Coarse-grain simulations on NMR conformational ensembles highlight non catalytic functional residues in proteins

Sophie Sacquin-Mora

Laboratoire de Biochimie Théorique, CNRS UPR9080, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France

E-mail: sacquin@ibpc.fr

Running title : Functional residues in proteins from NMR ensembles and coarse-grain simulations.

Keywords: NMR conformational ensembles; proteins mechanics; coarse-grain simulations; elastic network model.

Abstract

Dynamics are a key feature of protein function, and this is especially true of gating residues, which occupy cavity or tunnel lining positions in the protein structure, and will reversibly switch between open and closed conformations in order to control the diffusion of small molecules within a protein's internal matrix. Earlier work on globins and hydrogenases have shown that these gating residues can be detected using a multiscale scheme combining all atom classic molecular dynamics simulations and coarse grain calculations of the resulting conformational ensemble mechanical properties. Here we show that the structural variations observed in the conformational ensembles produced by NMR spectroscopy are sufficient to induce noticeable mechanical changes in a protein, which in turn can be used to identify residues important for function and forming a *mechanical nucleus* in the protein core. This new approach, which combines experimental data and rapid coarse-grain calculations and no longer needs to resort to time-consuming all-atom simulations, was successfully applied to five different protein families.

1. Introduction

Far from being static, rigid objects, proteins are highly flexible molecules, and understanding this flexibility is essential for understanding protein function[1-2]. Interestingly, protein conformational diversity and dynamics do not always involve large-scale protein motions, and function can arise from small structural changes[3]. In their work investigating the conformational changes between the apo and ligand-bound state in 60 enzymes, Gutteridge and Thornton showed that over 90% of the studied proteins presented root mean square deviations (RMSDs) below 2 Å[4]. These small conformational variations are nonetheless sufficient to support enzymatic activity, and inversely, slightly disrupting a protein's structure can greatly affect its biological function[5-6]. As a consequence, current *de novo* protein design strategies do not only involve the production of a particular fold, but also need to take into account protein dynamics, if one wants to achieve a specific function[7-9].

Molecular gates represent a particular class of dynamic structures that play an important role in protein activity. They can be defined as a system, involving individual residues, loops or secondary structure elements, that can reversibly switch between open and closed conformations, and that control the passage of small molecules into and out of the protein structure[10-11]. Gating residues, lining internal channels connecting a buried active site to the surface, or two distant active sites within a single protein, are a common feature in the protein world, with more than 60% of the enzymes in the Catalytic Site Atlas[12] presenting channels at least 15 Å long[13], and as shown by the 70 examples discussed in a review by Gora et al.[11]. The dynamic properties of these tunnel lining residues play a key role in controlling catalytic efficiency and enzymes substrate specificity[14-15], and they can be seen as a new category of functional residues, that take a

part in regulating biological processes without being directly involved in the catalytic reaction. As a consequence they have attracted growing interest among the drug design community over the recent years[16].

From a theoretical point of view, numerous tools are nowadays available for investigating protein dynamics[14, 17]. Classic all-atom Molecular Dynamics (MD) simulations represent a common strategy, sometimes combined with experimental approaches[18-19], which can however be computationally costly when working on large systems or modeling events taking place on a long time-scale[20]. Another option is to use coarse-grain models, that combine simplified protein representations and energy functions[21-23]. In that perspective, we initially developed the ProPHet (Probing Protein Heterogeneity) program, which models proteins as an elastic network, to investigate protein mechanical properties on the residue level[24]. The rigidity profiles produced by ProPHet then turned out to be a remarkable tool for linking a protein structure to its dynamic properties and its function, showing how small conformational changes within a protein could induce large, non trivial, mechanical effects and in turn impact its activity[25-27]. In earlier work on globins[28] and [NiFe]-hydrogenases[29], we also showed that the mechanical variations (calculated with ProPHET) observed in the conformational ensemble produced by classical equilibrium MD simulations can be used to identify a *mechanical nucleus* (MN) in the protein, which is a cluster formed by cavity lining residues occupying conserved positions along the protein sequence. In both cases, MN residues appear to be gating residues, oscillating between rigid and flexible states along the MD trajectory and controlling ligand migration in the vicinity of the active site.

Since its initial developments a few decades ago, NMR spectroscopy has turned out to be a powerful tool for exploring the structure and dynamics of proteins in solution[18, 30-32]. In particular, proteic entries in the Protein Data Bank[33] (PDB) produced by solution NMR often provide a conformational ensemble containing between 10 and 30 structural models instead of a single structure. In this work, we used these conformational ensembles to investigate the mechanical properties in five globular protein families using CG simulations performed by ProPHet. The conformational variations in the experimental NMR ensembles are sufficient to induce large mechanical changes between models from a single PDB entry. Combined with sequence alignements, the coarse-grain calculations based on the NMR structural models permit us to identify MN residues in the proteins without resorting to computationally costly all-atom MD simulations, thus highlighting potential mutational targets in a drug-design perspective.

2. Material and Methods

NMR conformational ensembles: We searched the PDB for proteins where NMR conformational ensembles, comprising a minimum of ten models, are available for at least three different species. This led to the selection of 19 PDB entries belonging to five protein families with various folds listed in table 1: Globins, dihydrofolate reductase (DHFR), cyclophilin A (CPA), peptidyl-tRNA hydrolase (PTH), and parvulin. For each NMR structure, the most representative models were selected using the OLDERADO webserver, http://www.ebi.ac.uk/pdbe/nmr/olderado/[34]. If the NMR structure was not listed online, we used the *Ensemble cluster* tool, which employs the same method as OLDERADO[35], in the UCSF-Chimera program, https://

www.cgl.ucsf.edu/chimera/[36], to determine the most representative models. Both the ChannelsDB database, https://webchemdev.ncbr.muni.cz/ChannelsDB/index.html[37], and the CASTP webserver, http://sts.bioe.uic.edu/castp/index.html?2r7g[38], were used to identify tunnel and cavity lining residues within all these protein structures.

Protein name	PDB NMR structures	number of NMR models	Residues number	Representative NMR models (starting with the most representative one)	Maximum C_α RMSD between two representative models (in Å)	MS residues
Globins	1myf	12	153	2,5,1,6,8	2.10	Val17, Ile28, Ala53, Met55, Leu69, **Leu72**, Glu83, Leu104, **Ile107**, **Ile111**, Leu135
Sequences identity range : 15-45 %	2m6z-A	20	141	16,2,13,15,18	2.35	**Val62**, **Leu66**, Phe98, **Leu101**, Ser102, **Leu105**, Val132
MN with 3MS positions/4	2m6z-B	20	146	16,2,13,15,18	1.60	Trp15, Ala27, Gly64, **Val67**, **Phe71**, Asn102, **Leu106**, Gly107, **Leu110**, Val134
	1mwb	20	123	17,15,19	2.65	Ser2, Phe21, Val25, His46, **Phe50**, **Ala54**, **Val87**, His117
Dihydrofolate Reductase	2l28	25	162	1,6,22,23,25	2.00	**Phe3**, **Leu4**, **Trp5**, Phe30, Gln33, Thr34, Val95, **Ala97**, **Gly98**, **Phe103**
Sequences identity range : 23-38 %	2kgk	15	162	1,5,6,7,9,12	2.40	**Phe5**, **Met6**, **Val7**, Asn19, Leu21, **Phe96**, **Gly97**, Gly98, **Tyr102**
MN with 3MS positions/4	2ith	20	162	7,9,12,20	1.52	**Ser5**, **Ala7**, Glu20, Val99, **Ile100**, **Gly101**
	1yho	25	186	3,6,10,13,19,21,24	0.90	**Cys6**, **Ile7**, **Val8**, Ile114, **Val115**, **Gly116**, **Tyr121**
Cyclophilin A	2mvz	20	146	5,7,8,14,15	1.80	**Phe31**, **Gly57**, **Met87**, **Phe101**, Ile103, **Phe118**
Sequences identity range : 34-41 %	1oca	20	165	6,18	0.74	**Phe36**, **Gly65**, **Leu98**, Ser99, **Phe112**, **Phe113**
MN with 2MS positions/3	1clh	12	166	1,2,8,9	2.30	Gln56, Gly57, Gly59, **Ile88**, Ala91, Gln102, **Phe103**, **Phe104**, **Phe125**
Peptidyl-tRNA hydrolase	2lgj	10	191	10, 8, 7, 6, 9	2.13	Gly9, Leu24, **Gly25**, Val28, Met69, **Val91**, Val92, **Ile93**, His94, His131, Arg134
Sequences identity range : 39-82 %	2jrc	40	191	5, 26, 29	1.62	Val8, **Gly25**, Ala29, **Ile91**, Val92, **Ile93**, His94, Ala169
MN with 3MS positions/3	2mjl	10	197	6, 4, 2	1.34	Leu9, Val10, Gly11, Ala26, **Gly27**, **Met93**, **Ala95**, Leu136, Leu138
Parvulin	1jns	18	92	15, 12, 11, 16, 3	1.64	**His8**, **Ile9**, Phe30, Gly46, Gly47, Met57, Val58, Phe61, Val64, **Ile85**, **Ile86**
Sequences identity range : 31-55 %	1nmw	10	114	7, 3, 4, 10	1.52	**His59**, **Leu60**, Leu61, Cys113, His157, **Ile158**, **Ile159**
MN with 4MS positions/5	1j6y	20	120	19, 18	2.60	**His12**, **Ile13**, Leu14, Ser68, Ser72, Asp105, His114, **Ile115**
	2rqs	20	97	3, 20, 8, 19	2.91	Cys12, **His14**, Phe36, Leu56, Phe59, Phe68, **Val92**, **Ile93**
	2jzv	25	111	6, 11, 12	3.27	**His146**, **Ile147**, Leu148, Ile149, Ala170, Ala188, Phe216, Ser229, His239, **Ile240**, **Ile241**

Table 1: List of the NMR structural ensembles used in this work. In the last column, cavity or tunnel lining residues are underlined and mechanical nucleus residues are highlighted in bold red.

Coarse-grain Simulation: Coarse-grain Brownian Dynamics (BD) simulations were performed using the ProPHet program, http://bioserv.rpbs.univ-paris-diderot.fr/services/ProPHet/ [26-27, 39], on all the representative models listed in table 1. Diverging from most common CG models, where each residue is described by a single pseudo-atom[40], ProPHet uses a more detailed model enabling different residues to be distinguished, in which the amino acids are represented by one pseudo-atom located at the Cα position,

and either one or two (for larger residues) pseudo-atoms replacing the side-chain (with the exception of Gly)[41]. Interactions between the pseudo-atoms are treated according to the standard Elastic Network Model (ENM)[40], that is, pseudoatoms closer than the cutoff parameter, Rc = 9 Å, are joined by gaussian springs that all have identical spring constants of γ = 0.42 N m^{-1} (0.6 kcal mol^{-1} Å^{-2}). The springs are taken to be relaxed for the starting conformation of the protein, i. e. its crystallographic structure. Mechanical properties are obtained from 200,000 BD steps at 300 K. The simulations are analyzed in terms of the fluctuations of the mean distance between each pseudo-atom belonging to a given amino acid and the pseudo-atoms belonging to the remaining protein residues. The inverse of these fluctuations yields an effective force constant k_i that describes the ease of moving a pseudo-atom i with respect to the overall protein structure:

$$k_i = \frac{3k_B T}{\left\langle \left(d_i - \langle d_i \rangle \right)^2 \right\rangle}$$

where $\langle\rangle$ denotes an average taken over the whole simulation and $d_i = \langle d_{ij} \rangle_{j*}$ is the average distance from particle i to the other particles j in the protein (the sum over $j*$ implies the exclusion of the pseudo-atoms belonging to the same residue as i). The distances between the Cα pseudo-atom of residue k and the Cα pseudo-atoms of the adjacent residues k-1 and k+1 are excluded, since the corresponding distances are virtually constant. The force constant for each residue k in the protein is the average of the force constants for all its constituent pseudo-atoms i. We will use the term *rigidity profile* to describe the ordered set of force constants for all the residues in a given protein. Note that, following earlier studies which showed how prosthetic groups, such as hemes, or small ligands had little influence on calculated force constants[24, 42], we chose not to

include these in the protein representations. This enables us to study the proteins intrinsic mechanical properties independently of the nature and position of any bound ligand.

Following a procedure developed during earlier work on globins[28] and hydrogenases[29], we compared the rigidity profiles obtained for the selected representatives models. The models extracted from the NMR conformational ensembles do not present important structural variations, with Cα rmsds between two models from the same PDB entry being always inferior to 3Å (except for the structure of parvulin domain from *S. aureus*, 2jzv, which presents both a flexible N-terminal tail and a flexible loop). However, these small structural changes are nonetheless sufficient to induce noticeable variations in the mechanical properties of a limited number of residues. We made a systematic pairwise comparison of all the rigidity profiles produced for the representative models from a single PDB entry and, for each residue in the protein, kept the maximum value that could be observed for its force constant variation. We then use the resulting max(Δk) profiles to determine *mechanically sensitive (MS) residues*, i. e. residues presenting the largest force constant variations between two NMR models. This procedure leads to the selection of roughly 10% of the residues for each NMR structure that are listed in Table 1.

Sequence alignments: Sequences from pdb entries belonging to the same protein family were aligned using the online MUSCLE tool, https://www.ebi.ac.uk/Tools/msa/muscle/[43], with the default parameters. Despite the important mechanical similarity that can be observed in the rigidity profiles within a single protein family, the corresponding sequences display relatively low identities ranging from 15% to 82% (see table 1).

Mechanically sensitive residues that were identified using the protocol described in the previous section are highlighted in bold and color along the protein sequences in Figure 3a and all throughout the electronic supplementary material.

3. Results

Globins

Four NMR structures were selected in the PDB for the globin family: Myoglobin (Mgb) from *P. catodon* (PDB entry 1myf), truncated hemoglobin (Tr. Hb) from *S. sp. PCC 6803* (1mwb), and the α and β chains from human hemoglobin, Hb (2m6z). The force constant profiles for the main representative model of each protein are shown in Figure 1. Similar to what has been observed in previous studies [28, 39], the analogous aspects of the profiles reflect the common globin fold, with α-helices appearing as grouped rigidity

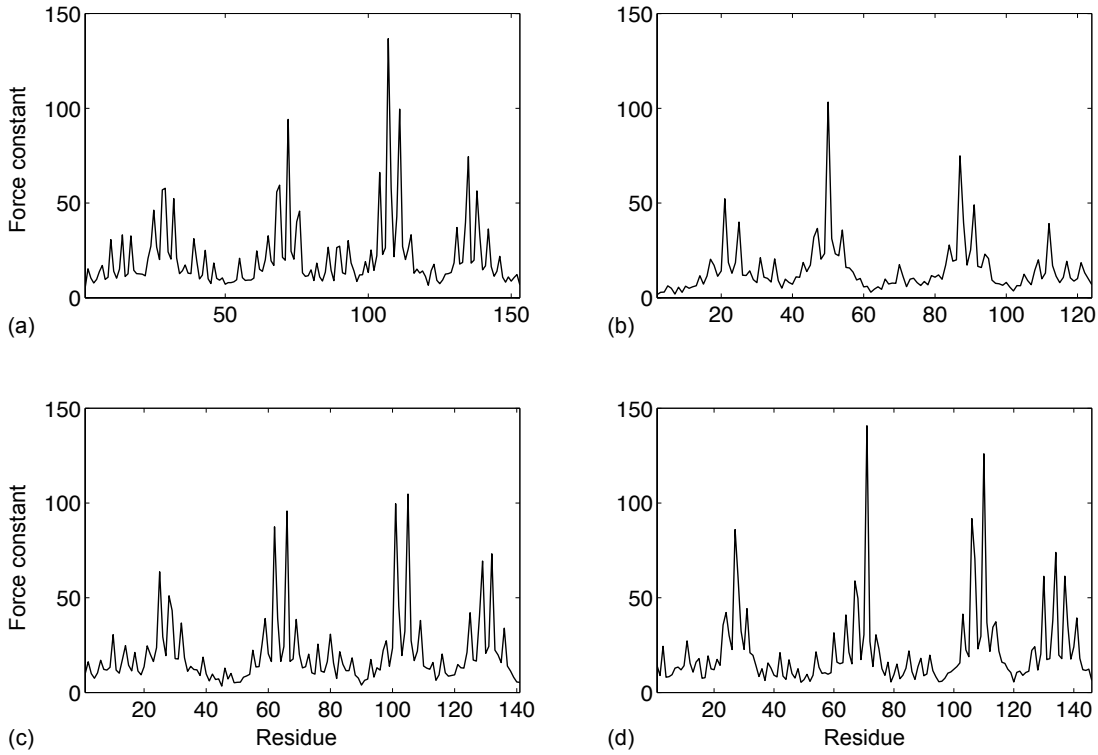


Figure 1: Rigidity profiles (in kcal.mol^{-1}.Å^{-2}) of the first NMR representative model for the four globin chains under study. (a) Mgb, (b) Tr. Hb, (c) αHb, (d) βHb.

peaks along the protein sequence. The max(Δk) profiles resulting from the systematic pairwise comparison of the force constants calculated for different representative models for a given NMR conformational ensemble are shown in Figure 2 and residues with max(Δk) values above a given threshold (indicated by the red horizontal line) are defined as mechanically sensitive residues, since their mechanical properties are the most impacted by the small conformational variations observed between the NMR models. The MS residue positions are highlighted in red on the multiple sequence alignement shown in Figure 3a and we can see that most of these positions are common to one or more globins. Four positions in particular, that are framed and highlighted in bold green in Figure 3a, E11, E15, G8 and G12 (using the standard numbering for globins that indicates the residue position along helices E and G) correspond to MS

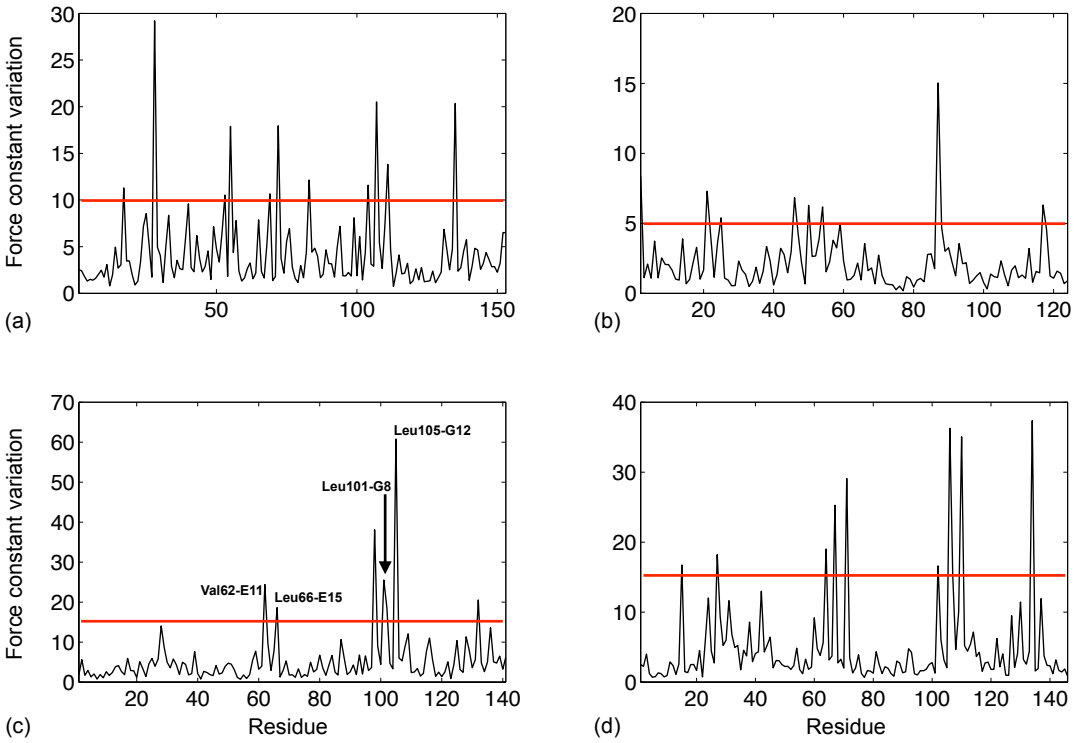


Figure 2: Maximum variation (in kcal.mol^{-1}.Å^{-2}) of the force constant observed upon pairwise comparison of the NMR representative models. (a) Mgb, (b) Tr. Hb, (c) αHb, (d) βHb. The red horizontal line indicates the threshold value chosen for the selection of mechanically sensitive residues listed in table 1.

residues in three chains out of four. We can see from Figure 3b that these residues form a central cluster right at the heart of the globin fold, with their side chain tightly interacting, and they all correspond to tunnel or pocket lining residues. These four positions also exactly concurs with the mechanical nucleus of globins which was determined in our earlier work using clusterized representative structures from a trajectory produced by all-atom classical Molecular Dynamic simulations[28]. A search in the literature shows that the mechanical nucleus residues play an important part in protein function by controlling the ligand diffusion process. Positions E11 and G8 were the object of numerous mutational studies in myoglobin [44-46], E11 and E15 are thought to regulate ligand migration in tr. Hb by acting as gate-opening molecular switches[47-50],

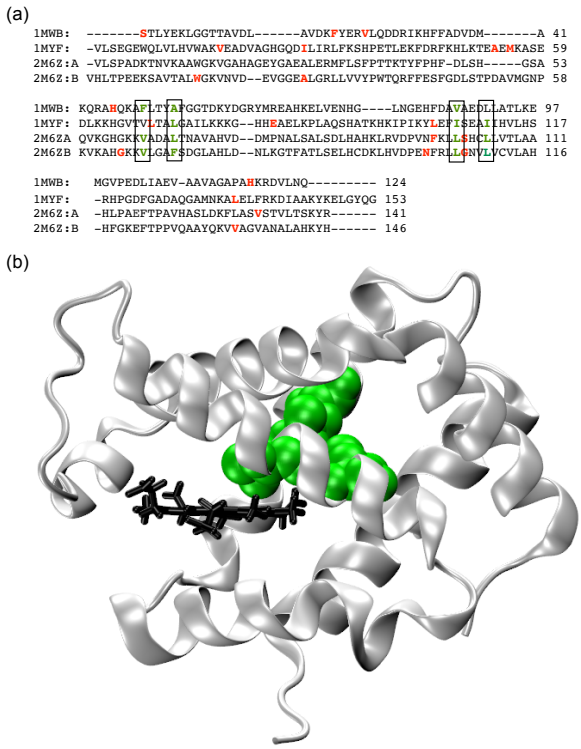


Figure 3: (a) Multiple sequence alignment for the globin structures; mechanically sensitive residues are shown in bold red, and residues belonging to the mechanical nucleus are framed and highlighted in bold green. (b) Cartoon representation of myoglobin with the heme group shown in black and the mechanical nucleus residues as green van der Waals spheres. Figure 3a and Figures 4-7 were prepared using Visual Molecular Dynamics[78].

and in human Hb, gating movements of the leucine residue in G12 governs the hopping of gaseous ligand from and to different binding sites[51]. In neuroglobin, the mechanical properties for all four MN residues have also been shown to be particularly sensitive to pressure changes in high-pressure crystallography[52].

Dihydrofolate reductase

Four NMR conformational ensembles were found in the PDB for dihydrolate reductases (DHFR) from *H. sapiens* (1yho), *L. casei* (2l28), *B. anthracis* (2kgk), and *H. volcanii* (2ith). The combination of coarse-grain mechanical calculations and multiple sequence alignement leads to the identification, when keeping only MS residues that are present in at least 3 out of 4 sequences (see the alignment in Figure SI1), of a six-residues mechanical nucleus comprising Cys6, Ile7, Val8, Val115, Gly116 and Try121 (using the human sequence numbering) shown in Figure 4. All residues from the mechanical nucleus are listed as tunnel lining or cavity lining in ChannelsDB or CASTP and lie on the interface between the subdomains defined in ref. (53). A search in the experimental literature shows residues 6-8 belong to the protein initial β-strand, that contains residues critical for ligand and cofactor binding[54-55], Phe96 from the BaDHFR (corresponding to Val115 in the human DHFR) is also involved in ligand biding[55] and so is Ile94, its equivalent in the *E. coli* DHFR variant[56]. Interestingly, the F98Y mutation in DHFR from *S. aureus* (which corresponds to a mutation of Tyr121 in human DHFR) is responsible for an important loss in TMP affinity, due to Tyr98 forming a hydrogen bond with Leu5 (Ile7 in human), thus preventing it from binding the TMP ligand[57]. The same mutation has also been investigated in DHFR from *M. tuberculosis*, with the Y100F mutant displaying low affinity for classical DHFR inhibitors[58].

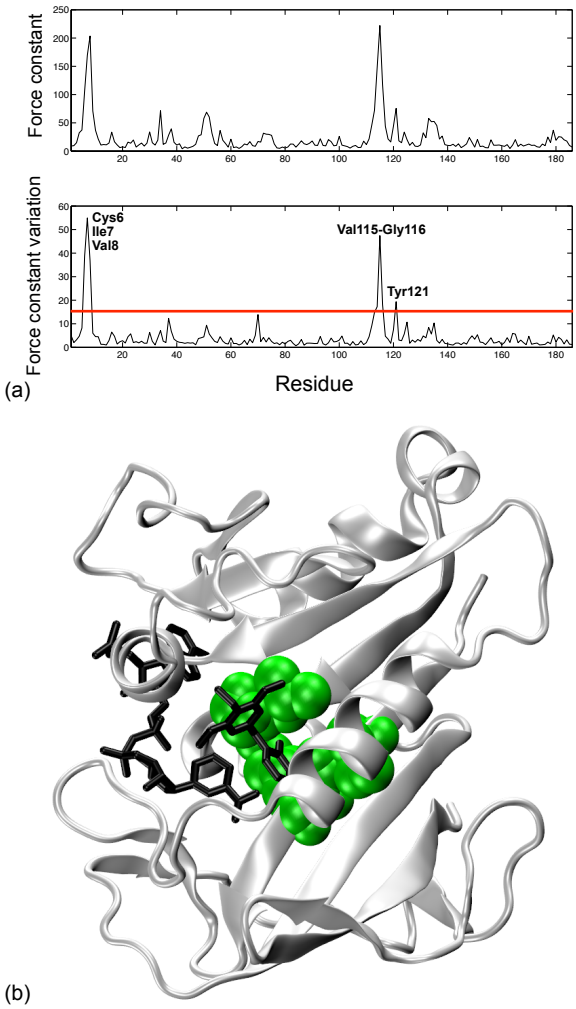


Figure 4: (a) Upper black line, rigidity profile of the first NMR representative model of human DHFR. Lower black line, maximum variation of the force constant observed upon pairwise comparison of the NMR representative models for human DHFR, the red horizontal line indicates the threshold value chosen for the selection of mechanically sensitive residues listed in table 1. (b) Cartoon representation of human DHFR with the NDP and TRR ligands shown in black and the mechanical nucleus residues as green van der Waals spheres.

Cyclophilin A

Coarse-grain calculations were made on three NMR conformational ensembles for cyclophilin A from *H.sapiens* (1oca), *G. kaustophilus* (2mvz) and *E. coli* (1clh). The consensus mechanical nucleus is formed by six residues (Phe36, Gly65, Leu98, Phe112, Phe113 and Phe129, using the human sequence numbering) that correspond to MS positions in at least 2 out of the 3 sequences (see Figure SI1 for the alignment). The

ChannelsDB entry for 3rdd, which corresponds to a crystallographic structure for human CPA, indicates that Phe113 is a tunnel lining residues, and CASTP calculations identify all the MS residues in the *E. coli* CPA as cavity lining residues. From the experimental literature, we find that the ligand binding pocket in human CPA comprises residues Gly65, Leu98, Phe112 and Phe113, and Phe36, Phe112 and Phe129 belong to the protein aromatic core[59]. In a study using NMR relaxation experiments[60],

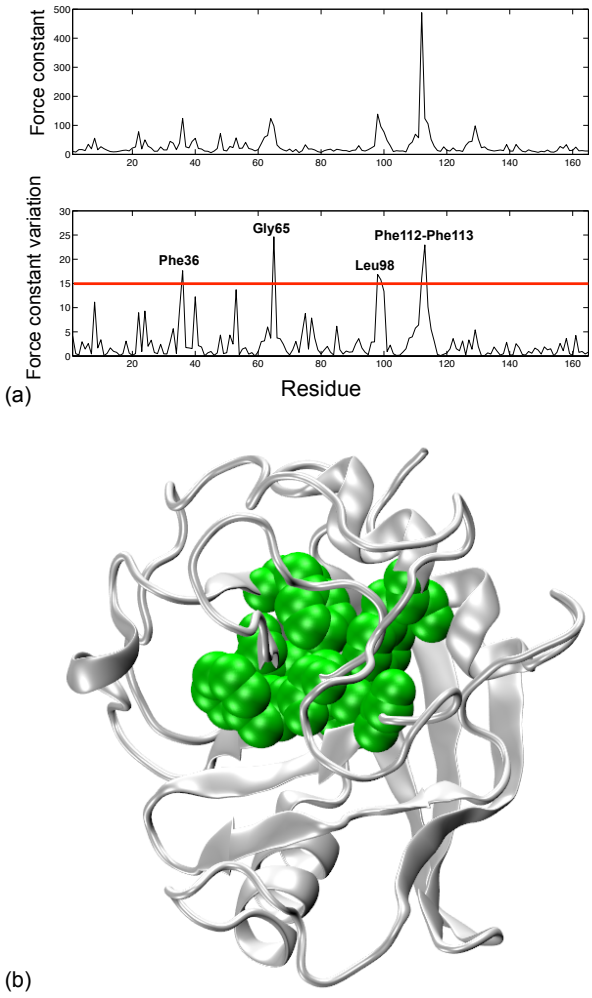


Figure 5: (a) Upper black line, rigidity profile of the first NMR representative model of human CPA. Lower black line, maximum variation of the force constant observed upon pairwise comparison of the NMR representative models for human CPA, the red horizontal line indicates the threshold value chosen for the selection of mechanically sensitive residues listed in table 1. (b) Cartoon representation of human CPA with the mechanical nucleus residues as green van der Waals spheres.

Eisenmesser et al. showed how the protein motions play an essential part in its catalytic activity, and this catalytic activity is severely diminished for the F113W mutant. In addition, the large variations in rigidity observed for Leu98, Ser99 and Phe113 in human CPA concur with the fact that these residues display the largest differences in NMR chemical shifts between the enzyme's major and minor conformers. A mutational study on Ser99 (which lies outside of the catalytic site), designed to stabilized the previously hidden minor conformation, causes a large decrease in the catalytic rate[61]. In the E. coli CPA, residues Ile88, Phe103 and Phe125 belong to the protein hydrophobic core, but are not directly involved in its isomerase activity[62]. Interestingly, MS residues Ala91 and Phe103 are located on both extremities of the B5-B6 surface loop, which overlooks the active site. As a consequence, mechanical variations in these two residues are likely to impact the loop conformational variability, and therefore the ligand's access to the catalytic site. A similar effect is also observed in the CPA from *G. kaustophilus*, with residues Gly57, Met87 and Phe101 lying on the edge of active site loops[63].

Peptidyl-tRNA hydrolase

Three NMR conformational ensembles were found in the PDB for peptidyl-tRNA hydrolase (Pth) from *M. tuberculosis* (2jrc), *M. smegmatis* (2lgj) and *V. cholerae* (2mjl), which lead to the identification of a three-residues mechanical nucleus formed by Gly25, Ile91 and Ile93 (using the *M. tuberculosis* sequence numbering), see Figure 6. All the mechanically sensitive residues listed in table 1 for the 2jrc and the 2lgj structures, are either tunnel lining or cavity lining residues in ChannelsDB or CASTP. The α/β globular fold for Pth comprises an active site channel connecting the surface of the molecule to the central β-sheet core[64]. The mechanical nucleus residues are not

directly involved in the Pth catalytic activity (unlike the neighboring His22 and Asp95), however, Gly25 is strictly conserved in this proteic family, while Ile91 and Ile93 occupy hydrophobic positions, and all residues are lining the substrate binding channel[65]. Ile93 lies at the base of the *gate loop*, which controls the entrance to the substrate binding cleft[66-67], as has to move away to permit ligand binding in the enzyme's active site, which means that a perturbation of its mechanical properties is likely to disrupt the

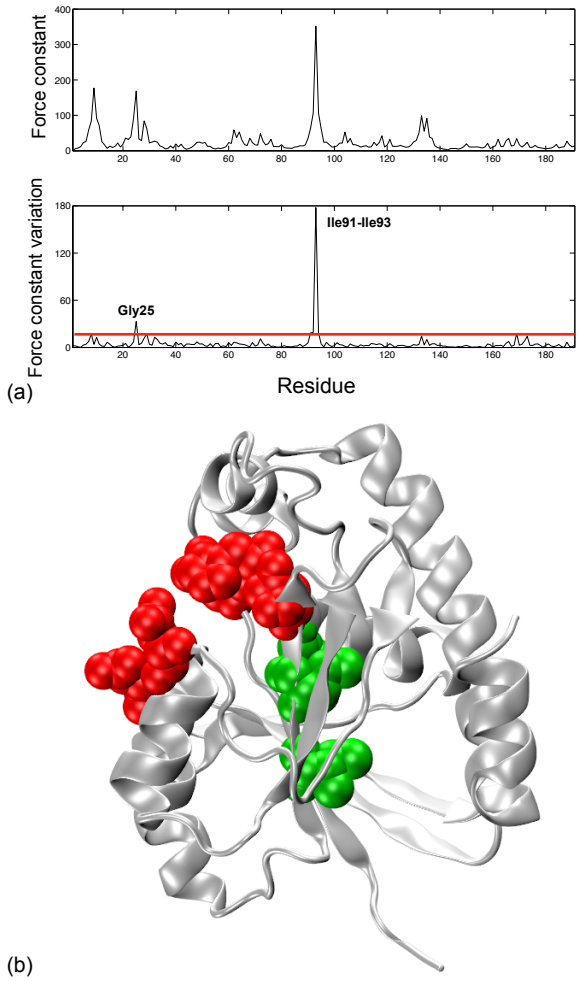


Figure 6: (a) Upper black line, rigidity profile of the first NMR representative model of Pth from *M. tuberculosis*. Lower black line, maximum variation of the force constant observed upon pairwise comparison of the NMR representative models for Pth from *M. tuberculosis*, the red horizontal line indicates the threshold value chosen for the selection of mechanically sensitive residues listed in table 1. (b) Cartoon representation of Pth from *M. tuberculosis* with the mechanical nucleus residues as green van der Waals spheres and the catalytic residues (at the entrance of the peptide binding cleft) as red van der Waals spheres.

protein catalytic activity. In addition, NMR studies show that Gly25 is affected by RNA binding to Pth, despite its central position in the protein core[68], due to the interplay between the tRNA binding region and the peptide binding channel[64].

Parvulin

Five NMR conformational ensembles were found in the PDB for human parvulin (1nmw), and variants from *E. coli* (1jns), *A. thaliana* (1j6y), *C. symbiosum* (2rqs) and *S. aureus* (2jzv), thus leading to the identification of a four-residues mechanical nucleus formed by His59, Ile60, Il158 and Ile159 (using human sequence numbering) shown in

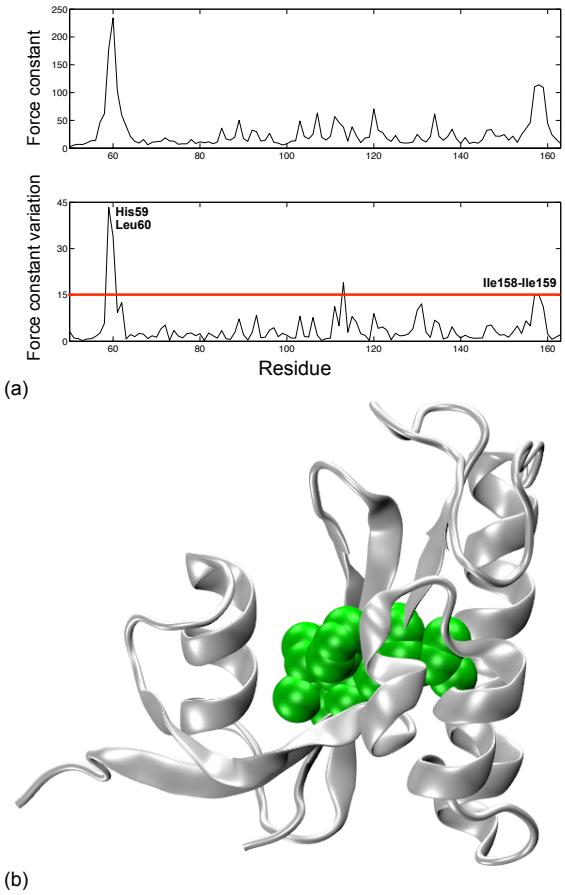


Figure 7: (a) Upper black line, rigidity profile of the first NMR representative model of human parvulin. Lower black line, maximum variation of the force constant observed upon pairwise comparison of the NMR representative models for human parvulin, the red horizontal line indicates the threshold value chosen for the selection of mechanically sensitive residues listed in table 1. (b) Cartoon representation of human parvulin with the mechanical nucleus residues as green van der Waals spheres.

Figure 7. In parvulin from *E. Coli*, all the mechanical nucleus residues are listed as tunnel binding by ChannelsDB and cavity lining by CasTP (see table1). Structural studies on human parvulin highlight the functional importance of the two conserved histidines His59, His157 and Ile159, which are located in the enzyme's active center[69-71]. Interestingly, later mutational studies on the His59/His157 pair showed that theses residues are not directly involved in the parvulin isomerase activity, but more likely contribute to the enzyme's stability[71]. His8 in *E. coli* parvulin, and His14 from the *C. symbiosum* parvulin (which both correspond to His59 in human parvulin) are located in the protein hydrophobic core and belong to the putative binding pocket[72-73], in parvulin from *A. thaliana*, His12 displays no chemical shift after binding of the peptide ligand, thus suggesting that this residue is not directly involved in the catalytic cycle[74]. The NMR conformational ensemble obtained for *S. aureus* parvulin highlights the structural variability of the catalytic histidines (His 146 and His239), and the protein hydrophobic core also comprises residue Ile137, Ile240 and Ile241[75]. A later study combining Molecular Dynamics Simulation and experimentally available S^2 order parameter for three parvulins suggests that the conserved histidines play a key role in the modulation of the enzyme's internal dynamics[76].

4. Discussion and Conclusion

As shown in earlier work on globins and hydrogenases, protein mechanical properties are a dynamic feature[28-29]. Due to the robustness of the Elastic Network Model approach[77], a protein's rigidity profile remains qualitatively the same along time, with its main peaks associated with a given set of residues (that are often likely to be catalytic residues). However, it will also present noticeable variations from one

equilibrium structure to the other for another limited number of residues (termed as *mechanically sensitive*, MS, residues), even though the conformational variations between these structures remains negligible. We have shown previously that one could use conformational ensembles produced by classic all-atom Molecular Dynamics trajectories to identify MS residues in a protein, and repeating this approach on several proteins from the same family will lead to the definition of this family's *mechanical nucleus*, i. e. a set of conserved positions along the protein sequence that are occupied by MS residues. In the present work, we use protein conformational ensembles obtained by NMR spectroscopy for five distinct protein families, and show that the structural sampling from these ensembles is sufficient to identify a mechanical nucleus comprising between 3 and 6 residues in each family. In the case of globins, this mechanical nucleus concurs perfectly with the globin MN determined using trajectories from MD simulations. Its four constitutive residues have been highlighted as playing a key role in substrate diffusion in the protein matrix in numerous experimental approaches. The MN residues identified in the four other protein families also appear to be excellent candidates for playing a part in protein function as gating residues. Most of them are tunnel or cavity lining residues (underlined residues in Table 1), and for each protein family, experimental evidence, from point mutation studies or evolutionary data, supports the hypothesis that, while not being directly involved in the chemical reaction taking place in the protein active site, the MN residues are nonetheless important for protein function. In particular, by modulating the shape and size of the cavities and tunnels leading to the protein active site, MN residues can impact an enzyme's substrate selectivity and catalytic efficiency. Therefore, they represent an interesting target for the fine tuning of protein function. Altogether, our use of NMR

structural ensembles and coarse-grain calculations provides us with a simple and efficient tool for identifying key spots in the protein structure, where point mutations could help modulate the protein enzymatic activity or specificity. While the current study has been limited to small proteins, comprising less than 200 residues, recent progress in NMR spectroscopy on large systems[32] should ensure that experimental conformational ensembles for larger proteins will be accessible in the near future.

Acknowledgments: This work was supported by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011-01).

Supporting Material: The multiple sequence alignments for the 5 protein families under study are available as supporting information.

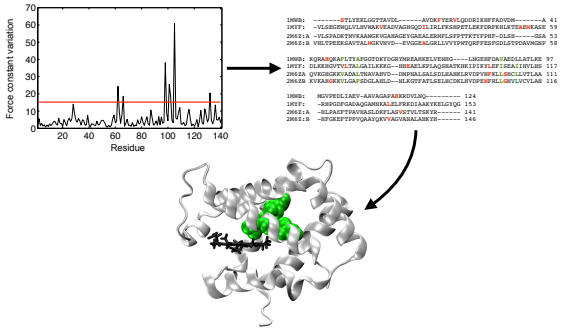
References

1. Micheletti, C., Comparing proteins by their internal dynamics: exploring structure-function relationships beyond static structural alignments. *Phys Life Rev* **2013,** *10* (1), 1-26.
2. Orozco, M., A theoretical view of protein dynamics. *Chem. Soc. Rev.* **2014,** *43* (14), 5051-66.
3. Monzon, A. M.; Zea, D. J.; Fornasari, M. S.; Saldano, T. E.; Fernandez-Alberti, S.; Tosatto, S. C.; Parisi, G., Conformational diversity analysis reveals three functional mechanisms in proteins. *PLoS Comput Biol* **2017,** *13* (2), e1005398.
4. Gutteridge, A.; Thornton, J., Conformational changes observed in enzyme crystal structures upon substrate binding. *Journal of Molecular Biology* **2005,** *346* (1), 21-8.
5. Mesecar, A. D.; Stoddard, B. L.; Koshland, D. E., Jr., Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. *Science* **1997,** *277* (5323), 202-6.
6. Koshland, D. E., Jr., Conformational changes: how small is big enough? *Nat Med* **1998,** *4* (10), 1112-4.
7. Boehr, D. D.; D'Amico, R. N.; O'Rourke, K. F., Engineered control of enzyme structural dynamics and function. *Protein science : a publication of the Protein Society* **2018,** *27* (4), 825-838.
8. Johansson, K. E.; Lindorff-Larsen, K., Structural heterogeneity and dynamics in protein evolution and design. *Curr. Opin. Struct. Biol.* **2018,** *48*, 157-163.
9. Maria-Solano, M. A.; Serrano-Hervas, E.; Romero-Rivera, A.; Iglesias-Fernandez, J.; Osuna, S., Role of conformational dynamics in the evolution of novel enzyme function. *Chem Commun (Camb)* **2018,** *54* (50), 6622-6634.
10. Zhou, H. X.; McCammon, J. A., The gates of ion channels and enzymes. *Trends Biochem. Sci.* **2010,** *35* (3), 179-185.
11. Gora, A.; Brezovsky, J.; Damborsky, J., Gates of enzymes. *Chem Rev* **2013,** *113* (8), 5871-923.
12. Furnham, N.; Holliday, G. L.; de Beer, T. A.; Jacobsen, J. O.; Pearson, W. R.; Thornton, J. M., The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.* **2014,** *42* (Database issue), D485-9.
13. Pravda, L.; Berka, K.; Svobodova Varekova, R.; Sehnal, D.; Banas, P.; Laskowski, R. A.; Koca, J.; Otyepka, M., Anatomy of enzyme channels. *BMC Bioinformatics* **2014,** *15*, 379.
14. Kingsley, L. J.; Lill, M. A., Substrate tunnels in enzymes: structure-function relationships and computational methodology. *Proteins* **2015,** *83* (4), 599-611.
15. Kaushik, S.; Marques, S. M.; Khirsariya, P.; Paruch, K.; Libichova, L.; Brezovsky, J.; Prokop, Z.; Chaloupkova, R.; Damborsky, J., Impact of the access tunnel engineering on catalysis is strictly ligand-specific. *The FEBS journal* **2018,** *285* (8), 1456-1476.
16. Marques, S. M.; Daniel, L.; Buryska, T.; Prokop, Z.; Brezovsky, J.; Damborsky, J., Enzyme Tunnels and Gates As Relevant Targets in Drug Design. *Med. Res. Rev.* **2017,** *37* (5), 1095-1139.
17. Maximova, T.; Moffatt, R.; Ma, B.; Nussinov, R.; Shehu, A., Principles and Overview of Sampling Methods for Modeling Macromolecular Structure and Dynamics. *PLoS Comput Biol* **2016,** *12* (4), e1004619.
18. Narayanan, C.; Bafna, K.; Roux, L. D.; Agarwal, P. K.; Doucet, N., Applications of NMR and computational methodologies to study protein dynamics. *Arch. Biochem. Biophys.* **2017,** *628*, 71-80.
19. Debiec, K. T.; Whitley, M. J.; Koharudin, L. M. I.; Chong, L. T.; Gronenborn, A. M., Integrating NMR, SAXS, and Atomistic Simulations: Structure and Dynamics of a Two-Domain Protein. *Biophys. J.* **2018,** *114* (4), 839-855.
20. Seyler, S. L.; Beckstein, O., Sampling large conformational transitions: adenylate kinase as a testing ground. *Molecular Simulation* **2014,** *40* (10-11), 855-877.
21. Poma, A. B.; Cieplak, M.; Theodorakis, P. E., Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. *J Chem Theory Comput* **2017,** *13* (3), 1366-1374.
22. Tiwari, S. P.; Reuter, N., Conservation of intrinsic dynamics in proteins-what have computational models taught us? *Curr. Opin. Struct. Biol.* **2017,** *50*, 75-81.
23. Zheng, W.; Wen, H., A survey of coarse-grained methods for modeling protein conformational transitions. *Curr. Opin. Struct. Biol.* **2017,** *42*, 24-30.
24. Sacquin-Mora, S.; Lavery, R., Investigating the local flexibility of functional residues in hemoproteins. *Biophys. J.* **2006,** *90* (8), 2706-17.
25. Sacquin-Mora, S.; Sebban, P.; Derrien, V.; Frick, B.; Lavery, R.; Alba-Simionesco, C., Probing the flexibility of the bacterial reaction center: the wild-type protein is more rigid than two site-specific mutants. *Biochemistry* **2007,** *46* (51), 14960-8.
26. Sacquin-Mora, S., Motions and mechanics: investigating conformational transitions in multi-domain proteins with coarse-grain simulations. *Mol. Simul.* **2014,** *40* (1-3), 229-236.
27. Sacquin-Mora, S., Bridging Enzymatic Structure Function via Mechanics: A Coarse-Grain Approach. *Methods Enzymol.* **2016,** *578*, 227-48.
28. Bocahut, A.; Bernad, S.; Sebban, P.; Sacquin-Mora, S., Frontier residues lining globin internal cavities present specific mechanical properties. *J. Am. Chem. Soc.* **2011,** *133* (22), 8753-61.

29. Oteri, F.; Baaden, M.; Lojou, E.; Sacquin-Mora, S., Multiscale simulations give insight into the hydrogen in and out pathways of [NiFe]-hydrogenases from Aquifex aeolicus and Desulfovibrio fructosovorans. *J Phys Chem B* **2014,** *118* (48), 13800-11.
30. Wider, G.; Wuthrich, K., NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Current Opinion in Structural Biology* **1999,** *9* (5), 594-601.
31. Kay, L. E., New Views of Functionally Dynamic Proteins by Solution NMR Spectroscopy. *J. Mol. Biol.* **2016,** *428* (2 Pt A), 323-331.
32. Jiang, Y.; Kalodimos, C. G., NMR Studies of Large Proteins. *J. Mol. Biol.* **2017,** *429* (17), 2667-2676.
33. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000,** *28* (1), 235-42.
34. Kelley, L. A.; Sutcliffe, M. J., OLDERADO: on-line database of ensemble representatives and domains. On Line Database of Ensemble Representatives And DOmains. *Protein science : a publication of the Protein Society* **1997,** *6* (12), 2628-30.
35. Kelley, L. A.; Gardner, S. P.; Sutcliffe, M. J., An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* **1996,** *9* (11), 1063-5.
36. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004,** *25* (13), 1605-12.
37. Pravda, L.; Sehnal, D.; Svobodova Varekova, R.; Navratilova, V.; Tousek, D.; Berka, K.; Otyepka, M.; Koca, J., ChannelsDB: database of biomacromolecular tunnels and pores. *Nucleic Acids Res.* **2018,** *46* (D1), D399-D405.
38. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J., CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **2018,** *46* (W1), W363-W367.
39. Sacquin-Mora, S., Mechanical variations in proteins with large-scale motions highlight the formation of structural locks. *J Struct Biol* **2018,** *203* (3), 195-204.
40. Tozzini, V., Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **2005,** *15* (2), 144-50.
41. Zacharias, M., Protein-protein docking with a reduced protein model accounting for side-chain flexibility. *Protein Sci.* **2003,** *12* (6), 1271-82.
42. Sacquin-Mora, S.; Laforet, E.; Lavery, R., Locating the active sites of enzymes using mechanical properties. *Proteins* **2007,** *67* (2), 350-9.
43. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004,** *32* (5), 1792-7.
44. Quillin, M. L.; Li, T. S.; Olson, J. S.; Phillips, G. N.; Dou, Y.; Ikedasaito, M.; Regan, R.; Carlson, M.; Gibson, Q. H.; Li, H. Y.; Elber, R., STRUCTURAL AND FUNCTIONAL-EFFECTS OF APOLAR MUTATIONS OF THE DISTAL VALINE IN MYOGLOBIN. *J. Mol. Biol.* **1995,** *245* (4), 416-436.
45. Ishikawa, H.; Uchida, T.; Takahashi, S.; Ishimori, K.; Morishima, I., Ligand migration in human myoglobin: Steric effects of isoleucine 107(G8) on O-2 and CO binding. *Biophys. J.* **2001,** *80* (3), 1507-1517.
46. Dantsker, D.; Roche, C.; Samuni, U.; Blouin, G.; Olson, J. S.; Friedman, J. M., The position 68(E11) side chain in myoglobin regulates ligand capture, bond formation with heme iron, and internal movement into the xenon cavities. *J. Biol. Chem.* **2005,** *280* (46), 38740-38755.
47. Bidon-Chanal, A.; Marti, M. A.; Crespo, A.; Milani, M.; Orozco, M.; Bolognesi, M.; Luque, F. J.; Estrin, D. A., Ligand-induced dynamical regulation of NO conversion in Mycobacterium tuberculosis truncated hemoglobin-N. *Proteins* **2006,** *64* (2), 457-464.
48. Bidon-Chanal, A.; Marti, M. A.; Estrin, D. A.; Luque, F. J., Dynamical regulation of ligand migration by a gate-opening molecular switch in truncated hemoglobin-N from Mycobacterium tuberculosis. *J. Am. Chem. Soc.* **2007,** *129* (21), 6782-6788.
49. Lama, A.; Pawaria, S.; Bidon-Chanal, A.; Anand, A.; Gelpi, J. L.; Arya, S.; Marti, M.; Estrin, D. A.; Luque, F. J.; Dikshit, K. L., Role of Pre-A Motif in Nitric Oxide Scavenging by Truncated Hemoglobin, HbN, of Mycobacterium tuberculosis. *J. Biol. Chem.* **2009,** *284* (21), 14457-14468.
50. Mishra, S.; Meuwly, M., Nitric Oxide Dynamics in Truncated Hemoglobin: Docking Sites, Migration Pathways, and Vibrational Spectroscopy from Molecular Dynamics Simulations. *Biophys. J.* **2009,** *96* (6), 2105-2118.
51. Savino, C.; Miele, A. E.; Draghi, F.; Johnson, K. A.; Sciara, G.; Brunori, M.; Vallone, B., Pattern of Cavities in Globins: The Case of Human Hemoglobin. *Biopolymers* **2009,** *91* (12), 1097-1107.
52. Colloc'h, N.; Sacquin-Mora, S.; Avella, G.; Dhaussy, A. C.; Prange, T.; Vallone, B.; Girard, E., Determinants of neuroglobin plasticity highlighted by joint coarse-grained simulations and high pressure crystallography. *Scientific reports* **2017,** *7* (1), 1858.
53. Sawaya, M. R.; Kraut, J., Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence. *Biochemistry* **1997,** *36* (3), 586-603.

54. Kovalevskaya, N. V.; Smurnyy, Y. D.; Polshakov, V. I.; Birdsall, B.; Bradbury, A. F.; Frenkiel, T.; Feeney, J., Solution structure of human dihydrofolate reductase in its complex with trimethoprim and NADPH. *J Biomol NMR* **2005,** *33* (1), 69-72.
55. Beierlein, J. M.; Deshmukh, L.; Frey, K. M.; Vinogradova, O.; Anderson, A. C., The solution structure of Bacillus anthracis dihydrofolate reductase yields insight into the analysis of structure-activity relationships for novel inhibitors. *Biochemistry* **2009,** *48* (19), 4100-8.
56. Bhabha, G.; Lee, J.; Ekiert, D. C.; Gam, J.; Wilson, I. A.; Dyson, H. J.; Benkovic, S. J.; Wright, P. E., A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* **2011,** *332* (6026), 234-8.
57. Frey, K. M.; Liu, J.; Lombardo, M. N.; Bolstad, D. B.; Wright, D. L.; Anderson, A. C., Crystal structures of wild-type and mutant methicillin-resistant Staphylococcus aureus dihydrofolate reductase reveal an alternate conformation of NADPH that may be linked to trimethoprim resistance. *J. Mol. Biol.* **2009,** *387* (5), 1298-308.
58. Dias, M. V.; Tyrakis, P.; Domingues, R. R.; Paes Leme, A. F.; Blundell, T. L., Mycobacterium tuberculosis dihydrofolate reductase reveals two conformational states and a possible low affinity mechanism to antifolate drugs. *Structure* **2014,** *22* (1), 94-103.
59. Ottiger, M.; Zerbe, O.; Guntert, P.; Wuthrich, K., The NMR solution conformation of unligated human cyclophilin A. *J. Mol. Biol.* **1997,** *272* (1), 64-81.
60. Eisenmesser, E. Z.; Millet, O.; Labeikovsky, W.; Korzhnev, D. M.; Wolf-Watz, M.; Bosco, D. A.; Skalicky, J. J.; Kay, L. E.; Kern, D., Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **2005,** *438* (7064), 117-21.
61. Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T., Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **2009,** *462* (7273), 669-73.
62. Clubb, R. T.; Ferguson, S. B.; Walsh, C. T.; Wagner, G., Three-dimensional solution structure of Escherichia coli periplasmic cyclophilin. *Biochemistry* **1994,** *33* (10), 2761-72.
63. Holliday, M. J.; Camilloni, C.; Armstrong, G. S.; Isern, N. G.; Zhang, F.; Vendruscolo, M.; Eisenmesser, E. Z., Structure and Dynamics of GeoCyp: A Thermophilic Cyclophilin with a Novel Substrate Binding Mechanism That Functions Efficiently at Low Temperatures. *Biochemistry* **2015,** *54* (20), 3207-17.
64. Pulavarti, S. V.; Jain, A.; Pathak, P. P.; Mahmood, A.; Arora, A., Solution structure and dynamics of peptidyl-tRNA hydrolase from Mycobacterium tuberculosis H37Rv. *J. Mol. Biol.* **2008,** *378* (1), 165-77.
65. Kabra, A.; Fatma, F.; Shahid, S.; Pathak, P. P.; Yadav, R.; Pulavarti, S. V.; Tripathi, S.; Jain, A.; Arora, A., Structural characterization of peptidyl-tRNA hydrolase from Mycobacterium smegmatis by NMR spectroscopy. *Biochim. Biophys. Acta* **2016,** *1864* (10), 1304-14.
66. Singh, A.; Gautam, L.; Sinha, M.; Bhushan, A.; Kaur, P.; Sharma, S.; Singh, T. P., Crystal structure of peptidyl-tRNA hydrolase from a Gram-positive bacterium, Streptococcus pyogenes at 2.19 A resolution shows the closed structure of the substrate-binding cleft. *FEBS Open Bio* **2014,** *4*, 915-22.
67. Zhang, F.; Song, Y.; Niu, L.; Teng, M.; Li, X., Crystal structure of Staphylococcus aureus peptidyl-tRNA hydrolase at a 2.25 A resolution. *Acta Biochim Biophys Sin (Shanghai)* **2015,** *47* (12), 1005-10.
68. Giorgi, L.; Bontems, F.; Fromant, M.; Aubard, C.; Blanquet, S.; Plateau, P., RNA-binding site of Escherichia coli peptidyl-tRNA hydrolase. *The Journal of biological chemistry* **2011,** *286* (45), 39585-94.
69. 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* **1997,** *89* (6), 875-86.
70. Bayer, E.; Goettsch, S.; Mueller, J. W.; Griewel, B.; Guiberman, E.; Mayr, L. M.; Bayer, P., Structural analysis of the mitotic regulator hPin1 in solution: insights into domain architecture and substrate binding. *The Journal of biological chemistry* **2003,** *278* (28), 26183-93.
71. 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* **2008,** *47* (44), 11481-9.
72. Kuhlewein, A.; Voll, G.; Hernandez Alvarez, B.; Kessler, H.; Fischer, G.; Rahfeld, J. U.; Gemmecker, G., Solution structure of Escherichia coli Par10: The prototypic member of the Parvulin family of peptidyl-prolyl cis/trans isomerases. *Protein science : a publication of the Protein Society* **2004,** *13* (9), 2378-87.
73. Jaremko, L.; Jaremko, M.; Elfaki, I.; Mueller, J. W.; Ejchart, A.; Bayer, P.; Zhukov, I., Structure and dynamics of the first archaeal parvulin reveal a new functionally important loop in parvulin-type prolyl isomerases. *The Journal of biological chemistry* **2011,** *286* (8), 6554-65.
74. Landrieu, I.; Wieruszeski, J. M.; Wintjens, R.; Inze, D.; Lippens, G., Solution structure of the single-domain prolyl cis/trans isomerase PIN1At from Arabidopsis thaliana. *J. Mol. Biol.* **2002,** *320* (2), 321-32.

75. Heikkinen, O.; Seppala, R.; Tossavainen, H.; Heikkinen, S.; Koskela, H.; Permi, P.; Kilpelainen, I., Solution structure of the parvulin-type PPIase domain of Staphylococcus aureus PrsA--implications for the catalytic mechanism of parvulins. *BMC Struct Biol* **2009,** *9*, 17.
76. Czajlik, A.; Kovacs, B.; Permi, P.; Gaspari, Z., Fine-tuning the extent and dynamics of binding cleft opening as a potential general regulatory mechanism in parvulin-type peptidyl prolyl isomerases. *Scientific reports* **2017,** *7*, 44504.
77. Lopez-Blanco, J. R.; Chacon, P., New generation of elastic network models. *Curr. Opin. Struct. Biol.* **2016,** *37*, 46-53.
78. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-8, 27-8.


Force constant variation
Residue
0
10
20
30
40
50
60
70
20
40
60
80
100
120
140

TOC image