

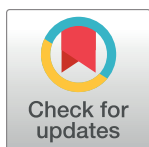
RESEARCH ARTICLE

# Conformational diversity analysis reveals three functional mechanisms in proteins

Alexander Miguel Monzon<sup>1</sup>, Diego Javier Zea<sup>2</sup>, María Silvina Fornasari<sup>1</sup>, Tadeo E. Saldaño<sup>1</sup>, Sebastian Fernandez-Alberti<sup>1</sup>, Silvio C. E. Tosatto<sup>3</sup>, Gustavo Parisi<sup>1\*</sup>

**1** Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes (CONICET), Bernal, Buenos Aires, Argentina, **2** Bioinformatics Unit, Fundación Instituto Leloir (CONICET), Buenos Aires, Argentina, **3** Department of Biomedical Sciences, University of Padua, Padua, Italy

\* [gusparisi@gmail.com](mailto:gusparisi@gmail.com)



**OPEN ACCESS**

**Citation:** Monzon AM, Zea DJ, Fornasari MS, Saldaño TE, Fernandez-Alberti S, Tosatto SCE, et al. (2017) Conformational diversity analysis reveals three functional mechanisms in proteins. *PLoS Comput Biol* 13(2): e1005398. doi:10.1371/journal.pcbi.1005398

**Editor:** Christine A. Orengo, University College London, UNITED KINGDOM

**Received:** September 10, 2016

**Accepted:** February 2, 2017

**Published:** February 13, 2017

**Copyright:** © 2017 Monzon et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** Funding for this research was provided by COST Action (BM1405) Non-Globular Proteins-net (SCET) and Universidad Nacional de Quilmes (PUNQ 1004/11) (GP). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

Protein motions are a key feature to understand biological function. Recently, a large-scale analysis of protein conformational diversity showed a positively skewed distribution with a peak at 0.5 Å C-alpha root-mean-square-deviation (RMSD). To understand this distribution in terms of structure-function relationships, we studied a well curated and large dataset of ~5,000 proteins with experimentally determined conformational diversity. We searched for global behaviour patterns studying how structure-based features change among the available conformer population for each protein. This procedure allowed us to describe the RMSD distribution in terms of three main protein classes sharing given properties. The largest of these protein subsets (~60%), which we call “rigid” (average RMSD = 0.83 Å), has no disordered regions, shows low conformational diversity, the largest tunnels and smaller and buried cavities. The two additional subsets contain disordered regions, but with differential sequence composition and behaviour. Partially disordered proteins have on average 67% of their conformers with disordered regions, average RMSD = 1.1 Å, the highest number of hinges and the longest disordered regions. In contrast, malleable proteins have on average only 25% of disordered conformers and average RMSD = 1.3 Å, flexible cavities affected in size by the presence of disordered regions and show the highest diversity of cognate ligands. Proteins in each set are mostly non-homologous to each other, share no given fold class, nor functional similarity but do share features derived from their conformer population. These shared features could represent conformational mechanisms related with biological functions.

## Author summary

Protein motions are commonly quantified measuring structural differences between conformers. The extension of these differences are called conformational diversity. These motions are essential to understand protein biology. We have found that the distribution of conformational diversity in a large dataset of proteins could be explained in terms of three sets sharing structure-based features emerging from the conformer population for each protein. The first set, which we called rigid, involve proteins showing almost no

backbone movements but with important changes in tunnels. In order of increasing conformational diversity, the other sets are called partially disordered and malleable, showing disordered regions and important cavities but with different behaviour to each other. Shared features in each set could represent conformational mechanisms related with biological functions.

## Introduction

Early crystallization studies on myoglobin found no apparent way the oxygen could possibly enter the molecule and bind to heme [1]. It took more than a decade to discover that protein motions were essential for myoglobin to be biologically active [2,3]. After these early findings, an overwhelming amount of information has accumulated relating protein motion with biological function. A wide range of movements have been explored in proteins, from large relative domain movements [4], secondary and tertiary element rearrangements [5] and loop displacements [6] to small residue rearrangements [7]. The upper limit in this scale of protein movements may certainly involve intrinsically disordered regions (IDRs) or proteins (IDPs) characterized by their high flexibility and mobility and clearly related with well-established disorder-based biological functions [8] although other notion of disorder role has been proposed [9].

A large-scale survey of protein motion degrees, studying the extension of protein conformational diversity using a redundant collection of crystallized structures for the same protein was recently published [10]. Since the early determination of haemoglobin conformers, it is generally accepted that different crystallographic structures for the same protein (i.e. with and without substrate or post-translational modifications) could represent putative instances of the conformational space of a protein [11]. To measure the structural differences between putative conformers, Burra and co-workers used C-alpha root mean square deviation (RMSD). Clearly, RMSD or other structural similarity scores measure the differences in ordered parts of the proteins and stress the importance of protein motions in the known protein structure space. From this distribution, it is possible to infer how a great majority of proteins have RMSD values compatible with the accepted error in estimating a structure using X-ray crystallography (about 0.4 Å). The obtained distribution is consistent with other works reporting low degrees of conformational diversity in proteins. In a study of conformational changes in 60 enzymes between their apo and substrate-bound forms, 75% of the data had an RMSD less than 1 Å, and 91% less than 2 Å, with an average of 0.7 Å [12]. Interestingly, comparisons of apo structures of the same protein show an RMSD of 0.5 Å, a value slightly below the observed apo and substrate-bound average. In agreement with these results, large-scale protein motions are not necessary to sustain biological function in the majority of proteins studied. This observation is supported with the finding that even small changes between conformers could greatly affect catalytic parameters and biological behaviour of enzymes [13,14]. Also, it has been suggested that key and widely extended biological properties in proteins [15], such as allostery and cooperativity, could arise from changes in the width of conformational distributions without any appreciable change in the average structure of the protein [16]. However, the distribution of Burra et al. shows also a large skew towards higher RMSD (observed maximum RMSD = 23.7 Å) indicating that a minor fraction of the analysed dataset does require large conformational changes to be functionally active [10]. High RMSD values, are commonly observed in multidomain proteins where hinge motions produce relative movements of domains as rigid bodies [17].

As pointed out by Fraunfelder, the search for general concepts between dynamics and protein function is a key issue in structural biology [2]. Knowing the degree of conformational changes could contribute highly to our understanding of how these motions modulate protein function. Large-scale analysis of different degrees and types of protein motions could also allow us to infer general rules about different structure-function relationships in protein structural space. In this work, we have used a large and well curated set of proteins derived from the CoDNaS (Conformational Diversity of Native State) database [18] to reproduce Burra et al distribution. Using several structural and dynamical analyses over the available population of conformers for each protein, we found that three main structure-function relationships emerge. Proteins in each group are mostly non-homologous and do not show any fold preference or functional similarity. However, proteins in each set share similar features emerged from conformational analysis, which as a global behaviour could represent conformational mechanisms related with protein function.

## Results

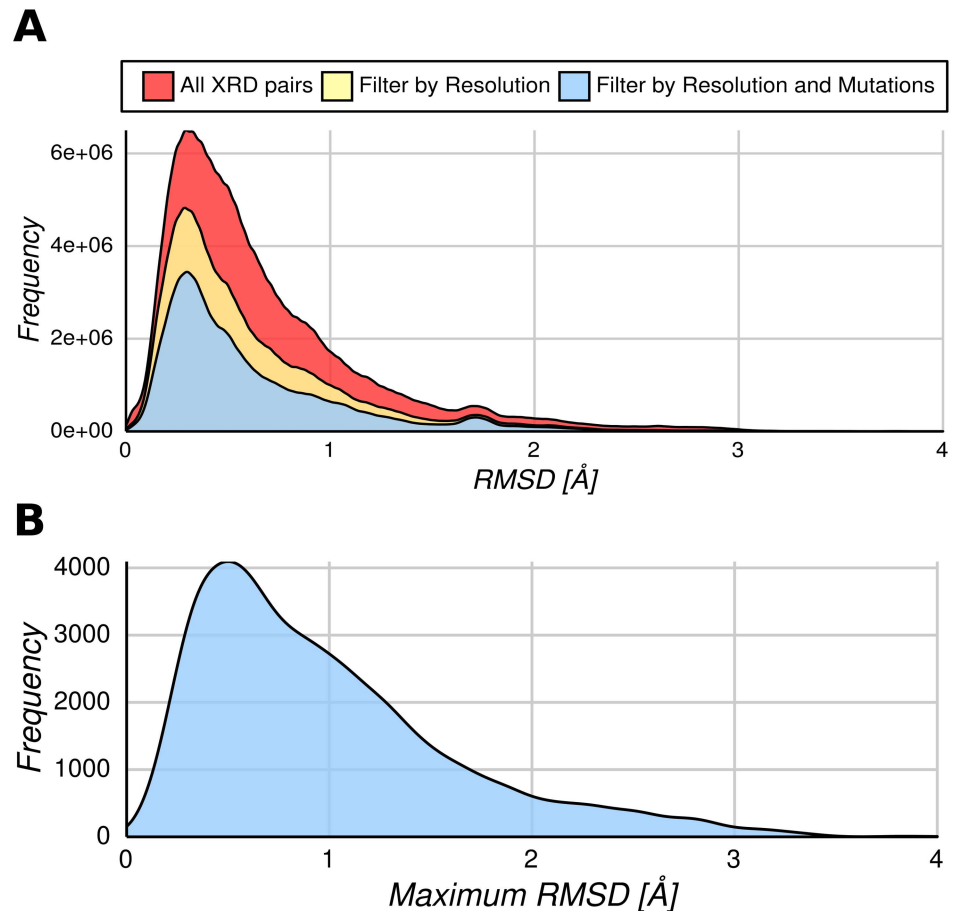
### General distribution and modulating factors

CoDNaS [18] is a database which contains a redundant collection of three-dimensional structures for the same sequence where each structure could be taken as snapshot of the conformational ensemble of the protein [11]. Essentially, for each protein in CoDNaS, we performed an extensive pairwise structural comparison between all conformers and derived different structural similarity scores linked with the corresponding information of the structure determination. In this way, it is easy to study structural differences between conformers and relate them to a given chemical and/or biological property characterizing conformers.

In Fig 1A, we show the general distribution of RMSD between conformers for all protein chains contained in CoDNaS obtained by X-Ray crystallography (~16,000 different protein chains). However, the presence of mutations, low resolution structures and the number of structures could affect the assessment of the degree of conformational diversity [19]. Additionally, as less than 6% of CoDNaS entries have NMR estimated structures and due to their differential flexibility behaviour compared with X-ray crystallographic structures [20], we removed conformers for each protein determined by this method in order to avoid bias produced by mixing NMR and X-Ray structures. In this sense, we used a reduced dataset (4,791 different protein chains with 74,417 conformers and 1,186,312 pairs of conformers) in order to avoid conformational diversity biases (see Methods). All distributions in Fig 1A show the previously identified trend [10]. In our analysis, it is also possible to study the maximum conformational diversity shown by a given chain or protein (the pair of conformers showing the maximum RMSD between all pair-wise comparisons for a given protein). This distribution (Fig 1B) again follows the general trend, with a median of 0.83 Å, an average of 0.99 Å and a large peak near 0.4 Å. It is possible to infer from this figure that most proteins require small movements between conformers to accomplish their biological functions.

### Mining the distribution

We have recently found that proteins showing order-disorder transitions between conformers show higher RMSD values than proteins showing no transitions [21]. In that study, the pair of *apo* and *holo* forms with the largest transition was selected and the RMSD was estimated using their common folded regions. This finding induced us to explore the role of disordered regions in the conformational diversity of folded regions in proteins. Using our working definition of disordered regions (see Methods) mapping IDRs on all conformers of a given protein, allowed us to split the maximum RMSD distribution (Fig 1B) into two new distributions shown in

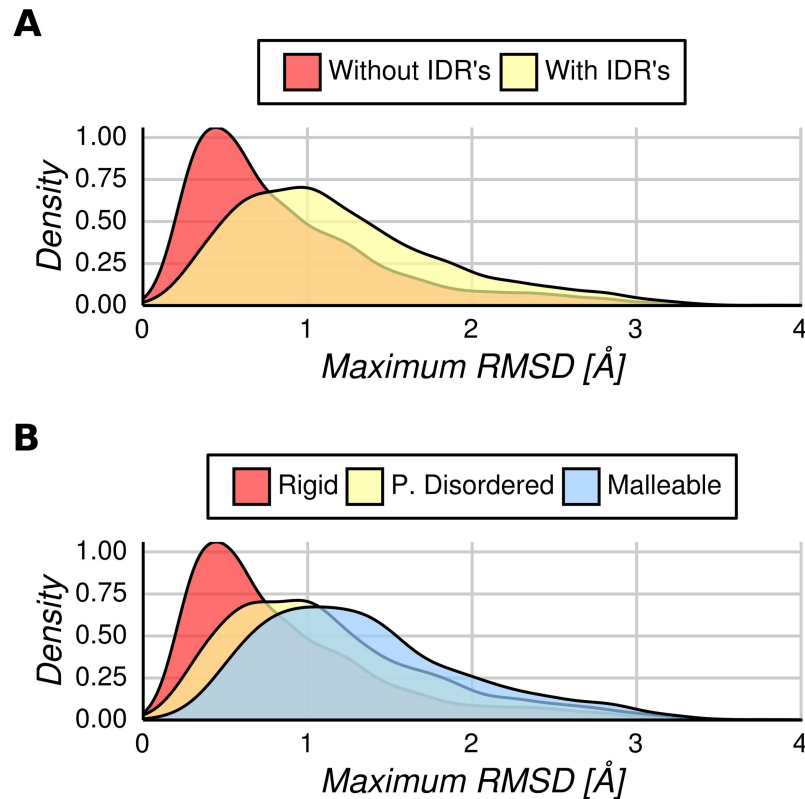


**Fig 1. Global conformational diversity distributions of proteins in CoDNaS.** (A) All pairwise RMSDs values between proteins with conformers obtained by X-ray diffraction (XRD) (light red distribution). The light yellow distribution considered only pairs of conformers which were obtained at resolution less than 2.5 Å and the light red distribution just conformers without mutations. (B) Maximum conformational diversity distribution of pairs of conformers with the maximum RMSD value per protein used in the rest of the manuscript.

doi:10.1371/journal.pcbi.1005398.g001

**Fig 2A.** One of the maximum RMSD distributions (light red in Fig 2A) with an average RMSD of 0.87 Å contains ordered proteins with no IDRs (in any of their available conformers). While the other has an average RMSD of 1.20 Å (light yellow in Fig 2A), and contains proteins with IDRs (in at least one of their available conformers). These distributions are statistically different in shape and median (Kolmogorov-Smirnov test and Wilcoxon rank-sum test respectively,  $P \ll 0.001$ ).

As proteins with IDRs have higher maximum RMSD of alpha carbons between conformers, this value could be used to predict whether that protein is disordered (at least one conformer with IDRs) or ordered (Fig 2A). The Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) for maximum RMSD as a predictor of disordered proteins is 0.67. Also, logistic regression for predicting disordered proteins using maximum RMSD values, shows a statistically significant coefficient ( $P < 0.001$ ). For a cutoff of 0.9 Å maximum RMSD, the accuracy to predict presence of disorder is 64% and for  $RMSD > 1.2$  Å the accuracy is 66%. Above 0.9 Å of maximum RMSD we found 3.03 times more proteins with disordered regions than below it. These results show that proteins with IDRs have higher RMSD values in their structured part than fully ordered proteins.



**Fig 2. Maximum conformational diversity distributions.** (A) Pairs of conformers showing maximum RMSD for proteins with/without IDRs considering all the available conformers per protein. In light red, we show the distribution without IDRs and in light yellow proteins with IDRs in at least one of their conformers. The distribution of proteins without IDRs had significantly lower overall RMSD values compared with proteins with IDRs (Kolmogorov—Smirnov test,  $P < 0.001$ ). (B) The maximum conformational diversity distribution can be represented by three main sets of proteins: rigid (all conformers per protein without IDRs), partially disordered (with IDRs at least in one conformer and in the maximum RMSD pair of conformational diversity) and malleable (with IDRs in at least one conformer but the maximum pair of conformational diversity without IDRs). These three distributions are significantly different in their median values (Kruskal—Wallis rank sum test,  $P < 0.001$  with a Nemenyi post-hoc test).

doi:10.1371/journal.pcbi.1005398.g002

Furthermore, the presence of disordered regions could still be used to separate the distribution of proteins with IDRs. Pairs of conformers showing maximum RMSD could contain or not disordered regions. We found that proteins with IDRs in the maximum RMSD pair have an average RMSD of 1.14 Å, while those without IDRs in the maximum RMSD pair have an average RMSD of 1.35 Å (Kolmogorov-Smirnov test and Wilcoxon rank-sum test,  $P$ -value  $< 0.01$ ). In increasing order of higher conformational diversity, we call these three distributions “rigid” with average RMSD = 0.85 Å (light red in Fig 2B), “partially disordered” with average RMSD = 1.1 Å (light yellow in Fig 2B) and finally “malleable” with average RMSD = 1.3 Å (light blue in Fig 2B). These three distributions are significantly different in their median values (Kruskal—Wallis rank sum test,  $P < 0.001$  with a Nemenyi post-hoc test). We will show later that, based on their structural and dynamical features, each of these distributions could represent different global conformational mechanisms.

Importantly, maximum RMSD and the number of conformers available for each protein have a negligible Spearman’s correlation coefficient ( $\rho = 0.094$ ), nor with protein length [10]. Also, the percentage of disordered conformers characterizing the partially disordered

and malleable sets with the number of conformers per protein have a very weak Spearman's correlation coefficient ( $\rho = 0.038$ ).

## Structural characterization of the distributions

It has been suggested that temperature used in structure estimation as well as the presence of crystallographic contacts can introduce biases in protein conformation. We found that the maximum RMSD distributions of the rigid, partially disordered and malleable sets are robust to the temperature used in structure estimation (room or cryogenic temperatures) and to the mean number of crystallographic contacts (please see supplementary Figure A-C in [S1](#) and [S2](#) Figs).

[S1 Table](#) summarizes several structural characterizations and composition of the rigid, partially disordered and malleable protein sets. It contains the mean, median and standard deviation of the compared distributions. The rigid protein distribution has the lowest mean RMSD and no disordered regions in all their available conformers. This is the most populated of the three groups (64.18% of the analysed proteins). We found that the partially disordered and malleable sets differ in the number of conformers with IDRs. The partially disordered set have 69.1% of their conformers with at least one disordered region, while the malleable set has only 24.86% of conformers showing IDRs. This difference is also evident in the amino acid composition of these regions. Thus, we compared the amino acid composition of disordered regions in our sets with the DisProt [\[22\]](#) database (a database of proteins with experimentally determined disorder regions) relative to PDB Select 25 [\[23\]](#) (proteins amenable to crystallization studies based on ordered residues). We found that IDRs in partially disordered and malleable sets are rich in amino acids characterizing flexible regions and depleted in amino acids related with globularity or foldability. This observation can be derived from the distribution shown in [S3 Fig](#), where amino acids proportions are ordered using the Vihinen amino acid flexibility distribution [\[24\]](#). It is also possible to observe that the partially disordered and malleable sets show similar amino acid composition, but with certain differences. Malleable proteins show a higher proportion of His, Thr and Lys and are depleted in Ala when compared with the partially disordered set, which in turn show an oddly higher proportion of Pro compared with malleable proteins and DisProt.

The differences in IDRs compositions of these sets directly impact the corresponding RMSD distributions. As derived from [S1 Table](#), partially disordered proteins have more IDRs showing a higher average percent of disorder and also longer IDRs compared to the malleable set. Malleable proteins show no IDRs in the maximum RMSD pair, mainly because their IDRs are mostly ordered in their conformers. However, these regions are highly flexible, introducing higher RMSD values when become ordered. On the contrary, highly flexible regions in partially disordered proteins are mostly disordered and do not impact the RMSD estimation. A direct consequence of this observation is observed when RMSD is estimated using only loop regions. We found that again malleable proteins show higher values than partially disordered proteins (see Figure A in [S4 Fig](#)) while no statistically significant differences were found using other secondary structures (see Figure B-C in [S4 Fig](#)). Furthermore, when removing all regions that appear disordered in at least one conformer available for each protein and recalculating the RMSD, we found no difference between partially disordered and malleable populations (Wilcoxon rank-sum test  $P = 0.74$  and Kolmogorov-Smirnov test  $P = 0.83$ ). In fact, the average RMSD of the three sets are very similar to each other with an average re-estimated RMSD of 0.86 Å, 0.993 Å and 0.995 Å for the rigid, partially disordered and malleable sets respectively. This supports the vision that conformational diversity of the three main groups mainly differs



in the conformation of ordered regions involved in order/disorder transitions but with different tendencies to transit between these states.

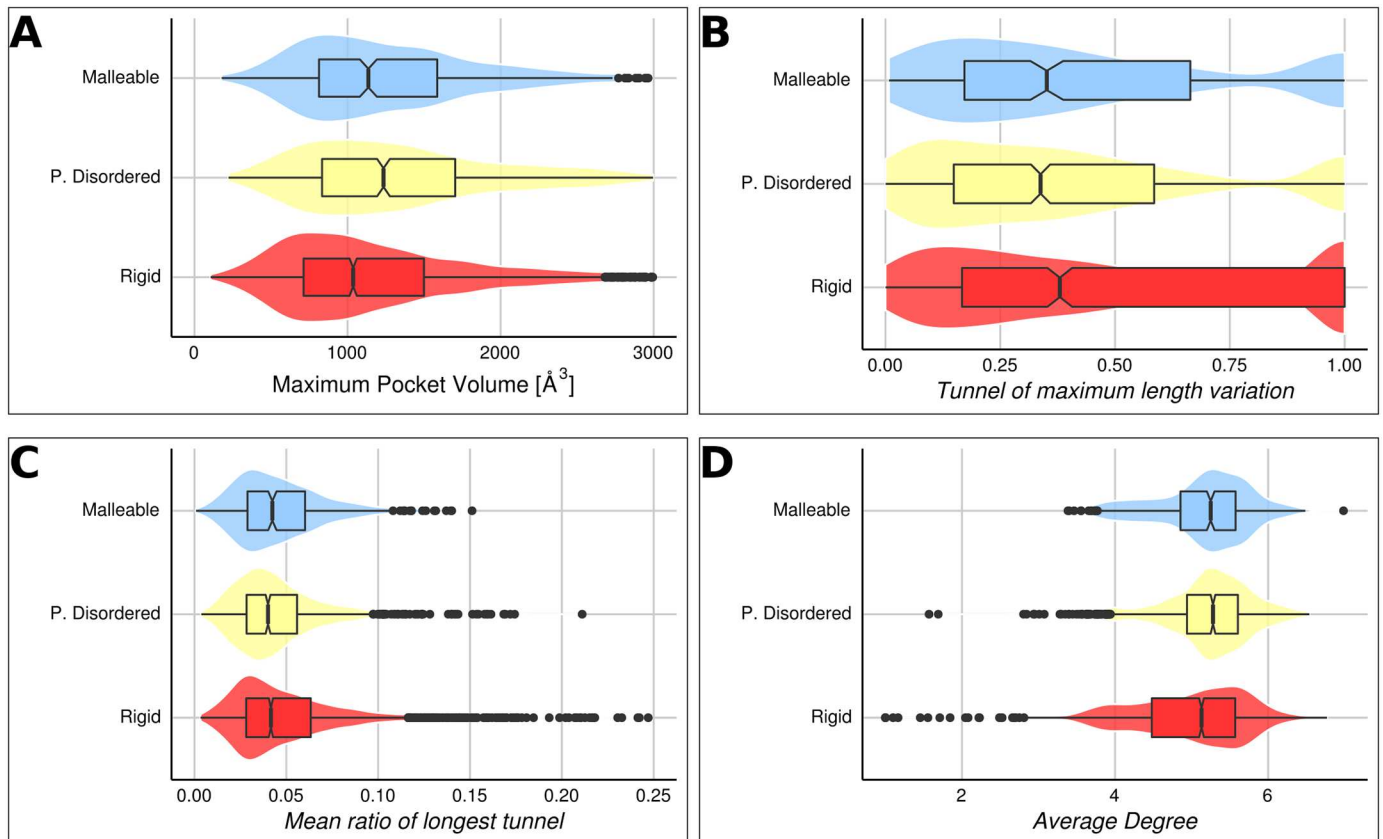
Other structural differences were also found between the sets. For example, partially disordered proteins contain more than twice as many hinges as malleable and rigid proteins (S5 Fig) and also higher normalized radius of gyration (S6 Fig). The presence of hinges may be related with an increase in conformational diversity due to relative domain movements, which could be supported by a slightly higher average of domains for partially disordered proteins (1.61 compared to 1.50 in malleable and 1.49 in rigid). On the other hand, the increased normalized radius of gyration in partially disordered proteins could evidence larger fold volumes and less compactness than in the other sets.

Conformers in our dataset could differ in several conditions associated to the corresponding crystallization conditions (for example post-translational modifications, different oligomeric structure, presence of ligands). Although it is difficult to establish an exact correlation between structural differences among conformers and protein function, we studied how the presence of the three sets are represented using just *apo* and *holo* conformers in the maximum pair of RMSD for each protein (see the composition of this subset in S2 Table). We found (S7 Fig) that using this subset of *apo* and *holo* conformer pairs we obtained the three main populations of rigid, partially disordered and malleable again significantly different in their median and they follow the same tendency as the original dataset (Kruskal—Wallis rank sum test,  $P < 0.001$  with a Nemenyi post-hoc test).

Finally, mapping protein domains with CATH showed that more than 84% of the folds for rigid proteins correspond to different superfamilies. The same behaviour is found in partially disordered and malleable proteins with 74% and 78%. The most populated cluster represents 2%, 6% and 3.4% of each set for rigid, partially disordered and malleable sets. Sequence clustering at 30% identity produce 57%, 72% and 80% of one-member clusters for the rigid, partially disordered and malleable sets respectively. These results are consistent with the idea that the majority of the proteins in each group are not homologous.

## Backbone-independent conformational diversity characterization

We have mentioned that large conformational changes are not required for proteins to sustain biological activity. Furthermore, low RMSD values as those found in rigid proteins (average RMSD = 0.85 Å) does not mean they lack conformational changes which may be involved in biological processes as previously indicated for allosteric proteins [25]. We study the presence of cavities and tunnels and their variation among conformers in order to study conformational diversity almost independent of carbon-alpha displacements. Cavities and tunnels are structures that connect the protein surface with buried active or binding sites in proteins and are essential for biological activity in most proteins [7,26]. Using Fpocket [27] to characterize the presence of cavities in each conformer (usually those with higher volume values), we observe how cavities are larger in partially disordered and malleable proteins (Fig 3A). Figures were represented as boxplots and violin superposed plots to facilitate evaluation and interpretation of the compared distributions (this information is also included in S1 Table). This estimation was performed using ordered pairs of conformers of the corresponding sets. While partially disordered and malleable proteins have similar volume cavities, they have important differences in their behaviour. Highly flexible regions commonly found in their ordered form in malleable proteins could define larger cavities and higher volume variation than in other sets. When IDRs in any of their conformers are removed, cavities in malleable proteins greatly reduce their volume while partially disordered cavities remain almost unaltered (see S1 Table). IDRs and highly flexible regions form and define, at least partially, the cavities of malleable



**Fig 3. Backbone-independent conformational diversity characterization.** On each boxplot, bottom and top of the box correspond to the first and third quartile, the vertical bar inside the box is the median (Second quartile) and the notches displays the median absolute deviation (M.A.D). Also, the violin plot under the boxplot shows the probability density of given variable. (A) Maximum pocket volume distribution between conformers in maximum pair of conformational diversity. Cavities are significantly greater in partially disordered and malleable proteins (Kruskal—Wallis rank sum test shows significant differences between this groups  $P < 0.001$ , with a Nemenyi post-hoc test shows that the volume of cavities in rigid proteins are significantly different of the cavity volumes in partially disordered or malleable proteins with  $P < 0.001$ ). (B) Maximum tunnel length variation distribution between conformers in maximum pair of conformational diversity. The parameter expresses the proportion of variation between largest tunnels in each conformer and it is calculated as  $|L_1 - L_2| / \max(L_1, L_2)$ , where  $L_i$  is the length of the largest tunnel on the corresponding conformer in the maximum pair. The mean value in rigid proteins is significantly greater than partially disordered proteins (Wilcoxon rank sum test,  $P < 0.001$ ). (C) Mean ratio of longest tunnel distribution between conformers in maximum pair of conformational diversity. The tunnel length is normalized by the conformer length. (D) Average degree distributions between conformers in maximum pair of conformational diversity. Partially disordered proteins show greater mean values than rigid proteins (Kruskal—Wallis rank sum test shows significant differences between this groups  $P < 0.001$  with a Nemenyi post-hoc test).

doi:10.1371/journal.pcbi.1005398.g003

proteins possibly modulating interaction with ligands which are more diverse in these proteins. Cavities in partially disordered and malleable sets are also more solvent exposed and more hydrophilic than those observed in rigid proteins in terms of pocket density and ASA values.

The size and variation of tunnels between conformers were also studied using MOLE 2.0 [28]. Considering the tunnel of maximum length and its variation among conformers, larger values are obtained in rigid proteins (Fig 3B). Opening and closing of main tunnels between conformers are unequivocally related with biological function [7]. Interestingly, we found that RMSD per position for tunnel lining residues is larger when compared with residues outside tunnels. This difference is only statistically significant when analysed rigid proteins (Kolmogorov-Smirnov test and Wilcoxon rank sum tests,  $P < 0.01$ ). Furthermore, rigid proteins have statistically more important tunnels than partially disordered and malleable proteins, as measured as the ratio of the longest tunnel (normalized for protein length) in the maximum



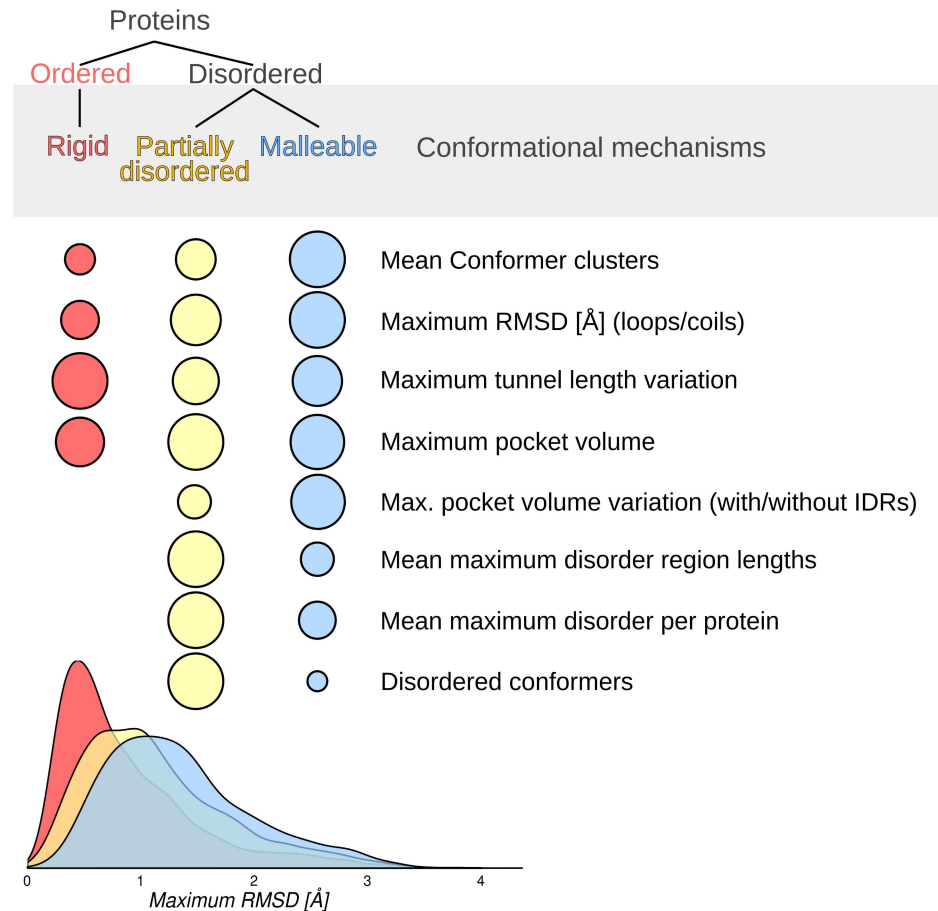
RMSD conformer pair (Kruskal—Wallis rank sum test  $P = 0.01$  with Nemenyi post-hoc test) (Fig 3C). Interestingly, these measures are not statistically different when partially disordered and malleable proteins are compared to each other. Additionally, we found that ligand molecular weights are bigger in malleable and partially disordered sets when compared with rigid proteins (average molecular weights: 414.12, 429.27 and 388.92 respectively).

Slight conformational changes were also studied as the average degree (contacts between residues) of the residue interaction network using RING [29]. We found (Fig 3D) that rigid proteins have lower average contacts per residue than partially disordered and malleable proteins (average degree: 5, 5.21 and 5.16 respectively, Kruskal-Wallis rank sum with Nemenyi post-hoc test  $P \ll 0.001$ ). Presence of more connected networks in the proteins with order-disorder transitions may be associated with an adaptive evolutionary process to compensate the presence of highly flexible regions in the ordered part of the protein.

## Discussion

We have found that the distribution of protein conformational diversity can be explained in terms of three main protein sets. The differences between them do not rely on structural classification, biological activity nor common evolutionary history. The average behaviour of the proteins in each set results from shared structural and dynamical features when their conformational populations are considered. The features shared between proteins in the same set could be seen as conformational mechanisms associated with protein function (Fig 4).

All ordered proteins have in general low conformational diversity and represents the most populated set (~60% of our dataset). This group is characterized by slight differences between conformers (mean maximum RMSD = 0.83 Å) evidencing how no large movements are required to accomplish the biological functions. We named the proteins in this set rigid. However, these proteins have larger tunnels when compared with the other sets and, importantly, the largest tunnel length variation among conformers. Apparently, backbone-independent conformational mechanisms such as the opening and closing of tunnels are required to sustain function. Tunnels are structures that allow the transit of ligands from the surface to the buried binding or active site of a protein. They can also connect buried cavities with each other and promote the exit of products [7]. As described above, residues lining these tunnels have larger RMSD than the rest of the positions in the protein. This indicates that the minimum movements for rigid proteins, are associated with the movement of functional structures such as tunnels. Interestingly, rigid proteins have on average ligands with lower molecular weight than the other two sets which also display larger cavities when compared with rigid proteins. These observations could indicate how the use of tunnels to interact with ligands could have a limitation handling larger ligands. The existence of open and closed tunnels or their size variation when conformers are compared, e.g. bottleneck dynamics [30] or conformational gating [31], may define different binding constants that could explain biological functions [32], enzyme specificity [26] and important regulatory processes such as allostery [15]. Over the last years, allostery without conformational change has been explained in terms of entropy effects arising from changes in the frequency and amplitude of thermal fluctuations around the mean conformation [33–35], as an alternative explanation to the classical consideration of changes in mean conformation [36,37]. The contribution of slight movements in opening and closing tunnels, pockets and cavities to this so-called entropic allostery [38] remains to be further quantified. Several proteins where tunnels and cavities define protein function are human P450 cytochrome [39], *A. aeolicus* lumazine synthase [26] with more than five channels longer than 15 Å, imidazole glycerol phosphate synthase from *T. maritima* [40] where gating regulates the activity of two active sites, cellulose cel48F with a 65 Å long tunnel (see Fig 5A and 5B)



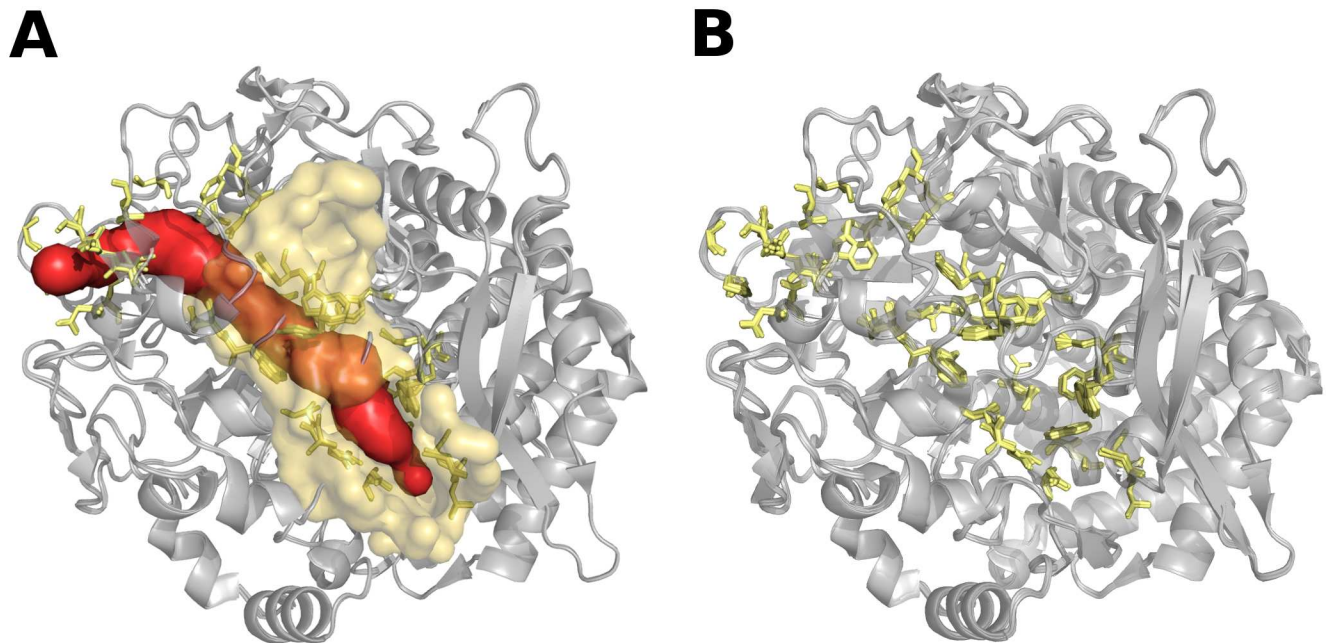
**Fig 4. Visual comparison of main structural features in each of the three protein sets described in this work.** Area of each circle is proportional to the average corresponding quantitative measure.

doi:10.1371/journal.pcbi.1005398.g004

[41,42], as well as very well studied proteins such as myoglobin and haemoglobin are all included in the rigid set derived from our study.

Disordered proteins are divided into two sets: partially disordered and malleable. Both sets have an increased conformational diversity compared to the rigid set described above. We have previously described how presence of IDRs and order-disorder transitions increases conformational diversity [21]. As protein function relates with motion, presence of IDRs could influence biological functions relying on the presence of ordered regions [43]. These two sets differ from each other in the propensity of their IDRs to appear as ordered or disordered as reflected in three main aspects. Composition of IDRs between sets is different, and possibly related with the different IDRs conformer content (see S1 Table). As previously indicated, the main reason for malleable proteins to show greater conformational diversity than partially disordered ones, is the fact that highly flexible regions (mainly loops) appear ordered and in different arrangements in malleable proteins, introducing higher RMSD values when conformers are compared.

Partially disordered and malleable proteins have larger cavities and shorter tunnels compared with rigid proteins. This could represent different ways to interact with ligands, as cognate ligands are larger in partially disordered and malleable proteins (average number of different ligands is 4.67, 5.64 and 9.30 for rigid, partially disordered and malleable sets).

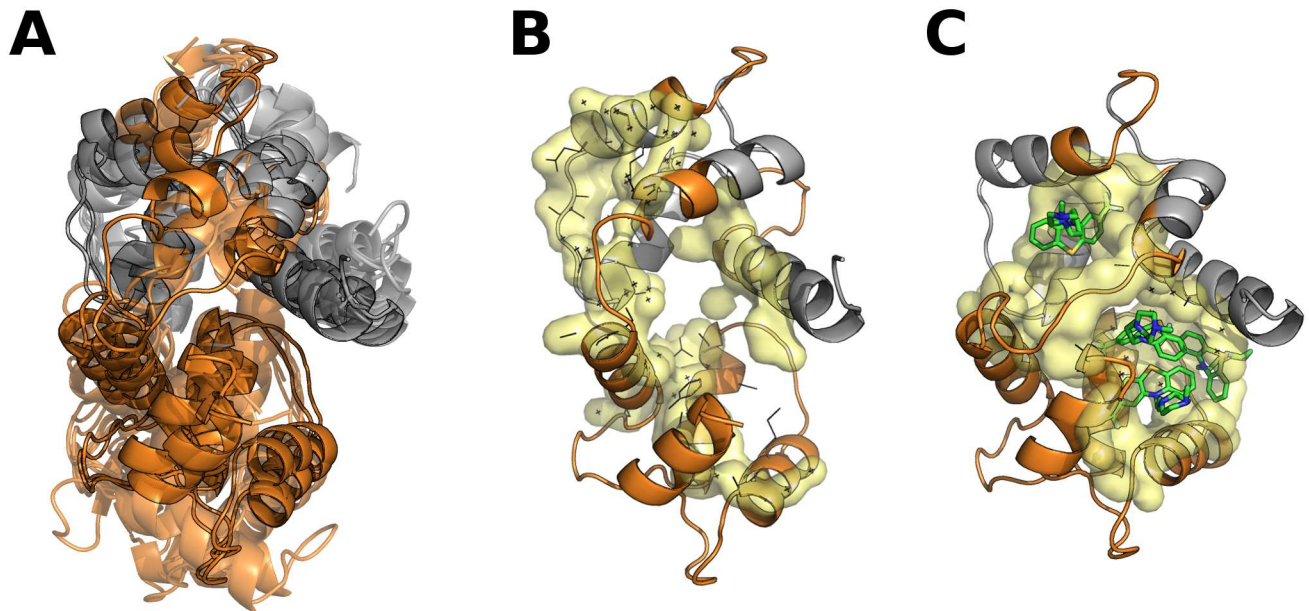


**Fig 5. Structural superposition of the cellulose cel48F conformations from *Clostridium cellulolyticum*.** This enzyme is a processive endo-cellulase with a large active and binding site to locate a cellulose chain which enters to the protein through a cleft located in the surface of the protein [41]. In our dataset, this protein contains 8 conformers (PDB codes: 1F9D\_A, 1F9O\_A, 1G9G\_A, 1FAE\_A, 1FBO\_A, 1FBW\_A, 1FCE\_A, 2QNO\_A) with a maximum RMSD of 0.21 Å. (A) We can see that there is almost no significant structural difference in the carbon-alpha trace between conformers in the pair of maximum RMSD (PDB codes: 1F9O\_A, 1G9G\_A), however, the tunnels (in red) as large as 65 Å long (predicted by MOLE) appeared in one conformer (PDB code: 1F9O\_A) which contains different ligands while in another conformer this tunnel is absent. (B) The superposition of all conformers only shows slight rotations and minimal movements in lining residues (in yellow) of this main tunnel, producing the opening and closing of the tunnel. Besides our conformers comparison, molecular dynamic simulations have also confirmed the rigidity of this protein [42].

doi:10.1371/journal.pcbi.1005398.g005

However, these two protein sets have additional differences in their corresponding cavities which are also more solvent exposed and hydrophilic than the rigid set. When IDRs in any conformer are removed, cavities in malleable proteins reduce their volume to the level of rigid proteins, while partially disordered cavities remain almost unaltered. In this sense, IDRs form and define, at least partially, cavities of malleable proteins, possibly modulating the ligand interaction. This higher flexibility defining cavities, appearing disordered in a minority of conformers, is the basic structural property characterizing the set that we called malleable. A typical example is calmodulin, a hub protein that can interact with 350 partners, which does not have a high disorder level and is commonly considered an ordered protein [44,45] (Fig 6A, 6B and 6C). Other well characterized proteins with flexible binding sites include the *E. coli* undecapryl-pyrophosphate synthase where flexible loops regulate the product size [46], Cyclin-dependent kinase 2 [47], Prothrombin [48] and Trypsin.

Contrary to malleable proteins, cavities are not affected by IDRs in the partially disordered set. This set shows the largest number of hinges and domains, hence we assume that IDRs could form high flexible linkers and connectors between domains [49]. This set also has the longest IDRs (see S1 Table) and about 5% of the proteins have a maximum disorder length which could involve entire domains. Due to the higher disorder content, a higher proportion of disordered conformers comprising about 70% of the IDPs in our dataset, we think that partially disordered proteins are mainly formed by very well characterized canonical IDPs which mostly agree with the current zeitgeist in IDPs [8,44,50]. In this set, we found several examples of well characterized IDPs such as the 50S ribosomal proteins L15, L11, L10 and L32e, guanine



**Fig 6. Calmodulin, a Ca<sup>2+</sup> sensor protein.** Calmodulin is a hub protein that can interact with 350 partners and it is commonly considered an ordered protein [44]. In our dataset, CaM has 79 conformers (8 with intrinsically disordered regions) with a maximum RMSD = 3.2 Å and a maximum disordered region of 34.5%. Calmodulin contains four EF-hand motifs where a pair of them form a globular domain. Each of this globular domains contains a pocket flexible enough to accommodate different target proteins through where four essential Met located in the pocket play a crucial role [45]. Additionally, each globular domain could adapt their relative orientation through flexible connectors increasing calmodulin capacity to interact with hundreds of partners. (A) We show several conformers of calmodulin superimposed (PDB codes: 1LIN\_A, 1NIW\_E, 3G43\_A, 4DCK\_B, 2FOT\_A, 2X0G\_B, 2BE6\_A, 2O60\_A, 1CDL\_A, 3GP2\_A). The disordered regions which are ordered in some conformers are in orange. (B-C). Whilst, in the pair of conformers with maximum RMSD, one conformation (B, PDB code: 1NIW\_E) is more extent with a not well defined pocket respect to the conformation bound to trifluoperazine (C, PDB code: 1LIN\_A), which is more compact and has higher cavity volume. We can show on one hand the dependence of malleable protein cavities volume with disordered regions, and on the other that the presence of these regions increase the pocket plasticity.

doi:10.1371/journal.pcbi.1005398.g006

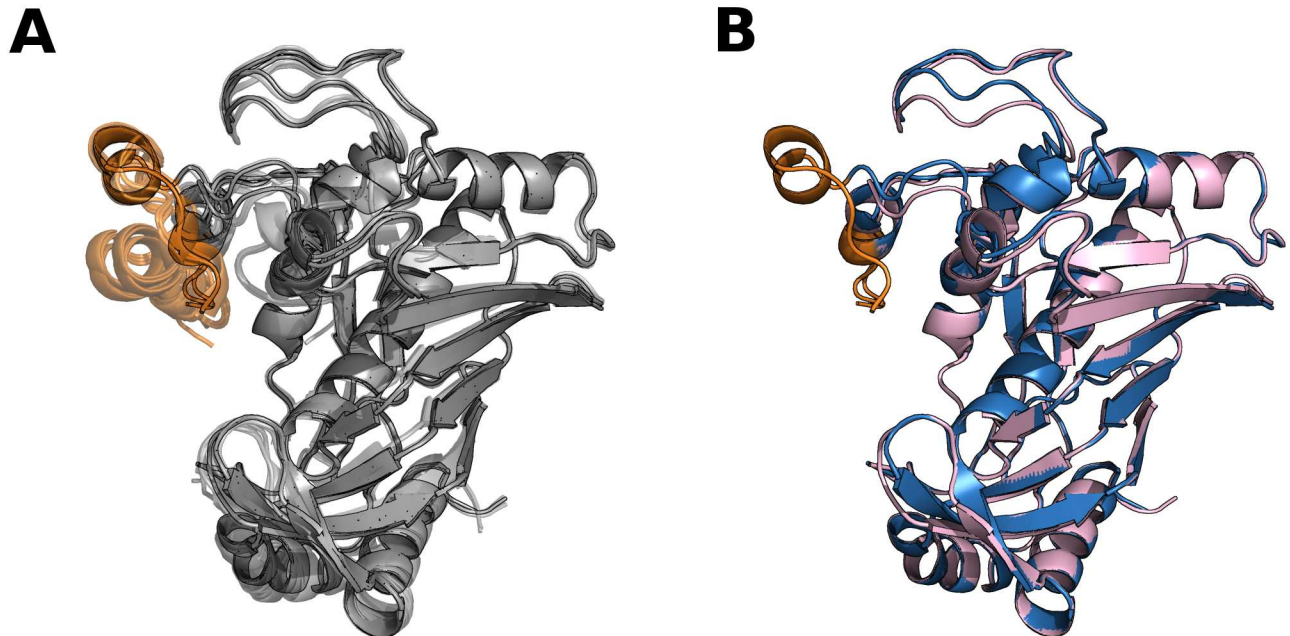
nucleotide-binding protein G and Thymidylate synthase [51,52] (see Fig 7A and 7B) are examples of order/disorder transitions related with function. Histo-blood group A transferase and Endonuclease VIII are examples of IDPs where function relates to the presence of a disorder state. An entropic chain in Serine/threonine-protein kinase PLK1 and Xanthine dehydrogenase/oxidase proteins serve just to mention some examples.

Historically, conformational diversity has been associated with large conformational changes thanks to the pioneering work by Chothia, Lesk and Gerstein [4,17]. Several examples were later compiled and extended by Gerstein in his protein motion database [5]. Where are these proteins in the sets described above? Most of the proteins show no IDRs, so we can find them in the higher RMSD tail of the rigid set. Just to measure the importance of such large motions, the 75 percentiles of rigid proteins is below 1.13 Å of RMSD. Above that number, we found 769 proteins without IDRs and with large conformational changes. Most of the changes comprise domains and/or fragments such as loops, normally as rigid bodies with hinge, shear or more complex motions [53]. The main point to stress is that this kind of motions represent about 15% of the total proteins studied here.

### Concluding remarks

The sustained advances in structural biology of the last years have been focused on the establishment of structure-function relationships to improve functional annotation of proteins displaying similar structures and/or sequences (for example [54]). This practical and useful trend





**Fig 7. Human thymidylate synthase (TS).** TS is responsible to catalyze the reductive methylation of dUMP to dTMP. This reaction is essential to maintain the nucleotide pools during cell growth. This enzyme is a valuable target for cytotoxic drugs used in cancer chemotherapy, because of its importance for DNA replication in actively dividing cells [51,52]. TS has disorder regions which plays an important role for dimer stabilization, assembly and regulation by order to disorder transitions. The TS has 15 conformers (11 with IDR's) in our data set with a maximum RMSD = 1.34 Å and a maximum percentage of disorder (according to our definition) of 9%. The disordered region of this protein is composed by residues from one of the inserts that presents human TS differs to bacterial TS. (A) We show all conformers of the TS superimposed (PDB codes: 1HW3\_A, 1HZW\_A, 1HZW\_B, 1100\_A, 1100\_B, 1YPV\_A, 3EDW\_X, 3EHI\_X, 3GH2\_X, 3N5E\_A, 3N5E\_B, 3N5G\_A, 4G2O\_X, 4G6W\_X, 4GD7\_A). The disorder region changes from disorder to order in many conformers (residues in orange), for example, in the active conformation (PDB code = 1HZW\_A) this region is ordered while in the inactive (PDB code = 1YPV\_A) is disordered. (B) Pair of conformers with maximum RMSD (1YPV\_A in pink, 4GD7\_A in light blue). The disorder region appears partially ordered in the two conformers, the only difference is that the region in the structure 4GD7\_A has eleven more ordered residues (which forms an alpha-helix) than the same region in the other conformation.

doi:10.1371/journal.pcbi.1005398.g007

hides the concept that the structure-function relationship is almost an evolutionary accident in view of the vast redundancy of similar functions sustained by different protein structures [55,56] or the contrary where similar structures achieve completely different functions (i.e. superfolds [57,58]). We have found three main conformational mechanisms expressed as structure-dynamics relationships representing different ways to achieve a vast diversity of biological functions. They are general enough to describe proteins behaviour, as early suggested by H. Frauenfelder in his search for general concepts in protein dynamics [2]. But certainly, the search for unique functional classes using low-level descriptors inside these categories appears to be a useless task.

## Methods

### Dataset generation

The CoDNaS database [59], containing a redundant collection of three dimensional structures for the same protein, was used to recruit proteins exhibiting conformational diversity. The final dataset contains 4,791 different protein chains with a total of 74,417 conformers (corresponding to 1,186,312 conformer pairs), all structures belonging to this dataset were obtained by X-Ray Diffraction at a resolution equal or less than 2.5 Å. The maximum conformational diversity for each protein is the pair of conformers with maximum C-alpha RMSD. In order to

obtain a reliable and comparable estimation of conformational diversity of each protein, our dataset only contains proteins with a minimum of 5 conformers (average 15.53 conformers per protein). It is supported by the work of Best et. al which showed that is necessary to take into account a given number of structures (around 5 structures) to estimate conformational features of the native state of proteins [11] (up to 5 conformers to study backbone flexibility and around 20 to study side-chain heterogeneity). These conformers mainly differ in the presence of bound ligands (~70% on average of the conformers).

### Finding intrinsic disorder regions

We defined intrinsic disorder regions (IDRs) for each protein conformer using MobiDB [60]. If a residue has missing electron density coordinates in a structure obtained with X-ray crystallography, it assumed to be disordered [61]. To define an IDR, only five or more consecutive missing residues that were not in the amino or carboxyl terminal of the protein sequence (the first or last twenty residues) were considered. The amino acid composition was obtained from composition profile [62] using 10,000 bootstrap iterations.

### Structural characterizations

Hinges were detected with FlexProt [63], the number of hinges identified maximizes the structural superposition between conformers. The Radius of gyration ( $R_g$ ) of each conformer was calculated using the MMTSB Tool Set (available from <http://blue11.bch.msu.edu/mmtsbt/>). Due to the dependence of  $R_g$  with protein size, a normalized  $R_g$  was calculated from the  $R_g$  of the ideal sphere of the same volume of the protein structure [64].

Pockets were detected with FPocket [27], using the highest scored pocket. The percentage of maximum variation between maximum pocket volumes (with/without IDRs residues) in each conformer was calculated as  $|\max(pV_{1,IDR}, pV_{2,IDR}) - \max(pV_1, pV_2)| / \max(\max(pV_{1,IDR}, pV_{2,IDR}), \max(pV_1, pV_2))$ , where  $pV_i$  is the pocket volume (with/without IDRs residues) on the corresponding conformer. MOLE 2.0 [65] was used for tunnel identification and characterization using the following parameters. Probe Radius and Origin Radius 3 Å, Interior Threshold 1.25 Å and default values for all the others. We parsed the XML file output to identify tunnel length, number of tunnels, residues lining each tunnel and so on. The proportion of variation between largest tunnels in each conformer was calculated as  $|L_1 - L_2| / \max(L_1, L_2)$ , where  $L_i$  is the length of the largest tunnel on the corresponding conformer in the maximum pair. The protein volume was estimated by 3vee [66] using a probe radius of 0.5 Å and a grid resolution of 1.5 Å. Residue interactions networks (RINs) were built with RING [29].

### Supporting information

**S1 Table. Summary of different descriptors reported on the 'Results' section.** The values correspond to the mean (with the confidence interval at 95%), standard deviation and median in each compared distribution. See [Methods](#) section for details about descriptors. (XLSX)

**S2 Table. Composition of apo/holo subset.** The values correspond to the mean (confidence interval at 95%), standard deviation and median in each compared distribution. (XLSX)

**S1 Fig. Maximum RMSD distributions at different crystallization temperatures.** (A) Representative maximum pair of conformers for each protein obtained at room-temperature (above 200 K) and Cryogenic temperature (100 K). (B) The three sets in a subset of proteins in which the conformers of the maximum pair of RMSD has been crystallised at cryogenic temperature



(100 K). This subset represents the 67% of the total proteins in our dataset. (C) Idem (A) which pairs obtained at room-temperature. This subset represents the 9% of the total proteins in our dataset.

(TIF)

**S2 Fig. RMSD vs. mean number of crystal contacts, calculated for pairs of conformers with maximum conformational diversity for each monomeric protein in our dataset.** We have used a subset of monomeric proteins from our dataset, with only one protein chain in the asymmetric unit (392 pairs of conformers) in order to remove hetero biological complexes. For each conformer in the maximum pair of RMSD, we estimated the average number of crystallographic contacts (at 4.5 Å of distance) between each atom of the protein chain in the asymmetric unit and the neighbour molecules in other unit cells. We obtained a negligible Spearman's correlation coefficient of 0.048.

(TIF)

**S3 Fig. Amino acid composition of IDRs.** Amino acid composition of IDRs presents in all conformers of malleable (light blue) and partially disordered (yellow) proteins relative to PDB Select 25. DisProt is used as a reference of experimental protein disorder.

(TIF)

**S4 Fig. (A–C) Distributions of RMSD values for maximum pairs of conformational diversity in each set discriminated by secondary structure.**

(TIF)

**S5 Fig. Number of hinges distribution in maximum pairs of conformational diversity in each set.** A single bar represents the relative frequency of a given number of hinges.

(TIF)

**S6 Fig. Mean normalized Radius of gyration distributions between conformers in maximum pair of conformational diversity.** Rigid proteins show an average significantly lower than partially disordered proteins (Wilcoxon rank-sum test  $P \ll 0.001$ ).

(TIF)

**S7 Fig. Maximum conformational diversity distribution of the *apo/olo* subset.**

(TIF)

**S1 Dataset. Datasets used in this study.**

(TAR)

## Author Contributions

**Conceptualization:** AMM GP.

**Data curation:** AMM.

**Formal analysis:** AMM DJZ GP.

**Funding acquisition:** SCET GP.

**Investigation:** SFA SCET GP.

**Methodology:** AMM GP.

**Software:** AMM.

**Visualization:** AMM.

**Writing – original draft:** AMM GP.

**Writing – review & editing:** AMM MSF TES SFA SCET GP.

## References

1. Perutz MF, Mathews FS. An X-ray study of azide methaemoglobin. *J Mol Biol.* 1966; 21: 199–202. PMID: [5969763](#)
2. Frauenfelder H, McMahon B. Dynamics and function of proteins: the search for general concepts. *Proc Natl Acad Sci U S A.* 1998/06/06. 1998; 95: 4795–4797. PMID: [9560180](#)
3. Case DA, Karplus M. Dynamics of ligand binding to heme proteins. *J Mol Biol.* 1979; 132: 343–368. PMID: [533895](#)
4. Gerstein M, Lesk AM, Chothia C. Structural mechanisms for domain movements in proteins. *Biochemistry.* 1994; 33: 6739–6749. PMID: [8204609](#)
5. Gerstein M, Krebs W. A database of macromolecular motions. *Nucleic Acids Res.* 1998/09/02. 1998; 26: 4280–4290. PMID: [9722650](#)
6. Gu Y, Li D-W, Brüschweiler R. Decoding the Mobility and Time Scales of Protein Loops. *J Chem Theory Comput.* 2015; 11: 1308–1314. doi: [10.1021/ct501085y](#) PMID: [26579776](#)
7. Gora A, Brezovsky J, Damborsky J. *Gates of Enzymes.* Chem Rev. 2013;
8. van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. *Chem Rev.* 2014; 114: 6589–631. doi: [10.1021/cr400525m](#) PMID: [24773235](#)
9. Janin J, Sternberg MJE. Protein flexibility, not disorder, is intrinsic to molecular recognition. *F1000 Biol Rep.* 2013; 5: 2. doi: [10.3410/B5-2](#) PMID: [23361309](#)
10. Burra P V, Zhang Y, Godzik A, Stec B. Global distribution of conformational states derived from redundant models in the PDB points to non-uniqueness of the protein structure. *Proc Natl Acad Sci U S A.* 2009; 106: 10505–10. doi: [10.1073/pnas.0812152106](#) PMID: [19553204](#)
11. Best RB, Lindorff-Larsen K, DePristo MA, Vendruscolo M. Relation between native ensembles and experimental structures of proteins. *Proc Natl Acad Sci U S A.* 2006; 103: 10901–6. doi: [10.1073/pnas.0511156103](#) PMID: [16829580](#)
12. Gutteridge A, Thornton J. Conformational changes observed in enzyme crystal structures upon substrate binding. *J Mol Biol.* 2005/01/25. 2005; 346: 21–28. doi: [10.1016/j.jmb.2004.11.013](#) PMID: [15663924](#)
13. Mesecar AD, Stoddard BL, K DE Jr. Orbital Steering in the Catalytic Power of Enzymes : Small Structural Changes with Large Catalytic Consequences. *Science (80-).* 1997; 277.
14. Koshland DE. Conformational changes: How small is big enough? *Nat Med.* 1998; 4: 1112–1114. doi: [10.1038/2605](#) PMID: [9771734](#)
15. Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? *Proteins.* 2004; 57: 433–43. doi: [10.1002/prot.20232](#) PMID: [15382234](#)
16. Tsai C-J, del Sol A, Nussinov R. Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play. *J Mol Biol.* 2008; 378: 1–11. doi: [10.1016/j.jmb.2008.02.034](#) PMID: [18353365](#)
17. Lesk AM, Chothia C. Mechanisms of domain closure in proteins. *J Mol Biol.* 1984/03/25. 1984; 174: 175–191. PMID: [6371249](#)
18. Monzon AM, Rohr CO, Fornasari MS, Parisi G. CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state. *Database (Oxford).* 2016; 2016: baw038-.
19. Parisi G, Zea DJ, Monzon AM, Marino-Buslje C. Conformational diversity and the emergence of sequence signatures during evolution. *Curr Opin Struct Biol.* Elsevier Ltd; 2015; 32: 58–65.
20. Sikic K, Tomic S, Carugo O. Systematic comparison of crystal and NMR protein structures deposited in the protein data bank. *Open Biochem J.* 2010; 4: 83–95. doi: [10.2174/1874091X01004010083](#) PMID: [21293729](#)
21. Zea D, Monzon AM, Gonzalez C, Fornasari MS, Tosatto SCE, Parisi G. Disorder transitions and conformational diversity cooperatively modulate biological function in proteins. *Protein Sci.* 2016;
22. Piovesan D, Tabaro F, Mičetić I, Necci M, Quaglia F, Oldfield CJ, et al. DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.* 2016; 45: gkw1279.
23. Hobohm U, Sander C. Enlarged representative set of protein structures. *Protein Sci.* 1994; 3: 522–4. doi: [10.1002/pro.5560030317](#) PMID: [8019422](#)

24. Vihinen M, Torkkila E, Riikonen P. Accuracy of protein flexibility predictions. *Proteins*. 1994; 19: 141–9. doi: [10.1002/prot.340190207](https://doi.org/10.1002/prot.340190207) PMID: [8090708](https://pubmed.ncbi.nlm.nih.gov/8090708/)
25. Nussinov R, Tsai C-J. Allostery without a conformational change? Revisiting the paradigm. *Curr Opin Struct Biol*. Elsevier Ltd; 2015; 30: 17–24.
26. Pravda L, Berka K, Svobodová Vařeková R, Sehnal D, Banáš P, Laskowski R a, et al. Anatomy of enzyme channels. *BMC Bioinformatics*. 2014; 15: 1–8.
27. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics*. 2009; 10: 168. doi: [10.1186/1471-2105-10-168](https://doi.org/10.1186/1471-2105-10-168) PMID: [19486540](https://pubmed.ncbi.nlm.nih.gov/19486540/)
28. Sehnal D, Svobodová Vařeková R, Berka K, Pravda L, Navrátilová V, Baná P, et al. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J Cheminform*. 2013; 5: 39. doi: [10.1186/1758-2946-5-39](https://doi.org/10.1186/1758-2946-5-39) PMID: [23953065](https://pubmed.ncbi.nlm.nih.gov/23953065/)
29. Martin AJ, Vidotto M, Boscaroli F, Di Domenico T, Walsh I, Tosatto SC. RING: networking interacting residues, evolutionary information and energetics in protein structures. *Bioinformatics*. 2011/04/16. 2011; 27: 2003–2005. doi: [10.1093/bioinformatics/btr191](https://doi.org/10.1093/bioinformatics/btr191) PMID: [21493660](https://pubmed.ncbi.nlm.nih.gov/21493660/)
30. Chovancova E, Pavelka A, Benes P, Strnad O, Brezovsky J, Kozlikova B, et al. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput Biol*. 2012; 8: e1002708. doi: [10.1371/journal.pcbi.1002708](https://doi.org/10.1371/journal.pcbi.1002708) PMID: [23093919](https://pubmed.ncbi.nlm.nih.gov/23093919/)
31. Zhou HX, Wlodek ST, McCammon J a. Conformation gating as a mechanism for enzyme specificity. *Proc Natl Acad Sci USA*. 1998; 95: 9280–3. PMID: [9689071](https://pubmed.ncbi.nlm.nih.gov/9689071/)
32. Biedermannová L, Prokop Z, Gora A, Chovancová E, Kovács M, Damborsky J, et al. A single mutation in a tunnel to the active site changes the mechanism and kinetics of product release in haloalkane dehalogenase LinB. *J Biol Chem*. 2012; 287: 29062–74. doi: [10.1074/jbc.M112.377853](https://doi.org/10.1074/jbc.M112.377853) PMID: [22745119](https://pubmed.ncbi.nlm.nih.gov/22745119/)
33. Tsai C-J, del Sol A, Nussinov R. Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play. *J Mol Biol*. 2008; 378: 1–11. doi: [10.1016/j.jmb.2008.02.034](https://doi.org/10.1016/j.jmb.2008.02.034)
34. del Sol A, Tsai C-J, Ma B, Nussinov R. The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure*. Elsevier Ltd; 2009; 17: 1042–50.
35. Cui Q, Karplus M. Allostery and cooperativity revisited. *Protein Sci*. 2008; 17: 1295–1307. doi: [10.1110/ps.03259908](https://doi.org/10.1110/ps.03259908) PMID: [18560010](https://pubmed.ncbi.nlm.nih.gov/18560010/)
36. Monod J, Wyman J, Changeux JP. On the Nature of Allosteric Transitions: a Plausible Model. *J Mol Biol*. 1965; 12: 88–118. Available: <http://www.ncbi.nlm.nih.gov/pubmed/14343300> PMID: [14343300](https://pubmed.ncbi.nlm.nih.gov/14343300/)
37. Koshland DE, Némethy G, Filmer D. Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry*. 1966; 5: 365–85. Available: <http://www.ncbi.nlm.nih.gov/pubmed/5938952> PMID: [5938952](https://pubmed.ncbi.nlm.nih.gov/5938952/)
38. Cooper A., Dryden DTF. Allostery without conformational change—A plausible model. *Eur Biophys J*. 1984; 11: 103–109. PMID: [6544679](https://pubmed.ncbi.nlm.nih.gov/6544679/)
39. Skopalík J, Anzenbacher P, Otyepka M. Flexibility of human cytochromes P450: molecular dynamics reveals differences between CYPs 3A4, 2C9, and 2A6, which correlate with their substrate preferences. *J Phys Chem B*. 2008; 112: 8165–73. doi: [10.1021/jp800311c](https://doi.org/10.1021/jp800311c) PMID: [18598011](https://pubmed.ncbi.nlm.nih.gov/18598011/)
40. Myers RS, Amaro RE, Luthey-Schulten ZA, Davisson VJ. Reaction coupling through interdomain contacts in imidazole glycerol phosphate synthase. *Biochemistry*. 2005; 44: 11974–11985. doi: [10.1021/bi050706b](https://doi.org/10.1021/bi050706b) PMID: [16142895](https://pubmed.ncbi.nlm.nih.gov/16142895/)
41. Parsiegla G, Reverbel-Leroy C, Tardif C, Belaich JP, Driguez H, Haser R. Crystal structures of the cellulase Cel48F in complex with inhibitors and substrates give insights into its processive action. *Biochemistry*. 2000; 39: 11238–11246. PMID: [10985769](https://pubmed.ncbi.nlm.nih.gov/10985769/)
42. Vital de Oliveira O. Molecular Dynamics and Metadynamics Simulations of the Cellulase Cel48F. *Enzyme Res*. Hindawi Publishing Corporation; 2014; 2014: 1–7.
43. Schulenburg C, Hilvert D. Protein Conformational disorder and enzyme catalysis. *Top Curr Chem*. 2013; 337: 41–68. doi: [10.1007/128\\_2012\\_411](https://doi.org/10.1007/128_2012_411) PMID: [23536241](https://pubmed.ncbi.nlm.nih.gov/23536241/)
44. Tompa P. Structure and Function of Intrinsically Disordered Proteins. Chapman & Hall/CRC Taylor & Francis Group; 2010.
45. Ikura M, Ames JB. Genetic polymorphism and protein conformational plasticity in the calmodulin superfamily: two ways to promote multifunctionality. *Proc Natl Acad Sci U S A*. 2006; 103: 1159–64. doi: [10.1073/pnas.0508640103](https://doi.org/10.1073/pnas.0508640103) PMID: [16432210](https://pubmed.ncbi.nlm.nih.gov/16432210/)
46. Ko TP, Chen YK, Robinson H, Tsai PC, Gao YG, Chen APC, et al. Mechanism of Product Chain Length Determination and the Role of a Flexible Loop in Escherichia coli Undecaprenyl-pyrophosphate Synthase Catalysis. *J Biol Chem*. 2001; 276: 47474–47482. doi: [10.1074/jbc.M106747200](https://doi.org/10.1074/jbc.M106747200) PMID: [11581264](https://pubmed.ncbi.nlm.nih.gov/11581264/)

47. Li Y, Zhang J, Gao W, Zhang L, Pan Y, Zhang S, et al. Insights on Structural Characteristics and Ligand Binding Mechanisms of CDK2. *Int J Mol Sci*. 2015; 16: 9314–40. doi: [10.3390/ijms16059314](https://doi.org/10.3390/ijms16059314) PMID: [25918937](https://pubmed.ncbi.nlm.nih.gov/25918937/)
48. Adams TE, Huntington JA. Structural transitions during prothrombin activation: On the importance of fragment 2. *Biochimie*. Elsevier B.V; 2015; 122: 235–242.
49. Mittag T, Kay LE, Forman-Kaya JD. Protein dynamics and conformational disorder in molecular recognition. *J Mol Recognit*. 2010; 23: 105–116. doi: [10.1002/jmr.961](https://doi.org/10.1002/jmr.961) PMID: [19585546](https://pubmed.ncbi.nlm.nih.gov/19585546/)
50. Tompa P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci*. 2012; 37: 509–16. doi: [10.1016/j.tibs.2012.08.004](https://doi.org/10.1016/j.tibs.2012.08.004) PMID: [22989858](https://pubmed.ncbi.nlm.nih.gov/22989858/)
51. Peña MMO, Xing YY, Koli S, Berger FG. Role of N-terminal residues in the ubiquitin-independent degradation of human thymidylate synthase. *Biochem J*. 2006; 394: 355–63. doi: [10.1042/BJ20051479](https://doi.org/10.1042/BJ20051479) PMID: [16259621](https://pubmed.ncbi.nlm.nih.gov/16259621/)
52. Lovelace LL, Minor W, Lebioda L. Structure of human thymidylate synthase under low-salt conditions. *Acta Crystallogr D Biol Crystallogr*. 2005; 61: 622–7. doi: [10.1107/S0907444905005895](https://doi.org/10.1107/S0907444905005895) PMID: [15858273](https://pubmed.ncbi.nlm.nih.gov/15858273/)
53. Gatzeva-topalova PZ, Warner LR, Pardi A, Carlos M. Diversity of function-related conformational changes in proteins: coordinate uncertainty, fragment rigidity and stability. *Biochemistry*. 2010; 49: 5683–5704. doi: [10.1021/bi100110x](https://doi.org/10.1021/bi100110x) PMID: [20469886](https://pubmed.ncbi.nlm.nih.gov/20469886/)
54. Loewenstein Y, Raimondo D, Redfern OC, Watson J, Frishman D, Linial M, et al. Protein function annotation by homology-based inference. *Genome Biol*. 2009; 10: 207. doi: [10.1186/gb-2009-10-2-207](https://doi.org/10.1186/gb-2009-10-2-207) PMID: [19226439](https://pubmed.ncbi.nlm.nih.gov/19226439/)
55. Bork P, Sander C, Valencia A. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci*. 1993; 2: 31–40. doi: [10.1002/pro.5560020104](https://doi.org/10.1002/pro.5560020104) PMID: [8382990](https://pubmed.ncbi.nlm.nih.gov/8382990/)
56. Elias M, Tawfik DS. Divergence and convergence in enzyme evolution: parallel evolution of paraoxonases from quorum-quenching lactonases. *J Biol Chem*. 2012; 287: 11–20. doi: [10.1074/jbc.R111.257329](https://doi.org/10.1074/jbc.R111.257329) PMID: [22069329](https://pubmed.ncbi.nlm.nih.gov/22069329/)
57. Orengo CA, Jones DT, Thornton JM. Protein superfamilies and domain superfolds. *Nature*. 1994/12/15. 1994; 372: 631–634. doi: [10.1038/372631a0](https://doi.org/10.1038/372631a0) PMID: [7990952](https://pubmed.ncbi.nlm.nih.gov/7990952/)
58. Liu X, Fan K, Wang W. The number of protein folds and their distribution over families in nature. *Proteins*. 2004; 54: 491–9. doi: [10.1002/prot.10514](https://doi.org/10.1002/prot.10514) PMID: [14747997](https://pubmed.ncbi.nlm.nih.gov/14747997/)
59. Monzon AM, Juritz E, Fornasari S, Parisi G. CoDNaS: a database of conformational diversity in the native state of proteins. *Bioinformatics*. 2013; 1–3.
60. Potenza E, Domenico TD, Walsh I, Tosatto SCE. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res*. 2014; 43: D315–D320. doi: [10.1093/nar/gku982](https://doi.org/10.1093/nar/gku982) PMID: [25361972](https://pubmed.ncbi.nlm.nih.gov/25361972/)
61. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. *J Mol Graph Model*. 2001; 19: 26–59. Available: <http://www.ncbi.nlm.nih.gov/pubmed/11381529> PMID: [11381529](https://pubmed.ncbi.nlm.nih.gov/11381529/)
62. Vacic V, Uversky VN, Dunker A K, Lonardi S. Composition Profiler: a tool for discovery and visualization of amino acid composition differences. *BMC Bioinformatics*. 2007; 8: 211. doi: [10.1186/1471-2105-8-211](https://doi.org/10.1186/1471-2105-8-211) PMID: [17578581](https://pubmed.ncbi.nlm.nih.gov/17578581/)
63. Shatsky M, Nussinov R, Wolfson HJ. Flexible protein alignment and hinge detection. *Proteins Struct Funct Genet*. 2002; 48: 242–256. doi: [10.1002/prot.10100](https://doi.org/10.1002/prot.10100) PMID: [12112693](https://pubmed.ncbi.nlm.nih.gov/12112693/)
64. Lobanov MI, Bogatyreva NS, Galzitskaia O V. [Radius of gyration is indicator of compactness of protein structure]. *Mol Biol (Mosk)*. 42: 701–6.
65. Berka K, Hanák O, Sehnal D, Banás P, Navrátilová V, Jaiswal D, et al. MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res*. 2012; 40: W222–7. doi: [10.1093/nar/gks363](https://doi.org/10.1093/nar/gks363) PMID: [22553366](https://pubmed.ncbi.nlm.nih.gov/22553366/)
66. Voss NR, Gerstein M. 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res*. 2010; 38: W555–62. doi: [10.1093/nar/gkq395](https://doi.org/10.1093/nar/gkq395) PMID: [20478824](https://pubmed.ncbi.nlm.nih.gov/20478824/)