Biochimica et Biophysica Acta 1850 (2015) 983–995

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbagen

The key to predicting the stability of protein mutants lies in an accurate description and proper configurational sampling of the folded and denatured states ☆

Andreas P. Eichenberger [a], Wilfred F. van Gunsteren [a,*], Sereina Riniker [a], Lukas von Ziegler [a], Niels Hansen [a,b]

[a] Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, CH-8093 Zürich, Switzerland
[b] Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, D-70569 Stuttgart, Germany

ARTICLE INFO

Article history:
Received 16 June 2014
Received in revised form 3 September 2014
Accepted 5 September 2014
Available online 18 September 2014

Keywords:
Protein stability
Molecular dynamics
Free energy
Hydrogen bonds

ABSTRACT

Background: The contribution of particular hydrogen bonds to the stability of a protein fold can be investigated experimentally as well as computationally by the construction of protein mutants which lack particular hydrogen-bond donors or acceptors with a subsequent determination of their structural stability. However, the comparison of experimental data with computational results is not straightforward. One of the difficulties is related to the representation of the unfolded state conformation.
Methods: A series of molecular dynamics simulations of the 34-residue WW domain of protein Pin1 and 20 amide-to-ester mutants started from the X-ray crystal structure and the NMR solution structure are analysed in terms of backbone–backbone hydrogen bonding and differences in free enthalpies of folding in order to provide a structural interpretation of the experimental data available.
Results: The contribution of the different β-sheet hydrogen bonds to the relative stability of the mutants with respect to wild type cannot be directly inferred from experimental thermal denaturation temperatures or free enthalpies of chaotrope denaturation for the different mutants, because some β-sheet hydrogen bonds show sizeable variation in occurrence between the different mutants.
Conclusions: A proper representation of unfolded state conformations appears to be essential for an adequate description of relative stabilities of protein mutants.
General significance: The simulations may be used to link the structural Boltzmann ensembles to relative free enthalpies of folding between mutants and wild-type protein and show that unfolded conformations have to be treated with a sufficient level of detail in free energy calculations of protein stability. This article is part of a Special Issue entitled Recent developments of molecular dynamics.

1. Introduction

The prediction of the stable, dominant fold of a protein in aqueous solution at a particular thermodynamic state point of interest from its amino-acid sequence is still one of the long-standing challenges in biochemistry [1,2]. Although one can determine stable folded conformations by X-ray diffraction in a crystalline state or by NMR spectroscopy in solution, these stable conformations tell little about the driving forces of protein folding or about the types of interactions and motions that stabilise a particular fold under particular thermodynamic conditions of temperature, pressure, pH, ionic strength or solvent composition. Different interactions such as electrostatic and van der Waals interactions and entropy differences between folded and unfolded conformations may contribute to different extents, depending on the amino-acid composition of the protein and the composition of the solvent. For example, in aqueous solution the hydrophobic effect may drive apolar side chains out of the aqueous phase, or the excellent hydrogen-bond donor and acceptor properties of water may stabilise unfolded conformations which allow for solute–solvent hydrogen bonding relative to folded conformations in which unpaired hydrogen-bond donors and acceptors are present in the interior of a folded conformation. Although solute–solute hydrogen bonding is unlikely to drive protein folding because of the small energetic differences between solute–solute, solute–solvent and solvent–solvent hydrogen bonding for a protein in water, the existence of unpaired, i.e. not hydrogen-bonded donors and acceptors in a particular folded structure of a protein will generally destabilise such a structure compared to unfolded, solvent-exposed structures. The contribution of particular hydrogen bonds to the stability of a protein fold can be investigated by the construction of protein mutants which lack particular hydrogen-bond donors or acceptors with a subsequent determination of the stability of the fold adopted by the mutant. Such an investigation can be carried

☆ This article is part of a Special Issue entitled Recent developments of molecular dynamics.

* Corresponding author. Tel.: +41 44 632 5501; fax: +41 44 632 1039.
E-mail address: wfvgn@igc.phys.chem.ethz.ch (W.F. van Gunsteren).

http://dx.doi.org/10.1016/j.bbagen.2014.09.014
0304-4165/

out experimentally as well as computationally, and a combination of these two approaches may lead to an interpretation of the experimental data in terms of the importance of particular hydrogen bonds and local structures to the stability of the protein fold.

One way to remove a hydrogen-bond donor from the backbone of a protein is to replace an α-amino-acid residue with the corresponding α-hydroxy-acid residue, an amide-to-ester backbone substitution [3]. Such a substitution does not involve the side chain and preserves the stereochemistry of the residue. It replaces an amide NH, a hydrogen-bond donor, with an ester O, a rather weak hydrogen-bond acceptor, and an amide carbonyl, a rather strong hydrogen-bond acceptor, with an ester carbonyl, a weaker hydrogen-bond acceptor. Such amide-to-ester substitutions have been investigated experimentally [4–6] and computationally [7,8] for different peptides and proteins.

In a systematic experimental study of the effect of solute–solute hydrogen bonding upon the stability of the three-stranded β-sheet fold of the 34-residue WW domain of the protein Pin1 (Fig. 1) 20 amide-to-ester substitutions were carried out and the stability of the resulting mutants was measured in two ways, with respect to thermal denaturation and with respect to chaotrope denaturation [9]. Since the mutants need not necessarily adopt the same dominant folded structure as the wild-type WW domain, four experimental techniques, far-UV CD, fluorescence and 1D ^{1}H NMR spectroscopy, and a ligand-binding assay, were used to establish the similarity between the stable folds of the WW domain and its 20 mutants [9]. Four mutants, R14ρ, F25ϕ, N26ν, and Q33θ, where the Greek characters are used to denote ester-linked residues [9], appeared to adopt structures different from the wild-type fold. They could only appear properly folded upon addition of trimethylamine N-oxide (TMAO), an osmolyte that can stabilise protein folds [10,11]. The thermal denaturation experiments yielded midpoint temperatures T_m for 16 mutants and the wild-type domain, while chaotrope denaturation experiments yielded free enthalpies of chaotrope denaturation, using guanidine hydrochloride and TMAO, for 20 mutants and the wild type [9].

The processes of thermal or chaotrope denaturation of a protein are as yet inaccessible to computational methods, because the folding equilibrium of proteins cannot sufficiently be sampled in computer simulations covering 0.1 to 1 μs. Even MD simulations of 100 μs yield only a handful of (un)folding transitions [12], which is insufficient to obtain reliable values for the free enthalpy of folding based on the relative occurrence of folded versus unfolded configurations in the MD trajectory. Using current computer power much better (un)folding statistics can be obtained for short poly-peptides [13,14]. However, the relative stability of a particular fold for a protein and a mutant can be calculated through the use of the so-called free-energy perturbation techniques in combination with a thermodynamic cycle linking the folded and unfolded conformations of the protein (Fig. 2). The free enthalpy differences

$$\Delta G_x^{fu} = G_x^f - G_x^u \tag{1}$$

between the folded (f) conformation and the unfolded (u) conformations for a mutant (x = m) and the wild-type (x = w) protein can be expressed in terms of free enthalpy differences

$$\Delta G_{mw}^{y} = G_m^y - G_w^y \tag{2}$$

between the mutant (m) and the wild-type (w) protein in the folded conformation (y = f) and the unfolded conformations (y = u),

$$\Delta\Delta G_{mw}^{fu} = \Delta G_m^{fu} - \Delta G_w^{fu} = \Delta G_{mw}^{f} - \Delta G_{mw}^{u}, \tag{3}$$

because the free enthalpy [15] or Gibbs energy [16] is a thermodynamic state function. As mentioned above, the quantities ΔG_x^{fu} are as yet computationally inaccessible for proteins, but the quantities ΔG_{mw}^{y} can be calculated using molecular simulation and free-energy perturbation or other techniques to compute free-energy differences between slightly different systems [2,17,18]. However, this approach has several

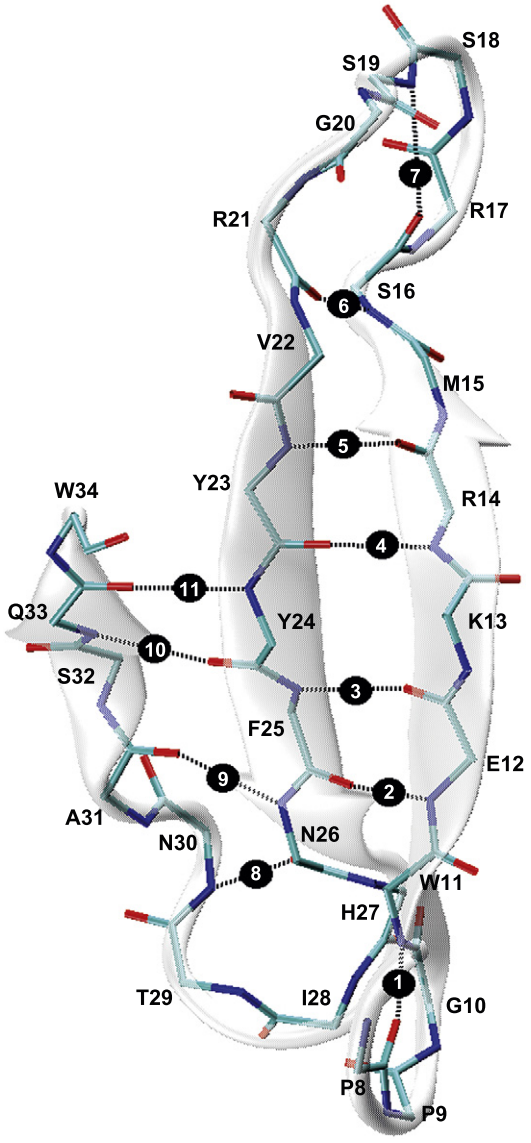

Fig. 1. Position of the 11 backbone–backbone hydrogen bonds within the β-sheet region of the Pin1 WW domain (crystal structure [32]). The sequence numbers 1–11 of the hydrogen bonds are used in Table 1.

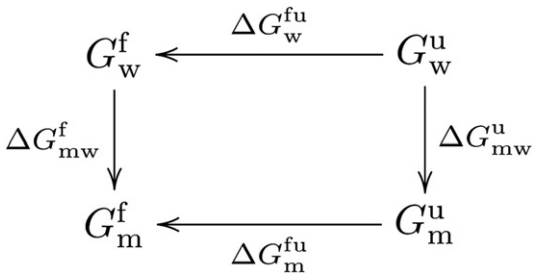

Fig. 2. Thermodynamic cycle for the Pin1 WW domain changing from unfolded (u) to folded (f) and wild-type (w) to mutant (m) states. The computation of the two free enthalpies changing the protein from the wild-type to the mutant state, ΔG_{mw}^{f} and ΔG_{mw}^{u}, is feasible using EDS, while the two free enthalpy differences involving (un)folding of the protein cannot be estimated directly in silico. The difference in folding free enthalpy differences, $\Delta\Delta G_{mw}^{fu}$, can be obtained using Eq. (3).

pitfalls that may render seemingly correct values for the relative free enthalpy of folding unreliable [2,19,20]. For example, short sampling times, e.g. 200 ps [21] may lead to insufficient sampling of the configurational space accessible to the protein. Sampling can be considerably enhanced by applying enhanced sampling techniques [18], of which enveloping distribution sampling (EDS) [22,23] has been shown to be effective when applied to peptide folding [24–26] and protein mutations [27]. In addition, EDS offers the opportunity to obtain an estimate of the extent of adequate sampling underlying the calculated relative free enthalpy values from monitoring different indicators [28], see Section 2.5.

A comparison of calculated $\Delta\Delta G^{fu}_{mw}$ values with differences in midpoint temperatures T^m between mutants (m) and wild-type (w) protein,

$$\Delta T^{m}_{mw} = T^{m}_{m} - T^{m}_{w}, \tag{4}$$

or with differences in chaotrope renaturation free enthalpies,

$$\Delta\Delta G^{c}_{mw} = \Delta G^{c}_{m} - \Delta G^{c}_{w}, \tag{5}$$

is not straightforward because Eq. (3) contains free enthalpy differences ΔG^{fu}_{x} between folded and unfolded conformations at one thermodynamic state point, whereas Eqs. (4) and (5) represent or contain free enthalpy differences of a system at different thermodynamic state points, differing in temperature in Eq. (4) or differing in solvent composition in Eq. (5) [29]. So, although all three quantities $\Delta\Delta G^{fu}_{mw}$, ΔT^{m}_{mw} and $\Delta\Delta G^{c}_{mw}$ in one way or the other reflect the relative stability of the folded form of a mutant compared to wild-type, their values need not show strong correlation. This is illustrated in Fig. 3 where the 16 measured ΔT^{m}_{mw} values of [9] are correlated with the corresponding measured $\Delta\Delta G^{c}_{mw}$ values of Table 2 of [9]. The correlation coefficient is 0.87.

In the present study we report the calculation of 20 $\Delta\Delta G^{fu}_{mw}$ values for the 20 mutants of the Pin1 WW domain reported in [9]. To obtain ΔG^{y}_{mw} values the method of enveloping distribution sampling (EDS) [22,23,30] is used. It yields a free enthalpy difference, ΔG^{f}_{mw} between two proteins, in this case an amide-to-ester mutant of the WW domain and the wild-type protein both in their folded, if stable, conformations. If the unfolded conformations of the wild-type and mutant proteins are to a large extent water exposed, the ΔG^{u}_{mw} values for the 20 mutants may be very similar, because an amide-to-ester mutation does not involve an amino-acid side chain. Since it is impossible to sufficiently sample the unfolded state conformations of a 34-residue protein, we approximated ΔG^{u}_{mw} in four different ways:

i) The free enthalpy difference associated with an amide-to-ester mutation in the central residue of an Ala-tripeptide in water. This approximation yields a single ΔG^{u}_{mw} value for all 20 mutants.

ii) This single ΔG^{u}_{mw} value was treated as a parameter ΔG^{u}_{0} that can be obtained from a fit of 20 calculated and experimental $\Delta\Delta G_{mw}$ values,

$$\Delta\Delta G^{fu}_{mw} = \Delta G^{f}_{mw} - \Delta G^{u}_{0}. \tag{6}$$

iii) The free enthalpy ΔG^{u}_{mw} between the mutant and wild-type proteins in the unfolded conformations is assumed to be linearly related to the corresponding free enthalpy difference ΔG^{f}_{mw} in the folded conformations [31],

$$\Delta G^{u}_{mw} = \alpha\Delta G^{f}_{mw} + \Delta G^{u}_{1}, \tag{7}$$

which yields

$$\Delta\Delta G^{fu}_{mw} = (1-\alpha)\Delta G^{f}_{mw} - \Delta G^{u}_{1}. \tag{8}$$

In this case two parameters, α and ΔG^{u}_{1}, are to be obtained from a fit of 20 calculated and experimental $\Delta\Delta G_{mw}$ values.

iv) Because none of the approximations of ΔG^{u}_{mw} yielded a good correlation of $\Delta\Delta G^{fu}_{mw}$ with the measured ΔT^{m}_{mw} or $\Delta\Delta G^{c}_{mv}$ values, we approximated ΔG^{u}_{mw} more precisely with an amide-to-ester mutation in the central residue of a tripeptide in water consisting of the residues of the protein at the mutation site. In this way, the effect of side chains of the mutated residue and its neighbours along the polypeptide chain when exposed to aqueous solution is taken into account. This more detailed treatment of the unfolded conformations required separate ΔG^{u}_{mw} calculations for 20 tripeptides.

There exists not only uncertainty about the unfolded conformational ensemble, but also about the folded one of a protein in solution at a particular thermodynamic state point. Fig. 4 shows that the crystal structure (PDB code 1PIN) of the WW domain derived from X-ray diffraction data of the whole Pin1 protein [32] and the structure of the WW domain in aqueous solution (PDB code 2KCF) derived from NMR spectroscopic data [33] are not identical. This is not surprising because the X-ray diffraction data (2.05 Å resolution) were obtained at 277 K and pH = 7.5 from a solution containing ammonium sulfate, whereas the NMR spectroscopic data obtained at 278 K and pH = 6.2 from a solution containing sodium phosphate resulted in 311 medium and long-range NOE distance bounds for the 34-residue protein. For both structures the ratio of observables and degrees of freedom leaves ample room for conformational variability. The C_α backbone atom-positional root-mean-square difference (*RMSD*) and the all-atom *RMSD* between structures 1PIN and 2KCF are 0.34 nm and 0.42 nm, respectively. Therefore, we conducted molecular dynamics (MD) simulations of the wild type and a few mutant proteins, and the EDS simulations of the 20 mutants in the folded state starting from each of these structures. A comparison of the results of both sets of simulations offers the opportunity to obtain an impression of the importance of structural relaxation in regard to the calculated free enthalpy values.

The four experimental techniques used to determine whether the mutants would adopt a similar fold as the wild-type protein, i.e. CD, fluorescence and 1D ^{1}H NMR spectroscopy and ligand binding, do not yield precise structural information on the mutants because the four signals measured by these techniques are only sensitive to some average structural properties of the mutant. Thus, local structural

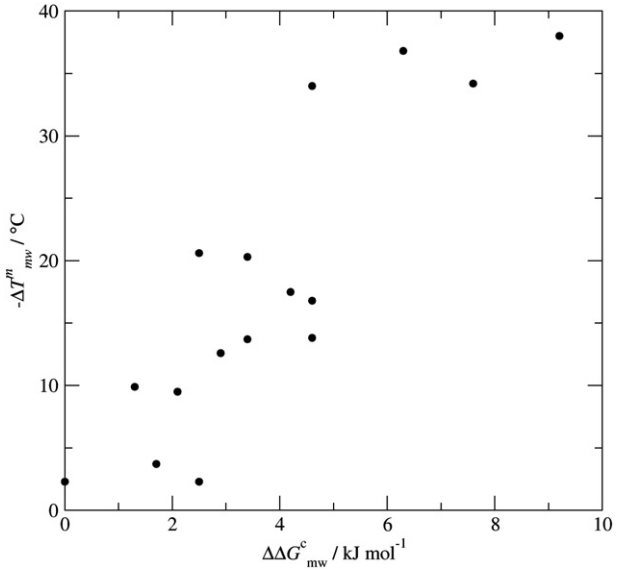

Fig. 3. Differences ΔT^{m}_{mw} between midpoints T^m of thermal denaturation curves, versus differences $\Delta\Delta G^{c}_{mw}$ of folding free energy differences as obtained by chaotrope denaturation experiments [9] for the twenty mutants (m) and the wild type (w) Pin1 WW domain (Table 1).

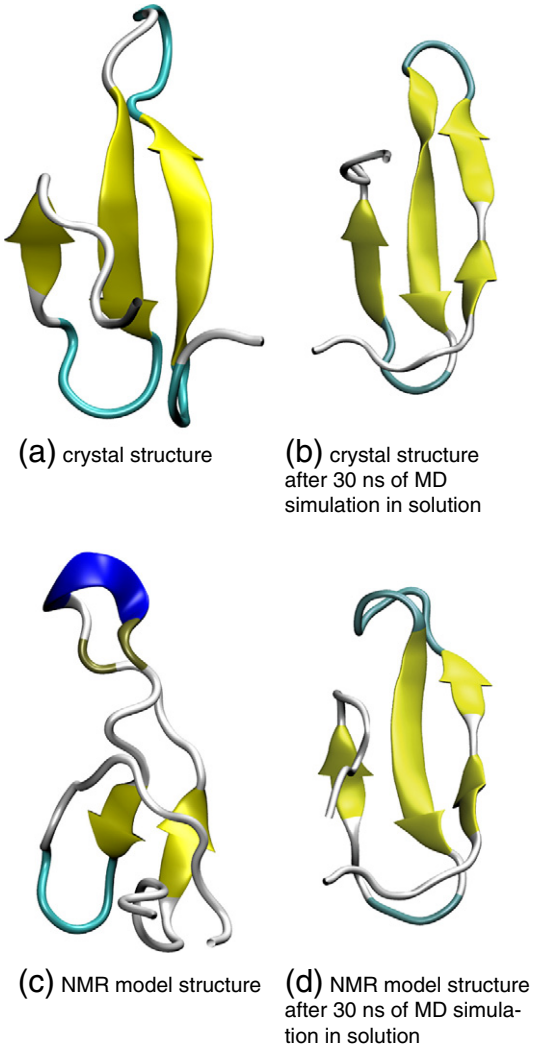

Fig. 4. Pin1 WW configuration of a) the crystal structure (PDB entry code 1PIN [32]) and b) the final configuration of a 30 ns MD simulation with initial configuration based on the crystal structure as well as c) the NMR model structure (PDB entry code 2KCF [33]) and d) the final configuration after 30 ns of MD simulation starting from the NMR model structure.

differences, in hydrogen bonding, may not be reflected in different signals. Therefore, the structural differences between the different mutants and the wild-type protein as they occur in the simulations of these proteins were analysed. They may also contribute to differences in $\Delta\Delta G^{fu}_{mw}$ values as obtained from the EDS simulations of the 20 mutants.

2. Computational methods

2.1. Simulation program and force field parameters

All simulations of the wild-type or mutant proteins, including the EDS simulations, have been carried out using the GROMOS simulation software [30,34,35] and the 53A6 force-field parameter set [36] for the wild type proteins, while for the ester linkages in the mutant proteins the parameter set $53A6_{OXY}$ [37] was used. Water was modelled using the simple point charge (SPC) model [38].

All analyses of the simulation trajectories have been done using the tools and programs contained in GROMOS++ [39] which is part of the GROMOS simulation software package [40].

2.2. Simulation protocol

Sixteen standard MD simulations were performed, namely for the wild-type PIN WW domain and for the mutants L7λ, R13ρ, F25ϕ, N26ν, N30ν and Q33θ starting from the X-ray (PDB code 1PIN [32]) and from the NMR model structure (PDB code 2KCF [33]) as well as for the Ala-tripeptide in its native and perturbed form, i.e. the central residue being an α-hydroxy amino acid. In addition, 81 EDS simulations were performed to obtain the quantities ΔG^{x}_{mw} for each of the 20 mutants (x = f) and their tripeptide analogues ($x = u$) starting from the X-ray and the NMR model structures and for the Ala-tripeptide (x = u). In this work, we name the simulations either 1PIN or 2KCF simulations, indicating which PDB structure was used to generate the initial positions of the solute atoms. The initial configuration of the Ala-tripeptide had −133° and 155° as values for the ϕ- and ψ-dihedral angles. The initial configurations of the 20 tripeptides were taken from the 1PIN and the 2KCF X-ray and NMR model structures.

2.2.1. Initial structures and MD end-state simulations

Initial coordinates of the protein atoms were derived from the PDB structures mentioned above. Missing hydrogen atom positions were generated based on geometric criteria using the GROMOS++ program gch. The protonation states of the amino acid residues were chosen corresponding to pH 7, leading to a total protein charge of +4e. Histidine 27 was protonated according to its hydrogen-bonding environment, based on the atom positions defined in the X-ray or NMR model structure: N_ε for the 1PIN simulations and N_δ for the 2KCF simulations. The different WW domain structures were energy minimised in vacuo and solvated in cubic boxes of water [38] molecules, such that the minimum distance of a non-hydrogen protein atom to the box wall was 1.4 nm. This led to box-edge lengths of 6.0/6.3 nm and 6876/7839 water molecules (1PIN/2KCF simulations). To yield an overall charge neutrality of the periodic boxes, four water molecules were replaced by chloride anions using the GROMOS++ program ion. The solvent was energy minimised with the positions of the heavy atoms within the protein kept fixed. Initial velocities for the simulations starting from the energy minimised PDB structures were generated based on a Maxwell–Boltzmann distribution at 1 K with position restraining of the protein atoms with an initial harmonic force constant of 25,000 kJ mol^{-1} nm^{-2}. The temperature was continuously raised up to 300 K, while the force constant of the position restraining was lowered, ending in a non-restrained protein simulation at 300 K after a simulation time of 1 ns. This procedure was followed by another 1 ns simulation at constant volume and 300 K and a 1 ns simulation at 300 K and at a constant pressure of 1 atm, using the weak coupling algorithm [41] with corresponding coupling times of $\tau_T = 0.1$ ps and $\tau_p = 0.5$ ps, respectively. The protein and solvent were separately coupled to the heat bath and an estimated isothermal compressibility of $4.575 \cdot 10^{-4}$ (kJ mol^{-1} nm^{-3})$^{-1}$ was used. All bond lengths and the bond angles of the water molecules were kept rigid by applying the SHAKE algorithm [42] with a relative geometric tolerance of 10^{-4}, allowing for an integration time step of 2 fs when solving Newton's equations of motion using the leap-frog algorithm [43]. Nonbonded interactions (van der Waals and electrostatic) were handled adopting triple-range cut-off radii: forces between atoms within the short-range cutoff of 0.8 nm were calculated every time step from a pair list that was generated every five steps, when also forces between atoms at a distance between 0.8 nm and 1.4 nm were computed and kept constant during the following four time steps. The long-range electrostatic forces were represented by a reaction field with a relative permittivity [44] of $\varepsilon_{RF} = 61$ outside the long-range cutoff of 1.4 nm. The centre of mass translation and rotation of the system were removed every 2 ps to avoid a flying ice cube [45].

The simulations were continued for at least 20 ns at constant temperature and pressure, using the same simulation parameters as

described above. The protein atom positions, atom velocities and system energies were saved every 5 ps for analysis.

The ester and amide end-state simulations of the Ala-tripeptide and the 20 tripeptides were performed using the same simulation procedure and parameters as described for the Pin1 WW mutants.

2.3. EDS reference-state simulations

A hybrid topology was used for the EDS simulations of the mutant WW domains, the Ala-tripeptide and the 20 tripeptides, bearing both states, ester and amide, in the so-called reference state. The most simple approach would have been to only perturb the peptidic N atom into an ester O atom and the backbone H atom into a dummy atom, i.e. a single topology approach [46], since only non-bonded solute–solute and solute–solvent energies have been used to compute the free enthalpies. However, differences in the bonded interaction parameters (force constants, bond lengths, bond angles, torsional dihedral angles and multiplicities) lead to small changes in flexibility and to different atom configurations of the ester linkage compared to the peptide linkage and are therefore not negligible. To account for these changes the topology was branched after the C_α atom of the previous residue and (re)unified at the carbonyl C atom, see Fig. S2.

The initial coordinates for the EDS simulations were generated from the final configurations of the corresponding (1PIN or 2KCF) wild-type or tripeptide MD equilibration, see above, by copying the necessary number of atom coordinates and velocities to cover the doubly branched reference state. To find the EDS smoothness parameter s and energy offset parameter ΔE^R a standard update scheme [28] was used and in some cases further optimised manually. The resulting parameters for the different mutant EDS simulations are listed in Table S1.

After the update procedure an EDS production simulation of 20 ns was performed with constant EDS parameters (Table S1). To optimise the sampling the EDS averages were based on these continuations of the EDS update simulations with frozen EDS parameters and the previously generated EDS configurations from the EDS update simulations. Energies were saved every 0.2 ps for further analysis and the computation of $\Delta\Delta G_{mw}$.

2.4. Computation of free-enthalpy differences

The free enthalpy [15] or Gibbs energy [16] of a system in the isothermal–isobaric ensemble (constant number of particles, pressure and temperature) is

$$G = -\frac{1}{\beta}\ln\ Z, \tag{9}$$

where β is the inverse of the temperature T multiplied by the Boltzmann constant k_B and Z denotes the partition function of the system. The difference in free enthalpy between two states B and A of a system can therefore be calculated as

$$\Delta G_{BA} = G_B - G_A = -\frac{1}{\beta}\ln\frac{Z_B}{Z_A}. \tag{10}$$

Instead of perturbing the system in its state A directly to its state B, a non-physical intermediate state R, a so-called reference state, can be defined, which allows the computation of the free enthalpy difference

$$\Delta G_{BA} = G_{BR} - G_{AR} = -\frac{1}{\beta}\ln\frac{Z_B}{Z_A}\frac{Z_R}{Z_R} = -\frac{1}{\beta}\ln\frac{\left\langle e^{-\beta(V_B - V_R)}\right\rangle_R}{\left\langle e^{-\beta(V_A - V_R)}\right\rangle_R} \tag{11}$$

from a single simulation of this reference state R, where $\langle \ldots \rangle_X$ denotes an ensemble average of state X. In enveloping distribution sampling (EDS) this reference state is chosen such that its configurational ensemble envelopes all important configurations of states A and B. The potential energy term of the Hamiltonian of the reference state R reads [18]

$$V_R\left(\mathbf{r}^N; s, \Delta E^R_{BA}\right) = -\frac{1}{\beta s}\ln\left[e^{-\beta s\left(V_A(\mathbf{r}^N) - E^R_A\right)} + e^{-\beta s\left(V_B(\mathbf{r}^N) - E^R_B\right)}\right] \tag{12}$$

where $\mathbf{r}^N$ denotes a configuration of the N particles in the system and V_X is the potential energy of state X. Since the equations of motion depend on the energy offset difference $\Delta E^R_{BA} = E^R_B - E^R_A$ only, and not on these quantities, E^R_A is set to zero for a two-state system. Therefore, only two parameters, the smoothness parameter s and the difference in energy offset $\Delta E^R_{BA} = E^R$, have to be determined such as to ensure sufficient sampling of both states A and B during the EDS MD simulation of the reference state R [23].

A computation of the free enthalpy difference is also possible in terms of energy difference distributions $\rho_A(\Delta V; \Delta V_{BA})$ and $\rho_B(\Delta V; \Delta V_{BA})$ sampled in the A and B ensembles, respectively [47–49],

$$\ln\frac{\rho_A(\Delta V; \Delta V_{BA})}{\rho_B(\Delta V; \Delta V_{BA})} = +\beta\Delta V - \beta\Delta G_{BA}, \tag{13}$$

with $\Delta V_{BA} = V_B - V_A$. The energy difference distributions ρ_A and ρ_B are not directly accessible from the EDS simulation of the reference state R,

$$\rho_R(\Delta V; \Delta V_{BA}) = \langle\delta[\Delta V - \Delta V_{BA}]\rangle_R, \tag{14}$$

but can be obtained by reweighting this distribution to the A and B ensembles [50,51],

$$\rho_X(\Delta V; \Delta V_{BA}) = \frac{\left\langle\delta[\Delta V - \Delta V_{BA}]e^{-\beta(V_X - V_R)}\right\rangle_R}{\left\langle e^{-\beta(V_X - V_R)}\right\rangle_R}, \tag{15}$$

where X denotes either of the two states A or B.

2.5. How to detect inadequate configurational sampling?

Using standard techniques to compute free enthalpy differences such as thermodynamic integration or perturbation, it is difficult to impossible to obtain a realistic estimate of the uncertainty in the calculated ΔG values due to inadequate sampling of the configurational space [2,19,20]. In contrast, the EDS technique to compute free enthalpy differences offers various indicators of inadequate sampling [28]:

1. A small value for the root-mean-square fluctuation of the energy difference $V_{BA}(t)$ between states A and B in an EDS simulation indicates undersampling.
2. A low number of occurrences or population of states A and B in an EDS simulation leads to insufficient statistics of the probabilities of these states in the ΔG calculation.
3. A low number of transitions between states A and B, even when both states show sizeable occurrences in an EDS simulation, leads to unreliable statistics of these occurrences and thus to unreliable ΔG values.

In addition to these quantities to be monitored in an EDS simulation to estimate the uncertainty in the obtained ΔG values, one may perform additional (standard, non-EDS) MD simulations of the end states A and B and compare the obtained energy distribution in states A and B to the corresponding distribution obtained by reweighting using Eq. (15) from the EDS simulation involving states A and B.

2.6. Fitting of the calculated $\Delta\Delta G^{fu}_{mw}$ values to experimental $\Delta\Delta G^{c}_{mw}$ values

The computation of the $\Delta\Delta G^{fu}_{mw}$ values according to Eqs. (6) and (8) was done by fitting these values obtained from the Pin1 WW mutant EDS simulations to the $\Delta\Delta G^{c}_{mw}$ values obtained from chaotrope

denaturation [9] using a least-squares fit varying one parameter ΔG_0^u when using Eq. (6) and two parameters α and ΔG_1^u when using Eq. (8). The fit was done for each computed data set separately, once for the $\Delta\Delta G_{mw}^{fu}$ values computed from the 1PIN EDS simulations and once for those computed from the 2KCF EDS simulations.

The parameters (Eqs. (6) and (8)) obtained from the fit are $\Delta G_0^u = 76.8$ kJ mol^{-1}, $\Delta G_1^u = -6.8$ kJ mol^{-1}, and $(1-\alpha) = -0.01$ for the fit using the $\Delta\Delta G_{mw}^{fu}$ values from the 1PIN EDS simulations, and $\Delta G_0^u = 75.8$ kJ mol^{-1}, $\Delta G_1^u = -7.1$ kJ mol^{-1}, and $(1-\alpha) = -0.01$ for the fit using the $\Delta\Delta G_{mw}^{fu}$ values from the 2KCF EDS simulations.

The corresponding plots $\Delta\Delta G_{mw}^{fu}$ versus $\Delta\Delta G_{mw}^{c}$ and $-\Delta T_{mw}^{m}$ are shown in Fig. 9.

2.7. Analysis

2.7.1. Atom-positional RMSD

The atom-positional root-mean-square deviation (*RMSD*) between two protein structures has been evaluated based on the C_α backbone atoms or all atoms according to the following formula,

$$RMSD\left(\mathbf{r}^{N_a}, \mathbf{r}_{ref}^{N_a}\right) = \sqrt{\frac{1}{N_a}\sum_{i=1}^{N_a}\left(\mathbf{r}_i - \mathbf{r}_{i,ref}\right)^2}, \quad (16)$$

where $\mathbf{r}^{N_a} = \left(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_a}\right)$ represents the positions of the atoms. In Eq. (16), N_a is the number of atoms considered, $\mathbf{r}_i$ the position of atom i in the first structure and $\mathbf{r}_{i,ref}$ the position of atom i in the second, reference structure. We used the X-ray structure [32] or NMR model structure [33] as reference structure if not mentioned differently. The *RMSD* at the beginning of the MD simulations is not zero because the X-ray and NMR model structures were energy minimised in vacuo prior to simulation. The calculation of *RMSD*s was done using the GROMOS++ program rmsd, which also performs a superposition of the centres of mass of the two structures to be compared as well as a rotational least-squares fit based on the C_α backbone atom positions.

2.7.2. Hydrogen bonds

Hydrogen bonds were defined according to a geometric criterion: a minimum donor–hydrogen-acceptor angle of 135° and a maximum hydrogen-acceptor distance of 0.25 nm. The hbond program of GROMOS++ [39] has been used to detect and monitor hydrogen bonds in the Pin1 WW domains.

The EDS simulation trajectories had to be transformed to a wild-type-like trajectory before analysing hydrogen bonds, since we are not interested in hydrogen bonds from the ester branch to the wild-type branch. This was done by filtering all ester branch atoms of the EDS simulation trajectory.

2.7.3. Detection of secondary structure elements

The rules of Kabsch and Sander [52] have been applied to detect and monitor secondary structure elements in the native and ester-linked HEWL simulations. In some cases one residue may be assigned to be part of two different secondary structure elements. In order to avoid ambiguous assignments in such cases, the following priority rules were applied: β-strand/β-bridge > α-helix > π-helix > 3_{10}-helix > hydrogen bonded turn > bend.

3. Results

3.1. Structures and stability of the wild-type WW domain

Two simulations of the wild-type Pin1 WW domain have been performed only differing in the initial configuration used. In one case, the initial protein atom positions were derived from an X-ray model structure (PDB code 1PIN [32]), while in the other simulation the initial configuration was based on a NMR model structure (PDB code 2KCF [33]). The two simulations are therefore referred to as 1PIN simulation or 2KCF simulation, respectively.

The X-ray model structure and NMR model structure clearly differ from each other as shown in Fig. 4 and mentioned in Section 1 (root-mean-square deviations (*RMSD*) of the C_α atoms and all atoms of 0.34 nm and 0.42 nm, respectively). Also the hydrogen-bond pattern for the backbone shows differences: Three hydrogen bonds present in the X-ray model structure are not present in the NMR model structure (Table 2). These differences may be due to the different thermodynamic state points, different environments of the protein and the procedures used to derive protein structure from measured values of observables such as X-ray diffraction intensities or spectroscopic absorption intensities [29]. Interestingly, the two protein configurations seem to converge to similar configurations in the MD simulations (Fig. 4, Table 2) but still show an *RMSD* of the C_α atoms of 0.29 nm between the two protein configurations after 30 ns of MD simulation.

Both, the 1PIN and 2KCF folds of the wild-type WW domain are stable during the MD simulations and the protein does not unfold as indicated by the positional *RMSD* of the C_α or all protein atoms as function of time in Fig. 5. Note that the *RMSD* during the first ns of the equilibration period is different from zero, although the protein atoms were positionally restrained to their initial positions. This is the result of the energy minimisation of the protein in vacuo which was performed before the initialisation of the MD simulations, compare Section 2.2.1. In the first half of the 30 ns 1PIN simulation the C_α atoms only deviate about 0.2 nm from the X-ray structure, while the deviation from the NMR model structure is about 0.35 nm. A configurational change within the 1PIN simulation after about 16 ns leads to a similar *RMSD* value of the 1PIN simulation from both PDB structures, ending up with a value of about 0.35 nm. The *RMSD* value in the 2KCF simulation with respect to the NMR model structure is rather constant around 0.3 nm and generally smaller, about 0.07 nm, than the deviation from the X-ray structure. As expected, the RMSD values of all protein atoms are somewhat larger (about 0.1 nm) than the corresponding deviations of the Cα atoms only.

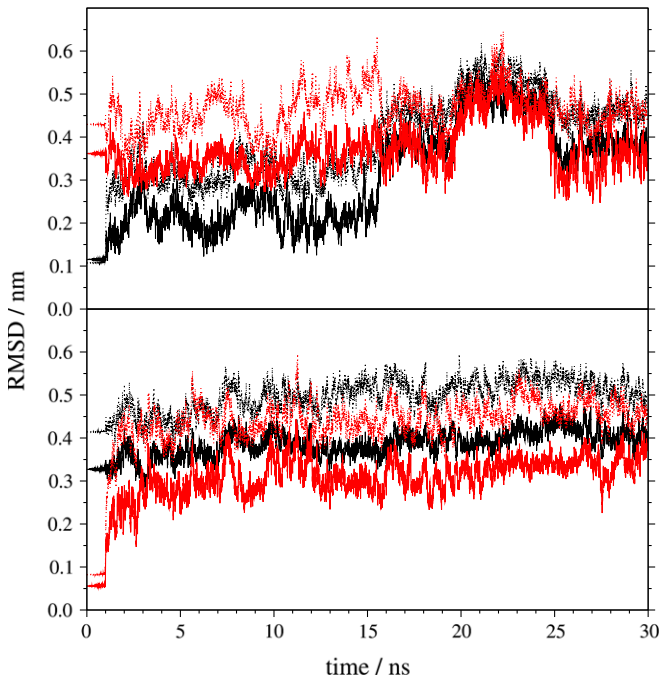

Fig. 5. Positional root-mean-square deviation of the backbone C_α atoms (solid lines) and all atoms (dotted lines) during the simulation of the native WW domain with respect to the crystal structure (black) [32] and the NMR model structure (red) [33]. Top graph: initial configuration of the MD simulation derived from the crystal structure [32]; bottom graph: initial configuration of the MD simulation derived from the NMR model structure [33].

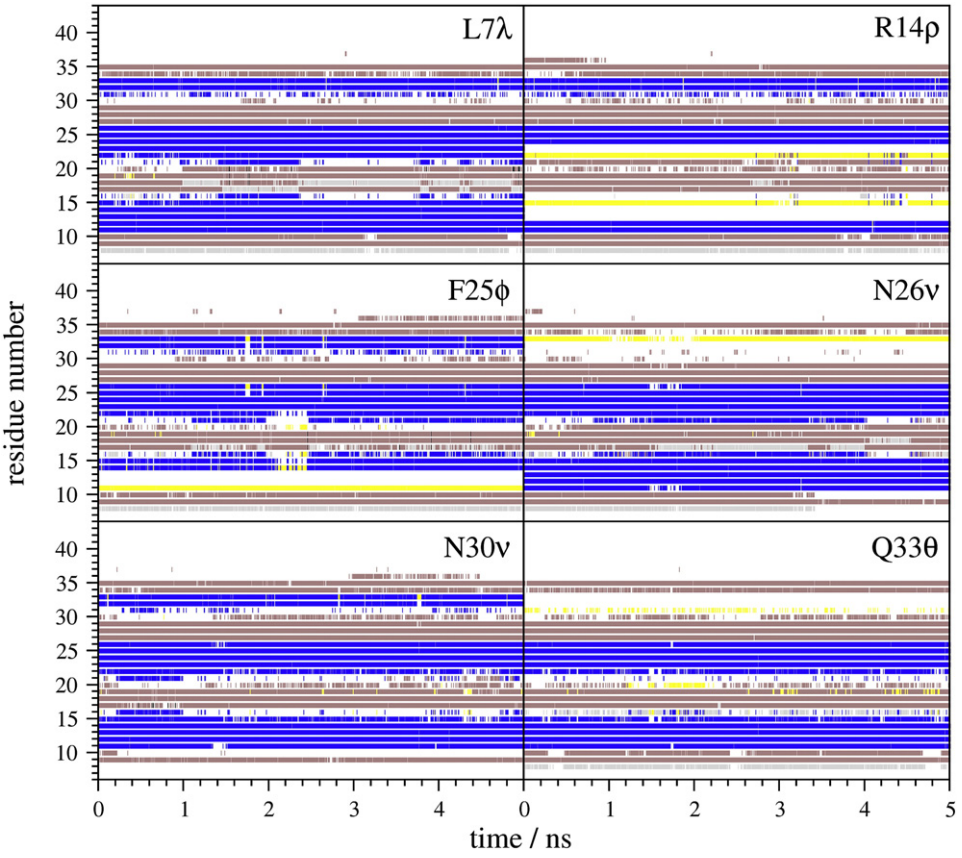

Fig. 6. Secondary structure of the six indicated mutant 1PIN end-state MD simulations starting from the 1PIN X-ray structure. Blue: β-strand; yellow: β-bridge; brown: bend. According to [52].

The evolution of secondary protein structure of the two wild-type WW domain MD simulations in time (Fig. S1) both show a triple-stranded β-sheet structure (residues 11–16, 21–26 and 31–33), but the 1PIN simulation maintains the structure slightly better for residues 16, 21 and 31. Residues 16 and 21 are both involved in the hydrogen bond No. 6 (Fig. 1). The β-sheet character of the protein is maintained in the simulations.

Fig. 6 shows the secondary structure of six selected mutant proteins (L7λ, R14ρ, F25ϕ, N26ν, N30ν and Q33θ) as a function of time. Only the two mutants L7λ and N30ν fully maintain the triple-stranded β-sheet as reported in the X-ray model structure [32] and seen in the end-state simulations of the wild-type WW domain (1PIN and 2KCF, see Fig. S1). The other four mutants either completely miss one β-strand (N26ν and Q33θ) or show a reduced β-sheet structure involving much fewer residues than in the wild-type protein (R14ρ and F25ϕ). This is in agreement with spectroscopic and ligand-binding experiments which indicated that the latter four mutants were misfolded [9]. No melting temperatures were reported in [9] for these mutants (see Table 1).

3.2. Hydrogen bonding

Table 2 shows the occurrence of the 11 backbone hydrogen bonds of Fig. 1 reported in [9] during the 1PIN and 2KCF MD simulations as well as within the corresponding X-ray or NMR model structures. Hydrogen bond Nos. 7 and 8 are not present in either of the experimentally derived protein structures and show a very low occurrence in both MD simulations. Hydrogen bond No. 6, which involves residues 16 and 21, and hydrogen bond No. 5, a close neighbour to hydrogen bond No. 6, show both a lower occurrence in the 2KCF simulation than in the 1PIN simulation of the wild-type, while hydrogen bond No. 1 is only moderately present in the 1PIN simulation and not present in the 2KCF simulation. Overall, only 7 out of the 11 backbone hydrogen bonds have an occurrence of at least 75% in both MD simulations, i.e. hydrogen bond Nos. 2, 3, 4, 5, 9, 10, and 11. In the NMR model structure there are 6 backbone hydrogen bonds present (Nos. 2, 3, 6, 9, 10, and 11), while the X-ray structure shows all but hydrogen bond Nos. 7 and 8.

We note that the hydrogen-bonding behaviour calculated from the wild-type branch of the topology of the hybrid mutant/wild-type residue (Fig. S2) in each of the 20 hybrid mutant/wild-type systems during the 1PIN and 2KCF EDS simulations (data not shown) should be and is similar to the hydrogen-bond pattern shown in the wild-type end-state simulations, indicating rather stable wild-type state structures of the WW domain during the mutant/wild-type EDS simulations.

3.3. Free enthalpy differences

Table 1 shows an overview of calculated differences in folding free enthalpies, $\Delta\Delta G^{fu}_{mw}$, as well as the experimentally derived chaotrope denaturation values $\Delta\Delta G^{c}_{mw}$ (Eq. (5)) together with the differences of midpoints of thermal denaturation curves ΔT^{m}_{mw} (Eq. (4)) reported in [9]. Calculated values have been obtained using Eqs. (3), (6) and (8).

As mentioned above, we calculated $\Delta G^{u}_{mw} = 39.8 \pm 0.4$ kJ mol^{-1} as the free enthalpy difference between an ester and an amide as the central residue of an Alatripeptide in water and used this value as an approximation of ΔG^{u}_{mw} in Eq. (3) for all (unfolded) mutants. The resulting $\Delta\Delta G^{fu}_{mw}$ values calculated from the EDS 1PIN and 2KCF simulation trajectories are reported in columns 7 and 8 of Table 1 and are generally much larger than the values (column 15) reported in the literature [9]. However, the $\Delta\Delta G^{fu}_{mw}$ values from the two individual EDS simulations (1PIN and 2KCF) are rather similar. Exceptions are the mutants R17ρ, S19σ, Y23ψ, Y24ψ and H27ν, which show differences larger than 2.5 kJ mol^{-1} (k_BT). These mutants can also be identified when looking at

Table 1
List of the twenty mutated Pin1 WW proteins indicating the eliminated and weakened backbone hydrogen bonds (compare Fig. 1), separated by a "/", in the EDS simulations based on the crystal (1PIN) [32] and NMR (2KCF) model [33] initial structures, the difference in experimentally derived midpoint of thermal denaturation curves [9] between the mutant and the native state, ΔT^{m}_{mw} Eq. (4), as well as experimentally ($\Delta\Delta G^{c}_{mw}$, Eq. (5)) [9] and computed ($\Delta\Delta G^{fu}_{mw}$, Eq. (3)) differences of folding free energy differences between mutant and native Pin1 WW domains.

System	Affected H-bond	Calculated ΔG_{mw} values [kJ mol^{-1}]				Calculated $\Delta\Delta G^{fu}_{mw}$ [kJ mol^{-1}]								Exp. $\Delta\Delta G^{c}_{mw}$	$-\Delta T^{m}_{mw}$
		ΔG^{f}_{mw}		ΔG^{u}_{mw}		From tri-Ala (Eq. (3))		From Eq. (6)		From Eq. (8)		From tripeptides		[kJ mol^{-1}]	[°C]
		1PIN	2KCF	1PIN	2KCF	1PIN	2KCF	1PIN	2KCF	1PIN	2KCF	1PIN	2KCF		
L7λ	–/–	−0.5	−0.3	38.7	35.8	−40.3	−40.1	−77.3	−76.1	6.8	7.1	−39.2	−36.1	0.0 ± 0.0	2.3
W11ω	1/–	113.5	112.9	108.3	109.9	73.7	73.1	36.7	37.1	6.2	6.2	5.2	3.0	4.6 ± 0.4	13.8
E12ε	2/–	92.3	89.9	75.6	76.2	52.5	50.1	15.5	14.1	6.4	6.3	16.7	13.7	6.3 ± 0.5	36.8
K13κ	–/3	84.1	85.9	80.0	80.6	44.3	46.1	7.3	10.1	6.4	6.4	4.1	5.3	2.9 ± 0.5	12.6
R14ρ	4/–	97.5[a,b]	95.9[a,c]	80.6[a,d]	80.3[a,e]	57.7	56.1	20.67	20.1	6.3	6.3	16.9	15.6	16.4 ± 0.4	n.d.
M15μ	–/5	48.6	50.6	45.5	44.3	8.8	10.8	−28.2	−25.6	6.6	6.7	3.1	6.3	1.7 ± 0.4	3.7
S16σ	6/–	146.1	144.2	142.4	142.1	106.3	104.4	69.3	68.4	6.1	5.9	3.7	2.1	4.6 ± 0.4	16.8
R17ρ	–/7	52.1	56.7	53.5	58.4	12.3	16.9	−24.7	−19.1	6.6	6.6	−1.4	−1.7	1.3 ± 0.3	9.9
S19σ	7/–	122.9	113.0	117.5	115.9	83.1	73.2	46.1	37.2	6.2	6.2	5.4	−2.9	2.5 ± 0.4	20.6
V22ϖ	–/6	81.7	82.3	78.3	78.9	41.9	42.5	4.9	6.6	6.4	6.4	3.4	3.4	2.5 ± 0.5	2.3
Y23ψ	5/–	110.1	85.9	81.0	76.0	70.3	46.1	33.3	10.1	6.3	6.4	29.1	9.9	9.2 ± 0.4	38
Y24ψ	11/4	93.9	90.9	79.5	82.9	54.1	51.1	17.1	15.1	6.3	6.3	14.4	8.0	4.6 ± 0.3	34
F25ϕ	3/10	90.1	91.4	77.1	82.4	50.3	51.6	13.3	15.6	6.4	6.3	13.0	9.0	17.6 ± 0.4	n.d.
N26ν	9/2	43.4	44.4	32.5	38.1	3.6	4.6	−33.4	−31.4	6.6	6.7	10.9	6.3	20.2 ± 0.5	n.d.
H27η	–/8	77.5	85.5	75.5	80.1	37.7	45.7	0.7	9.7	6.4	6.4	2.0	5.4	3.4 ± 0.4	20.3
N30ν	8/–	9.7	11.3	7.8	8.9	−30.1	−28.5	−67.1	−64.5	6.8	7.0	1.9	2.4	7.6 ± 0.4	34.2
A31α	–/–	84.4	83.9	80.7	80.7	44.6	44.1	7.6	8.1	6.4	6.4	3.7	3.2	3.4 ± 0.7	13.7
S32σ	–/9	146.6	146.4	142.6	141.5	106.8	106.6	69.8	70.6	6.1	5.9	4.0	4.9	4.2 ± 0.5	17.5
Q33θ	10/–	66.5	68.8	53.7	42.6	26.7	29.0	−10.3	−7.0	6.5	6.5	12.8	26.2	13.0 ± 0.3	n.d.
W34ω	11/–	104.2	103.6	108.2	105.1	64.4	63.8	27.4	27.8	6.3	6.2	−4.0	−1.5	2.1 ± 0.3	9.5

[a] A single distance restraint between the side chains of the ester and the native branch was used to improve convergence.
[b] A similar EDS simulation with $s = 1.0$, $\Delta E^{R} = 98.0$ kJ mol^{-1}, involving only atom type transformation instead of using the hybrid topology shown in Fig. S2, yields $\Delta G^{f}_{mw} = 99.1$ kJ mol^{-1}.
[c] A similar EDS simulation with $s = 1.0$, $\Delta E^{R} = 99.0$ kJ mol^{-1}, involving only atom type transformation instead of using the hybrid topology shown in Fig. S2, yields $\Delta G^{f}_{mw} = 97.4$ kJ mol^{-1}.
[d] A similar EDS simulation with $s = 1.0$, $\Delta E^{R} = 81.9$ kJ mol^{-1}, involving only atom type transformation instead of using the hybrid topology shown in Fig. S2, yields $\Delta G^{u}_{mw} = 81.4$ kJ mol^{-1}.
[e] A similar EDS simulation with $s = 1.0$, $\Delta E^{R} = 81.9$ kJ mol^{-1}, involving only atom type transformation instead of using the hybrid topology shown in Fig. S2, yields $\Delta G^{u}_{mw} = 81.0$ kJ mol^{-1}.

the convergence of the two ΔG^{f}_{mw} values during the 20 ns EDS simulations starting from the X-ray or NMR model structures as shown in Fig. 7. All other EDS simulations end up with generally rather converged ΔG^{f}_{mw} of similar size for both, the 1PIN EDS and the 2KCF EDS simulations. The convergence in the Ala-tripeptide EDS simulation is shown in Fig. S3. The convergence of the ΔG^{u}_{mw} values for the 20 tripeptide mutants in simulations starting from the tripeptide structure in the X-ray (1PIN) or NMR model (2KCF) structure is shown in Fig. 8. As

Table 2
Occurrence (%) of the 11 hydrogen bonds (compare Fig. 1) in the two Pin1 WW domain structures (1PIN/2KCF) as deposited in the PDB database, in the MD trajectories of the wild-type simulations and in the EDS reference state trajectories of the 20 mutant simulations. The values for hydrogen bonds involving weakened hydrogen-bond acceptors due to the amide-to-ester mutation are in italics.

Protein	Hydrogen bond number/occurrence in %										
	1	2	3	4	5	6	7	8	9	10	11
Exp. structure	100/0	100/100	100/100	100/0	100/0	100/100	0/0	0/0	100/100	100/100	100/100
wt	20/0	96/93	95/93	95/92	97/75	65/42	7/1	2/0	83/83	96/96	89/91
L7λ	42/0	93/96	96/95	95/88	76/38	58/10	2/0	1/13	90/76	95/91	97/94
W11ω	–/–	92/97	96/90	96/92	96/94	76/76	6/2	2/0	78/77	94/88	96/94
E12ε	37/43	–/–	*96/94*	97/94	94/80	48/41	3/2	2/0	80/80	98/92	94/95
K13κ	31/0	96/94	92/92	92/96	59/90	20/75	5/4	0/0	79/81	98/94	94/93
R14ρ	36/0	98/96	90/91	–/–	*97/90*	84/77	1/2	2/0	86/70	98/87	96/88
M15μ	33/0	98/92	94/94	96/96	90/80	75/50	1/6	0/4	85/67	94/86	94/88
S16σ	8/0	98/95	94/92	94/94	82/58	–/–	*0/0*	4/0	83/80	90/90	96/93
R17ρ	41/0	93/94	94/92	94/94	72/88	43/64	4/0	4/0	88/76	94/94	92/85
S19σ	8/0	96/94	91/91	93/95	94/80	93/30	–/–	4/0	80/68	94/87	96/98
V22ϖ	27/0	98/92	96/92	92/92	100/92	73/68	2/2	0/0	82/78	92/87	94/90
Y23ψ	34/0	96/98	95/95	*97/97*	–/–	80/51	7/7	0/0	88/76	98/92	95/96
Y24ψ	4/0	96/88	94/97	89/64	84/9	43/2	10/0	1/0	56/80	*68/92*	–/–
F25ϕ	3/0	*50/95*	–/–	96/89	86/87	61/76	6/2	6/0	84/72	74/82	94/92
N26ν	9/0	94/90	92/98	96/82	69/32	22/14	4/2	*0/2*	–/–	88/76	97/97
H27η	32/0	96/97	96/90	93/93	87/90	46/70	0/6	2/6	82/68	97/83	95/95
N30ν	22/14	92/98	94/94	94/97	92/90	54/70	11/6	–/–	82/84	86/92	94/97
A31α	10/0	79/96	98/90	96/94	90/90	66/80	6/8	9/3	*86/74*	82/96	98/96
S32σ	45/0	99/93	92/95	96/78	98/48	84/8	1/14	1/3	82/56	93/84	96/94
Q33θ	20/0	95/98	94/86	95/88	93/84	72/54	3/4	4/0	80/74	–/–	*93/88*
W34ω	26/0	94/96	92/93	96/90	100/83	74/48	7/10	0/0	84/68	94/94	94/87

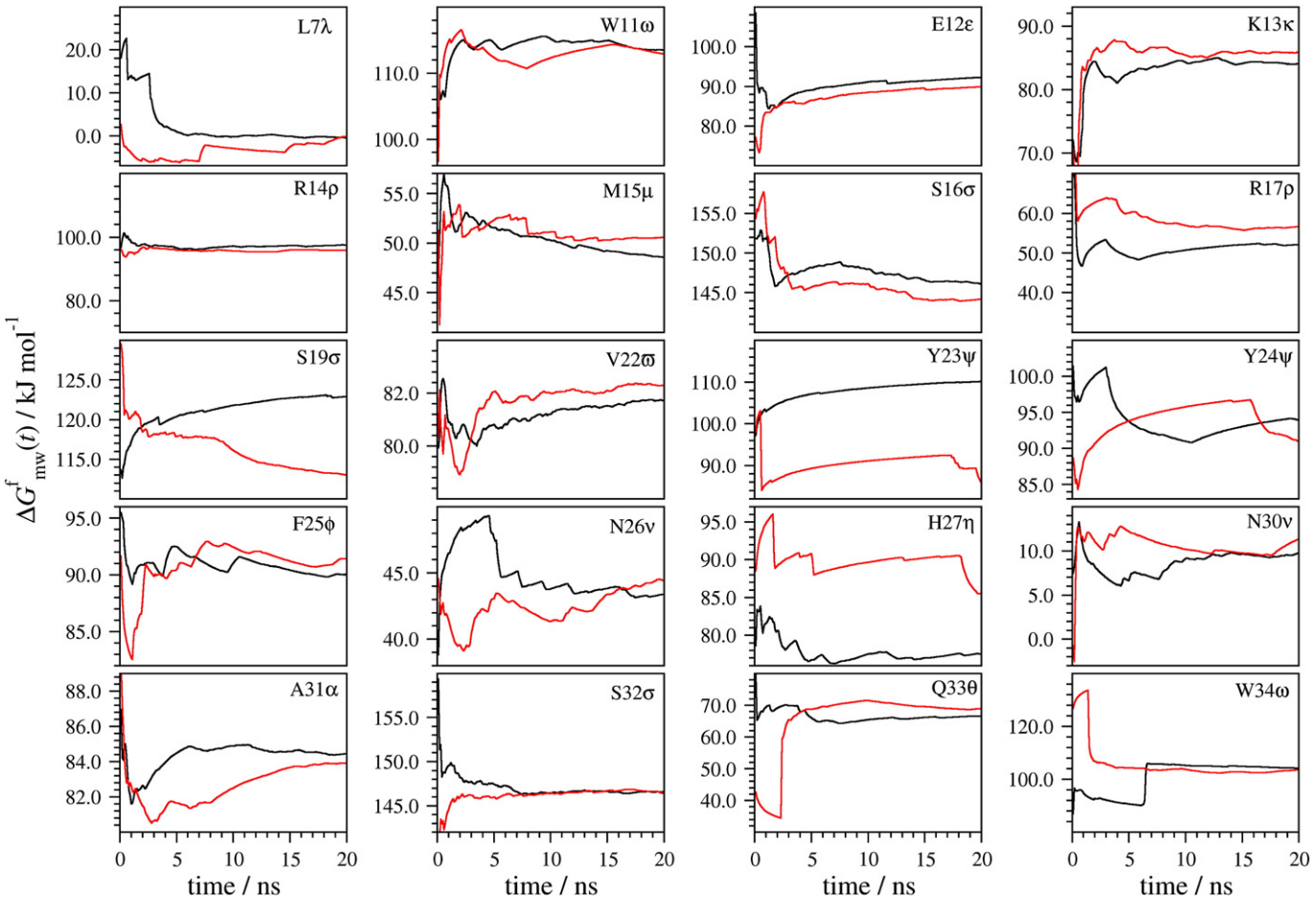

Fig. 7. ΔG^{f}_{mw} values (Eq. (3)) from the 1PIN (black) and 2KCF (red) EDS simulations as a function of simulation time for all 20 mutants of the Pin1 WW domain.

expected, the convergence is much better for the small tripeptide systems than for the corresponding protein mutants. The ΔG^{u}_{mw} values for the 20 tripeptides vary between 8 and 143 kJ mol^{-1}, which indicates that using the single value of 39 kJ mol^{-1} for the Ala tripeptide is a rather poor approximation of ΔG^{u}_{mw}.

Figs. S4 and S5 show the time series of the energy difference $\Delta V_{BA} = V_B - V_A$ of the various EDS reference state simulations and the subsequent Figs. S6 to S9 display the corresponding energy difference distributions. Figs. S6 and S7 show the energy difference distributions between the amide and ester states for the reference state and the two end states for the 1PIN and 2KCF EDS simulations. Except for the ester state of Y23ψ in the 1PIN EDS simulation (Fig. S6), the ester and amide states are sampled in the EDS simulations, although to a varying degree. The corresponding energy difference distributions for the tripeptides are shown in Figs. S8 and S9.

The reweighted end-state non-bonded solute–solute plus solute–solvent potential energy distributions are shown for all 1PIN and 2KCF EDS simulations in Figs. S10 and S11. Again, insufficient sampling of the ester state can be identified for the mutant Y23ψ (1PIN). In a number of cases the non-bonded solute–solute plus solute–solvent energy distribution as calculated from a standard (non-EDS) MD simulation of the wild-type protein (green curve) is not wholly congruent with the wild-type end-state energy distributions calculated from the EDS simulations (red curves).

The positional root-mean-square deviation of the C_α atoms from the X-ray and NMR model structures of the wild type protein during the EDS simulations of all 20 mutants is shown in Figs. S12 and S13. All but the N30ν mutant in the 1PIN simulation (Fig. S12) show stable RMSDs with respect to the crystal and NMR model structures with fluctuations between 0.2 nm and 0.5 nm, which is not larger than the RMSDs of the two wild-type MD simulations shown in Fig. 5. The N30ν mutant ends up with a non-converged RMSD value of about 0.7 nm after 23 ns, indicating that this mutant is unfolding during the 1PIN EDS simulation. In the 2KCF EDS simulations, all mutants are stable and converge to an RMSD value of about 0.35 nm from the NMR model structure.

4. Discussion

The correlation between the $\Delta\Delta G^{fu}_{mw}$ values calculated using Eqs. (3), (6) and (8) that use four different approximations of ΔG^{u}_{mw}, on the one hand and the $\Delta\Delta G^{c}_{mw}$ values inferred from chaotrope denaturation measurements or the $-\Delta T^{m}_{mw}$ values inferred from thermal denaturation measurements on the other hand is displayed in Fig. 9. Comparing these correlations with the correlation between $-\Delta T^{m}_{mw}$ and $\Delta\Delta G^{c}_{mw}$ shown in Fig. 3, it seems that the two experimental quantities $-\Delta T^{m}_{mw}$ and $\Delta\Delta G^{c}_{mw}$ correlate better with each other than with the calculated quantities $\Delta\Delta G^{fu}_{mw}$ based on a single value of ΔG^{u}_{mw} for all 20 mutants (Fig. 9a–c, e–g), which is obviously a too poor approximation of ΔG^{u}_{mw}. If the unfolded state of the protein mutants is represented by a tripeptide of the same composition as the mutated residue and its neighbours along the polypeptide chain, the correlation between simulation and experiment is much improved (Fig. 9d, h and Table 1). One should not expect perfect correlation between $-\Delta T^{m}_{mw}$ and $\Delta\Delta G^{c}_{mw}$ on the one hand and $\Delta\Delta G^{fu}_{mw}$ on the other hand, though, because the latter quantity is defined (Eq. (3)) in terms of free enthalpy differences ΔG^{fu}_{x} between folded and unfolded conformations of a protein or a mutant (x) at one thermodynamic state point, whereas the former two quantities

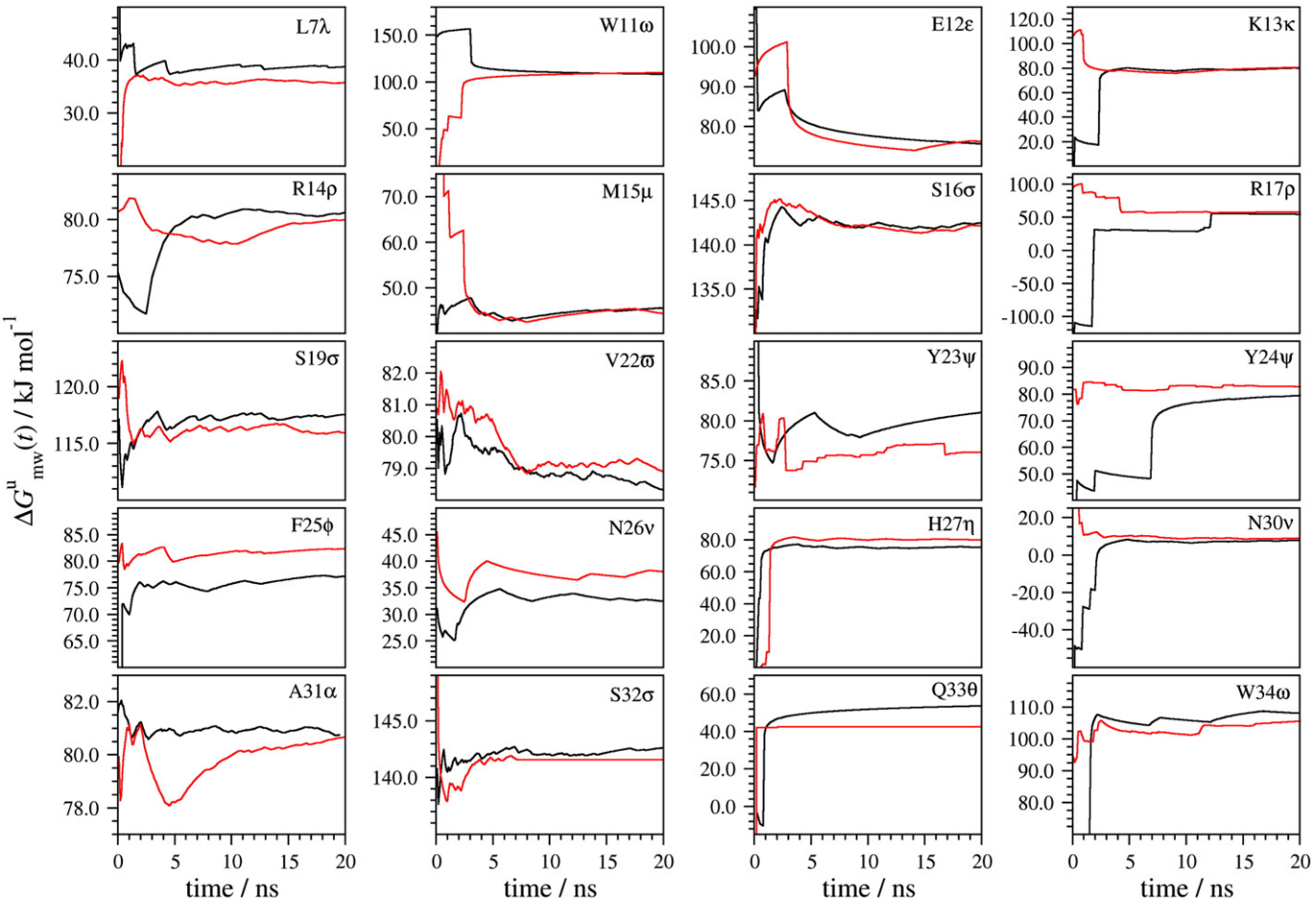

Fig. 8. ΔG^{u}_{mw} values from the 1PIN (black) and 2KCF (red) EDS simulations as a function of simulation time for all 20 tripeptide mutants.

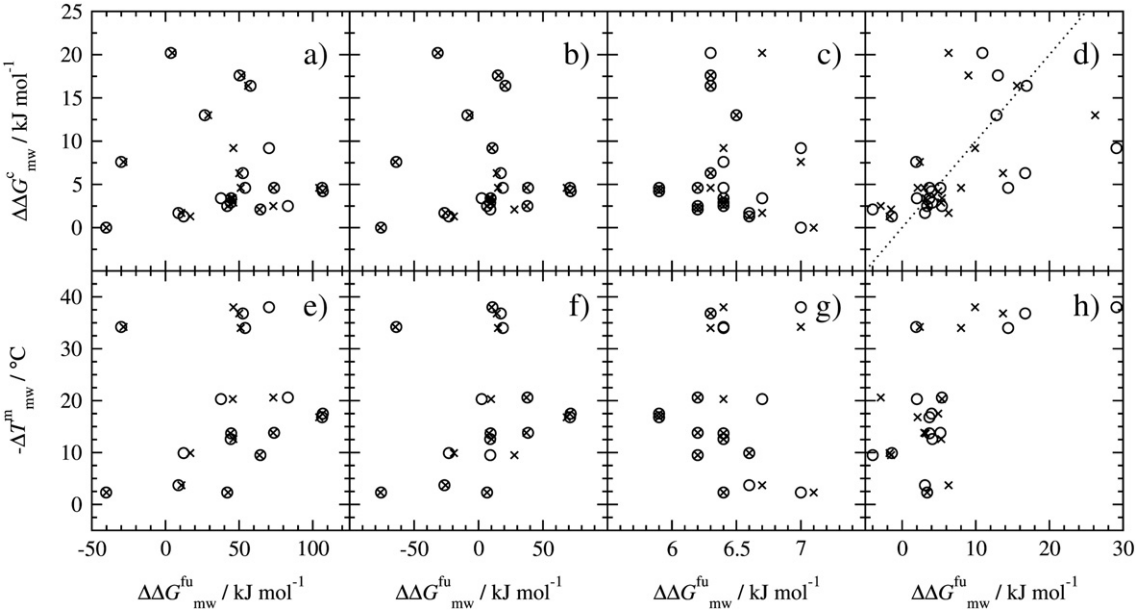

Fig. 9. Difference of folding free enthalpy differences $\Delta\Delta G^{c}_{mw}$ as obtained by chaotrope denaturation experiments [9] versus differences of folding free enthalpy differences $\Delta\Delta G^{fu}_{mw}$ (top graphs) and differences $-\Delta T^{m}_{mw}$ between midpoints T^{m} of thermal denaturation curves versus differences of folding free energy differences $\Delta\Delta G^{fu}_{mw}$ (bottom graphs). Calculations of the $\Delta\Delta G^{fu}_{mw}$ values have been done according to a)/e) Eq. (3), b)/f) Eq. (6), c)/g) Eq. (8) and d)/h) using separate ΔG^{u}_{mw} for 20 tripeptides. Circles and crosses distinguish between 1PIN and 2KCF simulations, respectively. The dotted line is intended to guide the eye along $\Delta\Delta G^{c}_{mw} = \Delta\Delta G^{fu}_{mw}$. The correlation coefficients are (from a) to h), data for 1PIN and 2KCF separated by a "/"): −0.04/−0.05, −0.04/−0.05, −0.05/0.02, 0.54/0.53, 0.19/0.11, 0.19/0.11, −0.16/−0.13, 0.67/0.52. The results for the mutation L7λ have been omitted from panels d) and h) to achieve similar ranges of the *x* and *y* axes.

are more similarly defined (Eqs. (4) and (5)) in terms of free enthalpy differences ΔG^{c}_{x} and midpoint temperatures T^{m}_{x} derived for a protein or mutant (x) at different thermodynamic state points of chemical potential and temperature. Although chaotrope and thermal denaturation may affect the protein or mutant stability in different ways, weighting the number of folded versus unfolded conformations at one thermodynamic state point makes ΔG^{fu}_{x} a more different quantity. But, as was mentioned in the Introduction, all three quantities shed light on the differences in stability of the folded conformations of the mutants and the wild-type protein.

The quality of the calculated $\Delta\Delta G^{fu}_{mw}$ values will depend on the quality of the biomolecular force field employed and on the extent of sampling of the relevant protein conformations in the MD simulations. The standard MD simulations for the wild type and six mutants show that the triple-stranded β-sheet fold is maintained for the wild type (Fig. S1) and mutants L7λ, N30ν (Fig. 6), whereas it is partly lost for the mutants R14ρ, F25ϕ, N26ν and Q33θ (Fig. 6) in agreement with the experimental observations of [9]. In all but one (N30ν starting from the X-ray structure) of the 42 EDS simulations of the wild type and mutants the atom-positional deviations from the X-ray and NMR structures (Figs. S12 and S13) is limited to 0.2–0.5 nm. This does not indicate any major force-field problem with respect to maintaining a proper fold, as indicated by the preservation of the backbone hydrogen bonds characterising the β-sheet fold of the protein and its mutants (Table 2). The analysis of the EDS simulations in terms of convergence of ΔG^{f}_{mw} values as function of time and between the pairs of simulations starting from X-ray and NMR structures (Fig. 7) shows converged values for 17 mutants, while for S19σ, Y23ψ, and H27η this convergence is not satisfactory. For the ΔG^{u}_{mw} values of the tripeptides (Fig. 8) differences between the 1PIN and 2KCF simulations are observed for Y23ψ, F25ϕ, N26ν, and Q33θ. The energy difference distributions (Figs. S6 and S7) show undersampling for Y23ψ in the EDS simulations of the proteins starting from the X-ray structure (Fig. S6). A comparison of non-bonded solute–solute plus solute–solvent end-state potential energy distributions for the wild type in the EDS simulations (red lines in Figs. S10 and S11) with that of the standard MD simulation of this end state (green lines in Figs. S10 and S11) shows rather good agreement.

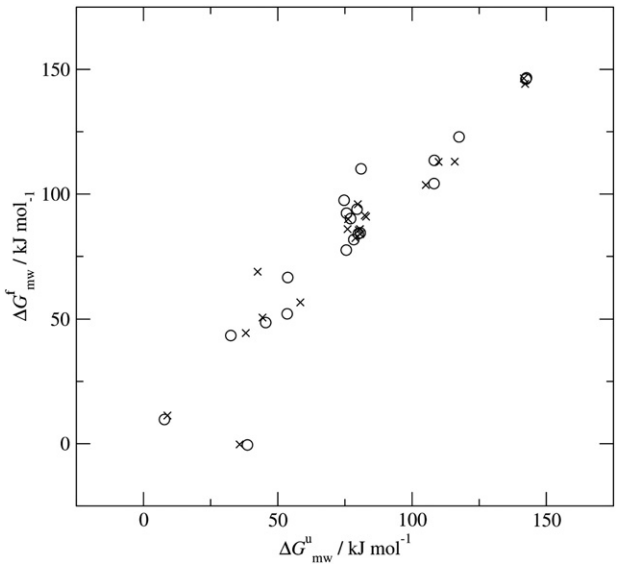

Fig. 10. Calculated free enthalpy difference ΔG^{f}_{mw} in the folded conformations of the 20 proteins versus calculated free enthalpy difference ΔG^{u}_{mw} in the unfolded conformations as computed from the 20 tripeptide simulations. Circles and crosses distinguish between 1PIN and 2KCF simulations. The correlation coefficients are 0.94 and 0.95 for the values calculated from the 1PIN and 2KCF simulations.

In summary, only a few of the 40 EDS simulations of the 20 mutants show some sampling and convergence problems.

The rather large differences between the $\Delta\Delta G^{fu}_{mw}$ values calculated using Eq. (3), (6) or (8), i.e. using three different assumptions for the value ΔG^{u}_{mw}, the free enthalpy difference between mutant and wild-type protein in the unfolded conformations, show that the assumption of an identical ΔG^{u}_{mw} value for all 20 mutants as used in Eqs. (3) and (6) is not justified. The assumption that ΔG^{u}_{mw} is proportional to $\Delta\Delta G^{f}_{mw}$ reduces the differences between the various $\Delta\Delta G^{fu}_{mw}$ values and brings them somewhat closer to the experimental $\Delta\Delta G^{c}_{mw}$ values, see Fig. 9. Indeed Fig. 10 shows that ΔG^{f}_{mw} is positively correlated with ΔG^{u}_{mw} as calculated for the 20 mutant tripeptides. The correlation coefficient is 0.94 and 0.95 for the values calculated from the 1PIN and 2KCF simulations, respectively. Using the tripeptide corresponding to the protein residue that is mutated and its neighbours along the poly-peptide chain as a representation of the unfolded state of the protein does improve the correlation between simulated and experimental results significantly, the correlation coefficients with the $\Delta\Delta G^{c}_{mw}$ and $-\Delta T^{m}_{mw}$ values being 0.54/0.67 and 0.53/0.52, respectively, calculated from the 1PIN/2KCF simulations. If we remove the four mutants R14ρ, F25ϕ, N26ν, and Q33θ, for which experimentally no thermal denaturation midpoint temperature could be determined and which could only be kept folded by adding TMAO to the solution, from the data set and in addition remove the $\Delta\Delta G^{fu}_{mw}$ values computed from S19ρ, Y23ψ, and H27η, the three mutants for which poor convergence between the EDS simulations started from the 1PIN and 2KCF structures was observed from the data set, the correlation coefficients for the remaining 13 mutants become 0.59/0.61 and 0.52/0.50 respectively. So no higher correlation is observed. This is due to the deviations between calculated and measured values for mutants L7λ, E12ε, and N30ν. In view of the different definitions of $\Delta\Delta G^{fu}_{mw}$, $\Delta\Delta G^{c}_{mw}$ and $-\Delta T^{m}_{mw}$ it is unclear what degree of correlation is to be expected between these different quantities.

The large variation in the values of ΔG_{mw} for the different mutants, not only in the folded state (ΔG^{f}_{mw}), but also in the unfolded state (ΔG^{u}_{mw}) modelled by the tripeptides, may at first sight be surprising, but is not considering the rather strong Coulomb interactions between (partial) atomic charges at a short distance from each other. When considering the corresponding $\Delta\Delta G^{fu}_{mw}$ values these are an order of magnitude smaller, at least where converged, and of a size comparable to the experimental $\Delta\Delta G^{c}_{mw}$ values. Yet this illustrates that the key to a precise calculation of folding free enthalpies lies in an accurate conformational description of the unfolded state [53,54].

5. Conclusion

Standard MD and enveloping distribution sampling (EDS) simulations of the wild-type Pin1 WW domain and 20 of its amide-to-ester mutants in aqueous solution have been used to shed light on the relation between protein folding free enthalpies and protein structure. Since experimental structural data were only available in the form of an X-ray crystal structure and an NMR solution structure of the wild-type protein showing roughly the same three-stranded β-sheet fold, but not for the 20 mutants, the simulated configurational ensembles complement the experimental data by providing a structural interpretation of these.

A direct comparison of experimental relative free enthalpies of folding upon chaotrope denaturation, $\Delta\Delta G^{c}_{mw}$, of experimental relative thermal denaturation temperatures, $-\Delta T^{m}_{mw}$, and of calculated relative free enthalpies of folding at a given thermodynamic state point, $\Delta\Delta G^{fu}_{mw}$, with each other is hampered by the different definitions of the corresponding free enthalpy differences. Moreover, for a few of the 20 mutants the configurational sampling within 20 ns appeared to be incomplete. Nevertheless, the simulations could identify the least stable mutants in agreement with the experimental data. The assumption of a single value for the relative free enthalpy ΔG^{u}_{mw} between mutant and wild type in the unfolded conformations turned out to be too crude.

Assuming that this ΔG^{u}_{mw} is proportional to ΔG^{f}_{mw}, the relative free enthalpy between mutant and wild type in the folded conformations, improved the agreement with the experimental data, but still seems to be a too crude assumption. Although the amide-to-ester mutation only involves the backbone, the presence of the nearby side chains and solvent molecules induces differences up to 135 kJ mol^{-1} in ΔG^{u}_{mw} values between the 20 tripeptides in solution. The use in Eq. (3) of ΔG^{u}_{mw} values obtained from simulations of tripeptide fragments greatly improved the correlation between simulated and experimental free enthalpy differences. This emphasises the importance of a proper sampling of protein configurations in both the folded and the unfolded states when calculating free enthalpies of folding. Reversely, the sensitivity of a $\Delta\Delta G^{fu}_{mw}$ value to both underlying conformational ensembles implies that an interpretation of measured $\Delta\Delta G^{c}_{mw}$ values or $-\Delta T^{m}_{mw}$ values in terms of particular protein conformations or structural features such as hydrogen bonding is a precarious endeavour.

The significance of local structural variations even in the folded conformations is illustrated by the variation in backbone–backbone hydrogen bonding between the X-ray and NMR structures of the wild type and between the different mutants. Hydrogen bonds 1, 5 and 6 (Fig. 1) involving residues 8, 11, 14, 16, 21 and 23 show sizeable variation in occurrence, and hydrogen bonds 7 and 8 involving residues 16, 19, 26 and 30 show rather low occurrence for the various mutants. This illustrates that an interpretation of the experimental $\Delta\Delta G^{c}_{mw}$ values and $-\Delta T^{m}_{mw}$ values in terms of the strength of particular hydrogen bonds is not straightforward. Such experimental data do not allow for a direct structural and mechanistic interpretation of the various forces stabilising the fold of a mutant or wild-type protein. EDS simulations do provide a means to establish the relation between the Boltzmann-weighted structural ensembles of the mutants and wild-type protein on the one hand and the relative free enthalpies of folding, $\Delta\Delta G^{fu}_{mw}$, between mutants and wild type on the other hand, which makes this recently proposed method suitable for a structural interpretation of free enthalpy data on proteins and their mutants.

Acknowledgements

We thank Philippe Hünenberger for helpful discussions and mentioning the work of U. Börjesson et al. [31].

This work was financially supported by the National Center of Competence in Research (NCCR) in Structural Biology and by grant number 200020-137827 of the Swiss National Science Foundation, and by grant number 228076 of the European Research Council, which is gratefully acknowledged. N.H. thanks the German Research Foundation (DFG) for financial support within the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbagen.2014.09.014.

References

[1] C. Anfinsen, Principles that govern folding of protein chains, Science 181 (4096) (1973) 223–230.
[2] Y.Y. Shi, A.E. Mark, C.X. Wang, F. Huang, H.J.C. Berendsen, W.F. van Gunsteren, Can the stability of protein mutants be predicted by free energy calculations? Protein Eng. 6 (1993) 289–295.
[3] C. Noren, S. Anthonycahill, M. Griffith, P. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins, Science 244 (4901) (1989) 182–188.
[4] J. Zhou, M. Case, J. Wishart, G. McLendon, Thermodynamic and structural effects of a single backbone hydrogen bond deletion in a metal-assembled helical bundle protein, J. Phys. Chem. B 102 (49) (1998) 9975–9980.
[5] G. Beligere, P. Dawson, Design, synthesis, and characterization of 4-ester CI2, a model for backbone hydrogen bonding in protein α-helices, J. Am. Chem. Soc. 122 (49) (2000) 12079–12082.
[6] T. Wales, M. Fitzgerald, The energetic contribution of backbone–backbone hydrogen bonds to the thermodynamic stability of a hyperstable P22 Arc repressor mutant, J. Am. Chem. Soc. 123 (31) (2001) 7709–7710.
[7] Z. Gattin, A. Glättli, B. Jaun, W.F. van Gunsteren, Simulation of beta-depsipeptides: the effect of missing hydrogen-bond donors on their folding equilibria, Biopolymers 85 (4) (2007) 318–332.
[8] A.P. Eichenberger, L.J. Smith, W.F. van Gunsteren, Ester-linked hen egg white lysozyme shows a compact fold in a molecular dynamics simulation — possible causes and sensitivity of experimentally observable quantities to structural changes maintaining this compact fold, FEBS J. 279 (2) (2012) 299–315.
[9] S. Deechongkit, P. Dawson, J. Kelly, Toward assessing the position-dependent contributions of backbone hydrogen bonding to β-sheet folding thermodynamics employing amide-to-ester perturbations, J. Am. Chem. Soc. 126 (51) (2004) 16762–16771.
[10] R. Kumar, I.V. Baskakov, G. Srinivasan, D.W. Bolen, J.C. Lee, E.B. Thompson, Interdomain signaling in a two-domain fragment of the human glucocorticoid receptor, J. Biol. Chem. 274 (35) (1999) 24737–24741.
[11] I. Baskakov, D.W. Bolen, Forcing thermodynamically unfolded proteins to fold, J. Biol. Chem. 273 (9) (1998) 4831–4834.
[12] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, W. Wriggers, Atomic-level characterization of the structural dynamics of proteins, Science 330 (2010) 341–346.
[13] W.F. van Gunsteren, Z. Gattin, Foldamers: structure, properties and applications, Ch. Simulation of Folding Equilibria, Wiley-VCH, 2007. 173–192.
[14] Z. Lin, W.F. van Gunsteren, Exploring the effect of side-chain substitutions upon the secondary structure preferences of beta-peptides, J. Phys. Chem. B 115 (2011) 12984–12992.
[15] IUPAP, Symbols, units and nomenclature in physics, Physica (Amsterdam) 93A (1978) 1–60.
[16] IUPAC, Quantities, Units and Symbols in Physical Chemistry, Blackwell Scientific Publications, Oxford, 1988.
[17] C. Chipot, A. Pohorille (Eds.), Free Energy Calculations: Theory and Applications in Chemistry and Biology, vol. 86, Springer-Verlag, 2007.
[18] C.D. Christ, A.E. Mark, W.F. van Gunsteren, Basic ingredients of free energy calculations: a review, J. Comput. Chem. 31 (8) (2010) 1569–1582.
[19] A.E. Mark, S.P. van Helden, P.E. Smith, L.H.M. Janssen, W.F. van Gunsteren, Convergence properties of free energy calculations: alpha-cyclodextrin complexes as a case study, J. Am. Chem. Soc. 116 (1994) 6293–6302.
[20] X. Daura, P.H. Hünenberger, A.E. Mark, E. Querol, F.X. Avilés, W.F. van Gunsteren, Free energies of transfer of trp analogs from chloroform to water: comparison of theory and experiment and the importance of adequate treatment of electrostatic and internal interactions, J. Am. Chem. Soc. 118 (1996) 6285–6294.
[21] Y.P. Pan, V. Daggett, Direct comparison of experimental and calculated folding free energies for hydrophobic deletion mutants of chymotrypsin inhibitor 2: free energy perturbation calculations using transition and denatured states from molecular dynamics simulations of unfolding, Biochemistry 40 (2001) 2723–2731.
[22] C.D. Christ, W.F. van Gunsteren, Enveloping distribution sampling: a method to calculate free energy differences from a single simulation, J. Chem. Phys. 126 (18) (2007) 184110.
[23] C.D. Christ, W.F. van Gunsteren, Multiple free energies from a single simulation: extending enveloping distribution sampling to nonoverlapping phase-space distributions, J. Chem. Phys. 128 (17) (2008) 174112.
[24] Z. Lin, T.A. Timmerscheidt, W.F. van Gunsteren, Using enveloping distribution sampling (EDS) to compute the free enthalpy difference between right- and left-handed helices of a β-peptide in solution, J. Chem. Phys. 137 (2012) 064108.
[25] Z. Lin, W.F. van Gunsteren, Combination of enveloping distribution sampling (EDS) of a soft-core reference-state Hamiltonian with one-step perturbation to predict the effect of side chain substitution on the relative stability of right- and left-helical folds of β-peptides, J. Chem. Theory Comput. 9 (2013) 126–134.
[26] Z. Lin, W.F. van Gunsteren, Enhanced conformational sampling using enveloping distribution sampling, J. Chem. Phys. 139 (2013) 144105.
[27] N. Hansen, J.R. Allison, F. Hodel, W.F. van Gunsteren, Relative free enthalpies for point mutations in two proteins with highly similar sequence but different folds, Biochemistry 52 (2013) 4962–4970.
[28] N. Hansen, J. Dolenc, M. Knecht, S. Riniker, W.F. van Gunsteren, Assessment of enveloping distribution sampling to calculate relative free enthalpies of binding for eight netropsin–DNA duplex complexes in aqueous solution, J. Comput. Chem. 33 (6) (2012) 640–651.
[29] W.F. van Gunsteren, J. Dolenc, A.E. Mark, Molecular simulation as an aid to experimentalists, Curr. Opin. Struct. Biol. 18 (2008) 149–153.
[30] S. Riniker, C.D. Christ, H.S. Hansen, P.H. Hünenberger, C. Oostenbrink, D. Steiner, W.F. van Gunsteren, Calculation of relative free energies for ligand–protein binding, solvation, and conformational transitions using the GROMOS software, J. Phys. Chem. B 115 (46) (2011) 13570–13577.
[31] U. Börjesson, P.H. Hünenberger, Effect of mutations involving charged residues on the stability of staphylococcal nuclease: a continuum electrostatics study, Protein Eng. 16 (11) (2003) 831–840.
[32] R. Ranganathan, K.P. Lu, T. Hunter, J.P. Noel, Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent, Cell 89 (6) (1997) 875–886.
[33] J.A. Kowalski, K. Liu, J.W. Kelly, NMR solution structure of the isolated Apo Pin1 WW domain: comparison to the X-ray crystal structures of Pin1, Biopolymers 63 (2) (2002) 111–121.
[34] N. Schmid, C.D. Christ, M. Christen, A.P. Eichenberger, W.F. van Gunsteren, Architecture, implementation and parallelisation of the GROMOS software for biomolecular simulation, Comput. Phys. Commun. 183 (4) (2012) 890–903.

[35] A.P.E. Kunz, J.R. Allison, D.P. Geerke, B.A.C. Horta, P.H. Hünenberger, S. Riniker, N. Schmid, W.F. van Gunsteren, New functionalities in the GROMOS biomolecular simulation software, J. Comput. Chem. 33 (3) (2012) 340–353.
[36] C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6, J. Comput. Chem. 25 (13) (2004) 1656–1676.
[37] B.A.C. Horta, P.F.J. Fuchs, W.F. van Gunsteren, P.H. Hünenberger, New interaction parameters for oxygen compounds in the GROMOS force field: improved pure-liquid and solvation properties for alcohols, ethers, aldehydes, ketones, carboxylic acids, and esters, J. Chem. Theory Comput. 7 (4) (2011) 1016–1031.
[38] H. Berendsen, J. Postma, W. van Gunsteren, J. Hermans, Ch. Interaction Models for Water in Relation to Protein Hydration, Reidel, Dordrecht, 1981. 331–342.
[39] A.P. Eichenberger, J.R. Allison, J. Dolenc, D.P. Geerke, B.A.C. Horta, K. Meier, C. Oostenbrink, N. Schmid, D. Steiner, D.Q. Wang, W.F. van Gunsteren, GROMOS++ software for the analysis of biomolecular simulation trajectories, J. Chem. Theory Comput. 7 (10) (2011) 3379–3390.
[40] The official GROMOS website, URL http://www.gromos.net.
[41] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular-dynamics with coupling to an external bath, J. Chem. Phys. 81 (8) (1984) 3684–3690.
[42] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical-integration of Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (3) (1977) 327–341.
[43] R. Hockney, J. Eastwood, Computer Simulation Using Particles, McGraw-Hill, New York, 1981.
[44] T.N. Heinz, W.F. van Gunsteren, P.H. Hünenberger, Comparison of four methods to compute the dielectric permittivity of liquids from molecular dynamics simulations, J. Chem. Phys. 115 (3) (2001) 1125–1136.
[45] S. Harvey, R. Tan, T. Cheatham, The flying ice cube: velocity rescaling in molecular dynamics leads to violation of energy equipartition, J. Comput. Chem. 19 (7) (1998) 726–740.
[46] D.A. Pearlman, A comparison of alternative approaches to free-energy calculations, J. Phys. Chem. 98 (5) (1994) 1487–1493.
[47] K.S. Shing, K.E. Gubbins, The chemical-potential in dense fluids and fluid mixtures via computer-simulation, Mol. Phys. 46 (5) (1982) 1109–1128.
[48] J.G. Powles, W.A.B. Evans, N. Quirke, Non-destructive molecular-dynamics simulation of the chemical-potential of a fluid, Mol. Phys. 46 (6) (1982) 1347–1370.
[49] G. Jacucci, N. Quirke, Free-energy calculations for crystals, Lect. Notes Phys. 166 (1982) 38–57.
[50] Z.W. Salsburg, J.D. Jacobson, W. Fickett, W.W. Wood, Application of the Monte-Carlo method to the lattice-gas model. I. Two-dimensional triangular lattice, J. Chem. Phys. 30 (1) (1959) 65–72.
[51] A.M. Ferrenberg, R.H. Swendsen, New Monte-Carlo technique for studying phase-transitions, Phys. Rev. Lett. 61 (23) (1988) 2635–2638.
[52] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (12) (1983) 2577–2637.
[53] W.F. van Gunsteren, R. Bürgi, C. Peter, X. Daura, The key to solving the protein-folding problem lies in an accurate description of the denatured state, Angew. Chem. Int. Ed. 40 (2) (2001) 351–355.
[54] W.F. van Gunsteren, R. Burgi, C. Peter, X. Daura, Comment on the communication "The key to solving the protein-folding problem lies in an accurate description of the denatured state" by van Gunsteren et al. — reply, Angew. Chem. Int. Ed. 40 (24) (2001) 4616.