

# SCIENTIFIC REPORTS



OPEN

## Fine-tuning the extent and dynamics of binding cleft opening as a potential general regulatory mechanism in parvulin-type peptidyl prolyl isomerases

András Czajlik<sup>1,\*</sup>, Bertalan Kovács<sup>1,\*</sup>, Perttu Permi<sup>2</sup> & Zoltán Gáspári<sup>1</sup>

Received: 15 November 2016

Accepted: 06 February 2017

Published: 16 March 2017

Parvulins or rotamases form a distinct group within peptidyl prolyl cis-trans isomerases. Their exact mode of action as well as the role of conserved residues in the family are still not unambiguously resolved. Using backbone  $S^2$  order parameters and NOEs as restraints, we have generated dynamic structural ensembles of three distinct parvulins, SaPrsA, TbPin1 and CsPinA. The resulting ensembles are in good agreement with the experimental data but reveal important differences between the three enzymes. The largest difference can be attributed to the extent of the opening of the substrate binding cleft, along which motional mode the three molecules occupy distinct regions. Comparison with a wide range of other available parvulin structures highlights structural divergence along the bottom of the binding cleft acting as a hinge during the opening-closing motion. In the prototype WW-domain containing parvulin, Pin1, this region is also important in forming contacts with the WW domain known to modulate enzymatic activity of the catalytic domain. We hypothesize that modulation of the extent and dynamics of the identified 'breathing motion' might be one of the factors responsible for functional differences in the distinct parvulin subfamilies.

Peptidyl prolyl isomerases (PPIases) catalyze the isomerization of the peptide bond preceding proline residues. PPIases play an important role not only in protein folding but also in the regulation of several of biological processes like chromatin remodeling, transcription and nuclear receptor signaling<sup>1</sup>. They can be divided into three non-homologous and structurally different families, known as cyclophilins, FKBP (FK506 binding protein and its relatives), and parvulins. The latter, highly conserved subfamily consists of small (~10 kDa) proteins that are present in both pro- and eukaryotes<sup>2</sup>. Their structure consists of a four-stranded antiparallel  $\beta$ -sheet surrounded by four  $\alpha$ -helices ( $\alpha\beta\beta\alpha\beta$ , parvulin fold)<sup>3</sup>. Parvulins play key roles in many important biological processes including the cell-cycle regulation, apoptosis and protein quality control<sup>4,5</sup>. Due to these functions, they are involved in neurodegenerative disorders such as Parkinson's and Alzheimer's disease<sup>6,7</sup> as well as various cancers<sup>8–11</sup>.

The two main classes of parvulins are the Pin1-type and non-Pin1-type parvulins. For the first group the isomerization reaction is phosphorylation-dependent, that is, Pin1-type parvulins selectively recognize either Ser-Pro or Thr-Pro sequences where the serine/threonine is phosphorylated. Most of them, like their archetype Pin1, contain an N-terminal WW domain responsible for a ligand recognition and a conserved C-terminal PPIase domain with a phosphate-binding loop. Interestingly, there are some known members of the Pin1 family that do not possess a WW domain, i.e. some plant Pin1 proteins<sup>12</sup> and TbPin1 from *Trypanosoma brucei*<sup>13</sup>. In contrast, the non-Pin1-type parvulins are single domain proteins and their isomerization mechanism is phosphorylation-independent. Thus, the phosphate-binding site is missing, which is the only significant structural and functional difference in the PPIase domain between the Pin1-type and non-Pin1-type parvulins.

<sup>1</sup>Pázmány Péter Catholic University, Faculty of Information Technology and Bionics, Budapest, Hungary. <sup>2</sup>University of Jyväskylä, Dept. of Biological and Environmental Science, Dept. of Chemistry, Nanoscience Center, Jyväskylä, Finland. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to Z.G. (email: gaspari.zoltan@itk.ppke.hu)

The exact mechanism of action of PPIases is not yet elucidated. It is unclear whether different PPIase families or distinct members within a family exhibit similar mechanisms. However, it is generally accepted that there is no breaking and reforming of the peptide bond, thus, the bond is converted from the *cis* to the *trans* form via rotation through a twisted amide intermediate<sup>14</sup>. In a recent study on cyclophilin A, dynamic structural ensembles were generated using chemical shift data for a structurally heterogeneous state where both the *cis* and *trans* isomer of the ligand are present. Analysis of the electrostatic field within the ligand binding site suggested an ‘electrostatic handle’ mechanism, speculated to be also valid for other PPIases<sup>15</sup>. Furthermore, although the active site of the parvulin-type PPIases is well-defined, the mechanistic role of the constituent residues is not yet fully clarified.

The two highly conserved histidines of parvulin-type PPIases have been suggested to be important for catalysis. However, many His mutants of Pin1 retained PPIase activity and, interestingly, the selectivity of Pin1 towards phosphorylated substrates was dependent on the identity of the replacing residues. Circular dichroism investigations together with proteolytic susceptibility data led to the suggestion that the mutations influenced the dynamics of Pin1 rather than causing substantial structural rearrangements<sup>16</sup>.

The high-resolution structure of human Par14 offered insights into a hydrogen-bonding network involving the two histidines as well as a threonine and an aspartate at the ‘outer edges’ of the imidazole side chains. Replacing the threonine with alanine in Pin1 resulted in 300-fold decrease in catalytic activity while not compromising structural integrity<sup>17</sup>. The aspartate is in a position occupied by a cysteine in Pin1, also suggested to be important in catalysis earlier<sup>18</sup>. Detailed theoretical studies hinted that this cysteine, through changes in its protonation state, can mediate dynamic changes in this network<sup>19</sup>. Indeed, replacing this cysteine with alanine or serine caused the disruption of the hydrogen bond between the histidines<sup>20</sup>.

NMR analysis of conformational exchange in Pin1 suggested a link between motional modes present in the catalytic domain and the rate of catalysis, leading to the hypothesis that the internal motions assisting catalysis are an intrinsic feature of Pin1<sup>21</sup>. Ligand binding has been shown to influence the internal dynamics of Pin1, leading to more extensive contact between the PPIase and WW domains proposed to be linked to the loss of flexibility at specific conserved hydrophobic sites<sup>22</sup>. Specifically, changes in side-chain mobility upon ligand binding highlighted the role of an internal conduit consisting of hydrophobic side-chains. These residues are conserved in Pin1 homologs and have been suggested to play an important role in inter-domain communication<sup>23</sup>. More recent studies showed that the role of the WW domain negatively regulates PPIase activity in Pin1<sup>20</sup>.

Molecular dynamics studies of Pin1 revealed allosteric pathways and suggested that substrate binding by the WW domain leads to preorganization of the catalytic site<sup>24</sup>. The range of identified residues participating in allosteric communication extends those revealed by NMR studies of side-chain flexibility<sup>23</sup>. Importantly, the preorganization was identified as a closure of the loop regions surrounding the substrate-binding cleft, and the presence of the WW domain enhances the flexibility of these loops<sup>24</sup>. A recent study combining NMR spectroscopy and molecular dynamics indicated that the WW domain undergoes structural changes upon ligand binding and these changes affect its association with the PPIase domain in full-length Pin1, a mechanism proposed to be responsible for different activity of Pin1 on ligands with single and multiple recognition sites<sup>25</sup>.

To get further insights to the differences between various parvulins, we have combined molecular dynamics simulations with experimentally available backbone S<sup>2</sup> order parameters to conduct a comparative analysis of three single-domain parvulins. One of these (TbPin1) is Pin1-type parvulin lacking a WW domain and the other two (SaPrsA, CsPinA) are non-Pin1-type parvulins.

SaPrsA from *Staphylococcus aureus* is responsible for folding of secreted proteins. Although member of a different subclass, its three-dimensional structure and active site arrangement proved to be almost the same as for human Pin1. The profound knowledge of histidine protonation states of His residues was investigated in detail experimentally obtained, revealing different tautomeric states for the two conserved histidines and the presence of a hydrogen bond between their side chains<sup>26</sup> and this is also reflected in the corresponding PDB structure (id: 2JZV).

TbPin1 from *Trypanosoma brucei* is considered as a putative Pin1-type parvulin despite it lacks the WW domain<sup>13</sup>. It was shown that replacing Cys65 (corresponding to Cys113 in Pin1) with Ala diminishes its catalytic activity, in accordance with other studies on the role of this residue (see above). In the structures deposited in PDB (id: 2LJ4) both His residues are protonated and there are no hydrogen bonds between them.

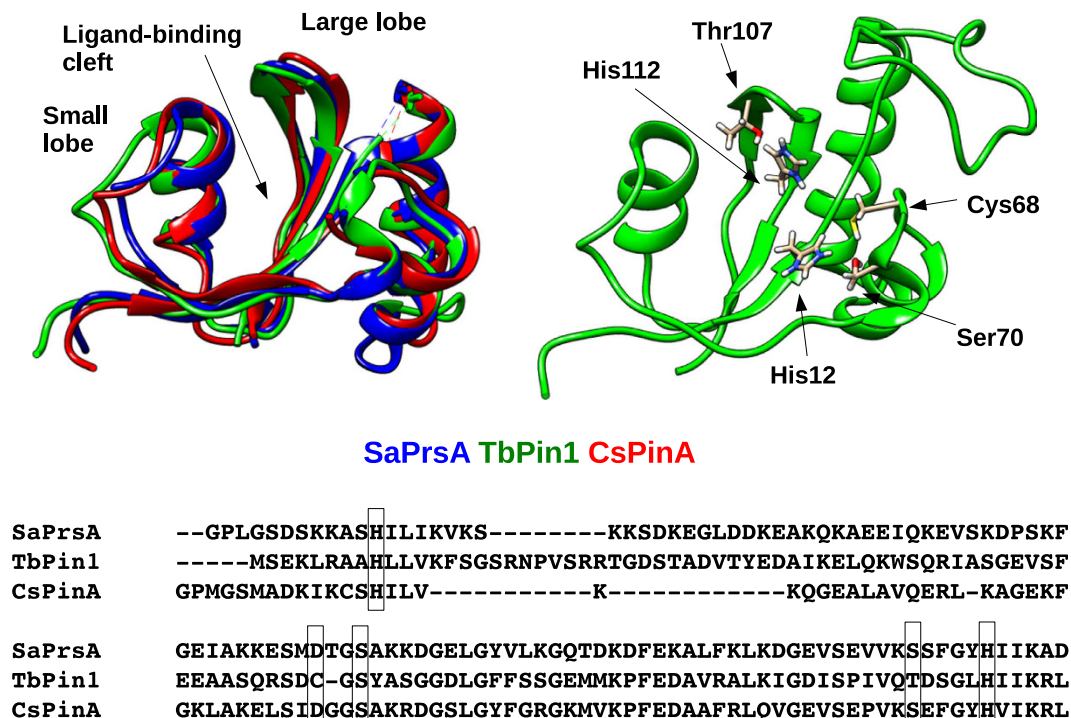
CsPinA from the psychrophilic archaeon *Cenarchaeum symbiosum* has been shown to possess an atypically large peptide-binding site. Similarly to TbPin1, the two histidines are protonated in the available PDB structure<sup>27</sup>.

The three investigated parvulins share a common structural core (Fig. 1) where the ligand-binding cleft is flanked by two lobes. The smaller lobe, shown in the left of the figure and closer to the N-terminus, consists of a short helix and a loop structure, whereas the larger one is formed by a four-stranded antiparallel  $\beta$ -sheet and two helices located opposite the cleft. The conserved histidines are located in the two central strands of the  $\beta$ -sheet. Notably, all residues forming the hydrogen-bonding network described above can be found in the large lobe (Fig. 1). Available backbone S<sup>2</sup> order parameters for the three molecules suggest some characteristic differences with TbPin1 showing the lowest average values (Fig. 2).

## Results and Discussion

**Restrained ensembles resemble the native structures and are restricted relative to unrestrained ones.** For all three parvulins, SaPrsA, TbPin1 and CsPinA, restrained ensembles were generated according to the MUMO protocol<sup>28</sup> using backbone S<sup>2</sup> order parameters and NOE distance restraints. As a control, unrestrained ensembles were also generated (see Methods for details). The MUMO and unrestrained ensembles contain 728 conformers each.

In the case of the MUMO ensembles, correspondence to S<sup>2</sup> order parameters could be achieved without compromising the agreement with C $\alpha$  and H $\alpha$  chemical shifts that are most sensitive to protein structure (Table 1).



**Figure 1.** Top left: superimposed C $\alpha$  trace of SaPrsA (blue), TbPin1 (green) and CsPinA (red) showing only the residues aligned by MAMMOTH-Mult. Top right: TbPin1 structure with the residues involved in the hydrogen-bonding network highlighted. Bottom: sequence alignment of the three parvulins with the residues involved in the hydrogen-bonding network highlighted.

The unrestrained ensembles, similarly to those deposited in the PDB, are not compatible with the backbone S<sup>2</sup> data, as can be expected<sup>29</sup>.

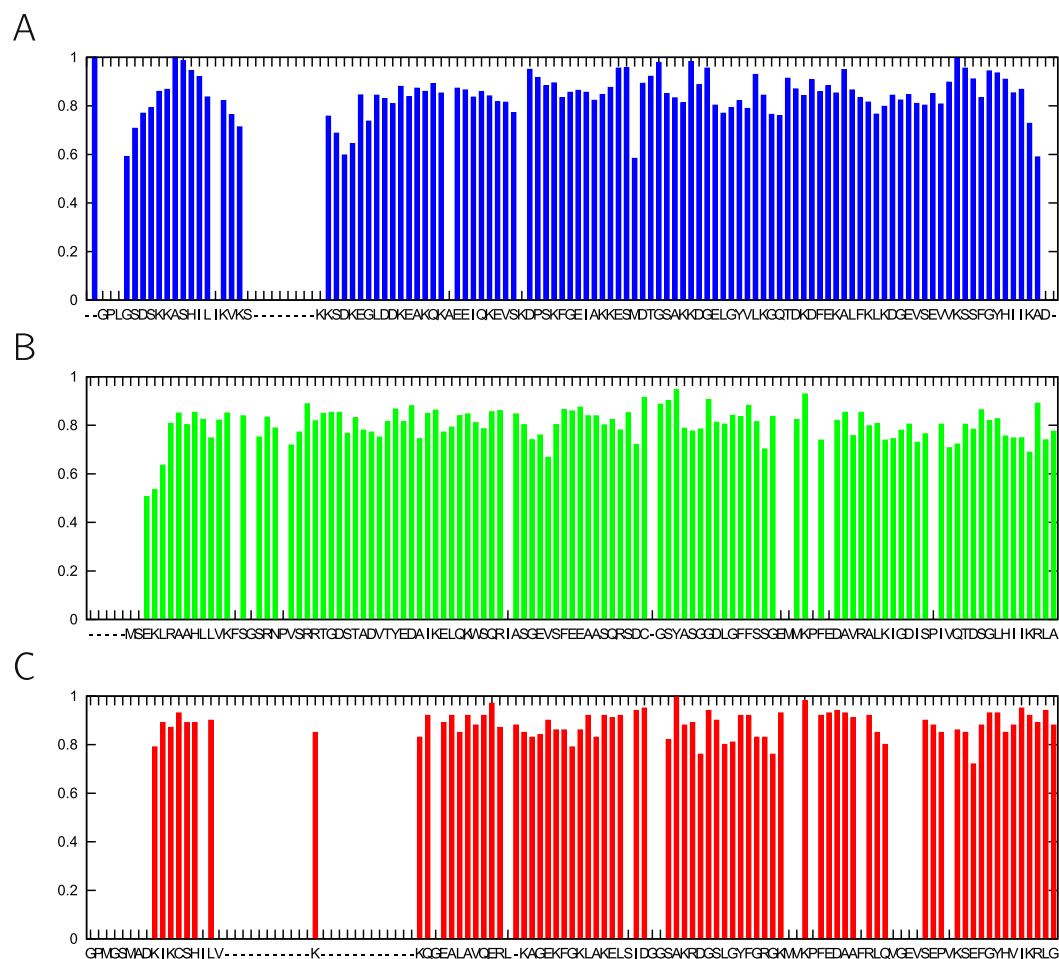
For CsPinA, the S<sup>2</sup> value of the C-terminal residue, Gly97, had to be excluded from the backbone S<sup>2</sup> correlation because of a conformational drift during the MD simulation resulting in two alternative orientations of this residue in the final ensemble. Closer analysis hints that this might be the consequence of the NOE restraints in the region (exclusion of all restraints violated in the PDB ensemble hinders the occurrence of the conformational drift). However, as this region is not included in any of the consensus mappings, this does not affect any of our conclusions below.

It should be noted that in our calculations NOE data were used to restrain the ensemble close to the native conformation, but, as in other ensembles reflecting multiple NMR-derived parameters, it can not be expected that all NOE restraints are fulfilled<sup>30,31</sup>. The ratio of violated restraints is below 1% for all three molecules in this treatment (Supplementary Table S3).

Both the MUMO and the unrestrained ensembles are clearly more diverse than the PDB-deposited ones, with the restrained ones being conformationally more restricted, as evidenced by PCA analysis (Fig. 3) and RMSD values in Table 1. In general, the PDB-deposited ensembles sample only a subset of the conformational space occupied by the MUMO ensembles and, with the notable exception of CsPinA, the MUMO ensembles correspond to a subset of the unrestrained ones (Fig. 3). This trend is more evident when only structurally equivalent residues, defined in the basis of a structural alignment of the three proteins (see Methods), are considered.

**The three parvulin ensembles differ in the extent of binding cleft opening.** The ensembles of the three different PPLase domains were compared using the set of residues that could be aligned in a multiple structural alignment (see Methods). The resulting mapping contains 89 residues including the substrate binding cleft and the two surrounding lobes (Fig. 1). PCA analysis of the combined ensembles reveals that they occupy distinct regions of the conformational space according to the first two modes covering 44 and 25% of the variability of the structures, respectively (Fig. 4A,B). The same remains mostly valid for the unrestrained ensembles (Fig. 4C,D, with the first two modes covering 36 and 24% of the variability) and even when the MUMO and unrestrained ensembles are analyzed together (Fig. 4E).

PCA mode 1 in the MUMO ensembles largely overlaps with PCA mode 2 of the unrestrained ones (Fig. 4F). Closer analysis of this mode reveals that mode 1 in the MUMO ensembles reflect a motion roughly corresponding to the opening and closing of the substrate binding cleft and can be approximated by measuring the distance between residues near the tip of the two flanking loops of the cleft (Fig. 4G,H). Along this coordinate, the TbPin1 ensemble occupies the largest region, thus, our analysis suggests that this motion is primarily present in the TbPin1 ensemble but is also clearly present in SaPrsA. Nevertheless, the deviations between the ensembles can be primarily attributed to the differences in binding cleft opening in parvulins, termed ‘breathing motion’

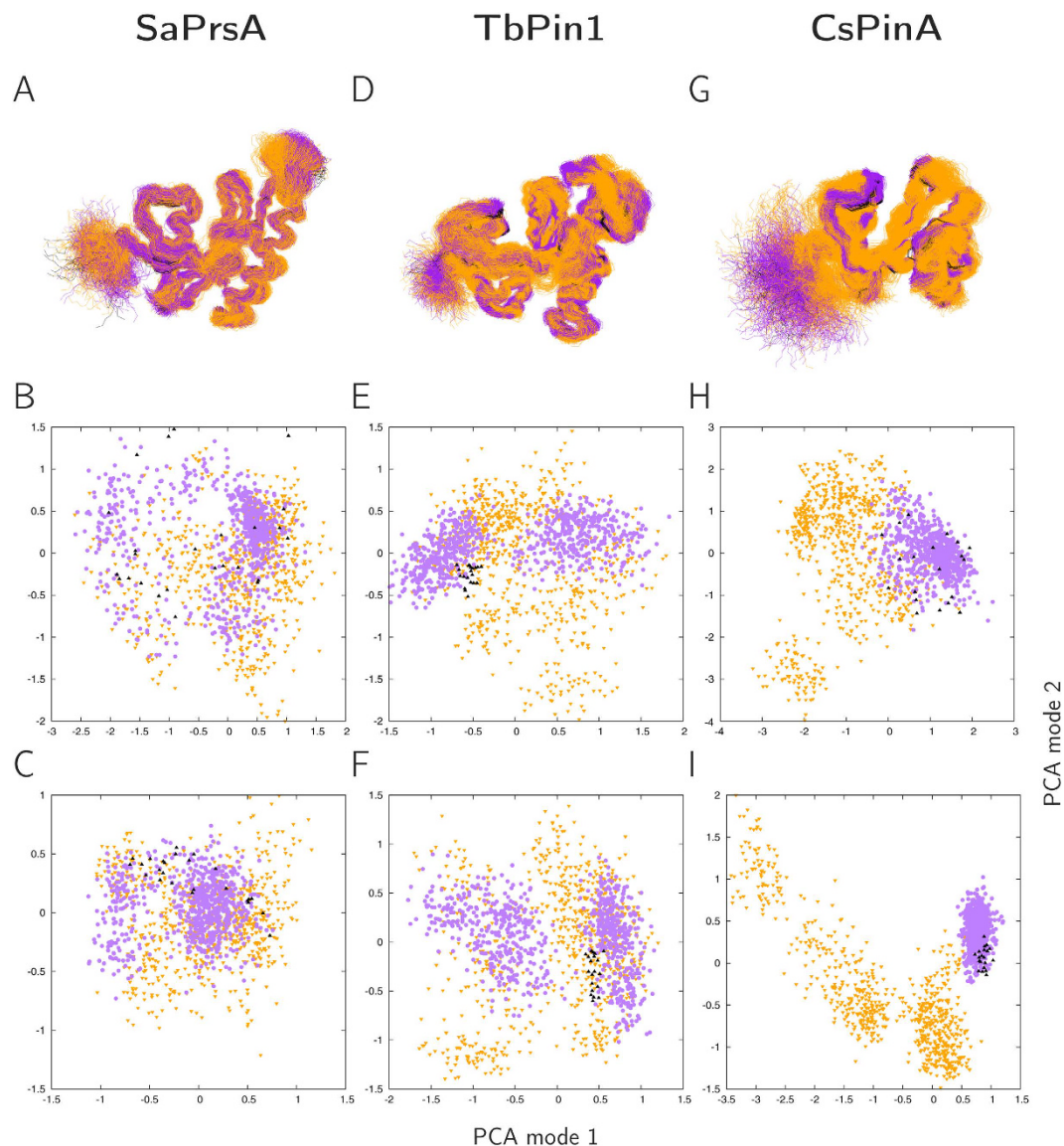


**Figure 2. Experimental  $S^2$  values for the three parvulins investigated.** Sequences are aligned to make comparison straightforward (based on the alignment produced by MAMMOTH-Mult, <https://ub.cbm.uam.es/software/online/mamothmult.php>). (A) SaPrsA, (B) TbPin1, (C) CsPinA.

	Ensemble size	Backbone RMSD (Å)		Backbone $S^2$ correlation		Chemical shift correlation	
		Full molecule	Fit with flexible parts excluded	Full molecule	$C\alpha$	$H\alpha$	
<b>SaPrsA</b>		<b>Residues 1–111</b>	<b>Residues 6–111</b>	<b>Residues 1–111</b>			
2JZV	25	2.01 ± 0.65	0.75 ± 0.22	0.59	0.98	0.91	
unrestrained	728	2.09 ± 0.57	1.43 ± 0.38	0.58	0.98	0.91	
MUMO	728	1.89 ± 0.65	1.12 ± 0.20	0.93	0.99	0.92	
<b>TbPin1</b>		<b>Residues 1–115</b>	<b>Residues 3–115</b>	<b>Residues 1–115</b>			
2LJ4	20	0.68 ± 0.18	0.43 ± 0.08	0.38	0.96	0.84	
unrestrained	728	2.12 ± 0.49	1.93 ± 0.46	0.16	0.97	0.86	
MUMO	728	1.71 ± 0.52	1.58 ± 0.53	0.92	0.97	0.86	
<b>CsPinA</b>		<b>Residues 1–97</b>	<b>Residues 6–97</b>	<b>Residues 1–97</b>			
2RQS	20	2.16 ± 0.72	0.64 ± 0.17	0.11	0.97	0.69	
unrestrained	728	3.35 ± 1.00	2.35 ± 0.80	0.23	0.98	0.74	
MUMO	728	2.44 ± 0.72	1.21 ± 0.26	0.78*	0.98	0.75	

**Table 1. Correspondence of the ensembles to experimental data.** Backbone RMSD values are calculated for all residues with MOLMOL. Note that the number of residues are not comparable and the large values for the deposited ensembles are the consequence of the inclusion of flexible terminal parts. \* $S^2$  value for GLY97 excluded.

hereafter. We have analyzed the residues involved in ligand binding and largely conserved in all three parvulins analyzed (see Methods). PCA analysis of selected heavy atoms is shown in Fig. 4I. PCA mode 1 largely describes the alterations of the distance of residues located at opposite sides of the binding cleft, most prominently those

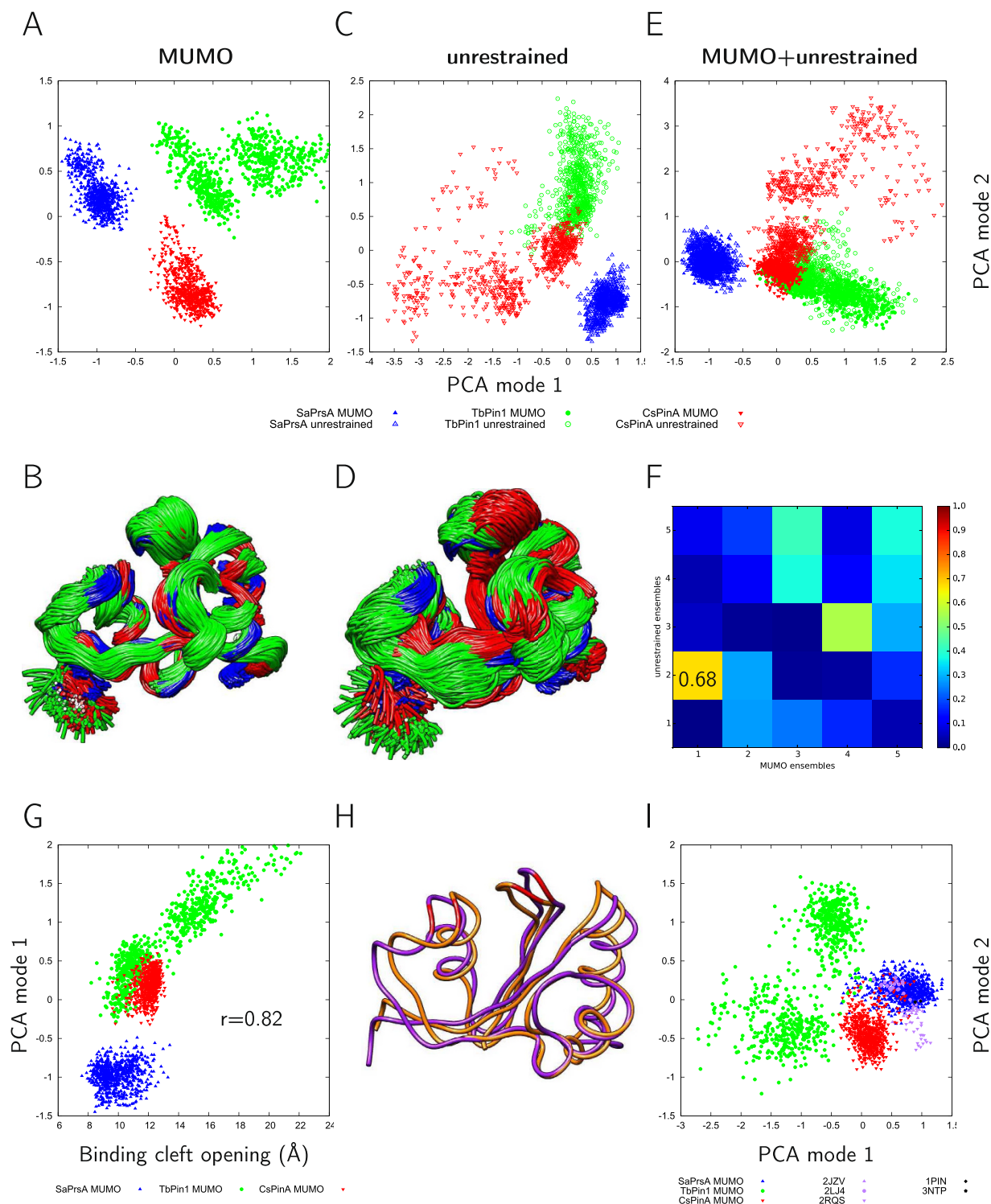


**Figure 3.** Top panels (A,D,G): structures superimposed with MOLMOL (black: ensembles deposited in the PDB, orange: unrestrained ensembles, purple: MUMO ensembles). Middle panels (B,E,H): PCA plots (first two modes) of the full structures. Bottom panels (C,F,I): PCA plots (first two modes) of the residues corresponding to the consensus mapping of the three proteins (see Methods for details). Black triangles: deposited PDB ensembles, orange triangles: unrestrained ensembles, purple circles: MUMO ensembles.

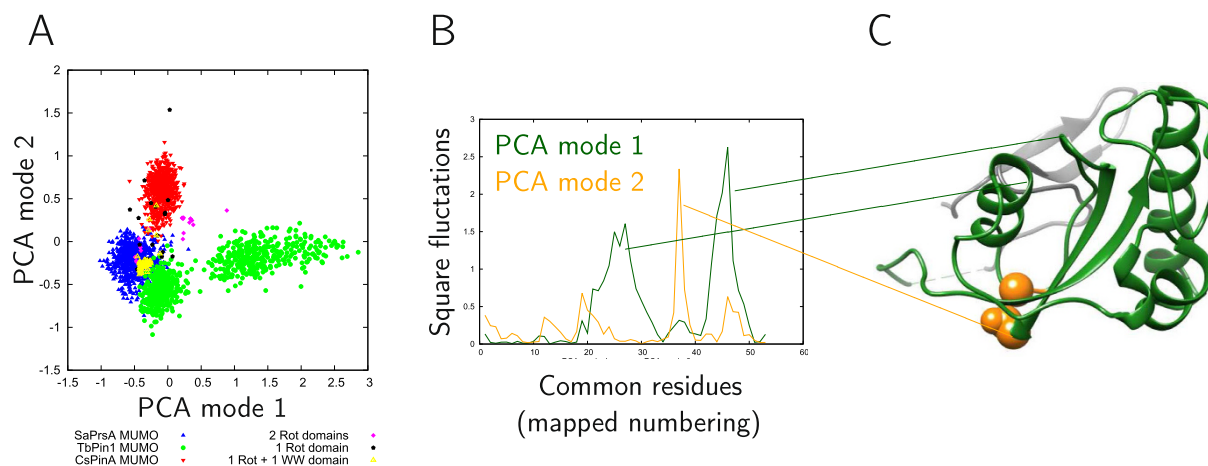
corresponding to Met130 and Cys113 in Pin1. Thus, the differences observed in the binding sites can also mostly be attributed to the opening-closing motion separating the full structures in the MUMO ensembles.

$S^2$  restraining yields a conformational ensemble consistent with the fast (ps-ns) internal motions, thus, it is expected that the resulting ensemble samples the conformational space around an average structure representing the native state. However, in the case of TbPin1, the nature of the conformational movements sampled, in particular the breathing motion, would be expected to occur on a slower time scale. Thus, we regard the generated ensembles as reflecting the upper limit of the conformational space sampled by the three parvulins during their fast motions. In this interpretation the ensembles do not necessarily reflect that the binding site opening - at least to the extent reflected by the TbPin1 ensemble - indeed occurs on such a fast time scale, although the correspondence to  $S^2$  order parameters strengthens the validity of larger motions in TbPin1 along this mode than in the other two parvulins.

In principle,  $S^2$  restraining does not necessarily restrict the extent of the motions sampled but limits primarily only their directions. Interestingly, RMSD values suggest that the TbPin1 ensemble is not more diverse than the SaPrsA or the CsPinA ensemble in general. Considering the results of PCA analysis it can be safely stated that its diversity is distributed along different internal motions than observed for the other two molecules. It should also be noted that for Pin1, conformational motions expected to be characteristic of slower time scales also occurred in a 100 ns simulation<sup>24</sup>. As both Pin1 and TbPin1 act on phosphorylated substrates, this observation - relatively



**Figure 4. Diversity of the ensembles using only the positions common to all three parvulins.** (A) PCA analysis (modes 1–2) and (B) structure superposition of the combined MUMO ensembles, (C) PCA analysis and (D) superposition of the combined unrestrained ensembles, (E) PCA analysis of the combined MUMO + unrestrained ensembles, (F) overlap of the first 5 PCA modes of the MUMO and unrestrained ensembles shown in (A and C); (G) correlation of the first PCA mode in the MUMO ensembles with the binding cleft opening defined by the distance between the C $\alpha$  atoms of residues 86 and 109 (2LJ4 numbering); (H) average conformer in the MUMO ensemble distorted along the PCA mode 1 to show the opening motion, residues 86 and 109 (2LJ4 numbering) highlighted; (I) PCA analysis of selected atoms in the binding site of the molecules (see Methods), PCA mode 1 here is also dominated by the opening-closing motion of the binding cleft.



**Figure 5.** (A) PCA analysis of all PPIase domains in the PDB using the positions common to all such domains. (B) Displacements along PCA mode 1 (dark green) and 2 (orange). Note that the residue numbering refers to the common positions comprising 53 residues only. (C) Location of the residues with largest displacements in PCA mode 2 projected to the full PIN1 (in orange, Gly144 and Glu145). Note that the parts retained based on the structure alignment contain only the two loops connected with the two peaks in PCA mode 1. The WW is domain is in the back, colored light gray.

large amplitude motions of the binding cleft at a fast time scale - might even have relevance for this subtype of parvulins. For TbPin1, NMR relaxation analysis revealed a group of residues with slow exchange located at the phosphate-binding loop, which might also indicate the presence of larger-scale opening-closing motions, although on a slower time scale.

**Comparison with other parvulins highlights diversity in the hinge region.** To compare the ensembles with other parvulin domains of known structure, we have generated a consensus residue mapping between 100 rotamase domains, including the representative structures of SaPrsA, TbPin1 and CsPinA, available in the PDB (Supplementary Fig. S6). Interestingly, this consensus mapping contains only 53 residues including only one of the conserved histidine residues, as the one closer to the N-terminus is not part of this consensus. We have performed PCA analysis on the three MUMO ensembles plus the 100 rotamase domains.

It is somewhat surprising that, contrary to expectations<sup>32,33</sup>, the diversity of the MUMO ensembles is higher than that of the different PDB-derived parvulins. However, at least for structures determined with crystallography it is expected that the crowded environment of a crystal does not favor open conformations.

Similar to the results obtained from comparing the MUMO ensembles only, PCA mode 1, covering 48% of the variability of the structures, corresponds to the opening and closing of the substrate binding site. Interestingly, the only group with a substantial distribution along this mode corresponds to the proteins with 2 rotamase domains (Fig. 5A). We note that from these, only one available structure, 1m5y (*E. coli* SurA, an outer membrane protein chaperone) contains both parvulin domains, and the two domains in all four chains of this PDB entry are well separated along the 1st PCA coordinate, with N-terminal domains having a positive first coordinate and C-terminal domains exhibiting a negative one in the plot (Fig. 5A). In this structure, the first (N-terminally located) parvulin domains are surrounded by an extension around the large lobe of the binding cleft.

Strikingly, in this analysis the MUMO ensembles of the three parvulins investigated are separated along PCA coordinate 2, which describes differences near the ‘bottom’ of the peptide binding cleft, regions acting as linkers between the two lobes around the binding cleft. The hinge region identified in PCA mode 1, defined by the minimum around residue 38 in the mapped numbering (Gly144 in Pin1), is mostly affected by displacements along PCA coordinate 2 (Fig. 5B). A similar motion is responsible for the separation of the MUMO ensembles along PCA mode 2 when analyzed without additional structures (Fig. 4A).

As the different parvulin ensembles are also separated along this motional mode, it is tempting to assume that they correspond to different states along a common motional mode occurring on a slower time scale and some of their functional differences can be explained by the differences required in their ligands to trigger proper binding and effective catalysis to occur. However, whether larger-scale motions of this type occur in all molecules and whether they can be linked to any aspect of catalysis remains to be shown. Our first-approximation estimate of the electrostatic field at the position of the carbonyl C of the isomerised amide bond of the substrate did not reveal any dependence on the extent of the opening of the binding cleft (not shown). However, the difference between the mode and extent of this opening motion between these molecules might still be valid and might influence their functional diversity.

In Pin1, the WW domain exerts a significant influence on the catalytic properties: Both WW-deletion mutants and point mutants influencing the interaction between the WW and parvulin domains show markedly different activity relative to wild-type Pin1. As the WW-parvulin interaction site is located on the side of the ligand-binding cleft, it can be speculated that the extent and mode of interaction with the WW domain modulates the breathing motion of the cleft, thereby contributing to the regulation of its enzymatic activity. This hypothesis is consistent with the recent molecular dynamics study of Pin1, where the presence of the WW domain enhanced the flexibility

of the loops around the binding site<sup>24</sup> and also with the separation of the WW-containing and WW-less parvulins in our comparative PCA analysis. These parvulins are separated along PCA mode 2 corresponding to differences distributed near the bottom of the binding cleft, ideal in a position to modulate the flexibility of the parvulin domain. This is supported by our observation that the hinge region identified for the opening-closing motion displays the highest displacement in the PCA analysis of all PPIase domains with known structure (Fig. 5C). Thus, we hypothesize that changes in the hinge region, either caused by mutations or the interaction with the WW domain, if present, can influence the opening of the substrate binding cleft. Indeed, residues adjacent to this putative hinge region have been shown to be influenced by mutations introduced into the WW domain of Pin1<sup>20</sup>. Moreover, NMR relaxation analysis identified two residues in this region, Ile 98 and Asp100 (according to residue numbering in 2LJ4) that are involved in slow conformational exchange<sup>13</sup>.

**Possible role of the conserved histidines and the hydrogen bond network.** In search for the factors that might influence the different preferences of the three parvulins with respect to the opening of the binding cleft, we performed detailed analysis of the conformation, possible pKa value distribution and hydrogen-bonding pattern of the conserved histidine residues. Both the conformation, as measured by the side-chain chi1 and chi2 torsions and the distribution of the predicted pKa values are different in all three ensembles. It is important to note here that both TbPin1 and CsPinA were calculated with fully protonated histidine side chains, consistent with the PDB-deposited structures, whereas for SaPrsA the results of the detailed analysis published along with the structure<sup>26</sup> were taken into account. Thus, in SaPrsA there is a hydrogen bond between the two His residues that can not form in the two other parvulins due to the protonation state and conformation of the imidazole rings (Fig. 6A). This is also reflected in the observation that the relative side-chain orientations of the histidine side chains are most restricted relative to each other in the SaPrsA ensemble (Fig. 6B).

It should be noted that although the pKa values of histidine residues are conformation-dependent<sup>34</sup>, the pKa values predicted here probably overestimate the actual variability occurring during fast motions (Fig. 6C).

Analysis of the CA positions of the residues involved in the putative hydrogen-bonding network suggest that there are motions confined within the large lobe that are at least weakly correlated with the breathing motion of the full parvulin molecules, again dominated by the motions TbPin1 (Fig. 6D). Although the two data sets are not independent (i.e. the residue-specific analysis uses a subset of CA atoms used for the global one) and correlation does by no means indicate causation on its own, our results are compatible with a scenario where the dynamic hydrogen bond proposed by Barman and Hamelberg<sup>19</sup> modulates the mobility of the  $\beta$ -strands of the large lobe relative to each other, precluding the small rearrangements coupled to the opening motion. Thus, changes in His protonation and thus the formation of the hydrogen bond might be one of the factors to be modulated by exact spatial context including substrate binding and interaction with the WW domain<sup>23,24</sup>. Besides the formation of a hydrogen bond between the two imidazole moieties, His protonation also affects the conformational preferences of His residues that might provide even more subtle ways of regulation in this respect. The lack of the hydrogen bond in TbPin1 and CsPinA is consistent with the observation that Pin1 histidine mutants were catalytically active<sup>16</sup> and supports the speculation that His protonation state can be a relevant factor in modulating activity. Testing of this hypothesis would require carefully designed *in vitro* experiments. Moreover, it was noted that the level of isomerase activity on phosphorylated substrates could be modulated by replacing the histidine residues. Among the parvulins investigated here, TbPin1 is a WW-domain-less parvulin exhibiting a Pin1-type PPIase acting on phosphorylated substrates. The above observations together with our findings are consistent with a model where the histidines play pivotal role is in modulating the dynamics of parvulin-type PPIases as suggested by Bailey and coworkers<sup>16</sup> and that this contributes to substrate binding. Thus, even if the histidine protonation states are not physiologically relevant as modeled here, the general mechanism, i.e. that the protonation state can influence the extent and nature of the breathing motion, might still be valid. Such a role of histidine residues would not be unprecedented as a recent study suggested the protonation state of a His side chain can modulate loop flexibility and ligand release in Langerin, a C-type lectin receptor<sup>35</sup>.

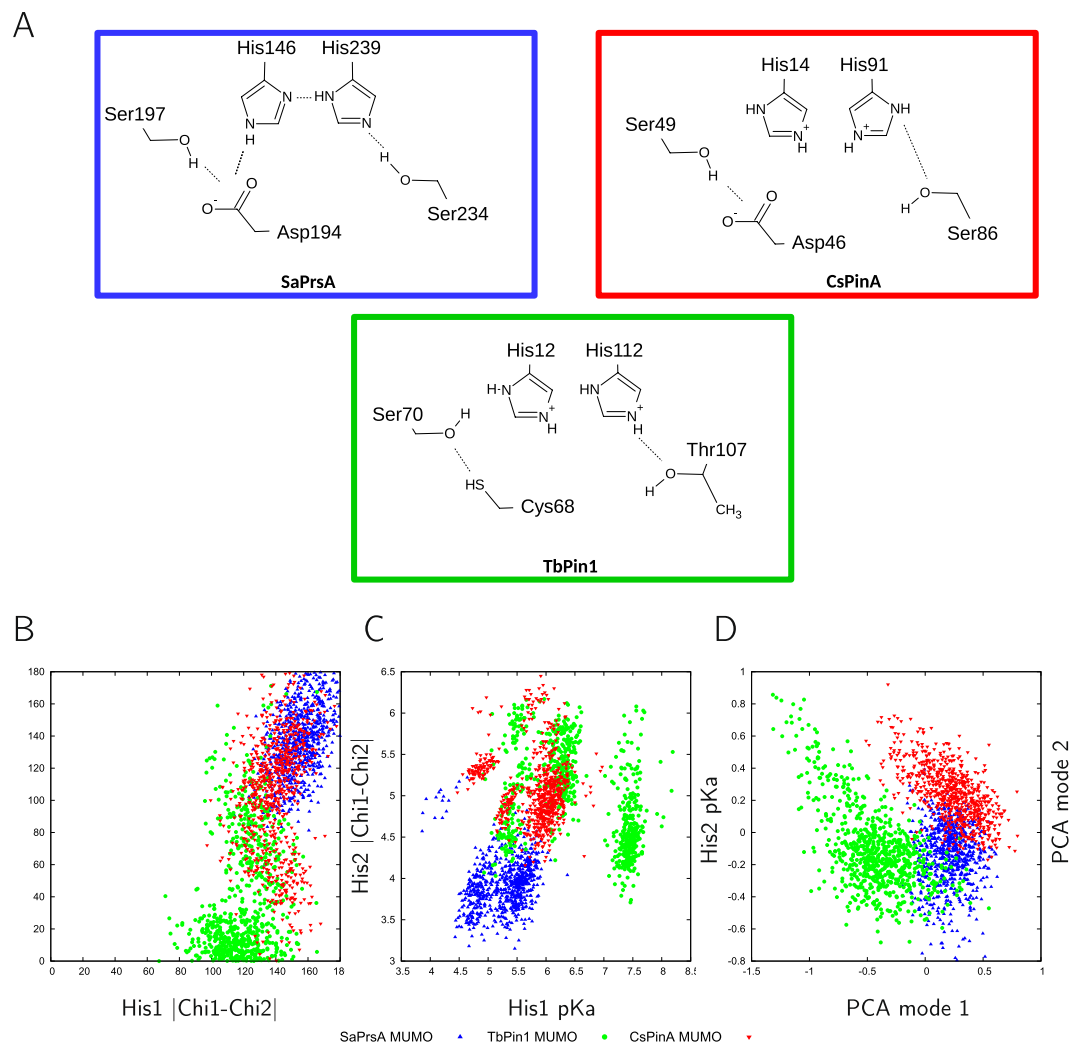
**A general model for parvulin specialization and regulation based on the extent and dynamics of the opening of the substrate binding cleft.** We propose a general model for parvulin-type PPIases where the extent of opening and the breathing motion of the substrate binding cleft plays a role in substrate selectivity and catalysis. Both of these factors, namely, the preferred geometry of the binding cleft and the dynamics of the breathing motion might be modulated by different factors, such as interactions with the substrate, a WW domain (if present) and the state of the hydrogen bond network connecting the strands of the large lobe. These factors might not be independent from each other, but the connections between the conduit formed by hydrophobic residues with the hydrogen bond network are yet to be established. Molecular dynamics simulations of Pin1 also suggested that the closure and flexibility of loops around the binding site are influenced by interactions with the WW domain and dependent on protein-substrate interactions at both the PPIase and the WW domain<sup>24,25</sup>.

Our model, however appealing, is a necessary simplification but provides a testable framework for future investigations. A recent paper has indicated the role of binding site dynamics in substrate binding and catalysis in an FKBP-like PPIase<sup>36</sup>. We note here that modulation of the ligand-binding site opening dynamics has been implicated in the evolution of GK domain proteins<sup>37</sup>, thus, similar mechanisms, if valid, might represent a general way to fine-tune protein function.

## Methods

**Generation of dynamic structural ensembles.** Calculations were performed using an in-house modified version of GROMACS 4.5.5<sup>38</sup> capable of handling S<sup>2</sup> order parameter restraints<sup>29,39</sup>. The coordinates





**Figure 6.** (A) Scheme of the residues involved in the proposed hydrogen bond network. The orientation of the histidines was taken from the first model of the PDB-deposited ensembles. Residue numbering corresponds to that of the PDB ensembles. Hydrogen bonds detected in the MUMO ensembles are shown. Note that the occurrence of the individual hydrogen bonds varies in the ensembles. (B) Chi1-Chi1 and Chi2-Chi2 differences between the two conserved histidines indicate different relative orientations of the side chains in the three ensembles. (C) PropKA-predicted pKa values of the histidines in the SaPrsA (blue), TbPin1 (green) and SaPrsA (red) ensembles. (D) PCA analysis of the CA positions of the 5 residues involved in the hydrogen bond network in the three MUMO ensembles. Mode 1 is dominated by relative displacements of the flanking Ser/Thr and Cys/Asp residues, whereas mode 2 describes the reorientation of the two central histidines. Overall, PCA coordinate 1 here shows a correlation of 0.61 with first PCA mode in the full MUMO ensembles corresponding to the opening-closing motion of the binding cleft.

and NOE restraint lists for structures 2LJ4, 2JZV and 2RQS were obtained from the RCSB web site.  $S^2$  order parameters were obtained from the original authors of the structures for 2JZV and 2LJ4, and from BMRB (entry 11080) for 2RQS. NOE lists were converted to a format similar to the one used in the distance restraint section of GROMACS topology files. In order to ensure standardized treatment of NOE lists and to ensure compatibility with the restraining scheme in GROMACS ( $r^{-6}$  averaging), a filtering procedure was used to select the NOE distances used for restraining. First, all stereospecific assignments were rewritten as ambiguous by listing all possible atom pairs. This ensures that no erroneous stereospecific assignments remain. After removing redundant restraints that might be generated with this step, the distances were checked against the original, PDB-deposited ensembles and those violated over 0.5 Å were removed. Finally, the remaining distances were converted to binned ones, corresponding to three categories, 2.5–3.5, 3.5–5.0 and 5.0–6.0 Ångstroms. This list was used in the following calculations to ensure that the structures remain close to the native state during the restrained simulations.

Using the prepared NOE and  $S^2$  restraints<sup>39</sup>, calculations were performed in a way similar to the MUMO protocol<sup>28</sup> with 8 replicas and NOE restraining over neighbouring ones. The AMBER99SB force field was used with the explicit water model TIP3P. Simulations were run for 10 ns (totaling to  $8 \times 10 = 80$  ns simulation time for each molecule), and after discarding the first 1 ns, structures were saved every 100 ps, resulting in a total of 728

conformers for all three molecules. As a control, unrestrained simulations were run with exactly the same setup but with  $S^2$  and distance restraint force constants set to zero. Correspondence of the ensembles to  $S^2$  parameters was checked with the CoNSEnsX server<sup>40</sup>. Chemical shifts were back-calculated with shiftx2<sup>41</sup> with parameter settings corresponding to the experimental conditions for each molecule as described in the original publications (using default settings resulted in only minor changes not affecting our conclusions). NOE distances were checked using an in-house script using the same calculation scheme as GROMACS.

**Comparative structural analysis of the ensembles.** To be able to do direct comparisons between the two ensembles, we have generated a structural alignment of the annotated representative models (MODEL 1 in each case) of the 2JZV, 2LJ4 and 2RQS structures with MAMMOTH-Mult<sup>42</sup>. Using this alignment we have prepared a consensus mapping of the residues and used this to generate PDB files containing only the residues aligned by MAMMOTH-Mult and have used a consensus residue numbering based on this alignment. The structural ensembles were merged with the original PDB-deposited conformers using this mapping (denoted three-way consensus thereafter) and the resulting ensemble was subjected to PCA analysis using ProDy<sup>43,44</sup>.

For a wider comparative analysis involving other PPIase domains, we searched sequences in the PDB<sup>45</sup> (pdb\_seqres.txt, downloaded on 6 May 2015) with hmmscan using the rotamase HMM profile (Pfam ID PF00639.16). The full protein sequences corresponding to the chains with a PPIase domain were obtained from UniProt and again scanned for the presence of PPIase and WW domains using hmmscan<sup>46</sup>. Coordinates of the PPIase domains in the protein chains as identified by the HMM search were superimposed using MAMMOTH-Mult and this alignment was used to generate a consensus residue mapping. We refer to this mapping as the core region, common to all parvulin-type PPIases analyzed here. This mapping provided the basis for generating PDB files for all structures containing only the aligned residues with consensus numbering. The structures were then superimposed with MOLMOL<sup>47</sup> and several outliers (all belonging to proteins with one PPIase and no WW domain) were removed. This structure set was then merged with PDB files corresponding to the calculated structural ensembles mapped according to this “wide consensus” and the resulting ensemble was subjected to PCA analysis using ProDy. Displacements along PCA modes were extracted from the nmd files generated by ProDy. Comparison of motional modes was performed as described by Meireles *et al.*<sup>48</sup>. The binding site openness was defined with the distance between the Ca atoms of residues Met86 and Ser109 in TbPin1, and the corresponding residues in the generated ensembles.

**Analysis of the ligand-binding site.** Atoms participating in ligand binding and common to all three investigated structures were identified as follows. Two X-ray structures of Pin1:ligand complexes were used, 1PIN<sup>18</sup> and 3NTP<sup>14</sup>. These two structures contain different ligand types, thus, a set of atoms that can be regarded as common between the two ligands had to be defined first. After visual inspection, 9 such atoms were chosen: from the part corresponding to the proline residue, the ring atoms including the alpha carbon, the amide N, carbonyl C, and from the part corresponding to the preceding residue (Ala in 1PIN), three C atoms corresponding to the C, CA and CB atoms.

Protein atoms closer than 7 Å to any of these identified 9 ligand atoms were identified and only those common in the two Pin1 structures were retained. In the next step, a MAMMOTH-MULT alignment of the 1PIN, 3NTP, 2JZV, 2LJ4 and 2RQS structures (using the first models from the original NMR ensembles) was generated and based on this mapping, atoms common in all 5 structures were retained. After inspecting the common set of atoms and residues, where the residue type did not match between the structures but the atom type was the same, the atom was retained. Atoms not common between all 5 structures were omitted, the only exception being Cys 113 (1PIN numbering), a residue proposed to be important in catalysis and replaced by Asp in 2JZV and 2RQS, in which cases the Asp CG atom was considered instead of the Cys SG. This resulted in 38 atoms altogether from 10 residues including the two conserved histidines and all residues that were used as a common set of binding site atoms shared by the 5 parvulins. Although this approach contains subjective elements, we expect that the size of the set and the included residues warrant that the results obtained have valid implications.

**Calculation of the electrostatic field.** With correspondence to the work of Camilloni *et al.*, electrostatic field was computed in the binding site of the generated ensembles<sup>15</sup>. According to the catalytic mechanism proposed by Camilloni *et al.*, a large Z-component of the electrostatic field in the position of the proline carbonyl carbon atom (given that the amide plane coincides with the x-y plane) facilitates cis-trans isomerisation of the proline peptide bond. In order to obtain the values of the electrostatic fields, all ensembles were rotated in a common frame. The common frame was given by pin-1 protein (PDB: 1PIN). Its ligand (Ala-Pro) was rotated in a way that the N, N-CD and N-CD-CA atoms of the proline coincide with the origin, the x axis and the x-y plane, respectively. Afterwards, all the replicas of the generated ensembles were superimposed to the properly rotated pin-1 structure. Prior to the superposition, multiple sequence and structure alignment was done with MAMMOTH-MULT for the original pin-1 structure and the generated structures of SaPrsA, TbPin1 and CsPinA, to identify the overlapping segments. For the superposition only Ca atoms were used that were close to the binding site (residues 115–163 in 1-pin, and corresponding residues in the other molecules). The electrostatic field was computed in the atomic positions of the ligand, based on the partial charges in the topology file generated by GROMACS. For calculating the electrostatic field, only atoms within the cutoff range and with partial charges higher than the charge cutoff were taken into account. To verify the robustness of the results, the electrostatic field was determined in two ways: (1) 15.00 Å cutoff range, 0.1 charge cutoff, and (2) 30.00 Å cutoff range, 0.05 charge cutoff. Also, the correlation of the binding site openness with the Z component of the electrostatic field for the carbonyl carbon atom was checked.

## References

- Hanes, S. D. Prolyl isomerases in gene transcription. *Biochim. Biophys. Acta* **1850**, 2017–34 (2015).
- Mueller, J. W. & Bayer, P. Small family with key contacts: par14 and par17 parvulin proteins, relatives of pin1, now emerge in biomedical research. *Perspect. Medicin. Chem.* **2**, 11–20 (2008).
- Fanghänel, J. & Fischer, G. Insights into the catalytic mechanism of peptidyl prolyl cis/trans isomerases. *Front. Biosci.* **9**, 3453–78 (2004).
- Ryo, A. *et al.* Prolyl-isomerase Pin1 accumulates in lewy bodies of parkinson disease and facilitates formation of alpha-synuclein inclusions. *J. Biol. Chem.* **281**, 4117–25 (2006).
- Rudrabhatla, P. *et al.* Pin1-dependent prolyl isomerization modulates the stress-induced phosphorylation of high molecular weight neurofilament protein. *J. Biol. Chem.* **283**, 26737–47 (2008).
- Butterfield, D. A. *et al.* Pin1 in Alzheimer's disease. *J. Neurochem.* **98**, 1697–706 (2006).
- Balastik, M., Lim, J., Pastorino, L. & Lu, K. P. Pin1 in Alzheimer's disease: multiple substrates, one regulatory mechanism? *Biochim. Biophys. Acta* **1772**, 422–9 (2007).
- Lu, K. P. *et al.* Targeting carcinogenesis: A role for the prolyl isomerase Pin1? in *Molecular Carcinogenesis* **45**, 397–402 (2006).
- Wulf, G. M. *et al.* Pin1 is overexpressed in breast cancer and cooperates with Ras signaling in increasing the transcriptional activity of c-Jun towards cyclin D1. *EMBO J.* **20**, 3459–72 (2001).
- Bao, L. *et al.* Prevalent overexpression of prolyl isomerase Pin1 in human cancers. *Am. J. Pathol.* **164**, 1727–37 (2004).
- Ayala, G. *et al.* The prolyl isomerase Pin1 is a novel prognostic marker in human prostate cancer. *Cancer Res.* **63**, 6244–51 (2003).
- Yao, J. L., Kops, O., Lu, P. J. & Lu, K. P. Functional conservation of phosphorylation-specific prolyl isomerases in plants. *J. Biol. Chem.* **276**, 13517–23 (2001).
- Sun, L. *et al.* Solution structural analysis of the single-domain parvulin TbPin1. *PLoS One* **7**, e43017 (2012).
- Xu, G. G., Zhang, Y., Mercedes-Camacho, A. Y. & Eitzkorn, F. A. A Reduced-Amide Inhibitor of Pin1 Binds in a Conformation Resembling a Twisted-Amide Transition State. *Biochemistry* **50**, 9545–50 (2011).
- Camilloni, C. *et al.* Cyclophilin A catalyzes proline isomerization by an electrostatic handle mechanism. *Proc. Natl. Acad. Sci. USA* **111**, 10203–8 (2014).
- Bailey, M. L., Shilton, B. H., Brandl, C. J. & Litchfield, D. W. The dual histidine motif in the active site of Pin1 has a structural rather than catalytic role. *Biochemistry* **47**, 11481–11489 (2008).
- Mueller, J. W. *et al.* Crystallographic proof for an extended hydrogen-bonding network in small prolyl isomerases. *J. Am. Chem. Soc.* **133**, 20096–20099 (2011).
- Ranganathan, R., Lu, K. P., Hunter, T. & Noel, J. P. Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **89**, 875–886 (1997).
- Barman, A. & Hamelberg, D. Cysteine-mediated dynamic hydrogen-bonding network in the active site of Pin1. *Biochemistry* **53**, 3839–3850 (2014).
- Wang, X., Mahoney, B. J., Zhang, M., Zintsmaster, J. S. & Peng, J. W. Negative Regulation of Peptidyl-Prolyl Isomerase Activity by Interdomain Contact in Human Pin1. *Structure* **23**, 2224–2233 (2015).
- Labeikovsky, W., Eisenmesser, E. Z., Bosco, D. A. & Kern, D. Structure and Dynamics of Pin1 During Catalysis by NMR. *J. Mol. Biol.* **367**, 1370–1381 (2007).
- Namanja, A. T. *et al.* Substrate Recognition Reduces Side-Chain Flexibility for Conserved Hydrophobic Residues in Human Pin1. *Structure* **15**, 313–327 (2007).
- Namanja, A. T. *et al.* Stereospecific gating of functional motions in Pin1. *Proc. Natl. Acad. Sci. USA* **108**, 12289–12294 (2011).
- Guo, J., Pang, X. & Zhou, H.-X. X. Two Pathways Mediate Interdomain Allosteric Regulation in Pin1. *Structure* **23**, 237–247 (2015).
- Olsson, S., Strotz, D., Vögeli, B., Riek, R. & Cavalli, A. The Dynamic Basis for Signal Propagation in Human Pin1-WW. *Structure* **24**, 1464–1475 (2016).
- Heikkinen, O. *et al.* Solution structure of the parvulin-type PPIase domain of *Staphylococcus aureus* PrsA—implications for the catalytic mechanism of parvulins. *BMC Struct. Biol.* **9**, 17 (2009).
- Jaremko, Ł. *et al.* Structure and dynamics of the first archaeal parvulin reveal a new functionally important loop in parvulin-type prolyl isomerases. *J. Biol. Chem.* **286**, 6554–6565 (2011).
- Richter, B., Gsponer, J., Várnai, P., Salvatella, X. & Vendruscolo, M. The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J. Biomol. NMR* **37**, 117–135 (2007).
- Montalvao, R. W., De Simone, A. & Vendruscolo, M. Determination of structural fluctuations of proteins from structure-based calculations of residual dipolar couplings. *J. Biomol. NMR* **53**, 281–92 (2012).
- Lange, R. B. *et al.* Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* **320**, 1471–1475 (2008).
- Gáspári, Z., Várnai, P., Szappanos, B. & Perczel, A. Reconciling the lock-and-key and dynamic views of canonical serine protease inhibitor action. *FEBS Lett.* **584**, 203–206 (2010).
- Best, R. B., Lindorff-Larsen, K., DePristo, M. A. & Vendruscolo, M. Relation between native ensembles and experimental structures of proteins. *Proc. Natl. Acad. Sci. USA* **103**, 10901–6 (2006).
- Friedland, G. D., Lakomek, N. A., Griesinger, C., Meiler, J. & Kortemme, T. A correspondence between solution-state dynamics of an individual protein and the sequence and conformational diversity of its family. *PLoS Comput. Biol.* **5**, e1000393 (2009).
- Hudáky, P. & Perczel, A. Conformation dependence of pKa: Ab initio and DFT investigation of histidine. *J. Phys. Chem. A* **108**, 6195–6205 (2004).
- Hanske, J. *et al.* Intradomain Allosteric Network Modulates Calcium Affinity of the C-Type Lectin Receptor Langerin. *J. Am. Chem. Soc.* **138**, 12176–12186 (2016).
- Quistgaard, E. M. *et al.* Molecular insights into substrate recognition and catalytic mechanism of the chaperone and FKBP peptidyl-prolyl isomerase SlyD. *BMC Biol.* **14**, 82 (2016).
- Johnston, C. A., Whitney, D. S., Volkman, B. F., Doe, C. Q. & Prehoda, K. E. Conversion of the enzyme guanylate kinase into a mitotic-spindle orienting protein by a single mutation that inhibits GMP-induced closing. *Proc. Natl. Acad. Sci.* **108**, E973–8 (2011).
- Hess, B., Kutzner, C., Van Der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
- Fizil, Á., Gáspári, Z., Barna, T., Marx, F. & Batta, G. 'Invisible' conformers of an antifungal disulfide protein revealed by constrained cold and heat unfolding, CEST-NMR experiments, and molecular dynamics calculations. *Chem. - A Eur. J.* **21**, 5136–5144 (2015).
- Ángyán, A. F., Szappanos, B., Perczel, A. & Gáspári, Z. CoNSENsX: an ensemble view of protein structures and NMR-derived experimental data. *BMC Struct. Biol.* **10**, 39 (2010).
- Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: Significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57 (2011).
- Lupyan, D., Leo-Macias, A. & Ortiz, A. R. A new progressive-iterative algorithm for multiple structure alignment. *Bioinformatics* **21**, 3255–3263 (2005).
- Bakan, A., Meireles, L. M. & Bahar, I. ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577 (2011).
- Bakan, A. *et al.* Evol and ProDy for bridging protein sequence evolution and structural dynamics. *Bioinformatics* **30**, 2681–2683 (2014).
- Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).

47. Koradi, R., Billeter, M., Wüthrich, K. & Wüthrich, K. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).
48. Meireles, L., Gur, M., Bakan, A. & Bahar, I. Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. *Protein Sci.* **20**, 1645–1658 (2011).

### Acknowledgements

The authors acknowledge the support of the National Research, Development and Innovation Office – NKFIH through grant no. NF104198.

### Author Contributions

Z.G. designed research, A.C. and B. K. performed calculations, all authors participated in data analysis and interpretation as well as in writing the manuscript.

### Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

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

**How to cite this article:** Czajlik, A. *et al.* Fine-tuning the extent and dynamics of binding cleft opening as a potential general regulatory mechanism in parvulin-type peptidyl prolyl isomerases. *Sci. Rep.* **7**, 44504; doi: 10.1038/srep44504 (2017).

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



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017