an additional data, a standard Gibbs free-energy change of 19.0 ± 0.9 kcal/mol has been reported by Campos and coworkers[86].
[ϕ] ΔCp_{unf} value estimated as follows: sum of ΔCp_{unf} of the two partial unfolding steps of apoFld (1.4 ± 0.3 and 1.6 ± 0.3 kcal/mol·K) plus the ΔCp of binding reported for FMN (−0.6 ± 0.02 kcal/mol·K)[81].

Table 2. Calculated thermal unfolding energetics from MD simulations

Protein (condition)[b]	Force Field[a]	ΔH_{unf} (kcal/mol)[c,d]	ΔCp_{unf} (kcal/mol·K)[e,d]	ΔG^0_{unf} (kcal/mol)[f,d]
Barnase (pH: ~4.1, T: 295, 315, 335 K)	Charmm22-CMAP	110.4 ± 3.1	1.0 ± 0.1	7.5 ± 1.2
nuclease (pH: 7.0, T: 307, 317, 327 K)	Charmm22-CMAP	75.1 ± 4.5	1.7 ± 0.3	4.8 ± 1.7
nuclease (pH: 5.0, T: 307, 317, 327 K)	Charmm22-CMAP	71.0 ± 4.5	1.5 ± 0.4	4.4 ± 2.8
nuclease (pH: 4.1, T: 307, 317, 327 K)	Charmm22-CMAP	86.3 ± 4.8	1.7 ± 0.1	4.8 ± 2.2
WT CI2 (pH: 3.0, T: 320, 335, 350 K)	Charmm22-CMAP	46.1 ± 1.9	0.4 ± 0.0	4.3 ± 0.4
WT CI2 (pH: 6.3, T: 335, 350, 365 K)	Charmm22-CMAP	57.1 ± 1.9	0.5 ± 0.1	6.9 ± 0.6
Ile76Ala CI2 (pH: 3.0, T: 320, 335, 350 K)	Charmm22-CMAP	27.7 ± 1.7	0.5 ± 0.0	1.0 ± 0.2
WT lysozyme (pH: 2.4, T: 305, 320, 335 K)	Charmm22-CMAP	184.6 ± 2.9	1.3 ± 0.2	10.6 ± 0.6
Ile3Glu lysozyme (pH: 2.4, T: 305, 320, 335 K)	Charmm22-CMAP	174.6 ± 2.9	1.5 ± 0.1	7.8 ± 0.6
pseudo-WT lysozyme (pH: 3.0, T: 305, 320, 335 K)	Charmm22-CMAP	205.1 ± 1.0	1.6 ± 0.2	16.2 ± 0.7
pseudo-WT lysozyme (pH: 3.7, T: 305, 320, 335 K)	Charmm22-CMAP	227.4 ± 1.2	1.7 ± 0.3	22.5 ± 2.0
apoFld (pH: 7.0, T: 305, 320, 335 K)	Charmm22-CMAP	35.6 ± 6.0 48.1 ± 4.1 (83.7 ± 10.1)	1.5 ± 0.1 1.0 ± 0.0 (2.5 ± 0.1)	1.3 ± 1.7 3.0 ± 0.9 (4.3 ± 2.6)
apoFld (pH: 7.0, T: 305, 320, 335 K)	Amber99SB-ILDN	49.5 ± 4.2 18.3 ± 3.3 (67.8 ± 7.5)	1.4 ± 0.1 1.1 ± 0.1 (2.5 ± 0.2)	2.1 ± 1.5 0.1 ± 0.9 (2.2 ± 2.4)
holoFld	Charmm22-CMAP / Par.-1	114.2 ± 7.8	3.0 ± 0.2	17.3 ± 2.6[h]

(pH: 7.0, T: 305, 320, 335 K)	Charmm22-CMAP / Par.-2	104.2 ± 5.8	2.9 ± 0.6
	Charmm22-CMAP / Par.-3	103.0 ± 6.5	2.6 ± 0.1

[a] Force fields tested for the calculation, and FMN parameterizations used in holoFld systems (see **Methods**). The water model used is always Tip3p, as described in **Methods**.

[b] ΔH_{unf} is calculated at the three indicated temperatures. ΔCp_{unf} is obtained as the slope of a ΔH_{unf} *vs.* T linear plot

[c] Calculated enthalpy change upon thermal unfolding (ΔH_{unf}) at T_m (see values in **Table 1**), obtained by extrapolation. When a given protein has been simulated at more than one pH value, the simulation pH is indicated in each case.

[d] For three-state apoFld, three calculated ΔH values are shown per row. The upper one corresponds to the enthalpy change of the Native-to-Intermediate transition; the intermediate one corresponds to the enthalpy change of the Intermediate-to-Unfolded transition; and the lower one (between parenthesis) corresponds to the total ΔH_{unf}, obtained by adding up the values calculated for each transition. Likewise, in the ΔCp_{unf} and ΔG^0_{unf} columns, the three values indicated correspond (from top to bottom) to the Native-to-Intermediate, Intermediate-to-Unfolded, and global (Native-to-Unfolded) heat capacity or Gibbs free-energy changes, respectively.

[e] Calculated ΔCp_{unf} obtained as the slope of a ΔH_{unf} *vs.* T linear plot. Fitting errors are given.

[f] Unfolding Gibbs free-energy changes at 298.15 K calculated using the Gibbs-Helmholtz equation[27] (or **Eq. S5** for HoloFld; see **SI Methods**). For nuclease, the temperature of reference used, 293.15 K, is the one at which most of the experimental data are reported (**Table 1**).

[g] No error reported.

[h] Standard Gibbs free-energy (at 1M FMN) calculated through **Eq. S5** (**SI Methods** and the footnote ‡ in **Table 1**).

Figures

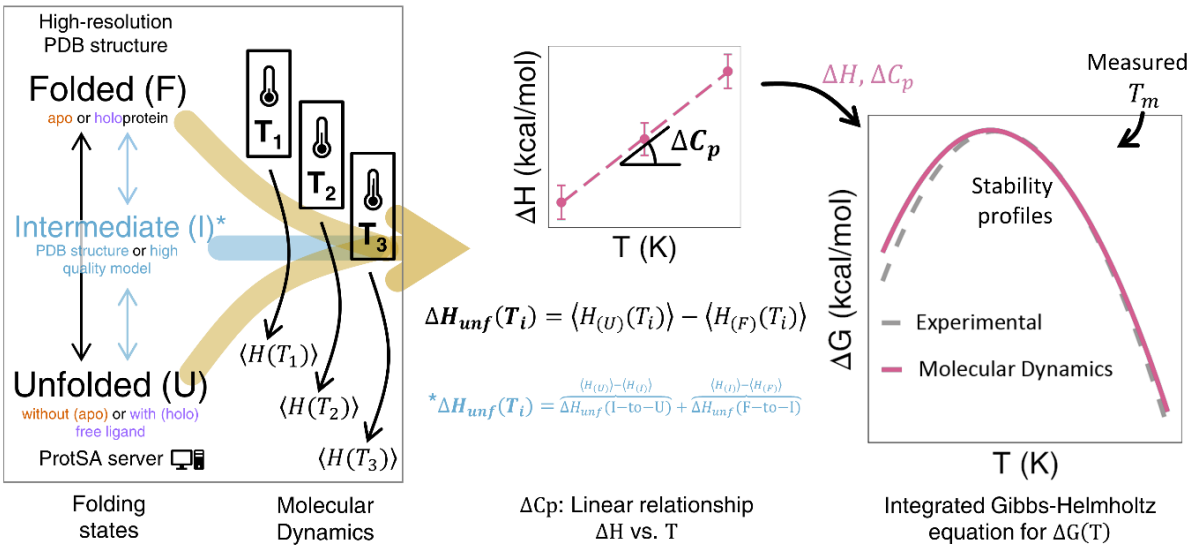


Figure 1. General workflow of the devised MD-based approach. The enthalpy of simulation boxes containing either folded ($H_{apo(F)}$ or $H_{holo(F)}$) or unfolded ($H_{apo(U)}$ or $H_{apo(U)+cofactor}$ in holoproteins) protein (or a structure representative of the intermediate state, $H_{apo(I)}$) are directly computed and averaged from Molecular Dynamics simulations. The unfolding enthalpy change (ΔH_{unf}) is obtained as the difference. The simulations are performed at three temperatures and the change in heat capacity (ΔCp_{unf}) is obtained as the slope of a linear plot of enthalpy change versus temperature. These two magnitudes are combined with the experimental T_m of the protein to calculate the conformational stability using the Gibbs-Helmholtz equation. For holoproteins, a similar equation, **Eq. S5** in the **Supplementary Material,** is used that applies a correction to Gibbs free-energy to account for the ligand concentration, and uses the van't Hoff approximation to describe the temperature dependence of the binding constant, $K_b(T)$. The number of water molecules and ions added to the folded, unfolded (or intermediate, if applicable) boxes must be identical. Forty (40) replicas of the folded box (normally built from a high-resolution PDB structure) and one hundred (100) of the unfolded box built from a filtered sample of completely unfolded conformations generated from ProtSA server[26] are simulated. For intermediate states, 100 simulation replicas can be built from a representative structural ensemble. For holoproteins, the unfolded box is built by placing an unfolded protein molecule generated with ProtSA and one molecule of cofactor at a given minimum distance of the protein. The rest of general details can be found in **Methods** and in panels **a** of **Figures 2-4** and **Extended Data Figures 1-4**.

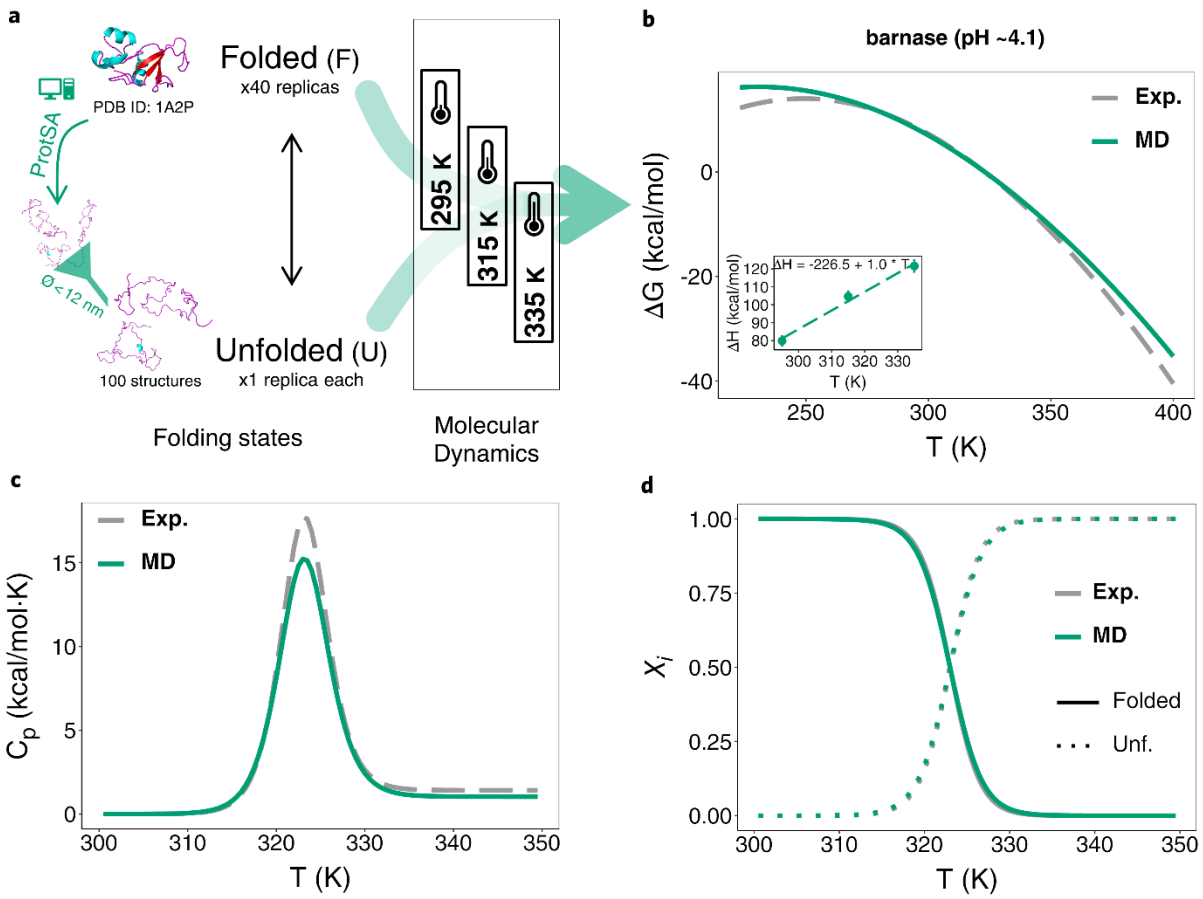


Figure 2. Simplified MD-based scheme and comparison with experimental results for a two-state protein example: barnase. a) The protein models, the number of structures (unfolded) and replicas (folded) simulated, the diameter cutoff used to filter too-elongated unfolded structures obtained from ProtSA[26] (left), and temperatures selected for the MD-based calculation (Charmm22-CMAP) of thermodynamics of barnase. **b-d)** Temperature profiles of ΔG_{unf}, Cp_{unf} and protein molar fractions (χ_i) (*in silico* vs. experimental), respectively, obtained for barnase simulated at pH ~4.1. Inset in **b** depicts the calculated ΔH_{unf} vs. T linear plot with the fitted equation (the slope being ΔCp_{unf}) obtained from the MD simulations. The color-coding is indicated in the legends of the panels.

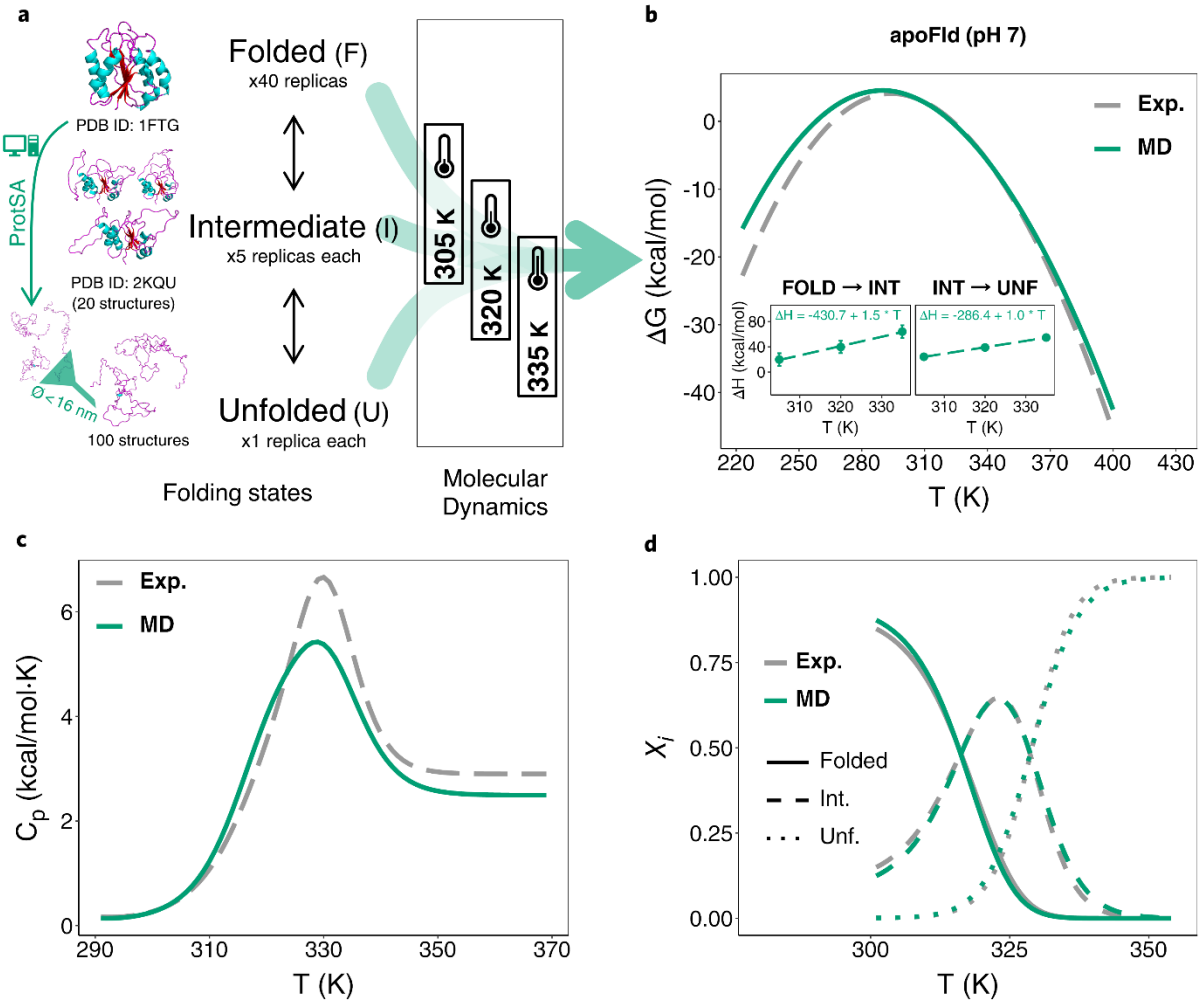


Figure 3. Simplified MD-based scheme and comparison with experimental results for a three-state protein example: apoFld. **a)** Protein models, number of structures (unfolded) and replicas (folded) simulated, diameter cutoff used to filter too-elongated unfolded structures obtained from ProtSA[26] (left), and temperatures selected for the MD-based calculation (Charmm22-CMAP) of apoFld thermodynamics. **b-d)** Temperature profiles of the global stability ($\Delta G_{unf} = \Delta G_{unf(F\text{-}to\text{-}I)} + \Delta G_{unf(I\text{-}to\text{-}U)}$), Cp_{unf} and χ_i, respectively (*in silico* vs. experimental). Inset in **b** depicts linear plots of calculated ΔH_{unf} from the MD simulations *vs.* T, with the fitted equation (the slope being ΔCp_{unf}) obtained. The color-coding is indicated in the legends of the panels.

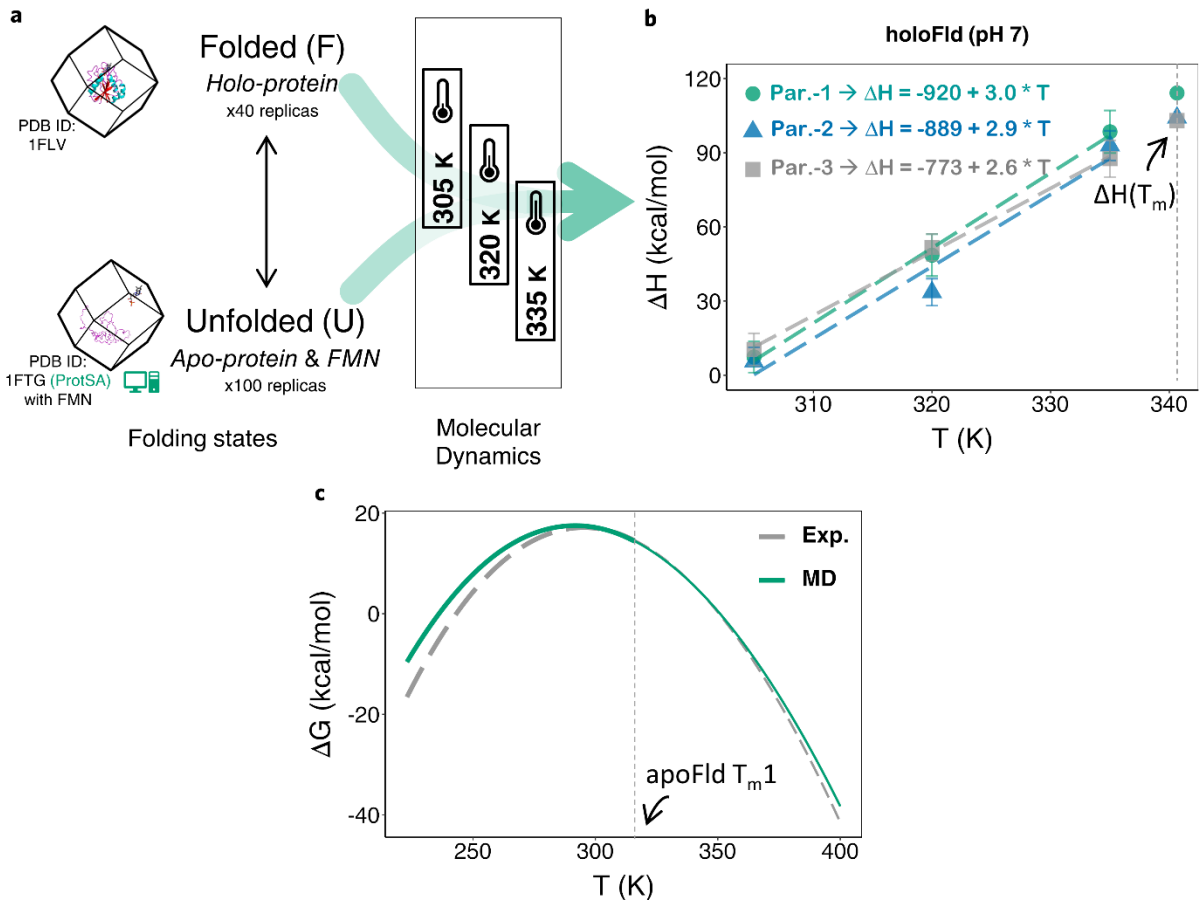


Figure 4. Simplified MD-based scheme and comparison with experimental results for a holoprotein example: holoFld. **a)** Protein and cofactor models placed in the simulation boxes, folding states, number of structures (unfolded) and replicas (folded) simulated, and temperatures selected for the MD-based calculation (Charmm22-CMAP) of holoFld thermodynamics. **b)** Calculated ΔH_{unf} vs. T linear plots, with the fitted equations (slopes are the respective ΔCp_{unf}) obtained for the three FMN parameterizations tested. Extrapolated ΔH_{unf} values at T_m (340.7 K) are indicated over the vertical dashed line placed at this temperature. **c)** Temperature profiles of ΔG_{unf} (*in silico* vs. experimental) obtained from **Eq. S5**. Curves appear depicted with finer lines beyond the first T_m of the apoprotein (316.2 K, **Table 1**, vertical dashed line) to indicate –as described in **Results**– that in this region the ΔG_{unf} values calculated are not reliable. The fact that at 298.15 K the calculated stability of HoloFld (17.3 ± 2.6 kcal/mol) agrees within error with the stability measured from thermal unfolding curves (19.0 ± 0.9 kcal/mol)[86] seems to validate the accuracy of the profiles in the range of temperatures below the apoprotein T_m1.

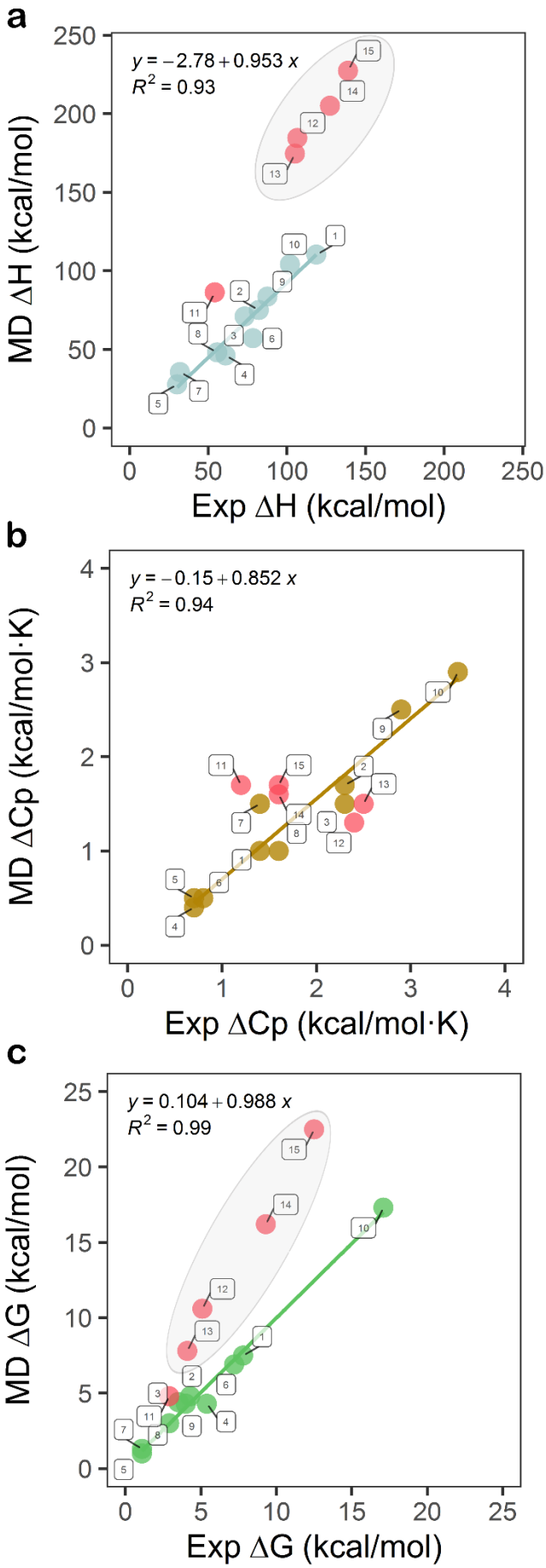
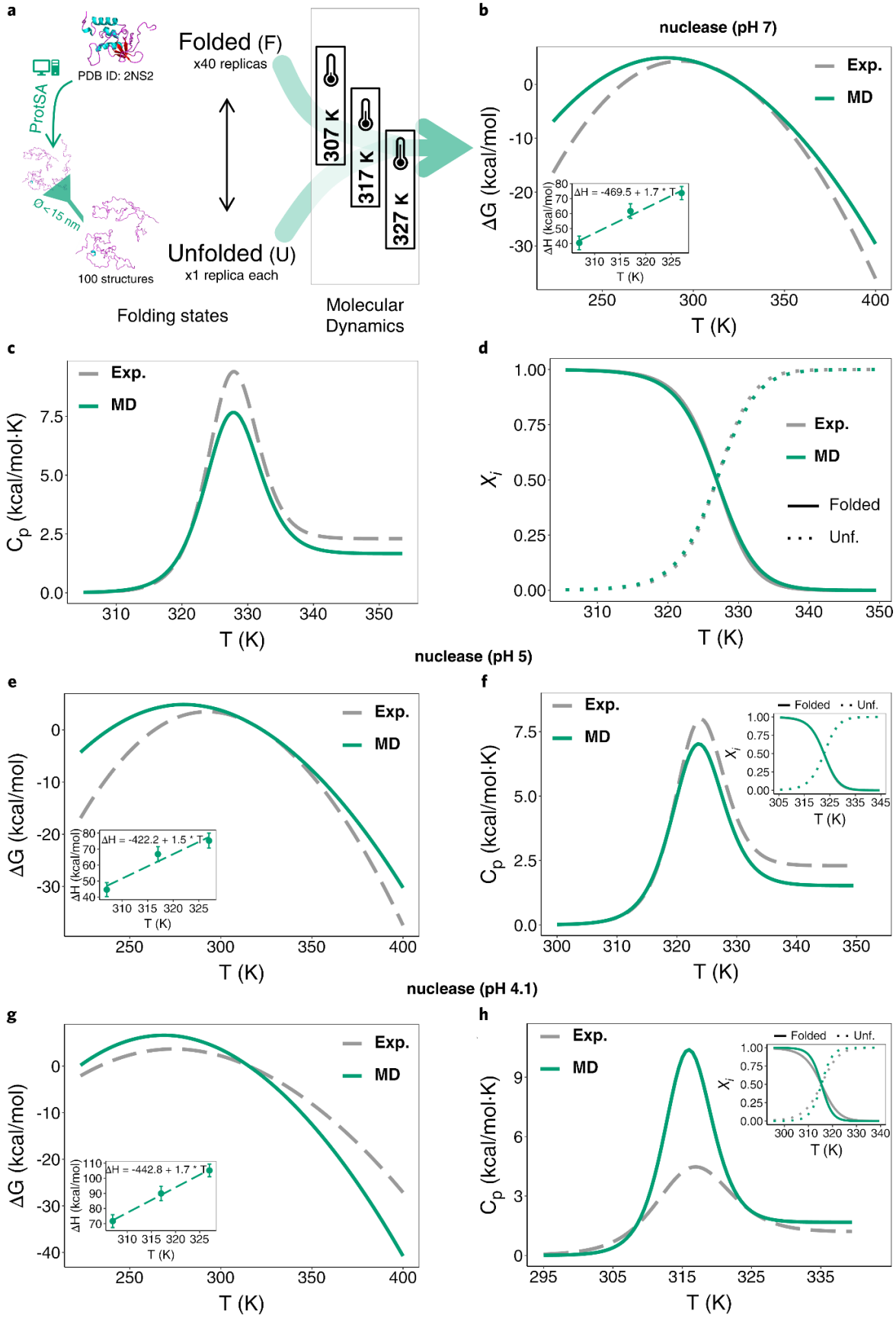
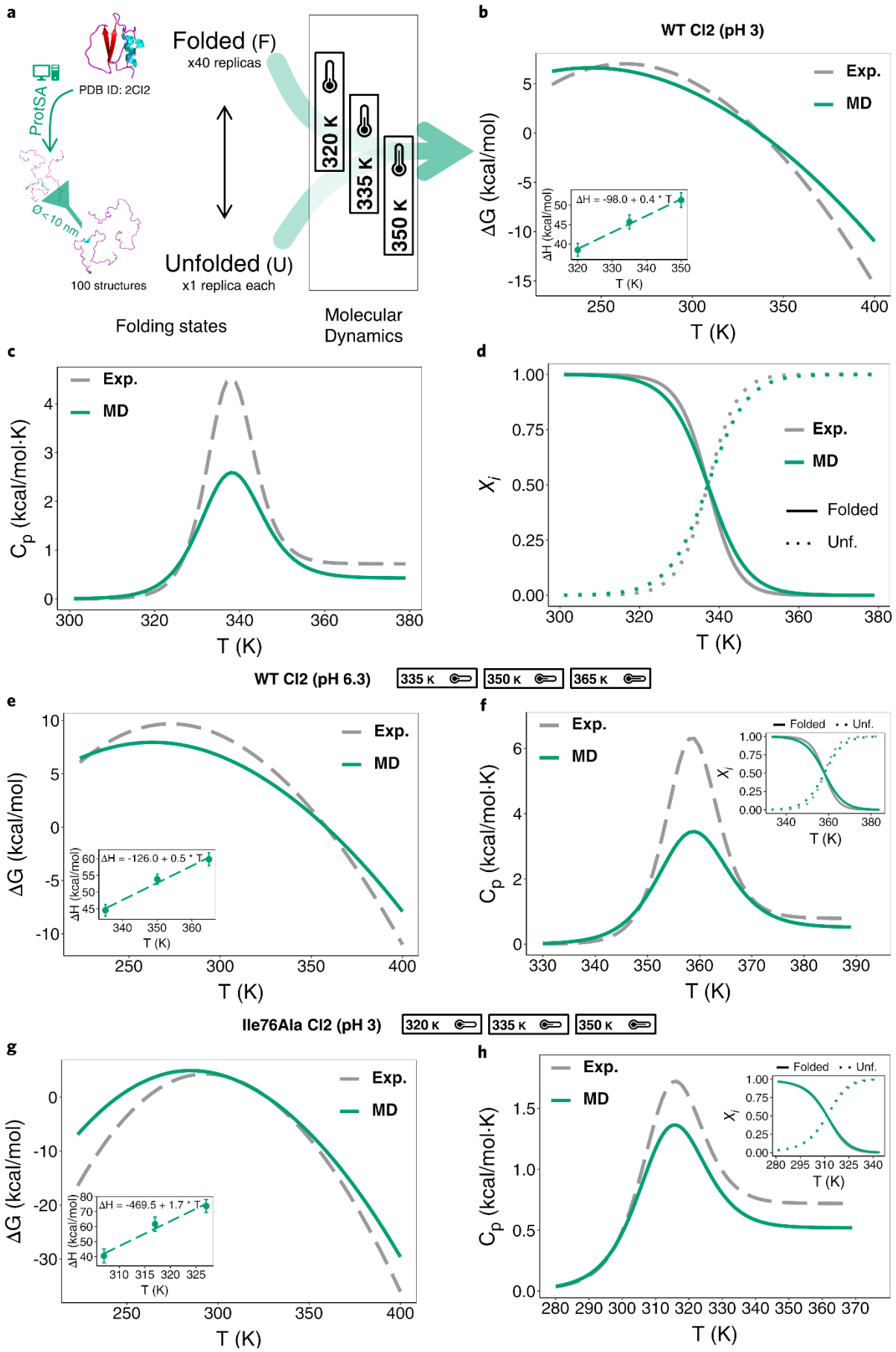

a
y = −2.78 + 0.953 x
R2 = 0.93
MD ΔH (kcal/mol)
Exp ΔH (kcal/mol)
b
y = −0.15 + 0.852 x
R2 = 0.94
MD ΔCp (kcal/mol·K)
Exp ΔCp (kcal/mol·K)
c
y = 0.104 + 0.988 x
R2 = 0.99
MD ΔG (kcal/mol)
Exp ΔG (kcal/mol)

Figure 5. Global assessment of the approach for calculation of unfolding thermodynamics with Charmm22-CMAP/Tip3p. a) Scatter plot of MD-calculated vs. experimental ΔH_{unf} for the set of proteins simulated (including different solvating conditions and variants). The linear fit shown in this panel (also in panels **b** and **c**) was performed over the following ten systems: barnase at pH ~4.1 (dot number 1 in legend), nuclease at pH 7.0 (2) and pH 5.0 (3), WT CI2 at pH 3.0 (4), Ile76Ala CI2 at pH 3.0 (5), WT CI2 at pH 6.3 (6), apoFld(F-to-I) (7), apoFld(I-to-U) (8), apoFld(F-to-U) (9) and holoFld(FMN Par.-2) (10). The fitting equation and the square Pearson correlation coefficient are given. **b)** Scatter plot and linear fit of MD-calculated vs. experimental ΔCp_{unf}. **c)** Scatter plot and linear fitting of MD-calculated vs. experimental protein stability (ΔG^0_{unf} at 298.15 K for all proteins except for nuclease that is compared at 293.15 K). Experimental values (*x*-axis) are the averages (or individual value in some cases) of data obtained from literature, as summarized in **Table 1**, while calculated values are those presented in **Table 2**. Red circles represent outliers (or cases treated as such, see the **Results and Discussion** section) not considered in the linear fitting, namely: nuclease at pH 4.1 (dot number 11 in legend), WT lysozyme at pH 2.4 (12), Ile3Glu lysozyme at pH 2.4 (13), pseudo-WT lysozyme at pH 3.0 (14) and pseudo-WT lysozyme at pH 3.7 (15). In panels **a** and **c**, the 4 outliers of lysozyme and pseudo-lysozyme systems are enclosed in a semi-transparent gray oval to visualize them as similar systems whose enthalpy change upon unfolding (ΔH_{unf}) and protein stability (ΔG_{unf}) are all overestimated by our simulations. Of the three setups tested for holoFld, the results obtained with FMN parameterization 2 (the most accurate one, see **Tables 1** and **2**) are depicted.

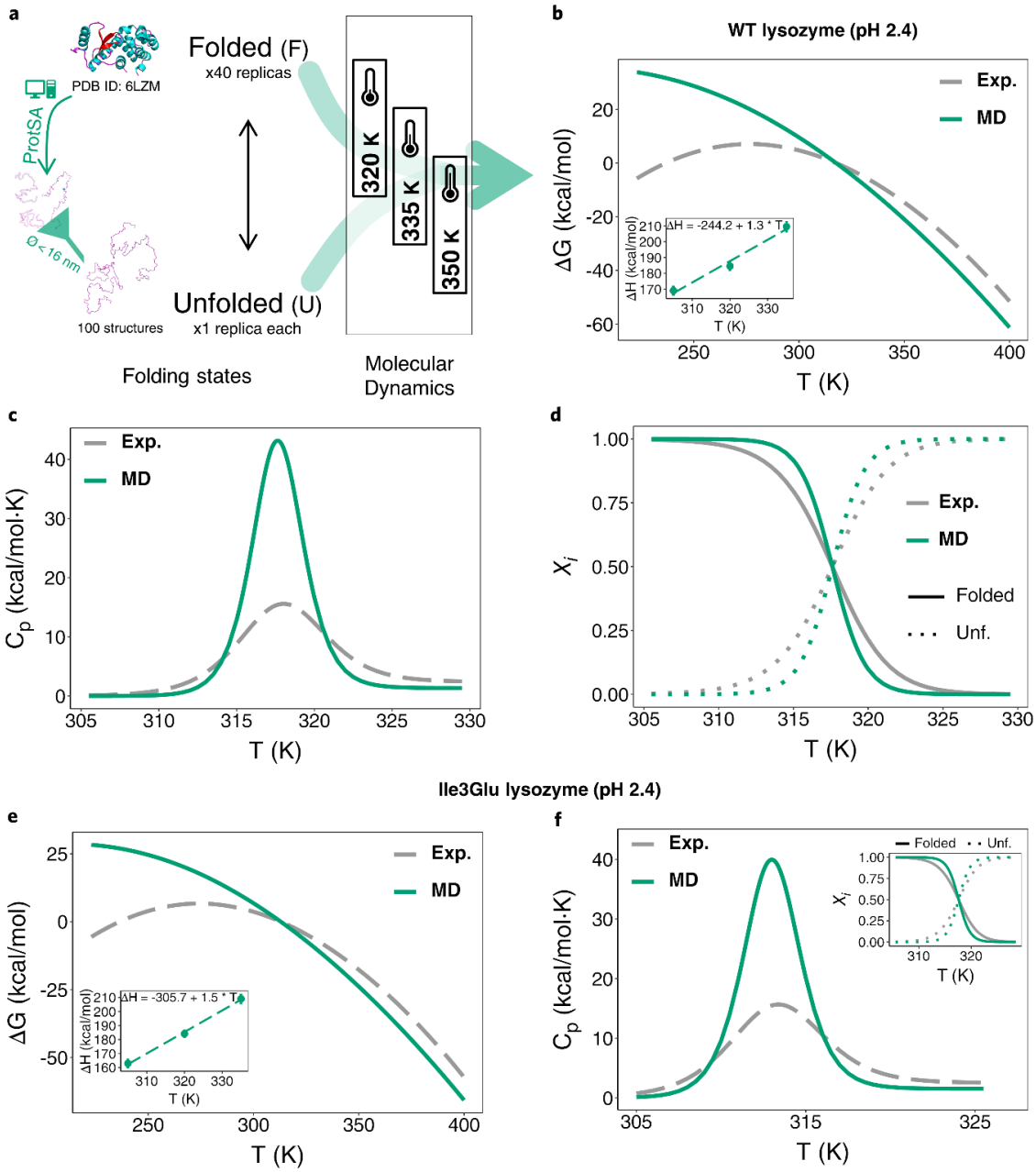
Extended Data Figures



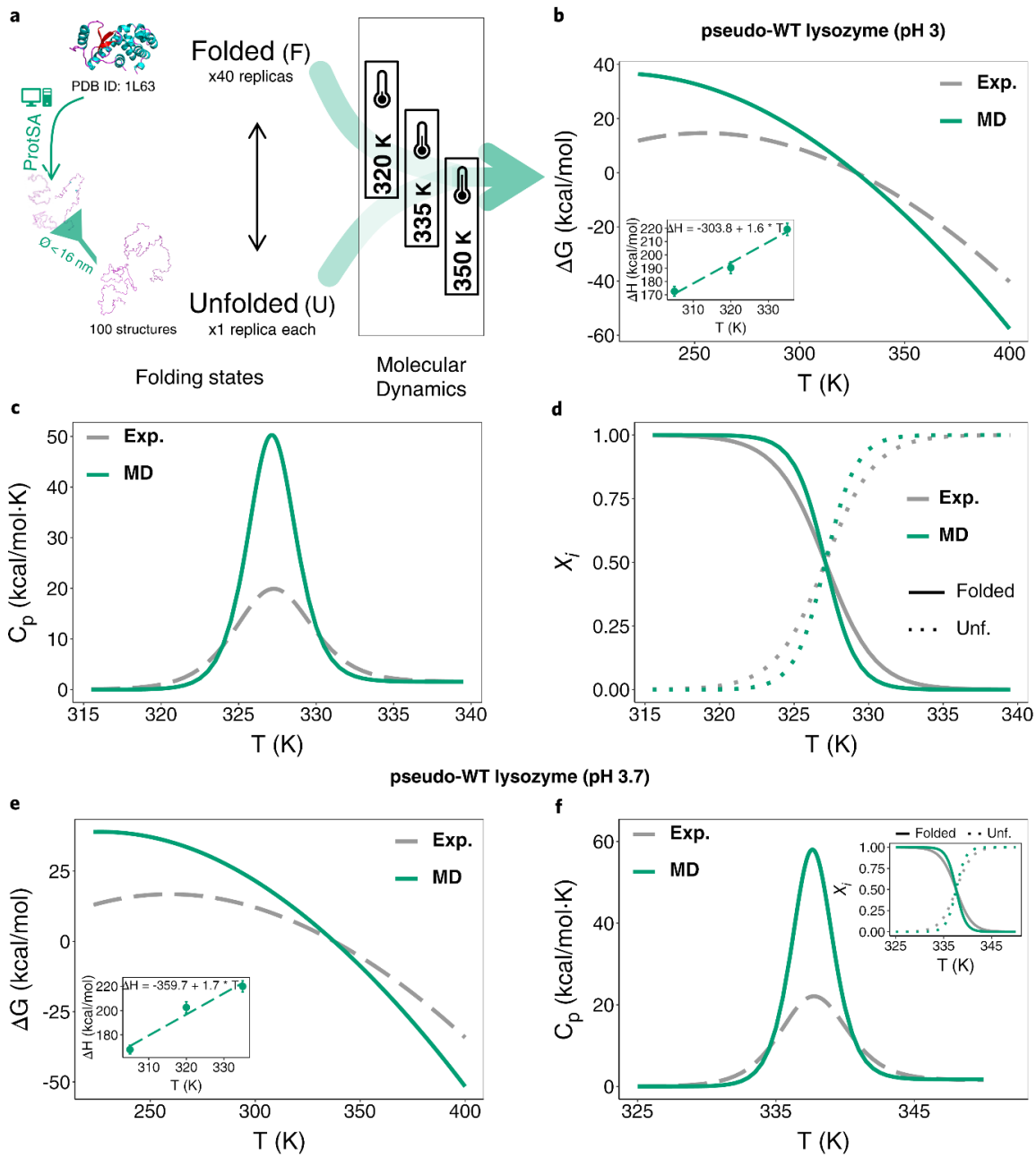
Extended Data Figure 1. Simplified MD-based scheme and comparison with experimental results for nuclease. a) Protein models and folding states, number of structures (unfolded) and replicas (folded) simulated, diameter cutoff used to filter too-elongated unfolded structures obtained from ProtSA[9] (left), and temperatures selected for the MD-based calculation (Charmm22-CMAP) of nuclease thermodynamics at pH 7.0, 5.0 and 4.1. **b-d)** Temperature profiles of ΔG_{unf}, Cp_{unf} and χ_i (computed vs. experimental), respectively, obtained for nuclease simulated at pH 7.0. **e** and **f)** Temperature profiles of ΔG_{unf} and Cp_{unf} (computed vs. experimental), respectively, obtained for nuclease simulated at pH 5.0. **g** and **h)** Temperature profiles of ΔG_{unf} and Cp_{unf} (computed vs. experimental), respectively, obtained for nuclease simulated at pH 4.1. The ΔH_{unf} *vs.* T linear plot for pH ~5.0 (and ~4.1) appears as an inset in **e** (and **g**). The temperature profile of χ_i for pH 5.0 (and 4.1) is included as an inset in **f** (and **h**). The color-coding is indicated in the legends of the panels.


a
PDB ID: 2CI2
ProtSA
Ø<10 nm
100 structures
Folded (F)
x40 replicas
Unfolded (U)
x1 replica each
320 K
335 K
350 K
Folding states
Molecular Dynamics
b
WT CI2 (pH 3)
Exp.
MD
ΔG (kcal/mol)
ΔH = -98.0 + 0.4 * T
ΔH (kcal/mol)
T (K)
c
Exp.
MD
Cp (kcal/mol·K)
T (K)
d
Exp.
MD
Folded
Unf.
Xi
T (K)
WT CI2 (pH 6.3)
335 K
350 K
365 K
e
Exp.
MD
ΔG (kcal/mol)
ΔH = -126.0 + 0.5 * T
ΔH (kcal/mol)
T (K)
f
Exp.
MD
Folded
Unf.
Cp (kcal/mol·K)
Xi
T (K)
Ile76Ala CI2 (pH 3)
320 K
335 K
350 K
g
Exp.
MD
ΔG (kcal/mol)
ΔH = -469.5 + 1.7 * T
ΔH (kcal/mol)
T (K)
h
Exp.
MD
Folded
Unf.
Cp (kcal/mol·K)
Xi
T (K)

Extended Data Figure 2. Simplified MD-based scheme and comparison with experimental results for CI2. a) Protein models and folding states, number of structures (unfolded) and replicas (folded) simulated, diameter cutoff used to filter too-elongated unfolded structures obtained from ProtSA[9] (left), and temperatures selected for the MD-based calculation (Charmm22-CMAP) of CI2 thermodynamics. **b-d)** Temperature profiles of ΔG_{unf}, Cp_{unf} and χ_i (*in silico* vs. experimental), respectively, obtained for WT CI2 at pH 3.0. Inset in **b** depicts the ΔH_{unf} vs. T linear plot with the fitted equation (the slope being ΔCp_{unf}) obtained from the MD simulations. **e** and **f)** Temperature profiles of ΔG_{unf} and Cp_{unf} (computed vs. experimental), respectively, obtained for Ile76Ala at pH 3.0. The ΔH_{unf} vs. T linear plot appears as an inset in **e** and the temperature profile of χ_i is included as an inset in **f**. **g** and **h)** Temperature profiles of ΔG_{unf} and Cp_{unf} (computed vs. experimental), respectively, obtained for WT CI2 at pH 6.3. Insets in **g** and **h** similar to those in **e** and **f**, respectively. The color-coding is indicated in the legends of the panels.



Extended Data Figure 3. Simplified MD-based scheme and comparison with experimental results for lysozyme. a) The protein models and folding states, the number of structures (unfolded) and replicas (folded) simulated, the diameter cutoff used to filter too-elongated unfolded structures obtained from ProtSA[9] (left), and temperatures selected for the MD-based calculation (Charmm22-CMAP) of thermodynamics of WT lysozyme and the variant Ile3Glu at pH 2.4. **b-d)** Temperature profiles of ΔG_{unf}, Cp_{unf} and χ_i (*in silico* vs. experimental), respectively, obtained for WT lysozyme at pH 2.4. Inset in **b** depicts the ΔH_{unf} vs. T linear plot with the fitted equation (the slope being ΔCp_{unf}) obtained from the MD simulations. **e** and **f)** Temperature profiles of ΔG_{unf} and Cp_{unf} (*in silico* vs. experimental), respectively, obtained for Ile3Glu at pH 2.4. Inset in **e** similar to that in **b**. The temperature profile of χ_i is included as an inset in **f**. The color-coding is indicated in the legends of the panels.



Extended Data Figure 4. Simplified MD-based scheme and comparison with experimental results for pseudo-lysozyme. a) protein models and folding states, number of structures (unfolded) and replicas (folded) simulated, diameter cutoff used to filter too-elongated unfolded structures obtained from ProtSA[9] (left), and temperatures selected for the MD-based calculation (Charmm22-CMAP) of pseudo-WT lysozyme thermodynamics. **b-d)** Temperature profiles of ΔG_{unf}, Cp_{unf} and χ_i (computed vs. experimental), respectively, obtained for pseudo-WT lysozyme at pH 3.0. Inset in **b** depicts the ΔH_{unf} vs. T linear plot with the fitted equation (the slope being ΔCp_{unf}) obtained from the MD simulations. **e** and **f)** Temperature profiles of ΔG_{unf} and Cp_{unf} (computed vs. experimental), respectively, obtained for pseudo-WT lysozyme at pH 3.7. Inset in **e** similar to that in **b**. The temperature profile of χ_i is included as an inset in **f**. The color-coding is indicated in the legends of the panels.

ASSOCIATED CONTENT

Supplementary Information

The Supplementary Information is available as a PDF document with the following sections:

SUPPLEMENTARY INFORMATION METHODS

Simulated Solvating Conditions.

Calculation of thermal stability (ΔGunf) of a three-state holoprotein (derivation of a Gibbs-Helmholtz-like equation).

SUPPLEMENTARY INFORMATION TABLES

Supplementary Information Tables 1-4.

SUPPLEMENTARY INFORMATION FIGURE

Supplementary Information Figure 1.

SUPPLEMENTARY INFORMATION REFERENCES

Author Information

ORCID

Juan J. Galano-Frutos: 0000-0002-1896-7805

Javier Sancho: 0000-0002-2879-9200

Corresponding Author

Javier Sancho, E-mail: jsancho@unizar.es, Tel: +34 976 761 286.

Author Contributions

J.S. conceived and directed the investigation. J.J.G-F. and F.N-F. carried out and analysed the Molecular Dynamics simulations. J.J.G-F. and J.S. analysed data and wrote the manuscript.

Funding Sources

This work was supported by grants PID2019-107293GB-I00 and PDC2021-121341-I00 (MICINN, Spain) and E45_20R (Gobierno de Aragón, Spain).

Notes

The authors declare no conflict of interest.

Acknowledgements

We thank the Biocomputation and Complex Systems Physics Institute (BIFI) of the University of Zaragoza and the Red Española de Supercomputación (RES) for computing facilities granted to perform Molecular Dynamics simulations.

Abbreviations

DSC, Differential Scanning Calorimetry; FMN, Flavin Mononucleotide; IS, Ionic Strength; LEM, Linear Extrapolation Method; MD, Molecular Dynamics; NMR, Nuclear Magnetic Resonance; NPT, Isothermal-isobaric ensemble in MD simulations; NVT, Canonical ensemble in MD simulations; PBC, Periodic Boundary Conditions; PME, Particle Mesh Ewald; R_g, Radius of gyration.