

# SCIENTIFIC REPORTS

There are amendments to this paper

OPEN

## The protonation state of an evolutionarily conserved histidine modulates domain swapping stability of FoxP1

Exequiel Medina<sup>1</sup>, Pablo Villalobos<sup>1</sup>, Ricardo Coñuecar<sup>1</sup>, César A. Ramírez-Sarmiento<sup>2</sup> & Jorge Babul<sup>1</sup>

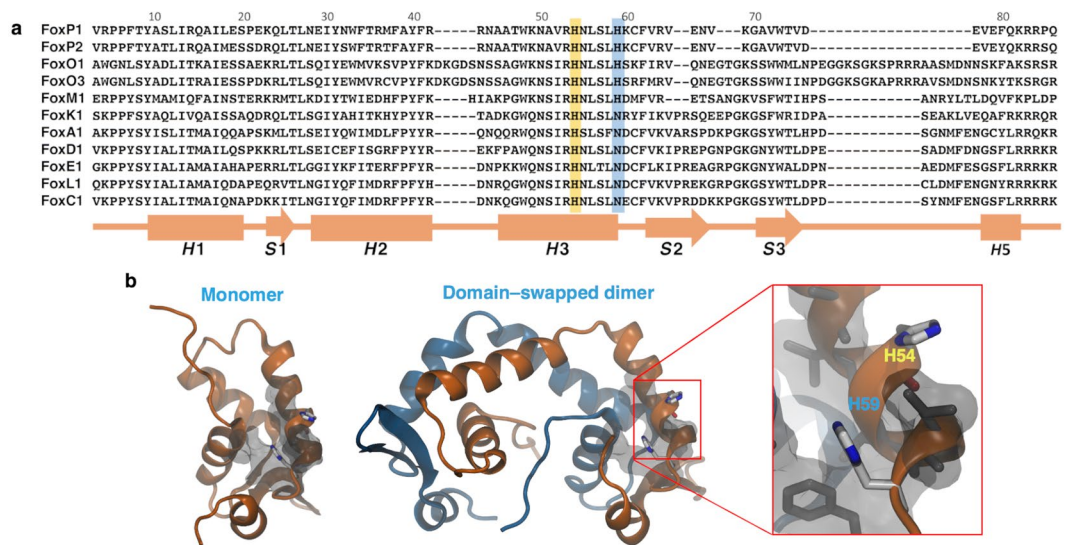
Forkhead box P (FoxP) proteins are members of the versatile Fox transcription factors, which control the timing and expression of multiple genes for eukaryotic cell homeostasis. Compared to other Fox proteins, they can form domain-swapped dimers through their DNA-binding –forkhead– domains, enabling spatial reorganization of distant chromosome elements by tethering two DNA molecules together. Yet, domain swapping stability and DNA binding affinity varies between different FoxP proteins. Experimental evidence suggests that the protonation state of a histidine residue conserved in all Fox proteins is responsible for pH-dependent modulation of these interactions. Here, we explore the consequences of the protonation state of another histidine (H59), only conserved within FoxM/O/P subfamilies, on folding and dimerization of the forkhead domain of human FoxP1. Dimer dissociation kinetics and equilibrium unfolding experiments demonstrate that protonation of H59 leads to destabilization of the domain-swapped dimer due to an increase in free energy difference between the monomeric and transition states. This pH-dependence is abolished when H59 is mutated to alanine. Furthermore, anisotropy measurements and molecular dynamics evidence that H59 has a direct impact in the local stability of helix H3. Altogether, our results highlight the relevance of H59 in domain swapping and folding stability of FoxP1.

Proteins are versatile macromolecules with exquisite equilibria between all different intermolecular forces to favour the native state. Understanding of the physical chemical properties of side chain interactions with side<sup>1</sup> or main chains<sup>2</sup> from other residues in the folded protein is relevant to describe their role in reaching the native state and also their involvement in several biological processes. This is particularly critical when only a few substitutions are responsible for the functional and biophysical differences between proteins, allowing their diversification throughout protein evolution in nature<sup>3,4</sup>.

One representative case is the extensive forkhead box (Fox) family of transcription factors, distributed among eukaryotic organisms<sup>5</sup> and responsible for pleiotropic processes during embryonic development and adult cellular homeostasis, such as immune system regulation, glucose metabolism and different tissue functions in humans and other mammals<sup>6,7</sup>. All members of the Fox family are composed by a structurally conserved DNA-binding domain (forkhead) of ~100 amino acid residues that was used to classify these proteins into 19 subfamilies, from A to S<sup>5</sup>. Most Fox members have been described to fold into monomers *in vitro* regardless of the presence of their cognate DNA. However, only the forkhead domain from the P subfamily (FoxP) has been characterized to form domain-swapped dimers in solution both in absence and presence of DNA<sup>8–11</sup>, despite the high sequence identity (~70–80%) with other Fox members. These intriguing aspects motivated the analysis of the structural and functional properties of FoxP proteins and the molecular basis of their unique behaviour.

Canonically, three-dimensional domain swapping is described as a folding-upon-binding mechanism where two or more proteins exchange identical structural segments<sup>12</sup>, breaking native monomeric contacts and

<sup>1</sup>Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Casilla 653, Santiago, 7800003, Chile. <sup>2</sup>Institute for Biological and Medical Engineering, Schools of Engineering, Medicine and Biological Sciences, Pontificia Universidad Católica de Chile, Av. Vicuña Mackenna 4860, Santiago, 7820436, Chile. Correspondence and requests for materials should be addressed to C.A.R.-S. (email: cesar.ramirez@uc.cl) or J.B. (email: jbabul@uchile.cl)



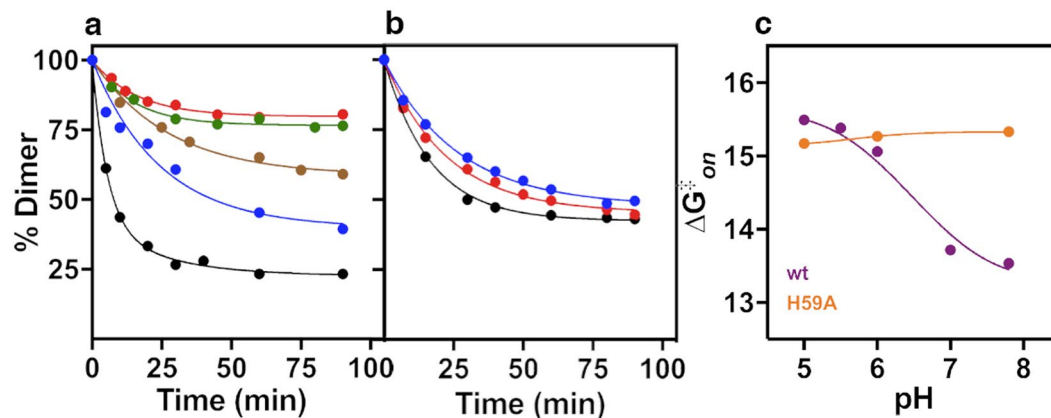
**Figure 1.** Evolutionary and structural context of H59 in the DNA-binding domain of human FoxP1. **(a)** Sequence alignment of the DNA-binding domain from representative members of the human Fox subfamilies. Histidines highlighted in yellow and blue are the family-conserved and the M/O/P subfamily-specific residues, respectively. A secondary structure topology scheme is shown at the bottom of the sequence alignment. **(b)** Cartoon representation of the structure of the A39P/C61Y monomeric mutant (PDB 2KIU) and the domain-swapped dimer (generated by homology modelling using FoxP2 as template) of the forkhead domain of human FoxP1, showing the position of the two histidine residues (H54 and H59) as sticks. Hydrophobic residues surrounding H54 and H59 at 5 Å are shown in grey using their van der Waals radius.

establishing them in an intermolecular fashion. The large number of interactions being reformed imposes a critical energy barrier that controls the kinetic and thermodynamic properties of this process, due to the need of monomer unfolding to enable domain swapping<sup>13,14</sup>. Phenotypically, these constraints are evidenced in the form of extremely slow monomer-dimer transitions and low association affinity for most studied proteins.

We previously characterized domain swapping of the forkhead domain of human FoxP1 (hereafter FoxP1), showing that its association equilibrium is reached without needing extensive protein unfolding in ~2 h at 37 °C under low  $\mu$ M protein concentrations<sup>11</sup>. This is nearly 1,000-fold faster and favourable than in other domain swapping proteins<sup>8–11</sup>, thus suggesting that FoxP1 exhibits a different molecular mechanism due to key structure and sequence features. Conversely, specific mutations affecting domain swapping of FoxP proteins are involved in severe diseases such as IPEX syndrome in humans<sup>9,15</sup>. Alas, the limited structural and biophysical characterization of these proteins impedes unravelling the kinetic mechanism of this process and its relationship with their biological role.

In this context, Blane and Fanucchi<sup>16</sup> recently showed the effect of changing the pH around the theoretical pKa of histidine residues on the tertiary structure and DNA binding of the forkhead domain of FoxP2. The pH dependence of these traits was related with protonation changes of a conserved histidine (H54) located in the helix H3 of all forkhead domains<sup>7</sup>, due to the inferred role of this residue in protein-DNA binding based on crystallographic<sup>9,10</sup> and functional<sup>17</sup> data. However, sequence comparison between several Fox members shows the presence of another histidine (H59) only in the M, O and P subfamilies<sup>7,17</sup>, which could also be relevant for protein function and stability (Fig. 1a). Moreover, the three-dimensional structure of FoxP1 shows that both residues are surrounded by a hydrophobic environment (Fig. 1b), where their interaction ability is limited due to the effective contact radius (~5 Å)<sup>18</sup>. In this scenario, the current structural and biophysical analysis is not sufficient to determine if the consequences observed in FoxP2 were due to changes in the protonation states of the strictly conserved H54 and/or the H59 that is only evolutionarily conserved in the M/O/P subfamilies.

In this study, we explore the impact of each histidine residue in the dimerization and folding stability of the forkhead domain of FoxP1 using both wild type protein (wt) and the H59A mutant. Kinetic analysis of domain swapping shows that only H59 is involved in dimer stability, with its deprotonation significantly increasing the association rate to favour the dimer conformation. This pH-dependent behaviour is also observed in equilibrium unfolding studies, where the global stability of the wt protein, and not the H59A mutant, depends on pH. Moreover, fluorescence anisotropy experiments monitoring local changes in helix H3 due to unfolding suggest that its stability is relevant to modulate the dimerization of FoxP1. Finally, molecular dynamics simulations reveal that the protonation state of H59 affects the persistence of a hydrogen bond between its imidazole ring and the carbonyl group of the backbone of residue N55, as well as having a long-range effect in the contact probability of several native interactions on the folding transition state. Altogether, these findings highlight the relevance of H59 in the stability of helix H3 and the modulation of the dimerization propensity of FoxP proteins, providing insights of how key residue changes during the evolution of the Fox family of transcription factors aid in sculpting their folding landscape.



**Figure 2.** Kinetic measurements of FoxP1 around the pKa of its H59 side chain. Changes in the dimer abundance upon time at different pH values for wt (a) and H59A mutant (b) after initial incubation of 5  $\mu$ M of the dimer species. Isolated dimers were incubated at pH 5.0 (black), 5.5 (blue), 6.0 (brown), 7.0 (green) and 7.8 (red) at 37  $^{\circ}$ C, and dimer population was quantified by SEC experiments. Data was fitted to a single exponential decay. (c) Kinetic properties of FoxP1 around the pKa of its H59 side chain. Comparison of  $\Delta G^{\ddagger}$  (association) of wt and H59A as a function of pH. Values were calculated using the Eyring equation.

pH	wt				H59A			
	$K_d$ ( $\mu$ M)	$\Delta G_d$ (kcal $\cdot$ mol $^{-1}$ )	$k_{off} \times 10^{-3}$ (min $^{-1}$ )	$k_{on} \times 10^3$ (M $^{-1}$ $\cdot$ min $^{-1}$ )	$K_d$ ( $\mu$ M)	$\Delta G_d$ (kcal $\cdot$ mol $^{-1}$ )	$k_{off} \times 10^{-3}$ (min $^{-1}$ )	$k_{on} \times 10^3$ (M $^{-1}$ $\cdot$ min $^{-1}$ )
5.0	31.2 $\pm$ 0.3	6.4 $\pm$ 0.1	140.9 $\pm$ 9.5	3.5 $\pm$ 0.7	7.9 $\pm$ 0.1	7.2 $\pm$ 0.1	59.3 $\pm$ 3.5	7.5 $\pm$ 0.8
5.5	13.1 $\pm$ 0.2	6.9 $\pm$ 0.1	40.5 $\pm$ 4.0	3.1 $\pm$ 0.7	—	—	—	—
6.0	3.7 $\pm$ 0.3	7.7 $\pm$ 0.1	33.2 $\pm$ 3.8	9.0 $\pm$ 1.6	5.9 $\pm$ 0.1	7.4 $\pm$ 0.1	37.7 $\pm$ 1.8	6.4 $\pm$ 0.3
7.0	0.8 $\pm$ 0.1	8.6 $\pm$ 0.1	68.1 $\pm$ 7.2	80.2 $\pm$ 16.4	—	—	—	—
7.8	0.6 $\pm$ 0.1	8.8 $\pm$ 0.1	63.8 $\pm$ 5.1	107.5 $\pm$ 13.5	7.5 $\pm$ 0.2	7.2 $\pm$ 0.2	43.4 $\pm$ 2.5	5.8 $\pm$ 0.5

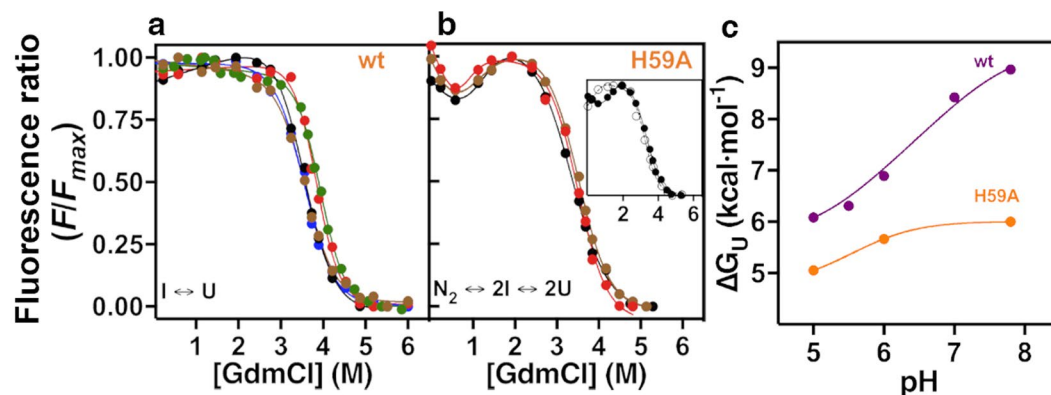
**Table 1.** Thermodynamic and kinetics parameters of domain swapping for wt FoxP1 and H59A mutant at 37  $^{\circ}$ C.

## Results

**Protonation state of H59 modulates the domain swapping kinetics of FoxP1.** To determine if pH changes around the pKa of histidine affect the dimerization and conformational stability of FoxP proteins, we analysed these parameters in a pH range from 5.0 to 7.8 using human FoxP1 as a model. The dependence of monomer-dimer transitions on the (de)protonation of histidine residues was ascertained using size exclusion chromatography (SEC), monitoring the kinetics of dimer dissociation by quantifying monomer and dimer fractions at different times at 37  $^{\circ}$ C until the reaction reached equilibrium. Results showed a strong effect of pH in the final dimer concentration under equilibrium conditions (Fig. 2a) and also in the kinetics of the monomer-dimer transition (Table 1), with increasing pH favouring the dimer.

This pH dependence can be attributed to protonation changes in either H59 and/or H54 residues. This was corroborated by *in silico* analysis of the monomeric structure of FoxP1 (PDB ID 2KUI) in the H++ server<sup>19</sup> to estimate the pKa values of titratable residues. We found that only His residues H54 and H59 are effectively titratable in the pH range used herein (data not shown). Thus, to experimentally elucidate the specific role of H59 protonation in the monomer-dimer transition, we generated the H59A mutant to interrupt any possible side chain interaction chemically related with the imidazole ring, and then we performed the same experiment as indicated for the wt protein. In this case, the H59A mutant lost the pH dependence of its monomer-dimer equilibrium, and also the kinetic decay was scarcely affected by pH changes when compared to the wt protein (Fig. 2b and Table 1).

In order to quantify how the protonation state of these residues affects the stability of the domain-swapped dimer, we calculated the equilibrium dissociation constant ( $K_d$ ) at each pH. For the wt protein,  $K_d$  values were strongly affected by pH, ranging from 31.2  $\pm$  0.3 to 0.6  $\pm$  0.1  $\mu$ M when pH is increased from 5.0 to 7.8 (Table 1). The free energy differences between monomer and dimer ( $\Delta G_d$ ) as a function of pH calculated from  $K_d$  evidence a dramatic transition between pH 5.5 and 7.0. These results suggest that changes in the protonation state of His residues occurring in this pH range modulate domain swapping stability by  $\sim$ 2.4 kcal $\cdot$ mol $^{-1}$ , with an increase in  $K_d$  occurring at pH values in which His residues are expected to be protonated. In contrast, the H59A mutant shows a steady  $\Delta G_d$  of  $\sim$ 7.3 kcal $\cdot$ mol $^{-1}$  (Table 1), with an increased dimer stability at pH values in which His residues are expected to be protonated (pH < 7) but a decreased stability when the His residue is expected to be deprotonated (pH > 7), when compared to the wt protein. These results suggest that the pH dependence of domain swapping with thermodynamics in the wt protein is due to the equilibrium between the protonated and



**Figure 3.** Equilibrium unfolding experiments of wt FoxP1 and H59A mutant. Protein unfolding as a function of denaturant (GdmCl) was monitored using intrinsic fluorescence for wt (a) and H59A (b) at a total protein concentration of  $15\ \mu\text{M}$ . (a) wt FoxP1 was incubated with different denaturant concentrations at pH 5.0 (black), 5.5 (blue), 6.0 (brown), 7.0 (green) and 7.8 (red), and the data obtained were fitted to a two-state  $I \rightleftharpoons U$  monomer unfolding. (b) The same unfolding experiment was performed with the H59A mutant at pH 5.0, 6.0 and 7.8, using the same colour pattern as for wt protein. Data was fitted to a three-state  $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$  unfolding mechanism. **Inset.** Unfolding curves of the H59A mutant at pH 5.0 using 3 (white circles) or  $15\ \mu\text{M}$  (black circles) of total protein concentration. (c)  $\Delta G_U$  dependence with pH for wt and the H59A mutant. Free energy values for intermediate unfolding ( $I \rightleftharpoons U$ ) obtained were fitted to Wyman-Tanford relationship (Eq. 2) to calculate the pKa changes between the intermediate and unfolded states.

deprotonated forms of H59 instead of H54, concluding that the latter is mostly involved in protein-DNA interactions<sup>16</sup> and does not participate, at least directly, in protein-protein interactions.

This pH dependence of the monomer-dimer equilibrium in FoxP1 could emerge from changes on either the association ( $k_{\text{on}}$ ) or dissociation ( $k_{\text{off}}$ ) rate constants, having a critical effect in the energetic barrier between monomer and dimer. Therefore, we fitted the dissociation kinetics of the wt and the H59A mutant proteins to single exponential decays to determine  $k_{\text{off}}$ , and then  $k_{\text{on}}$  was derived from the relationship between the equilibrium and kinetic dissociation constants as in Eq. 1 (see Methods). Data obtained from these analyses (Table 1) indicates that the main effect of pH is on  $k_{\text{on}}$ , showing a dramatic increase of  $\sim 50$ -fold when FoxP1 is changed from acidic to basic conditions with respect to the pKa of the imidazole ring.

To further elucidate the effect of H59 in domain swapping, we used this kinetic information to calculate the free energy changes in association ( $\Delta G_{\text{on}}^\ddagger$ ) and dissociation ( $\Delta G_{\text{off}}^\ddagger$ ) using the Eyring approximation (Fig. 2c). First, we observed that the  $\Delta G_{\text{on}}^\ddagger$  and  $\Delta G_{\text{off}}^\ddagger$  values described herein were significantly lower than those previously reported for other domain swapping models, such as the C-terminal domain of SARS virus protease ( $M_{\text{pro}}-C$ )<sup>20</sup>, despite the similar  $K_d$  (Fig. 2c and Supplementary Fig. S1). These results highlight the low kinetic barrier for domain swapping in FoxP1. Secondly,  $\Delta G_{\text{on}}^\ddagger$  decreased by  $\sim 2.5\ \text{kcal}\cdot\text{mol}^{-1}$  upon increasing the pH above 6.0. Altogether, these findings strongly suggest that deprotonation of H59 in the transition state favours domain swapping of FoxP1.

As expected, H59A shows only a marginal decrease in  $k_{\text{on}}$  and no significant changes in  $\Delta G_{\text{on}}^\ddagger$  and  $\Delta G_{\text{off}}^\ddagger$  (Table 1, Fig. 2c and Supplementary Fig. S1), further supporting the specific relevance of H59 in modulating the height of the energetic barrier for domain swapping. Taking into advantage the sole presence of H54 in this mutant, we used the Wyman-Tanford relationship (Eq. 2, see Methods), which relates the pH-dependent free energy changes with the pKa of any residue involved in the dimerization process<sup>21</sup>, to determine the pKa changes of H54 and H59 in the dimerization pathway. Given that no significant changes in  $\Delta G_{\text{on}}^\ddagger$  were observed for the H59A mutant upon changing the pH, the behaviour of H54 was explained by a pKa of  $\sim 5.3$ . Using this value as a fixed parameter for the analysis of the  $\Delta G_{\text{on}}^\ddagger$  values obtained for the wt protein, we obtained the pKa of the H59 residue. Our values estimate a decrease of the pKa of this residue by approximately 1.8 units upon reaching the transition state for domain swapping (calculated pKa (M) = 7.3; pKa ( $\ddagger$ ) = 5.5; pKa (D) = 5.5).

**Protonation state of H59 is involved in global folding stability.** Once the role of H59 in the dimerization of FoxP1 was determined, we evaluated the potential involvement of this residue in conformational stability. For these analyses, we compared the equilibrium folding of wt FoxP1 and the H59A mutant, since both folding and association events concurrently take place during its domain swapping<sup>11</sup>.

In order to analyse the equilibrium unfolding mechanism, we used the intrinsic fluorescence of tryptophan residues W33, W48 and W73 to monitor changes in tertiary structure as a function of the GdmCl concentration at different pH values (Fig. 3). Although we previously showed that the unfolding mechanism of FoxP1 is characterized by the presence of a monomeric intermediate, as ascertained by circular dichroism experiments ( $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$ )<sup>11</sup>, changes in tryptophan fluorescence for all pH values monitor only one transition. This transition occurs within the same range of denaturant concentrations in which the intermediate, seen by circular dichroism, is in equilibrium with the unfolded state ( $I \rightleftharpoons U$ )<sup>11</sup> (Fig. 3a). The absence of the  $N_2 \rightleftharpoons 2I$  transition could be explained by (1) masking of this transition due to the heterogeneous fluorescence signal of the multiple



pH	Wild-type		H59A					
	$\Delta G_u^*$	$m$ -value <sup>†</sup>	$\Delta G_1^*$	$m_1^†$	$\Delta G_2^*$	$m_2^†$	$\Delta G_U^*$	$m^†$
5.0	6.1 ± 0.5	1.7 ± 0.1	7.9 ± 0.9	3.0 ± 0.1	5.0 ± 0.4	1.4 ± 0.1	17.9	5.8
5.5	6.3 ± 0.2	1.8 ± 0.1	—	—	—	—	—	—
6.0	6.9 ± 0.5	2.0 ± 0.1	8.6 ± 0.9	3.2 ± 0.2	5.6 ± 0.7	1.6 ± 0.2	19.8	6.4
7.0	8.4 ± 0.4	2.2 ± 0.1	—	—	—	—	—	—
7.8	8.9 ± 0.4	2.3 ± 0.1	8.2 ± 1.2	3.0 ± 0.1	6.0 ± 1.0	1.8 ± 0.1	20.2	6.6

**Table 2.** Thermodynamic parameters of conformational stability of the wild-type and H59A mutant at 37°C.

\*Expressed as kcal·mol<sup>-1</sup>. †Expressed as kcal·mol<sup>-1</sup>·M<sup>-1</sup>.

tryptophan residues of FoxP1; and/or (2) the positions of the tryptophan residues are insensitive to dimer dissociation. For these reasons, unfolding data from wt FoxP1 was fitted to a two-state mechanism ( $I \rightleftharpoons U$ ).

Our equilibrium unfolding experiments revealed that the unfolding free energy ( $\Delta G_U$ ) of wt FoxP1 depends on pH in a similar fashion to what we observed for the domain swapping process (Table 2), with  $\Delta G_U$  varying from 6.1 to 8.9 kcal·mol<sup>-1</sup> (Fig. 3c). Although the data were analysed in terms of a two-state mechanism, we observed an increase in the  $m$ -value of unfolding in the same pH range as the dissociation measurements (Table 2); when pH is changed from pH 5.0 to 7.8, the  $m$ -value changed from 1.8 ± 0.1 to 2.2 ± 0.1 kcal·mol<sup>-1</sup>·M<sup>-1</sup>. This change accounts for a relative increase of ~29% in the cooperativity of the  $I \rightleftharpoons U$  transition.

This pH-dependence of the changes in  $m$ -value in apparent two-state systems has been previously rationalized to understand the unfolding features of staphylococcal nuclease, ribonuclease A and T1, and  $\alpha$ -lactalbumin. From these works, it was determined that increases in  $m$ -values concomitant with increases in pH can be explained by an increase in cooperativity due the destabilization of the I and U states relative to the N state<sup>22</sup>. Examination of the  $m$ -values obtained for the wt protein show this behaviour: upon lowering the pH, the apparent unfolding cooperativity of the  $I \rightleftharpoons U$  decreases. This could be an indicator of the presence of the previous characterized intermediate<sup>11</sup>, which would be stabilized at low pH.

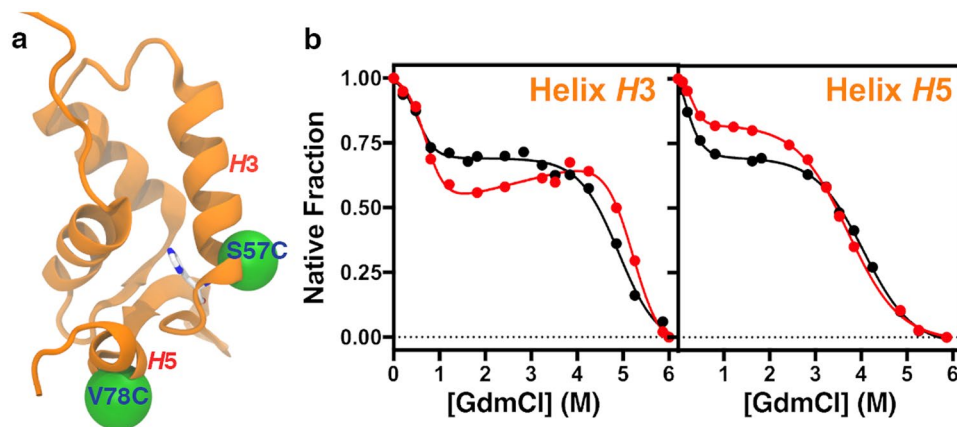
In contrast to the wt protein, the equilibrium unfolding of the H59A mutant followed by intrinsic fluorescence showed two clear transitions (Fig. 3b); the first, observed between 1–2 M of denaturant, whereas the second transition matches the  $I \rightleftharpoons U$  transition described for the wt protein (Fig. 3a). It is worth noting that the first transition is in good agreement with the  $N_2 \rightleftharpoons 2I$  transition ascertained for wt FoxP1 using circular dichroism<sup>11</sup>. This is consistent with the fact that a large proportion of dimer is expected at a protein concentration of 15  $\mu$ M, based on the  $K_d$  values throughout all the pH range analysed. In order to corroborate that the unfolding behavior is explained by a three-state  $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$  mechanism as in the case of the wt protein when monitored by circular dichroism, we performed an unfolding experiment at pH 5.0 using 3  $\mu$ M of total protein (Fig. 3b inset), 5-fold lower than the previous experiment. Data obtained at this protein concentration showed a clear shift in the denaturant concentration midpoint ( $C_m$ ) of the first transition ( $C_m[3 \mu M] = 0.6 M$ ;  $C_m[15 \mu M] = 1.3 M$ ), consistent with a three-state  $N_2 \rightleftharpoons 2I \rightleftharpoons 2U$  mechanism. These results suggest that changes in the local environment of H59 are possibly affecting the quantum yield of any of the three tryptophan residues in FoxP1, thus turning intrinsic fluorescence into a good observable property to monitor the local unfolding and dissociation events preceding the formation of the monomeric intermediate.

Data derived from unfolding curves were used to calculate the energetic parameters of the H59A mutant (Table 2), corroborating the relative insensitivity of  $m$ -values and  $\Delta G_1$  from the  $N_2 \rightleftharpoons 2I$  transition as in the kinetic analysis of the dimer–monomer reaction (Fig. 2 and Table 1). Despite that the  $N_2 \rightleftharpoons 2I$  event is largely unaffected by pH (Fig. 3b and Table 2), the protonation change in the remaining H54 residue has an implicit energetic cost of ~1 kcal·mol<sup>-1</sup> on  $\Delta G_2$  of the intermediate unfolding event ( $I \rightleftharpoons U$ ), which is moderate when compared to wt FoxP1 (Fig. 4c).

As previously done for the kinetic measurements with the H59A mutant and wt protein, we also analysed the  $\Delta G_U$  dependence of both proteins on pH changes using the Wyman–Tanford approach. With this analysis, we attempted to evaluate the possible changes in pKa values of the histidine(s) during unfolding of FoxP1 (Eq. 2), indicating differential protonation states between intermediate and unfolded conformations (Fig. 3c). The pKa values calculated for the H59A mutant showed only a moderate increase in the unfolded state ( $pK_{aI} = 5.3$ ;  $pK_{aU} = 5.9$ ), indicating that this residue is scarcely altered in the unfolded ensemble of the protein. However, pKa values derived from the wt protein using the pKa values from H54 as fixed parameters showed a dramatic change in the protonation behaviour of H59 residue ( $pK_{aI} = 5.7$ ;  $pK_{aU} = 7.7$ ).

Altogether, these results indicate that (i) protonation changes in residue H54 are mostly related to modulation of protein–DNA interactions<sup>9,10,16</sup>, and (ii) H59 is preferentially deprotonated in order to maintain a native-like environment and increase conformational stability.

**Protonation changes of H59 have a localized impact in folding stability.** Given that the protonation/deprotonation equilibrium of H59 is highly involved in the domain swapping stability of FoxP1, we analysed their local effects by creating two single-cysteine mutants within helix H3 (S57C) and H5 (V78C), both in the vicinity of H59, to attach an extrinsic fluorescent dye sensible to local structural changes (Oregon Green 488; OG488). These regions were chosen since they are exchanged with the adjacent monomer during domain swapping of FoxP1 (Fig. 4a).



**Figure 4.** Equilibrium unfolding experiments of helices *H3* (S57C) and *H5* (V78C). (a) Structural positions of the attached fluorophores onto the specific mutants S57C (*H3*) and V78C (*H5*). (b) Unfolding experiments as a function of the GdmCl concentration ascertained by measuring the changes in fluorescence anisotropy for S57C and V78C. Experiments were carried out using 0.1  $\mu\text{M}$  of labelled protein, incubated with different denaturant concentrations at pH 5.0 (black) and 7.8 (red). Data obtained in all experiments were fitted to a three-state monomer unfolding.

pH	S57C				V78C			
	$\Delta G_1^\ddagger$	$m_1^\ddagger$	$\Delta G_2^\ddagger$	$m_2^\ddagger$	$\Delta G_1^\ddagger$	$m_1^\ddagger$	$\Delta G_2^\ddagger$	$m_2^\ddagger$
5.0	$2.0 \pm 0.5$	$3.2 \pm 1.1$	$6.9 \pm 0.8$	$1.4 \pm 0.2$	$1.4 \pm 0.3$	$3.1 \pm 0.5$	$4.2 \pm 0.2$	$1.1 \pm 0.1$
7.8	$1.6 \pm 0.4$	$2.7 \pm 0.6$	$10.4 \pm 1.1$	$2.0 \pm 0.4$	$0.6 \pm 0.1$	$5.0 \pm 0.9$	$4.7 \pm 0.5$	$1.1 \pm 0.1$

**Table 3.** Thermodynamic parameters of local stability of S57C and V78C mutants at 37 °C.  $^\ddagger$ Expressed as  $\text{kcal}\cdot\text{mol}^{-1}$ .  $^\dagger$ Expressed as  $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{M}^{-1}$ .

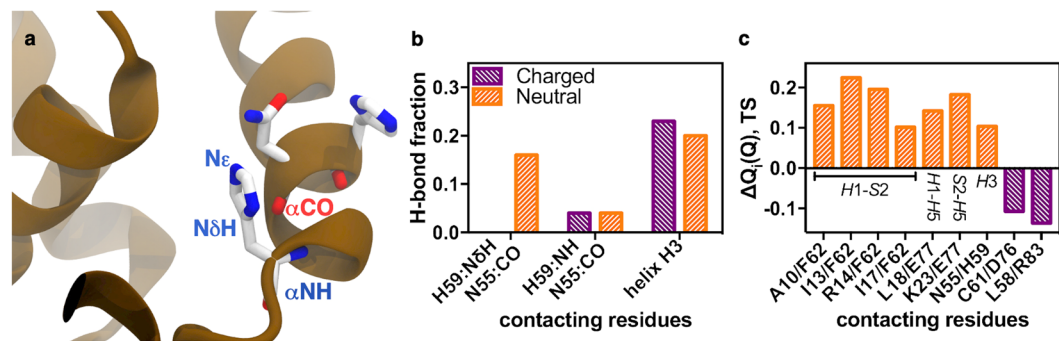
Freshly labelled, SEC-purified S57C and V78C proteins were used for equilibrium unfolding experiments in which local changes were monitored using fluorescence anisotropy at pH 5.0 and 7.8. In these conditions, the imidazole rings of H59 and H54 are expected to be fully protonated or deprotonated, respectively, based on the calculated pKa values. Moreover, we used protein concentrations of 0.1  $\mu\text{M}$ , in which the monomer fraction is the most populated state, to focus our measurements on folding stability effects. Results for both proteins clearly indicate two transitions at pH 5.0 and 7.8 (Fig. 4b), thus being consistent with a three-state  $\text{N} \rightleftharpoons \text{I} \rightleftharpoons \text{U}$  monomer unfolding mechanism. The first transition falls between 0.2–1 M of denaturant for both mutants, but the second transition occurs between 4–6 M of denaturant for S57C and 3–5 M for V78C (Fig. 4b), suggesting that unfolding of helix *H5* precedes that of *H3*.

By fitting the fluorescence anisotropy data from both proteins to a three-state  $\text{N} \rightleftharpoons \text{I} \rightleftharpoons \text{U}$  monomer unfolding mechanism, we determined that the first transition ( $\text{N} \rightleftharpoons \text{I}$ ) shows no significant differences in its  $\Delta G_1$  value regardless of the pH. Moreover, the changes in stability of helices *H3* and *H5* during the  $\text{N} \rightleftharpoons \text{I}$  transition, with  $\Delta G_1$  values ranging between 0.6–2  $\text{kcal}\cdot\text{mol}^{-1}$  (Table 3), are low when compared to the global value obtained for the wt and H59A mutant in conditions where the dimer is the most populated species (Table 2). Nevertheless, a significant increase in  $m$ -value for helix *H5* (Table 3) is observed when H54 and H59 are deprotonated (pH 7.8) whereas helix *H3* is unaffected, suggesting different impacts of the changes in the protonation state of H59 on the local cooperativity of the  $\text{N} \rightleftharpoons \text{I}$  transition.

The second transition ( $\text{I} \rightleftharpoons \text{U}$ ) reveals relevant energetic differences between helix *H3* and *H5*. First, the free energy change of unfolding of the intermediate state ( $\Delta G_2$ ) monitored from helix *H3* at pH 5.0 is clearly higher than for helix *H5* by 2  $\text{kcal}\cdot\text{mol}^{-1}$ . These differences highlight different stabilities that depend on the structural context of both helices. Concomitantly, deprotonation of H59 (pH 7.8) leads to a significant increase in stability of around 3.5  $\text{kcal}\cdot\text{mol}^{-1}$  of helix *H3* and not *H5*, which is similar to the energetic change ( $\Delta\Delta G_U$  [pH 7.8–pH 5.0]) obtained from global unfolding analysis with the wt protein. These results suggest that H59 is involved in the pH-dependent stabilization of helix *H3*.

**Protonation of H59 alters hydrogen bond networks and long-range interactions in FoxP1.** The experimental evidence above indicates the relevance of the protonation state of H59 in maintaining the folding stability and cooperativity of the region comprising helix *H3* in FoxP1. These results can be attributed to the ability of the imidazole ring to establish hydrogen bonds, which are lost in the protonated form<sup>23</sup>.

In order to decipher the specific H-bonds gained or lost due to changes in the protonation state of H59, we performed all-atom molecular dynamics (MD) simulations using either protonated (HSP) or deprotonated (HSD) H59 (Supplementary Fig. S2). It is worth noting that, although the H++ server estimates that both H54 and H59 are titratable under the experimentally evaluated pH range herein, we only modified the charge state



**Figure 5.** Hydrogen bond and contact probability changes due to protonation of H59 in MD simulations. (a) Structural representation of helix *H3*, depicting H59 and N55 interaction, and showing the side chain nitrogen atoms ( $\delta$  and  $\epsilon$ ) of the H59 around carbonyl group (CO) of N55. (b) Side chain backbone and backbone-backbone hydrogen bond quantifications between charged (purple) or neutral (orange) H59 and N55 from helix *H3*, as well as total hydrogen bonds for helix *H3*. (c) Differences in contact probability on the folding TS for monomeric FoxP1 between residues from helices *H1*, *H3* and *H5* and strand *S2*. Changes towards higher contact probability for the deprotonated H59 are shown in orange, whereas changes favouring the protonated state are shown in purple.

of H59 due to the negligible impact of H54 in protein folding and stability on FoxP1 (Fig. 3) and its side chain location on the protein surface (Fig. 1) to interact with the DNA<sup>16</sup>. Interestingly, the major differences observed between both simulations are located within helix *H3*. The introduction of a protonated histidine introduces a perturbation in the N-terminal region of this secondary structure element due to changes in hydrogen bonds between H59 and N55 from helix *H3* (Fig. 5). Quantification of hydrogen bonds using a distance (donor-acceptor distance  $\leq 3.0$  Å) and angle ( $\leq 20^\circ$ ) criterion show that the imidazole ring of H59 and the backbone carbonyl group of N55 interact during  $\sim 20\%$  of the simulation only in the deprotonated state, whereas distances between the backbone amide carbonyl groups of the same residues did not show relevant differences upon deprotonation (Fig. 5b). This change causes an increase of hydrogen bonding on helix *H3* (Fig. 5b). These results suggest that the interaction of the side chain of deprotonated histidine and the carbonyl group of asparagine is key to stabilize the N-terminal region of helix *H3*.

Given that typical explicit-solvent MD do not regularly sample protein folding transitions, except in cases where special-purpose supercomputers for MD simulations are employed on proteins that fold below the millisecond timescale<sup>24,25</sup>, and that charges also have a strong effect on protein folding and stability due to non-native interactions<sup>26</sup>, we performed folding simulations using coarse-grained structure-based models<sup>27</sup> augmented by the use of charges<sup>26</sup> in all E, D, K and R residues, while also manipulating the charge state of H59. Simulations were run at several temperatures above and below  $T_F$  allowed sampling several ( $>100$ ) folding-unfolding transitions, thus allowing to integrate the data and calculate the folding landscape of monomeric FoxP1 (Supplementary Fig. S3). From these simulations, we observed a slight increase in  $T_F$  upon deprotonation of H59. Then, we used folding equilibrium simulations at  $T_F$  of the protonated FoxP1 to determine which native contacts were responsible for this change in stability and whether these changes occurred in the transition state (TS) or the native state. As shown in Fig. 5c, 7 native interactions showed a significant ( $>10\%$ ) increase in contact probability in the TS upon deprotonation. Moreover, these changes in contact probability were not present in the native state (data not shown). Most of the interactions involved regions near H59 (Supplementary Fig. S4), including four interactions between helix *H1* and strand *S2* (residue F62), and the native contact between H59 and N55 (Fig. 5c). The latter is in line with the results from explicit-solvent MD simulations for monomeric FoxP1 (Fig. 5b).

Altogether, these computational results aid in understanding how the deprotonation of H59 leads to local stabilization of helix *H3* and global stabilization of the transition state and native-like intermediate state of FoxP1.

## Discussion

In this work, the dependence of FoxP1 folding and dimerization on the protonation state of evolutionary histidines were analysed using both wt and the H59A mutant. Dissociation measurements indicated that protonation of the FoxM/O/P-exclusive H59, and not the family-conserved H54, imposes an energetic modulation of  $\sim 2.5$  kcal·mol<sup>-1</sup> in the monomer-dimer equilibrium (Fig. 2). This effect is observed only in the association process (Table 1), indicating that monomer and dimer are not equally sensitive to the protonation of the H59 residue. Despite this behaviour, the energetic barrier calculated between monomer and dimer is dramatically lower than in other domain swapping models<sup>20,28,29</sup>, which explains why the dimerization equilibrium of FoxP proteins is reached in hours under physiological conditions<sup>11,30,31</sup>.

In the same line, global unfolding experiments also indicated the direct relationship between protonation equilibrium of residue H59 and folding stability, observing the presence of an intermediate whose population is increased when this residue is protonated. H59A mutant only showed a change in  $\sim 1$  kcal·mol<sup>-1</sup> in these conditions (Fig. 3), corroborating both the relevance of H59 in the structural properties of FoxP1 and the lesser impact of the chemical properties of H54 in this context. Anisotropy measurements indicated, firstly, differences in local stability between helices *H3* and *H5* and, secondly, that the stability of helix *H3* is affected by protonation changes of H59 (Fig. 4). These results strongly suggest that local structure changes modulated by changes in the

protonation of histidines have an impact in global protein stability of FoxP1. This is not surprising, considering the relevance of the imidazole ring in modulating a myriad of aspects of protein folding<sup>23,32–35</sup> and function<sup>23,36</sup> through hydrogen bond (donor and acceptor), cation– $\pi$  and cation–aromatic interactions<sup>23,37,38</sup>.

Our results are further explained by MD simulations that ascertain the protonation state of H59, suggesting that its stabilization effect is mediated by hydrogen bonds between the imidazole ring and the carbonyl group of the backbone of N55 (Fig. 5b). Simulations performed with the H59N mutant of FoxP1 (conserved in most of Fox members) corroborated the relevance of the persistence of this interaction (data not shown). Moreover, the protonation state of H59 has an impact in increasing the stability of specific residue–residue contacts within the folding transition state of FoxP1, as shown by folding simulations using SBM models (Fig. 5c).

The significance of the single–point change from asparagine to histidine observed only in specific Fox sub-families seems to be non-trivial when physicochemical properties between these residues are compared in terms of protein stability. Both residues possess the ability to interact with others as hydrogen bond donors and/or acceptors<sup>39</sup>. However, only histidine is affected by pH in a physiological range due to its pKa, acting as a molecular switch that modulates the microenvironment of several active sites and relevant regions of proteins<sup>23,36,40</sup>. This mutation could then represent an ancestral and specific physicochemical change to locally destabilize the monomer structure due to microenvironment changes, promoting the loosening of local regions and the association via domain–swapping, which in turn allows stabilizing these elements in an intermolecular context.

A possible explanation of our results is thus related with the evolutionary context of Fox proteins. We postulate that the presence of this evolutionary histidine, which is only conserved in M, O and P subfamilies, is one of the many accumulative transitions in the pathway towards the emergence of the dimerization ability of FoxP proteins, summative to the prerequisite of alanine (FoxP members) instead of proline (most Fox proteins) in the hinge loop region to enable domain swapping<sup>11</sup>.

## Methods

**Protein expression and purification.** Codon-optimized DNA sequence encoding the forkhead domain of human FoxP1 and its mutants were cloned into a modified pET-28a vector, containing a His6-tag, a TEV cleavage site and an S-tag sequence in the 5' end of the gene. Amino acid residues are numbered according to the sequence numbering in the deposited structure of the A39P/C61Y mutant of the forkhead domain of human FoxP1 (PDB ID 2KIU)<sup>8</sup>. Single-point mutants were generated by PCR using the QuickChange Site-directed Mutagenesis kit (Stratagene, La Jolla, CA, USA). Proteins were purified as described by Medina *et al.*<sup>11</sup> and dialysed into standard buffer (20 mM HEPES pH 7.8, 150 mM NaCl, 2 mM  $\beta$ -mercaptoethanol) prior to each experiment, unless otherwise indicated.

**Size-exclusion chromatography (SEC).** Monomer and dimer fractions of wt FoxP1 and H59A mutant were quantified and isolated on a Waters Breeze HPLC system (Waters Corporation, Milford, MA, USA) using a Superdex 75 10/30 column (GE Healthcare Biosciences, Pittsburgh, PA, USA) as described in Medina *et al.*<sup>11</sup>. Briefly, the column was equilibrated with 45 ml of mobile phase (standard buffer) at room temperature before SEC experiments. When required, isolated fractions were kept on ice until their use in subsequent experiments.

**Dissociation kinetics.** Isolated dimer populations of wt FoxP1 and H59A mutant were diluted into standard buffer at different pH values (5.0, 5.5, 6.0, 7.0 and 7.8) to a final protein concentration of 5  $\mu$ M and further incubated for 2 h at 37 °C in a water bath. Aliquots after different times of incubation were analysed using SEC, and dimer and monomer concentrations were quantified using the area under the curve for each species compared to the total area of the corresponding elution profile.

**Equilibrium unfolding experiments.** Unfolding experiments were performed in buffer at different pH values (5.0, 5.5, 6.0, 7.0 and 7.8) supplemented with GdmCl (molecular biology grade; Thermo Fischer Scientific, Waltham, MA) at concentrations ranging from 0 to 6 M. For all unfolding experiments, the proteins were exposed to different GdmCl concentrations for at least 2 h at 37 °C using a water bath. Unfolding curves were obtained using 3 or 15  $\mu$ M of protein in terms of monomer concentration. For extrinsically labelled proteins, unfolding experiments were performed using 0.1  $\mu$ M of protein in terms of monomer concentration. All curves correspond to experiments performed in triplicates.

**Fluorescence labelling of FoxP1.** We prepared FoxP1 constructs in which the native cysteine residue (at position 61) was replaced by serine (C61S). Then, single-cysteine mutants S57C and V78C were generated to allow specific labelling. Before labelling, all buffers were sterile filtered and degassed. FoxP1 was concentrated to ~100  $\mu$ M in buffer A (20 mM HEPES pH 7.8, 150 mM NaCl, 2 M GdmCl) supplemented with 0.5 mM Tris 2-carboxyethyl phosphine hydrochloride (TCEP). Then, 2.5 ml of concentrated protein were loaded onto a PD10 desalting column (GE Healthcare) and the protein was eluted with freshly degassed 3.5 ml of buffer A without TCEP. The eluted protein was immediately labelled with Oregon Green 488 (OG488) maleimide fluorophore (Thermo Fischer Scientific). Finally, SEC was performed to remove free fluorophore excess.

**Fluorescence measurements.** Both intrinsic fluorescence and anisotropy measurements were performed in a Jasco FP-8300 spectrofluorometer (Jasco Corp, Japan), employing cells with 1 cm path lengths. In intrinsic fluorescence determinations, proteins were incubated in the specified buffer and spectra were recorded between 305 and 450 nm after excitation at 295 nm. The fluorescence intensity at 324 nm was used to monitor tertiary changes in both wt FoxP1 and H59A mutant. For anisotropy measurements, parallel and perpendicular fluorescent emission at 525 nm, after excitation at 495 nm, was recorded using labelled proteins. The G factor was determined using free OG488 dye at 1  $\mu$ M in the specified buffer to calculate the observed anisotropy. No changes in G factor were observed upon changing the denaturant concentration (data not shown).



**Data analysis.** Dissociation kinetics at specified pH values were fitted to a single exponential decay model, obtaining the observed dissociation rate ( $k_{\text{off}}$ ) and the equilibrium dimer fraction. Dimer fractions obtained in SEC experiments were used to determine the dissociation constant ( $K_d$ ) at the specified pH, as previously described<sup>11</sup>. Then, the association rates ( $k_{\text{on}}$ ) at each pH were determined using both  $K_d$  and  $k_{\text{off}}$  according to Equation 1:

$$K_d = \frac{k_{\text{off}}}{k_{\text{on}}} \quad (1)$$

Dissociation rates were analysed in terms of Eyring equation<sup>41</sup> to determine the free energy change between monomers or dimers and the transition state,  $\Delta G_{\text{on}}^\ddagger$  and  $\Delta G_{\text{off}}^\ddagger$  respectively.

Data from unfolding curves was fitted in terms of two-state (wt) or three-state mechanisms (H59A, S57C-OG488 and V78C-OG488) to obtain the free energy changes between the native (N), intermediate (I) and unfolded (U) states<sup>11</sup>.

The dependence of the resulting data on pH was determined using the Wyman-Tanford relationship<sup>42</sup> (Equation 2):

$$\Delta G = -RT \sum_i \ln \left\{ \frac{[(1 + 10^{(\text{pKa}_i^A - \text{pH})}) \cdot (1 + 10^{(\text{pKa}_i^B - \text{pH}_{\text{ref}})})]}{[(1 + 10^{(\text{pKa}_i^B - \text{pH})}) \cdot (1 + 10^{(\text{pKa}_i^A - \text{pH}_{\text{ref}})})]} \right\} \quad (2)$$

where R is the gas constant (in kcal·mol<sup>-1</sup>·K<sup>-1</sup>), T is temperature (310 K),  $\text{pKa}_i^A$  and  $\text{pKa}_i^B$  represent the pKa value of residue  $i$  in states A and B, and  $\text{pH}_{\text{ref}}$  corresponds to 5.0.

For the dissociation analysis,  $\text{pKa}_i^A$  represents the pKa value of residue  $i$  in either the dimer or monomer states, whereas  $\text{pKa}_i^B$  is the pKa value of residue  $i$  in the transition state ( $\ddagger$ ). For the folding analysis,  $\text{pKa}_i^A$  represents the pKa value of  $i$  residue in the I state, and  $\text{pKa}_i^B$  is the pKa value of  $i$  residue in the U state. In the case of mutant H59A, only one residue (H54) was assumed to be sensitive to pH changes. To obtain pKa values for H59 from Eq. 2, data from H59A was used as fixed parameters to obtain the second residue (H59) in the wt protein.

Fitting procedures were performed using the software GraphPad Prism 6.0c ([www.graphpad.com](http://www.graphpad.com)).

**Molecular dynamics simulations.** All-atom molecular dynamics (MD) were performed in NAMD 2.9<sup>43</sup> using the CHARMM36m force field<sup>44</sup>. Initial models were taken from the monomeric structure of the A39P/C61Y mutant of human FoxP1 (PDB ID 2KIU)<sup>8</sup>. Mutated residues were reversed to the ones in the wt protein using VMD<sup>45</sup>. Two systems were prepared with this structure, one with neutral (HSD) and another with protonated (HSP) H59. Proteins were solvated in TIP3P water boxes with periodic boundary conditions, and charges were neutralized by the addition of Cl<sup>-</sup> counter ions using VMD. Systems were minimized for 4,000 steps using the conjugated gradient algorithm. After minimization, proteins were slowly heated from 0 to 300 K and the whole system was equilibrated in the NPT ensemble for 20 ns. Three independent production MD simulations for each system were performed for 40 ns, using a time step of 2 fs. Inter-residue distance and hydrogen bond analysis were performed in VMD.

**Folding simulations.** Coarse-grained structure-based models (SBM) of monomeric FoxP1 were generated using the SMOG web server<sup>27</sup> using the default parameters. In these models each residue is represented as a single bead centred at its corresponding C $\alpha$  atom. Bonded interactions are maintained by harmonic restraints, whereas only non-bonded residues separated in sequence by at least 3 residues and in contact in the native state (i.e. the distance between any pair of atoms from these residues is less than 6 Å) are given attractive Lennard-Jones interactions<sup>46</sup> (Equation 3):

$$V_{\text{SBM}} = \sum_{i < j - 3} \varepsilon_{\text{C}} \left[ 5 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 6 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{10} \right] + \sum_{ij \notin \text{contacts}} \varepsilon_{\text{NC}} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} \quad (3)$$

where  $\varepsilon_{\text{C}}$  and  $\varepsilon_{\text{NC}}$  are the energies for native and non-native contacts,  $\sigma_{ij}$  is the distance between residue pairs in the native state and  $r_{ij}$  is their distance during the simulation.

To account for charge effects, these SBM models were supplemented with non-native charge interactions via a Debye-Hückel (DH) potential (Equation 4) as in previous works<sup>26</sup>:

$$V_{\text{DH}} = k_{\text{elec}} B(\lambda_{\text{D}}) \sum_{i,j} \frac{q_i q_j e^{(-r_{ij}/\lambda_{\text{D}})}}{\varepsilon r_{ij}} \quad (4)$$

where  $k_{\text{elec}}$  is the Debye constant,  $\lambda_{\text{D}}$  is the Debye length that depends on the salt concentration (here, we used 0.05 M of monovalent salt, thus  $\lambda_{\text{D}}$  is 0.735<sup>47</sup>),  $B(\lambda_{\text{D}})$  is the Debye coefficient of the solution (corresponding to  $\sim 1$  for dilute solutions<sup>47</sup>) and  $\varepsilon$  is the dielectric constant (here, 80). Charges  $q_i$  and  $q_j$  were placed on the C $\alpha$  atoms corresponding to Asp, Glu (-1), Arg and Lys (+1), whereas H59 was either treated as neutral or protonated, in which case a positive charge was added.

Simulations at 30 different temperatures around the folding temperature  $T_{\text{f}}$ , corresponding to the protein being always folded to always unfolded, were performed on GROMACS 4.5.4<sup>48</sup> for 10<sup>7</sup> steps using a time step  $\tau$  of 0.0005. The weighted histogram analysis method<sup>49</sup> was used to combine the simulation data from different temperatures into single free-energy profiles as a function of the global fraction of native contacts (Q). For this

reaction coordinate, two residues are considered in contact if their distance is within 1.2 times the native distance. Also, the probability of a native contact being formed,  $P_{(i,j)}$ , was calculated from simulations at  $T = T_p$ .

## Data Availability

All experimental and computational data is available from the corresponding authors on request.

## References

- Zangi, R. Side-chain–side-chain interactions and stability of the helical state. *Phys. Rev. E* **89**, 012723 (2014).
- Yang, X., Vadrevu, R., Wu, Y. & Matthews, C. R. Long-range side-chain-main-chain interactions play crucial roles in stabilizing the ( $\beta\alpha$ ) 8 barrel motif of the alpha subunit of tryptophan synthase. *Protein Sci.* **16**, 1398–1409 (2007).
- Camps, M., Herman, A., Loh, E. & Loeb, L. A. Genetic Constraints on Protein Evolution. *Crit. Rev. Biochem. Mol. Biol.* **42**, 313–326 (2007).
- Rasoolizadeh, A., Goulet, M. C., Sainsbury, F., Cloutier, C. & Michaud, D. Single substitutions to closely related amino acids contribute to the functional diversification of an insect-inducible, positively selected plant cystatin. *FEBS J.* **283**, 1323–1335 (2016).
- Shimeld, S. M., Degnan, B. & Luke, G. N. Evolutionary genomics of the Fox genes: Origin of gene families and the ancestry of gene clusters. *Genomics* **95**, 256–260 (2010).
- Le Lay, J. & Kaestner, K. H. The Fox genes in the liver: from organogenesis to functional integration. *Physiol. Rev.* **90**, 1–22 (2010).
- Hannenhalli, S. & Kaestner, K. H. The evolution of Fox genes and their role in development and disease. *Nat. Rev. Genet.* **10**, 233–240 (2009).
- Chu, Y. P. *et al.* Solution structure and backbone dynamics of the DNA-binding domain of FOXP1: Insight into its domain swapping and DNA binding. *Protein Sci.* **20**, 908–924 (2011).
- Bandukwala, H. S. *et al.* Structure of a Domain-Swapped FOXP3 Dimer on DNA and Its Function in Regulatory T Cells. *Immunity* **34**, 479–491 (2011).
- Stroud, J. C. *et al.* Structure of the forkhead domain of FOXP2 bound to DNA. *Structure* **14**, 159–166 (2006).
- Medina, E. *et al.* Three-Dimensional Domain Swapping Changes the Folding Mechanism of the Forkhead Domain of FoxP1. *Biophys. J.* **110**, 2349–2360 (2016).
- Bennett, M. J., Choe, S. & Eisenberg, D. Domain swapping: entangling alliances between proteins. *Proc. Natl. Acad. Sci.* **91**, 3127–3131 (1994).
- Rousseau, F., Schymkowitz, J. W. H. & Itzhaki, L. S. The unfolding story of three-dimensional domain swapping. *Structure* **11**, 243–251 (2003).
- Liu, Z. & Huang, Y. Evidences for the unfolding mechanism of three-dimensional domain swapping. *Protein Sci.* **22**, 280–286 (2013).
- van der Vliet, H. J. J. & Nieuwenhuis, E. E. IPEX as a Result of Mutations in FOXP3. *Clin. Dev. Immunol.* **2007**, 1–5 (2007).
- Blane, A. & Fanucchi, S. Effect of pH on the Structure and DNA Binding of the FOXP2 Forkhead Domain. *Biochemistry* **54**, 4001–4007 (2015).
- Sanders, D. A. *et al.* FOXM1 binds directly to non-consensus sequences in the human genome. *Genome Biol.* **16**, 130 (2015).
- Salamanca Vilorio, J., Allega, M. F., Lambrugh, M. & Papaleo, E. An optimal distance cutoff for contact-based Protein Structure Networks using side-chain centers of mass. *Sci. Rep.* **7**, 2838 (2017).
- Anandkrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **40**, W537–W541 (2012).
- Kang, X. *et al.* Foldon unfolding mediates the interconversion between Mpro-C monomer and 3D domain-swapped dimer. *Proc. Natl. Acad. Sci.* **109**, 14900–14905 (2012).
- Horng, J. C., Cho, J. H. & Raleigh, D. P. Analysis of the pH-dependent folding and stability of histidine point mutants allows characterization of the denatured state and transition state for protein folding. *J. Mol. Biol.* **345**, 163–173 (2005).
- Whitten, S. T., Wooll, J. O., Razeghifard, R., García-Moreno, E. B. & Hilser, V. J. The origin of pH-dependent changes in m-values for the denaturant-induced unfolding of proteins. *J. Mol. Biol.* **309**, 1165–1175 (2001).
- Liao, S.-M., Du, Q.-S., Meng, J.-Z., Pang, Z.-W. & Huang, R.-B. The multiple roles of histidine in protein interactions. *Chem. Cent. J.* **7**, 44 (2013).
- Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science* **334**, 517–520 (2011).
- Piana, S., Lindorff-Larsen, K. & Shaw, D. E. Atomic-level description of ubiquitin folding. *Proc. Natl. Acad. Sci.* **110**, 5915–5920 (2013).
- Tzul, F. O., Schweiker, K. L. & Makhatadze, G. I. Modulation of folding energy landscape by charge–charge interactions: Linking experiments with computational modeling. *Proc. Natl. Acad. Sci.* **112**, E259–E266 (2015).
- Noel, J. K., Whitford, P. C., Sanbonmatsu, K. Y. & Onuchic, J. N. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* **38**, W657–W661 (2010).
- Moschen, T. & Tollinger, M. A kinetic study of domain swapping of Protein L. *Phys. Chem. Chem. Phys.* **16**, 6383–6390 (2014).
- Bucci, E. *et al.* On the thermal stability of the two dimeric forms of ribonuclease A. *Biophys. Chem.* **116**, 89–95 (2005).
- Liu, L., Byeon, I.-J. L., Bahar, I. & Gronenborn, A. M. Domain Swapping Proceeds via Complete Unfolding: A 19 F- and 1 H-NMR Study of the Cyanovirin-N Protein. *J. Am. Chem. Soc.* **134**, 4229–4235 (2012).
- Schymkowitz, J. W., Rousseau, F., Irvine, L. R. & Itzhaki, L. S. The folding pathway of the cell-cycle regulatory protein p13suc1: clues for the mechanism of domain swapping. *Structure* **8**, 89–100 (2000).
- Plesniak, L. A., Connelly, G. P., McIntosh, L. P. & Wakarchuk, W. W. Characterization of a buried neutral histidine residue in *Bacillus circulans* xylanase: Nmr assignments, pH titration, and hydrogen exchange. *Protein Sci.* **5**, 2319–2328 (1996).
- Langella, E., Improtta, R. & Barone, V. Checking the pH-Induced Conformational Transition of Prion Protein by Molecular Dynamics Simulations: Effect of Protonation of Histidine Residues. *Biophys. J.* **87**, 3623–3632 (2004).
- Hansen, J. E., McBrayer, M. K., Robbins, M. & Suh, Y. A pH dependence study on the unfolding and refolding of apoazurin: Comparison with Zn(II) azurin. *Cell Biochem. Biophys.* **36**, 19–40 (2002).
- Fernández-Recio, J., Romero, A. & Sancho, J. Energetics of a hydrogen bond (charged and neutral) and of a cation- $\pi$  interaction in apoflavodoxin. *J. Mol. Biol.* **290**, 319–330 (1999).
- Zhu, R. *et al.* Allosteric histidine switch for regulation of intracellular zinc(II) fluctuation. *Proc. Natl. Acad. Sci.* **114**, 13661–13666 (2017).
- Loewenthal, R., Sancho, J. & Fersht, A. R. Histidine-aromatic interactions in barnase. *J. Mol. Biol.* **224**, 759–770 (1992).
- Markley, J. L. & Westler, W. M. Protonation-state dependence of hydrogen bond strengths and exchange rates in a serine protease catalytic triad: Bovine chymotrypsinogen A. *Biochemistry* **35**, 11092–11097 (1996).
- De Vos, S., Doumen, J., Langhorst, U. & Steyaert, J. Dissecting histidine interactions of ribonuclease T1 with asparagine and glutamine replacements: analysis of double mutant cycles at one position. *J. Mol. Biol.* **275**, 651–661 (1998).
- Park, C. & Raines, R. T. Dimer formation by a ‘monomeric’ protein. *Protein Sci.* **9**, 2026–2033 (2000).
- Eyring, H. The Activated Complex and the Absolute Rate of Chemical Reactions. *Chem. Rev.* **17**, 65–77 (1935).
- Tollinger, M., Crowhurst, K. A., Kay, L. E. & Forman-Kay, J. D. Site-specific contributions to the pH dependence of protein stability. *Proc. Natl. Acad. Sci.* **100**, 4545–4550 (2003).
- Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005).

44. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
45. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
46. Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953 (2000).
47. Givaty, O. & Levy, Y. Protein Sliding along DNA: Dynamics and Structural Characterization. *J. Mol. Biol.* **385**, 1087–1097 (2009).
48. Pronk, S. *et al.* GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
49. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **13**, 1011–1021 (1992).

## Acknowledgements

This work was supported by Fondo de Desarrollo Científico y Tecnológico (FONDECYT grant No 1170701 and 11140601). P. Villalobos was supported by doctoral fellowship from Comisión Nacional de Investigación Científica y Tecnológica (CONICYT grant No 21151101). This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). Authors thank Victoria Muñoz and Leonardo Sáez for their collaboration in obtaining the H59A mutant and its preliminary analysis.

## Author Contributions

E.M., P.V., J.B. and C.A.R.-S. designed research; E.M., P.V., R.C. and C.A.R.-S. performed research; E.M., P.V. and C.A.R.-S. analysed data; E.M., J.B. and C.A.R.-S. wrote the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-41819-5>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019