

## **Bcl-x<sub>L</sub> dynamics and cancer-associated mutations under the lens of protein structure network and biomolecular simulations**

Valentina Sora<sup>1</sup>, Elena Papaleo<sup>1,2\*</sup>

<sup>1</sup> Computational Biology Laboratory, Danish Cancer Society Research Center, Strandboulevarden 49, 2100, Copenhagen, Denmark

<sup>2</sup> Translational Disease Systems Biology, Faculty of Health and Medical Sciences, Novo Nordisk Foundation Center for Protein Research University of Copenhagen, Copenhagen, Denmark

\*corresponding author: [elenap@cancer.dk](mailto:elenap@cancer.dk)

### **Abstract**

Understanding the finely orchestrated interactions leading to or preventing programmed cell death (apoptosis) is of utmost importance in cancer research since the failure of these systems could eventually lead to the onset of the disease. In this regard, the maintenance of a delicate balance between promoters and inhibitors of mitochondrial apoptosis is crucial, as demonstrated by the interplay among the Bcl-2 family members. Particularly, Bcl-x<sub>L</sub> is a target of interest due to its forefront role of its dysfunctions in cancer development. Bcl-x<sub>L</sub> prevents apoptosis by binding both the pro-apoptotic BH3-only proteins, as PUMA, and noncanonical partners such as p53 at different sites. An allosteric communication between the BH3-only proteins binding pocket and the p53 binding site has been postulated and supported by NMR and other biophysical data, mediating the release of p53 from Bcl-x<sub>L</sub> upon PUMA binding. The molecular details, especially at the residue level, of this mechanism remain unclear. In this work, we investigated the distal communication between these two sites in both Bcl-x<sub>L</sub> in its free state and bound to PUMA, and we evaluated how missense mutations of Bcl-x<sub>L</sub> found in cancer samples might impair the communication and thus the allosteric mechanism. We employed all-atom explicit solvent microsecond molecular dynamics simulations analyzed through a Protein Structure Network approach and integrated with calculations of changes in free energies upon cancer-related mutations identified by genomics studies. We found a subset of candidate residues responsible for both maintaining protein stability and for conveying structural information between the two binding sites and hypothesized possible communication routes between specific residues at both sites.

**Keywords:** PUMA, molecular dynamics, p53, apoptosis, cancer mutations, allostery, long-range communication

## Introduction

B-cell lymphoma extra-large (Bcl-x<sub>L</sub>) is one of the best-known proteins of the Bcl-2 family, whose members have been extensively studied over the past decades due to their role in apoptosis regulation, tumorigenesis and cellular responses to anti-cancer therapies <sup>1</sup>.

The members of this family can be subdivided in three classes, according both to their role in apoptosis and to the different functional domains shared with the parent Bcl-2. The first class includes proteins containing all the four *Bcl2-Homology* domains (named BH1-2-3-4) and eliciting an anti-apoptotic response, whilst the second consists of pro-apoptotic proteins containing only the BH3 domain (BH3-only proteins) and the third is formed by proteins including all the BH domains but performing a pro-apoptotic function <sup>2-4</sup>.

Bcl-x<sub>L</sub> belongs to the first class, and structurally consists of eight  $\alpha$ -helices connected by loops of different lengths (**Fig. 1**) and a transmembrane C-terminal region recently solved <sup>5</sup>. Bcl-x<sub>L</sub> carries out its anti-apoptotic function by binding a variety of different partners, such as p53 <sup>6,7</sup> or other Bcl-2-family members, including many BH3-only proteins such as PUMA <sup>3,8</sup>.

It has been demonstrated that the interaction of Bcl-x<sub>L</sub> with the BH3-only proteins could drive long-range structural perturbations within Bcl-x<sub>L</sub>, such as the one observed at the p53 binding site upon the binding of PUMA into the Bcl-x<sub>L</sub> binding pocket for BH3 domains <sup>9,10</sup>. On the other hand, this allosteric communication may work both ways, determining conformational changes into the BH3-only binding pocket when p53 is bound <sup>11</sup>.

Allosteric mechanisms involve the transmission of local structural perturbations in a cascade of couplings of motions or collisional contacts that can be described as paths of communication from one site to a distal one within the protein structure <sup>12-16</sup>. These mechanisms can involve large conformational rearrangements or subtle changes in dynamics and local conformations <sup>17,18</sup>, and allostery is likely to be an intrinsic feature of any protein <sup>17,19</sup>. In this context, it can be described as a shift in the distribution of pre-existing states that are present in both the free and bound conformational ensemble of the protein <sup>20-22</sup>.

The network of residue contacts and the intrinsic dynamics of a protein are thus key components to understand long-range communication and the related molecular and functional mechanisms. One useful framework is the so-called Protein Structure Network (PSN) paradigm, which has been extensively used to describe the structure, topology, and dynamics of proteins <sup>23-28</sup>. In the PSN, intramolecular noncovalent interactions between residues pairs in a protein can be collectively represented as a network <sup>28-30</sup>.

Nevertheless, protein structures are not static entities. Indeed, they undergo a plethora of different motions and conformational changes and are better described as an ensemble of different

conformational states in a dynamic equilibrium<sup>22,31–36</sup> that can be perturbed by post-translational modifications, mutations or interactions with binding partners<sup>33,37–39</sup>. These biologically meaningful dynamic processes can be investigated by computational methods such as molecular dynamics (MD) simulations, which allow an atom level description of the protein conformations<sup>40–42</sup>. The integration of PSN approaches and MD itself proved useful in the last decade to shed light over mechanisms of allosteric communication paths between residues in proteins<sup>13,26,43–55</sup>. The study of protein ensembles from the PSN perspective also revealed to be useful in identifying paths of long-range communication that are only activated upon perturbations within the ensemble (i.e. ligand binding or mutations), despite being present also in the unperturbed system.<sup>28,56,57</sup>

In light of the above observation, we applied a MD-PSN pipeline to the study of the conformational ensemble of Bcl-x<sub>L</sub>. At first, we aimed to verify if the MD-PSN approach can describe the allosteric mechanisms induced by PUMA binding. Then, we investigated how the distal structural communication is propagated and to pinpoint the key residues responsible for the communication between the p53 and the BH3 binding sites of Bcl-x<sub>L</sub>, so that a more profound understanding of this fundamental allosteric mechanism can be achieved. Furthermore, we evaluated if the paths of structural communication between these two distal binding regions were an intrinsic feature of Bcl-x<sub>L</sub> and, as a such, if they were present in Bcl-x<sub>L</sub> even in the free state and then reinforced or (de)activated upon binding of the BH3-only PUMA protein.

## RESULTS

### The interaction with PUMA causes the loss of edges in the p53 binding sites

We collected microsecond all-atom MD simulations for Bcl-x<sub>L</sub> in its free and bound state to the BH3-only protein PUMA and used this ensemble to infer a contact-based protein structure network of the two forms of the protein.

We extracted the edges common to all the PSNs (i.e., the three Bcl-x<sub>L</sub> free simulations and the simulation of the complex Bcl-x<sub>L</sub>-PUMA) and compared them to the ones only common to the three PSNs of the Bcl-x<sub>L</sub> free MD ensembles to appreciate the effect induced by the binding of PUMA on Bcl-x<sub>L</sub> structure and dynamics (Figure 2).

We noticed that the majority of the edges conserved in all PSNs were located in the core of the protein with a high persistence, suggesting a major architectural role for these interactions. We also noted that all the interactions with high occurrence in the Bcl-x<sub>L</sub> free MD ensembles between the loops  $\alpha$ 1- $\alpha$ 2,  $\alpha$ 3- $\alpha$ 4 and the N-terminal portion of  $\alpha$ 4 disappeared when PUMA was bound to the protein (Figure 2). This region largely overlap with the p53 binding site of Bcl-x<sub>L</sub> (Figure 1B and see below for more details), which is located approximately 21 Å apart from the BH3 binding site.

The massive loss of interactions at the p53 binding site upon PUMA interaction into the BH3-binding pocket of Bcl-x<sub>L</sub> may indicate a structural perturbation transmitted long range from one site to the other.

### **Intrinsic and ligand-induced hub as predictive of interfaces for recruitment of binding partners of Bcl-x<sub>L</sub>**

Hubs are the residues with highest connectivity in a network, i.e. nodes connected by more than three edges in a PSN<sup>26</sup>. They can play a role in both protein structural stability and protein function or allow a proper flux of information between distal sites<sup>13,25,28</sup>. Particularly, proteins are known to be built of a significant number of strongly and weakly interacting hub residues stabilizing the tertiary structure by providing resilience against random mutations<sup>20,58–60</sup>.

We identified 14 hubs conserved in all the MD replicates of Bcl-x<sub>L</sub> in its free state, which are mostly located in the core  $\alpha$ -helices. Five of them were also present in the MD ensemble of Bcl-x<sub>L</sub> in complex with PUMA (Fig. 3A), whilst the others were specific of the unbound Bcl-x<sub>L</sub> (Fig. 3B). Of particular interest are F146, experimentally observed as important for the binding of the BH3-only proteins to Bcl-x<sub>L</sub><sup>11,61</sup> and C151, whose mutation C151Y features a high pathogenicity score according to the REVEL ensemble predictor<sup>62</sup> (Table S1).

The interaction with PUMA modified the Bcl-x<sub>L</sub> PSN toward an enrichment of hubs located into the BH3-only proteins binding pocket which were absent in the free state of the protein (i.e., A104, A119, S122, Q125, E129, L130, N136, R139, A142, L194). Hagn and coworkers<sup>11</sup> have reported some of these residues undergoing significant structural perturbations when Bcl-x<sub>L</sub> binds p53 at a distal site with respect to the BH3-binding groove using NMR chemical shift perturbation and Paramagnetic Relaxation Enhancement (PRE) measurements. In particular, they reported as key residues sensitive to the allosteric effect induced by p53-binding A104, A119, S122, Q125, A142, L194<sup>11</sup>, enforcing the notion that the binding of p53 influences the binding mechanism of the BH3-only PUMA protein. Moreover, the work of Campbell and coworkers<sup>61</sup>, using a cell free split-luciferase assay, identified four of the PUMA-induced hubs (E129, L130, N136, R139) as crucial residues for the binding of several BH3-only proteins, including PUMA. Two of these residues, L130 and R139, are identified as hubs only in the ensembles of the Bcl-x<sub>L</sub>-PUMA complex, reinforcing the hypothesis of a prominent role in mediating the binding process.

To investigate how the mutations in the five conserved hubs and in those specific of Bcl-x<sub>L</sub> in free state and of the Bcl-x<sub>L</sub>-PUMA complex might impact the protein stability or binding free energy, we calculated the variation of the  $\Delta\Delta G$  upon all possible missense variations at each site for both the free Bcl-x<sub>L</sub> NMR ensemble (Supplementary File S1) or the NMR ensemble of the complex with

PUMA (Supplementary File S2). As shown in Fig. 4A, mutations in C151, S154 and V163 were predicted to significantly destabilize the protein structure but with minor impact of binding.

Our PSN results, together with the experimental data from NMR and biochemical assays mentioned above, suggest that the PUMA-induced hub residues are predictive of the key residues for the interaction with the BH3-only proteins.

### **Identification of paths of long-range communication between the p53 binding site and the BH3-only proteins binding pocket**

We selected the pairs of residues for shortest paths calculation on the basis of the work of Hagn and coworkers<sup>11</sup>, who experimentally identified the residues constituting the binding interface between p53 and Bcl-x<sub>L</sub> and postulated the existence of a distal communication between the p53-binding site and the BH3-only proteins binding groove of Bcl-x<sub>L</sub> (Fig. 5A). The interaction of p53 with Bcl-x<sub>L</sub> leads to a conformational change in the pocket, conversely resulting in the displacement of the BH3-protein possibly bound<sup>11</sup> and to the initiation of the apoptotic process. This work has been corroborated by Follis and coworkers<sup>9</sup>, who described a long-range communication between the two binding sites resulting from the binding of PUMA, a pro-apoptotic BH3-only protein, to the BH3-proteins binding pocket. Indeed, they noted that the formation of the Bcl-x<sub>L</sub>-PUMA complex was accompanied by the displacement of p53 from Bcl-x<sub>L</sub> and postulated a two-way long-range structural route driving this phenomenon<sup>9</sup>.

Therefore, we calculated the shortest paths between the residues determined as fundamental for the binding of p53 (S18, Y22, S23, Q26, S28, I114, V155, D156, E158) and those in the BH3-only proteins binding pocket (A93, F97, R102, A104, S106, L108, Q111, A119, S122, Q125, F131, A142, F146, L194, Y195) showing perturbed chemical shifts upon p53 binding according to Hagn's work (Fig.5B). We also included an additional residue identified by Kriwacri's group<sup>9</sup> as crucial for PUMA binding, H113. We aimed: i) to verify if any communication path exists that may transmit the perturbation to the p53 binding site, ii) if these paths are pre-existing routes in the free ensemble of Bcl-x<sub>L</sub> in solution and iii) what are the key residues involved in the allosteric communication.

We observed that shortest paths exist in all simulation replicates of Bcl-x<sub>L</sub> free for the following residue pairs (the first element of each pair is a residue belonging to the p53 binding site, the second lies into the BH3-only proteins binding pocket): Y22-L108, Y22-H113, Y22-F146, S23-L108, S23-H113, S23-F146, Q26-L108, Q26-H113, Q26-F146, S28-L108, S28-H113, S28-F146, V155-L108, V155-H113, V155-F146.

We noted that almost all paths having S28 as first residue shared three residues of the path (V10, A167, L13) but none of these residues appear in other paths, suggesting a specificity of those residues in conveying structural information from S28 to the BH3-only proteins binding pocket. The only exception is path S28-H113 of replicate 3, in which another route is identified, lacking both

V167 and L13. V10, the only one ubiquitously present, has been reported in COSMIC as cancer-related<sup>63</sup>.

Similarly, all paths reaching F146 share the three terminal residues (C151, I166, L150). We observed that the same residues are also part of the terminal portion of all paths ending in L108, meaning that L108 and F146 share part of their communication route with the p53 binding site. Moreover, excluding from the paths ending either in L108 or F146 the two ones having S28 as extreme, we noticed that the number of intermediate shared residues increased (P27, V123, C151, I166, L150), thus making these paths almost perfectly superposable with one another. Our results strongly indicates that there are pre-existing communication routes between the p53- and the BH3-binding sites of Bcl-x<sub>L</sub>. We also identified a group of residues that are crucial players in the modulation of the long-range communication between the two binding sites, which could be sensitive sites for mutations to disrupt the allosteric mechanism.

We calculated the same paths for the ensemble of the Bcl-x<sub>L</sub>-PUMA complex, and we observed that a path existed only for the pairs Y22-L108, Y22-F146, S28-L108, S28-F146, V155-L108, V155-F146. We found that paths ending in F146 were all equally long and shared 5 out of 6 intermediate residues, thus portraying a shared communication route from residues belonging to the p53 binding site to F146. Considering their counterparts in the replicates of Bcl-x<sub>L</sub> free, paths in the Bcl-x<sub>L</sub>-PUMA complex are shorter and share only three residues (C151, I166, L150) with the corresponding ones in the Bcl-x<sub>L</sub> free ensembles. The only exceptions are paths ending in L108 in the Bcl-x<sub>L</sub>-PUMA ensemble, they are considerably longer with respect to their analogous in the Bcl-x<sub>L</sub> free ensembles, encompassing 20 residues instead of 7-10. Furthermore, the topological connections between these common residues are different. Therefore, despite displaying few similarities, the routes connecting the p53 binding site to F146 when Bcl-x<sub>L</sub> is in its free state and when PUMA is bound are distinct, zig-zagging among different helices in the core of the protein. Thus, we can conclude that the binding of PUMA induces a weakening of the pre-existing routes.

### **C151, S154 and V163 are involved both in protein stability and in long-range communication between the p53 binding site and the BH3-only proteins binding pocket**

C151 and S154 are both localized on  $\alpha 5$ , with their side chains protruding toward  $\alpha 6$ . On the other hand, V163 is localized on  $\alpha 6$  and their side chains point toward  $\alpha 5$  (Fig. 4, A).

C151, identified as hub in all the ensembles, is ubiquitous in paths ending in either L108 or F146 and is also present in the path S28-H113, thus appearing in eleven out of fifteen identified paths in the Bcl-x<sub>L</sub> free ensembles, and in all six paths observed in the Bcl-x<sub>L</sub>-PUMA ensemble. This finding indicates a possible front role in conveying structural information between the BH3-only proteins binding pocket and the p53 binding site. The structural relevance of C151 is further supported by the results of the in-silico mutational scan displaying a wide range of mutations of C151 predicted to impact the stability of Bcl-xL (Fig. 4, B). Particularly, C151Y is predicted as strongly destabilizing

( $\Delta\Delta G$  around 20 kcal/mol), and has also been predicted as highly pathogenic by its REVEL score (0.567) (Table S1).

S154 is also a hub in all the networks of Bcl-x<sub>L</sub> in free state, and is found in all paths having H113 as extreme. As for C151, our mutational scan predicted nearly half of its possible mutations as highly destabilizing. Interestingly, mutations S154I and S154N, found in cancer samples but predicted as non-pathogenic by their REVEL scores, are predicted as weakly impacting on the protein stability by FoldX. Therefore, our scan corroborates the hypothesis that these particular mutations have only a weak impact on Bcl-x<sub>L</sub> and its functionality, but at the same time suggests that other mutations of S154 may influence significantly the protein stability.

Both C151 and S154 have been also proposed as core structural residues by a comparative study on the conservation of sequence and structure among the Bcl-2 family members<sup>64</sup>.

V163 is also identified as hub and it is present in 13 out of 15 identified paths in the Bcl-x<sub>L</sub> free ensembles. As for C151 and S154, its mutations are predicted as highly destabilizing the protein fold.

In summary, the identification of C151, S154 and V163 as strong hubs, the significant impact of their mutations on protein stability as predicted by the mutational scan and their localization into the protein core suggest a major involvement in both maintaining the protein core architecture and in transferring structural perturbations from the p53 binding site to the BH3-only proteins binding pocket and vice versa.

### **The communication route between the p53 binding site and the BH3-only binding pocket is interrupted when PUMA is bound**

In contrast to what we observed for the ensembles of Bcl-x<sub>L</sub> in free state, we found no paths having H113 as extreme in one of the replicates of the Bcl-x<sub>L</sub>-PUMA complex, whilst in the other replicate no paths at all were found between the selected residues. Furthermore, H113 forms a connected component on its own in the former network, meaning that it is isolated from the rest of the protein, whilst in latter one all residues belonging to the BH3-only proteins binding site cluster in one connected component, still isolated from the rest of the protein. On the contrary, H113 belongs to the most populated connected component in all Bcl-x<sub>L</sub> free ensembles thanks to the edge formed with T115. Being the only edge connecting H113 to the rest of the component, it is also part of all paths of communication with the p53 binding site. It is present in nearly half of the structures in all the Bcl-x<sub>L</sub> free ensembles (~46%), meaning that the communication route between H113 and the p53 binding site is not always active. This suggests that a binding phenomenon perturbing H113 may shift the equilibrium towards a permanent activation or deactivation, as we observed when PUMA was bound. Indeed, in this case the edge was present in 3.1% of the structures, being thus removed in the final PSN (built with cut-off 20%, as explained in the Methods section). This implies

an increased distance between H113 and T115 which interrupts the communication, as also shown by the calculation of the distance over time in all the simulations (Fig. 6).

Since PUMA has been the only BH3-only protein identified so far as capable of triggering the release of p53 from Bcl-x<sub>L</sub> and H113 has been postulated as crucial for the interaction of PUMA with Bcl-x<sub>L</sub> itself [23340338], the sudden disruption of the paths between H113 and the p53 binding site upon PUMA binding may indicate the loss of an internal structural communication that elicits the release of p53 (Fig. 7).

## METHODS

### Molecular Dynamics simulations

We carried out three replicates of one-microsecond all-atom explicit solvent molecular dynamics (MD) simulations with *GROMACS* version 5<sup>65</sup> with the CHARMM22\*<sup>66</sup> force field for Bcl-x<sub>L</sub> free and the Bcl-x<sub>L</sub>/PUMA complex.

For Bcl-x<sub>L</sub> free, we employed the first model of the NMR-derived ensemble of Bcl-x<sub>L</sub> in the PDB entry 2LPC<sup>67</sup> as starting structure for the MD simulations, whilst for the Bcl-x<sub>L</sub>/PUMA complex we used the first model of the PDB entry 2M04<sup>9</sup>. The Bcl-x<sub>L</sub> free structure spans residues 1-44 and 85-209 of Bcl-x<sub>L</sub> (when compared to the UniProt sequence, isoform 1), missing a highly flexible loop between the first two helices and the membrane-inserted helix at the C-terminus. Only one structure of full-length Bcl-x<sub>L</sub> was available in the PDB at the time of writing (NMR ensemble, PDB entry 1LXL<sup>68</sup>), but its predicted resolution according to ResProx was too high (4.932 Å) to be considered as starting structure for our simulations, and the flexible loop was too long to be reliably modeled. Furthermore, many experimental studies have been carried out on the deleted variant of Bcl-x<sub>L</sub>, which is also fully functional in cellular experiments<sup>68,69</sup>. Hence, in order to maintain a good balance between sequence coverage and resolution, we selected the first model of the NMR ensemble 2LPC as starting structure. We first removed the HIS tag at the C-terminus of 2LPC, therefore retaining residues 1:169 of the PDB structure.

We collected three replicates for the free form of Bcl-x<sub>L</sub> and two replicates of the Bcl-x<sub>L</sub>-PUMA complex to ensure the reproducibility of our results, testing the impact of changing either the solvent model (from TIP3P to TIPS3P) or the N- and C-terminal (from capped to charged).

In one of the replicates of Bcl-x<sub>L</sub> in free state and in one of those of the complex, the starting structure was capped at both termini to avoid effects due to terminal partial charges. The N-terminus and C-terminus were made neutral by removing a proton and by adding an amide capping group, respectively. In particular, the choice of amidating the C-terminus was justified by the actual continuation of the backbone beyond our C-terminus in the full-length protein.



The starting structure was soaked in a dodecahedral box of water molecules at 150 mM [NaCl], using the TIP3P solvent model <sup>70</sup>. The edges of the box were all at least 20 Å distant from the protein atoms in all the Bcl-x<sub>L</sub> free simulations, and 35 Å and 25 Å in the two replicates of the Bcl-x<sub>L</sub>/PUMA complex, respectively. We simulated each system under periodic boundary conditions. We equilibrated the systems according to a protocol previously applied to other cases of study <sup>50,51</sup>, and we performed productive MD simulations in the canonical ensemble at 300 K using velocity rescaling with a stochastic term. We used the LINCS algorithm to constrain the heavy atom bonds <sup>71</sup> to allow the usage of a 2 fs timestep. Long-range electrostatic interactions were calculated with the Particle-mesh Ewald (PME) summation scheme <sup>72</sup>; Van der Waals and short-range Coulomb interactions were truncated at 10 Å.

### **Structural Network Analysis with PyInteraph.**

We used the *PyInteraph* suite <sup>26</sup> to unveil the paths of long-range structural communication in Bcl-x<sub>L</sub> applying the principles of graph theory, focusing on the possible structural modifications which may be transmitted from the p53 binding site to the BH3-only proteins binding pocket upon binding of p53 and vice versa. We considered as interacting pairs any two residues whose side-chain centers of mass lied within 5.125 Å. This cut-off was selected on the basis of the results produced by the *PyInKnife* pipeline <sup>51</sup>, which calculates the average number and size of the connected components and hubs distribution of a PSN implementing a Jackknife resampling method on the ensemble, estimating the variation of the data across the different resamplings. We tested contact cutoffs in the range of 5.0-5.5 Å, as suggested from other studies on globular proteins and different MD force fields <sup>50,51</sup>. Specifically, we probed values every 0.125 Å. We noticed that a contact cut-off of 5.0 Å led to a quite sparse PSN with many poorly populated connected components (Fig. S1), whilst a tendency of the nodes to group in one unique connected component is observable already at 5.25 Å (Fig. S1). Since the two cutoffs described above underpin the construction of unrepresentative networks, we retained a contact cutoff of 5.125 Å for the analyses.

We also applied a cutoff to the persistence ( $p_{crit}$ ) of the interaction in our PSN to filter out transient and spurious interactions from our network. This cut-off was selected on the basis of the size of the largest connected component evaluated at different cut-offs (from 0-100% with step 10%), identifying a  $p_{crit}$  of 20% as previously observed for other cases of studies <sup>26</sup>.

### **Prediction of pathogenicity score with REVEL and collection of cancer-related mutations**

REVEL is an ensemble method which accounts for 18 different pathogenicity prediction scores and trained with random forest model <sup>62</sup>. Mutations with REVEL scores higher than 0.4 are predicted to be pathogenic, according to the original publication. Mutations with a REVEL score of 0 are those for which the predictor could not assign a score due to the fact that they do not account for single nucleotide changes and as a such not covered by this predictor. Missense mutations of Bcl-x<sub>L</sub> found by genomics studies have been collected, aggregated from Cbioportal <sup>73</sup> and COSMIC <sup>74</sup>.

## Structure-based estimation of the impact of mutations on protein stability

We employed the *FoldX* energy function from the *FoldX suite*<sup>75</sup> to carry out *in silico* saturation mutagenesis using a *Python* wrapper that we recently developed and previously applied to other case study<sup>50,76–78</sup>. We collected an average  $\Delta\Delta G$  for each mutation over the whole NMR ensemble of 20 conformers for both the free state of Bcl-x<sub>L</sub> (PDB entry 2LPC) and its complex with PUMA (PDB entry 2M04). The ensemble was used to account for flexibility in the protein, as we recently did on another protein system<sup>50</sup> since *FoldX* energy function only allows for local conformational changes. We calculated the  $\Delta\Delta G$  upon mutations associated to protein stability and binding of the interactor. We applied the *BuildModel* module from the *FoldX suite* and five independent runs for mutations in our scan. The typical prediction error of *FoldX* is about 0.8 kcal/mol<sup>79</sup>. We then use twice the prediction error (i.e., 1.6 kcal/mol) as a threshold to discriminate between neutral and deleterious mutations in the analyses.

## Acknowledgments

This project was supported by the LEO foundation grant number LF17006 and the Carlsberg Distinguished Fellowship CF18-0314 to EP group and three Erasmus Plus for Traineeship Fellowships to VS to visit EP group in 2016, 2017 and 2018 by the University of Milano-Bicocca and the University of Bologna, respectively. The calculations described in this paper were performed with the support of the DeiC National Life Science Supercomputer at DTU. the EU PRACE-DECI GRANT 14th CancerBH3 and the ISCRA-Cineca HPC Grants HP10C8YXRK and HP10CBLBWO. The authors would like to thank Matteo Tiberti for technical help and useful suggestions.

## Figure Legends

Fig.1 – Structure of Bcl-x<sub>L</sub> free and of the Bcl-x<sub>L</sub>-PUMA complex

a) Structure of Bcl-x<sub>L</sub>, missing the loop  $\alpha$ 1- $\alpha$ 2, from the PDB structure 2LPC. The eight helices are colored and labelled. b) Structure of Bcl-x<sub>L</sub> bound to PUMA from the PDB structure 2M04. The eight helices are color-coded as in the free structure, and PUMA is shown in black.

Fig.2 – Common edges

Common edges for all PSNs (a) for all Bcl-x<sub>L</sub> free PSNs (b) and for Bcl-x<sub>L</sub>-PUMA (c). Edges are displayed as sticks and mapped on the PDB structure 2M04 for the Bcl-x<sub>L</sub>-PUMA complex and on 2LPC for Bcl-x<sub>L</sub> free, respectively. The thickness of each edge is proportional to its persistence. The C $\alpha$  of residues connected by edges are displayed as balls.

Fig. 3 – Common hubs

Common hubs for all PSNs (a) for all Bcl-x<sub>L</sub> free PSNs (b) and for Bcl-x<sub>L</sub>-PUMA (c). Hub nodes are shown as sticks. Bcl-x<sub>L</sub> is showed in grey, PUMA in black.

Fig 4 – Mutational hotspots in Bcl-x<sub>L</sub>

Hotspot residues are displayed as sticks and labelled, color-coded according to the helix they belong ( $\alpha$ 5 is pink,  $\alpha$ 6 is orange). b) Structural impact of mutations at hotspots. The results from the FoldX scan are shown as a heatmap, with each mutation color-coded according to the value of the calculated  $\Delta\Delta G$  with respect to the wild-type variant.

Fig.5 – BH3-only proteins and p53 binding sites

The p53 binding site is colored in green, whilst the BH3-only proteins binding pocket is shown in pink. The residues crucial for the binding process and for which at least a shortest path exists are shown as sticks and labelled.

Fig. 6 – Distribution of the H113-T115 distance within the PSNs

Cumulative histograms showing the distribution of the H113-T115 distance within the five PSNs. The three replicas of Bcl-x<sub>L</sub> free are displayed in green, orange, pink, whilst the PSNs for the Bcl-x<sub>L</sub>-PUMA complex are shown in blue and red. The horizontal dashed lines indicate the persistence cut-off in the PSNs, meaning that if the distance is lower than 5.125 Å (distance cut-off in the PSNs) in less than 20% of frames, no edge is present between the two residues.

Fig. 7 – Paths from the p53 binding site to H113 in Bcl-x<sub>L</sub> free

Paths connecting the p53 binding site to H113. They are present in all the PSNs of Bcl-x<sub>L</sub> free, but for clarity purposes only those in the PSN computed from the TIP3P+capping ensemble are displayed here. The thickness of each edge is proportional to its persistence, and the C $\alpha$  of residues belonging to each path are displayed as balls.

Fig. S1 – Nodes distribution at different distance cut-offs

Distribution of nodes in the five most populated connected components when the distance cut-off was set to a) 5 Å b) 5.125 Å c) 5.5 Å.

## References

- 1 Youle RJ, Strasser A. The BCL-2 protein family: opposing activities that mediate cell death. *Nat Rev Mol Cell Biol* 2008; **9**: 47–59.
- 2 Moldoveanu T, Follis AV, Kriwacki RW, Green DR. Many players in BCL-2 family affairs. *Trends Biochem Sci* 2014; **39**: 101–111.
- 3 Czabotar PE, Lessene G, Strasser A, Adams JM. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat Rev Mol Cell Biol* 2014; **15**: 49–63.
- 4 Aouacheria A, Combet C, Tompa P, Hardwick JM. Redefining the BH3 Death Domain as a ‘Short Linear Motif’. *Trends Biochem Sci* 2015; **40**: 736–748.
- 5 Raltchev K, Pipercevic J, Hagn F. Production and Structural Analysis of Membrane-Anchored Proteins in Phospholipid Nanodiscs. *Chem - A Eur J* 2018; **24**: 5493–5499.
- 6 Follis AV, Llambi F, Ou L, Baran K, Green DR, Kriwacki RW. The DNA-binding domain mediates both nuclear and cytosolic functions of p53. *Nat Struct Mol Biol* 2014; **21**: 535–43.
- 7 Lee D-H, Ha J-H, Kim Y, Jang M, Park SJ, Yoon HS *et al.* A Conserved Mechanism for Binding of p53 DNA-Binding Domain and Anti-Apoptotic Bcl-2 Family Proteins. *Mol Cells* 2014; **37**: 264–9.
- 8 Kale J, Osterlund EJ, Andrews DW. BCL-2 family proteins: Changing partners in the dance towards death. *Cell Death Differ* 2018; **25**: 65–80.
- 9 Follis AV, Chipuk JE, Fisher JC, Yun MK, Grace CR, Nourse A *et al.* PUMA binding induces partial unfolding within BCL-xL to disrupt p53 binding and promote apoptosis. *Nat Chem Biol* 2013; **9**: 163–168.
- 10 Follis AV, Llambi F, Kalkavan H, Yao Y, Phillips AH, Park C-G *et al.* Regulation of apoptosis by an intrinsically disordered region of Bcl-xL. *Nat Chem Biol* 2018; **14**: 458–465.
- 11 Hagn F, Klein C, Demmer O, Marchenko N, Vaseva A, Moll UM *et al.* BclxL changes conformation upon binding to wild-type but not mutant p53 DNA binding domain. *J Biol Chem* 2010; **285**: 3439–3450.
- 12 Papaleo E, Saladino G, Lambrughli M, Lindorff-Larsen K, Gervasio FL, Nussinov R. The role of protein loops and linkers in conformational dynamics and allostery. *Chem Rev* 2016; **116**: 6391–6423.
- 13 Papaleo E. Integrating atomistic molecular dynamics simulations, experiments, and network analysis to study protein dynamics: strength in unity. *Front Mol Biosci* 2015; **2**: 28.
- 14 Tsai C-J, Nussinov R. A unified view of ‘how allostery works’. *PLoS Comput Biol* 2014; **10**: e1003394.
- 15 Dokholyan N V. Controlling Allosteric Networks in Proteins. *Chem Rev* 2016; **116**: 6463–87.
- 16 Nussinov R, Tsai C-J. Allostery without a conformational change? Revisiting the paradigm. *Curr Opin Struct Biol* 2015; **30**: 17–24.
- 17 Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins?

- Proteins* 2004; **57**: 433–43.
- 18 Tsai CJ, 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.
- 19 Clarkson MW, Gilmore SA, Edgell MH, Lee AL. Dynamic coupling and allosteric behavior in a nonallosteric protein. *Biochemistry* 2006; **45**: 7693–9.
- 20 del Sol A, Tsai C-J, Ma B, Nussinov R. The origin of allosteric functional modulation: multiple pre-existing pathways. *Structure* 2009; **17**: 1042–50.
- 21 Fraser JS, Clarkson MW, Degnan SC, Erion R, Kern D, Alber T. Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 2009; **462**: 669–673.
- 22 Fenwick RB, Esteban-Martín S, Salvatella X. Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles. *Eur Biophys J* 2011; **40**: 1339–1355.
- 23 Csermely P, Korcsmáros T, Kiss HJM, London G, Nussinov R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* 2013; **138**: 333–408.
- 24 Di Paola L, Giuliani A. Protein contact network topology: a natural language for allostery. *Curr Opin Struct Biol* 2015; **31**: 43–48.
- 25 Vuillon L, Lesieur C. From local to global changes in proteins: a network view. *Curr Opin Struct Biol* 2015; **31**: 1–8.
- 26 Tiberti M, Invernizzi G, Lambrughli M, Inbar Y, Schreiber G, Papaleo E. PyIntergraph: a framework for the analysis of interaction networks in structural ensembles of proteins. *J Chem Inf Model* 2014; **54**: 1537–1551.
- 27 Feher VA, Durrant JD, Van Wart AT, Amaro RE. Computational approaches to mapping allosteric pathways. *Curr Opin Struct Biol* 2014; **25**: 98–103.
- 28 Bhattacharyya M, Ghosh S, Vishveshwara S. Protein Structure and Function: Looking through the Network of Side-Chain Interactions. *Curr Protein Pept Sci* 2016; **17**: 4–25.
- 29 Atilgan AR, Akan P, Baysal C. Small-world communication of residues and significance for protein dynamics. *Biophys J* 2004; **86**: 85–91.
- 30 Böde C, Kovács IA, Szalay MS, Palotai R, Korcsmáros T, Csermely P. Network analysis of protein dynamics. *FEBS Lett* 2007; **581**: 2776–2782.
- 31 Ramanathan A, Savol A, Burger V, Chennubhotla CS, Agarwal PK. Protein conformational populations and functionally relevant substates. *Acc Chem Res* 2014; **47**: 149–56.
- 32 Bonomi M, Heller GT, Camilloni C, Vendruscolo M. Principles of protein structural ensemble determination. *Curr Opin Struct Biol* 2017; **42**: 106–116.
- 33 Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. *Nat Chem Biol* 2009; **5**: 789–96.
- 34 Vendruscolo M. Determination of conformationally heterogeneous states of proteins. *Curr Opin Struct Biol* 2007; **17**: 15–20.

- 35 Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. *Nature* 2014; **508**: 331–339.
- 36 Ma B, Nussinov R. Protein dynamics: Conformational footprints. *Nat Chem Biol* 2016; **12**: 890–891.
- 37 Sumbul F, Acuner-Ozbabacan SE, Haliloglu T. Allosteric Dynamic Control of Binding. *Biophys J* 2015; **109**: 1190–1201.
- 38 Wei G, Xi W, Nussinov R, Ma B. Protein Ensembles: How Does Nature Harness Thermodynamic Fluctuations for Life? the Diverse Functional Roles of Conformational Ensembles in the Cell. *Chem Rev* 2016; **116**: 6516–6551.
- 39 Bhabha G, Biel JT, Fraser JS. Keep on moving: discovering and perturbing the conformational dynamics of enzymes. *Acc Chem Res* 2015; **48**: 423–30.
- 40 Orozco M. A theoretical view of protein dynamics. *Chem Soc Rev* 2014; **43**: 5051–66.
- 41 Karplus M, Kuriyan J. Molecular dynamics and protein function. *Proc Natl Acad Sci U S A* 2005; **102**: 6679–85.
- 42 Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys* 2012; **41**: 429–52.
- 43 Fellingine A, Ghitti M, Musco G, Fanelli F. Dissecting intrinsic and ligand-induced structural communication in the  $\beta$ 3 headpiece of integrins. *Biochim Biophys Acta - Gen Subj* 2017; **1861**: 2367–2381.
- 44 Contreras-Riquelme S, Garate J-A, Perez-Acle T, Martin AJM. RIP-MD: a tool to study residue interaction networks in protein molecular dynamics. *PeerJ* 2018; **6**: e5998.
- 45 Marino V, Dell'Orco D. Allosteric communication pathways routed by Ca<sup>2+</sup>/Mg<sup>2+</sup> exchange in GCAP1 selectively switch target regulation modes. *Sci Rep* 2016; **6**: 34277.
- 46 Stetz G, Verkhivker GM. Probing Allosteric Inhibition Mechanisms of the Hsp70 Chaperone Proteins Using Molecular Dynamics Simulations and Analysis of the Residue Interaction Networks. *J Chem Inf Model* 2016; **56**: 1490–1517.
- 47 Verkhivker GM. Biophysical simulations and structure-based modeling of residue interaction networks in the tumor suppressor proteins reveal functional role of cancer mutation hotspots in molecular communication. *Biochim Biophys Acta - Gen Subj* 2019; **1863**: 210–225.
- 48 Papaleo E, Lindorff-Larsen K, De Gioia L. Paths of long-range communication in the E2 enzymes of family 3: a molecular dynamics investigation. *Phys Chem Chem Phys* 2012; **14**: 12515.
- 49 Papaleo E, Renzetti G, Tiberti M. Mechanisms of intramolecular communication in a hyperthermophilic acylaminoacyl peptidase: a molecular dynamics investigation. *PLoS One* 2012; **7**: e35686.
- 50 Nygaard M, Terkelsen T, Olsen AV, Sora V, Salamanca J, Rizza F *et al*. The mutational landscape of the oncogenic MZF1 SCAN domain in cancer. *Front Mol Biosci* 2016. doi:10.3389/fmolb.2016.00078.

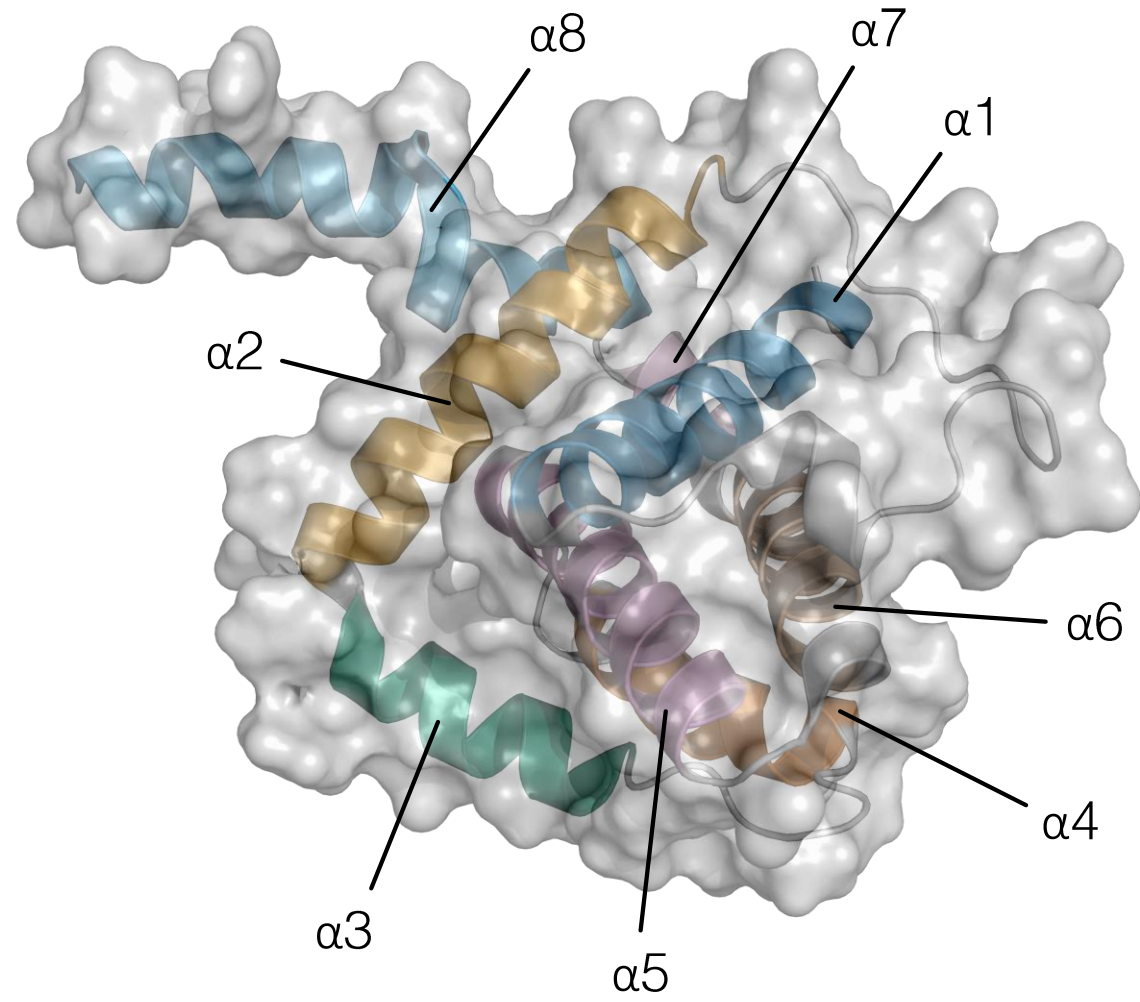
- 51 Salamanca Vilorio J, Allegra MF, Lambrughini M, Papaleo E. An optimal distance cutoff for contact-based protein structure networks using side chain center of masses. *Sci Rep* 2016; **7**: 2838.
- 52 Seeber M, Fellingine A, Raimondi F, Muff S, Friedman R, Rao F *et al.* Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. *J Comput Chem* 2011; **32**: 1183–94.
- 53 Mariani S, Dell’Orco D, Fellingine A, Raimondi F, Fanelli F. Network and Atomistic Simulations Unveil the Structural Determinants of Mutations Linked to Retinal Diseases. *PLoS Comput Biol* 2013; **9**. doi:10.1371/journal.pcbi.1003207.
- 54 Lambrughini M, Lucchini M, Pignataro M, Sola M, Bortolotti CA, Gupta VA *et al.* The dynamics of the  $\beta$ -propeller domain in Kelch protein KLHL40 changes upon nemaline myopathy-associated mutation. *RSC Adv* 2016; **6**: 34043–34054.
- 55 Angelova K, Fellingine A, Lee M, Patel M, Puett D, Fanelli F. Conserved amino acids participate in the structure networks deputed to intramolecular communication in the lutropin receptor. *Cell Mol Life Sci* 2011; **68**: 1227–39.
- 56 Invernizzi G, Tiberti M, Lambrughini M, Lindorff-Larsen K, Papaleo E. Communication routes in ARID domains between distal residues in helix 5 and the DNA-binding loops. *PLoS Comput Biol* 2014; **10**: e1003744.
- 57 Lambrughini M, De Gioia L, Gervasio FL, Lindorff-Larsen K, Nussinov R, Urani C *et al.* DNA-binding protects p53 from interactions with cofactors involved in transcription-independent functions. *Nucleic Acids Res* 2016; **44**: 9096–9109.
- 58 Brinda K V, Vishveshwara S. A network representation of protein structures: implications for protein stability. *Biophys J* 2005; **89**: 4159–70.
- 59 Amitai G, Shemesh A, Sitbon E, Shklar M, Netanel D, Venger I *et al.* Network analysis of protein structures identifies functional residues. *J Mol Biol* 2004; **344**: 1135–46.
- 60 Brinda K V, Vishveshwara S. Oligomeric protein structure networks: insights into protein-protein interactions. *BMC Bioinformatics* 2005; **6**: 296.
- 61 Campbell ST, Carlson KJ, Buchholz CJ, Helmers MR, Ghosh I. Mapping the BH3 Binding Interface of Bcl-x<sub>L</sub>, Bcl-2, and Mcl-1 Using Split-Luciferase Reassembly. *Biochemistry* 2015; **54**: 2632–2643.
- 62 Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 2016; **99**: 877–885.
- 63 Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 2019; **47**: D941–D947.
- 64 Lama D, Sankararamakrishnan R. Identification of core structural residues in the sequentially diverse and structurally homologous Bcl-2 family of proteins. *Biochemistry* 2010; **49**: 2574–84.

- 65 Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015; **2**: 19–25.
- 66 Piana S, Lindorff-Larsen K, Shaw DE. How robust are protein folding simulations with respect to force field parameterization? *Biophys J* 2011; **100**: L47-9.
- 67 Wysoczanski P, Mart RJ, Loveridge EJ, Williams C, Whittaker SB-M, Crump MP *et al.* NMR solution structure of a photoswitchable apoptosis activating Bak peptide bound to Bcl-xL. *J Am Chem Soc* 2012; **134**: 7644–7.
- 68 Muchmore SW, Sattler M, Liang H, Meadows RP, Harlan JE, Yoon HS *et al.* X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* 1996; **381**: 335–41.
- 69 Liu X, Dai S, Zhu Y, Marrack P, Kappler JW. The structure of a Bcl-xL/Bim fragment complex: implications for Bim function. *Immunity* 2003; **19**: 341–52.
- 70 Mark P, Nilsson L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J Phys Chem A* 2001; **105**: 9954–9960.
- 71 Hess B, Bekker H, Berendsen H, Fraaije J. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 1993; **12**: 1463–1472.
- 72 Li P, Roberts BP, Chakravorty DK, Merz KM. Rational design of particle mesh ewald compatible lennard-jones parameters for +2 metal cations in explicit solvent. *J Chem Theory Comput* 2013; **9**: 2733–2748.
- 73 Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA *et al.* The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012; **2**: 401–404.
- 74 Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* 2019; **47**: D941–D947.
- 75 Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res* 2005; **33**: W382-8.
- 76 Papaleo E, Parravicini F, Grandori R, De Gioia L, Brocca S. Structural investigation of the cold-adapted acylaminoacyl peptidase from *Sporosarcina psychrophila* by atomistic simulations and biophysical methods. *Biochim Biophys Acta - Proteins Proteomics* 2014; **1844**: 2203–2213.
- 77 Nielsen S V., Stein A, Dinitzen AB, Papaleo E, Tatham MH, Poulsen EG *et al.* Predicting the impact of Lynch syndrome-causing missense mutations from structural calculations. *PLOS Genet* 2017; **13**: e1006739.
- 78 Scheller R, Stein A, Nielsen S V., Marin FI, Gerdes A-M, Di Marco M *et al.* Toward mechanistic models for genotype-phenotype correlations in phenylketonuria using protein stability calculations. *Hum Mutat* 2019. doi:10.1002/humu.23707.
- 79 Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein

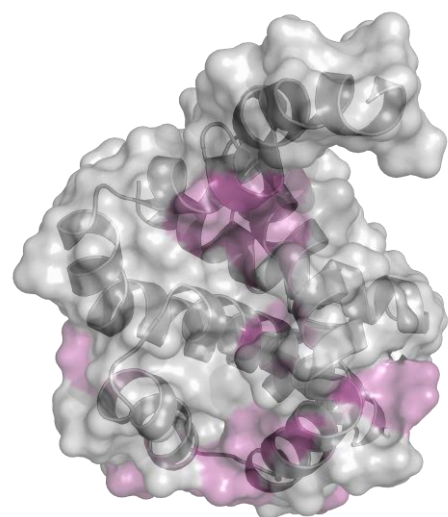
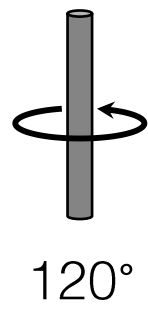
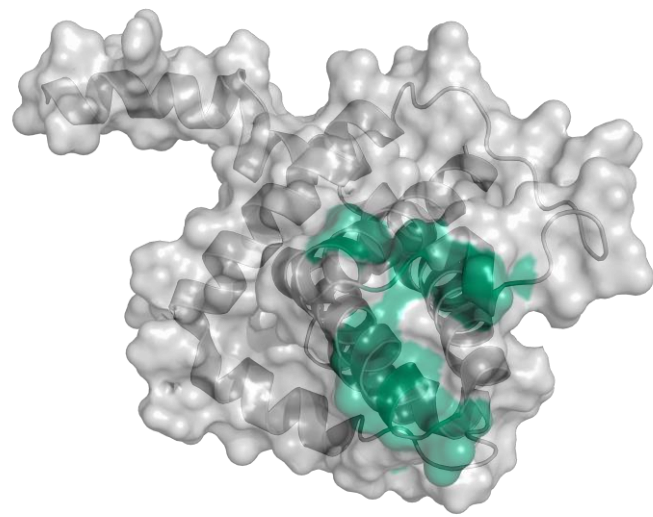
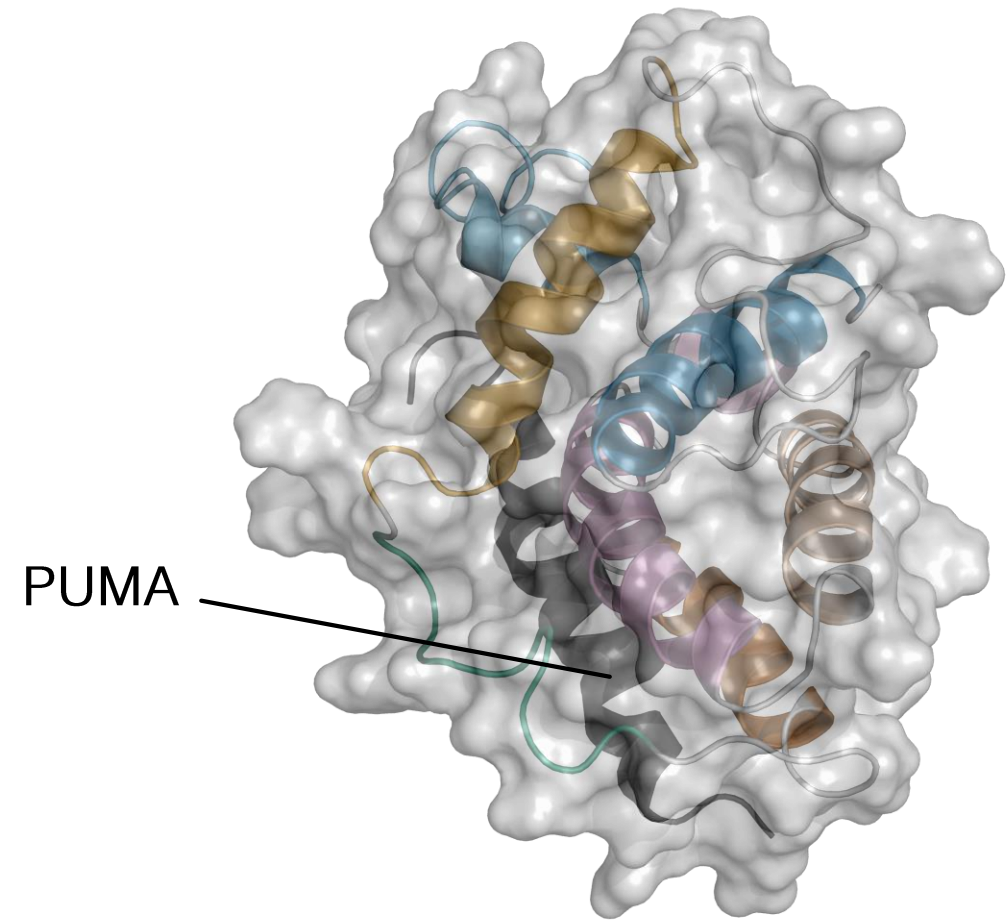


complexes: A study of more than 1000 mutations. *J Mol Biol* 2002; **320**: 369–387.

a)

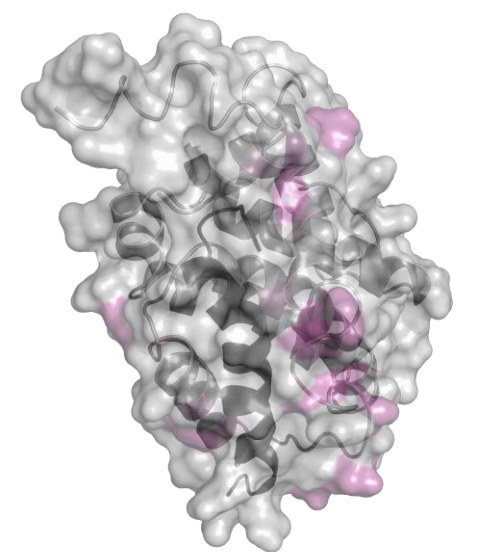
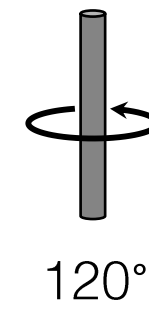
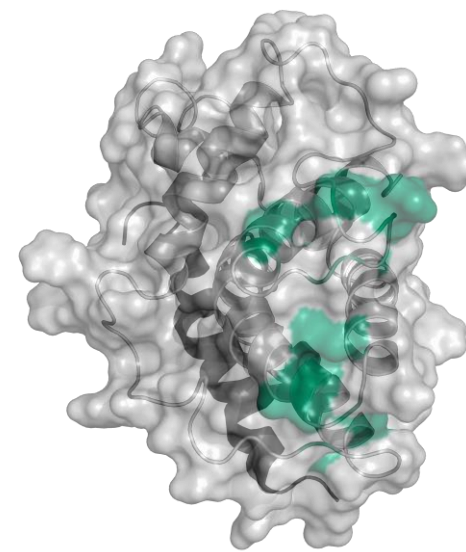


b)



BH3-only proteins  
binding site

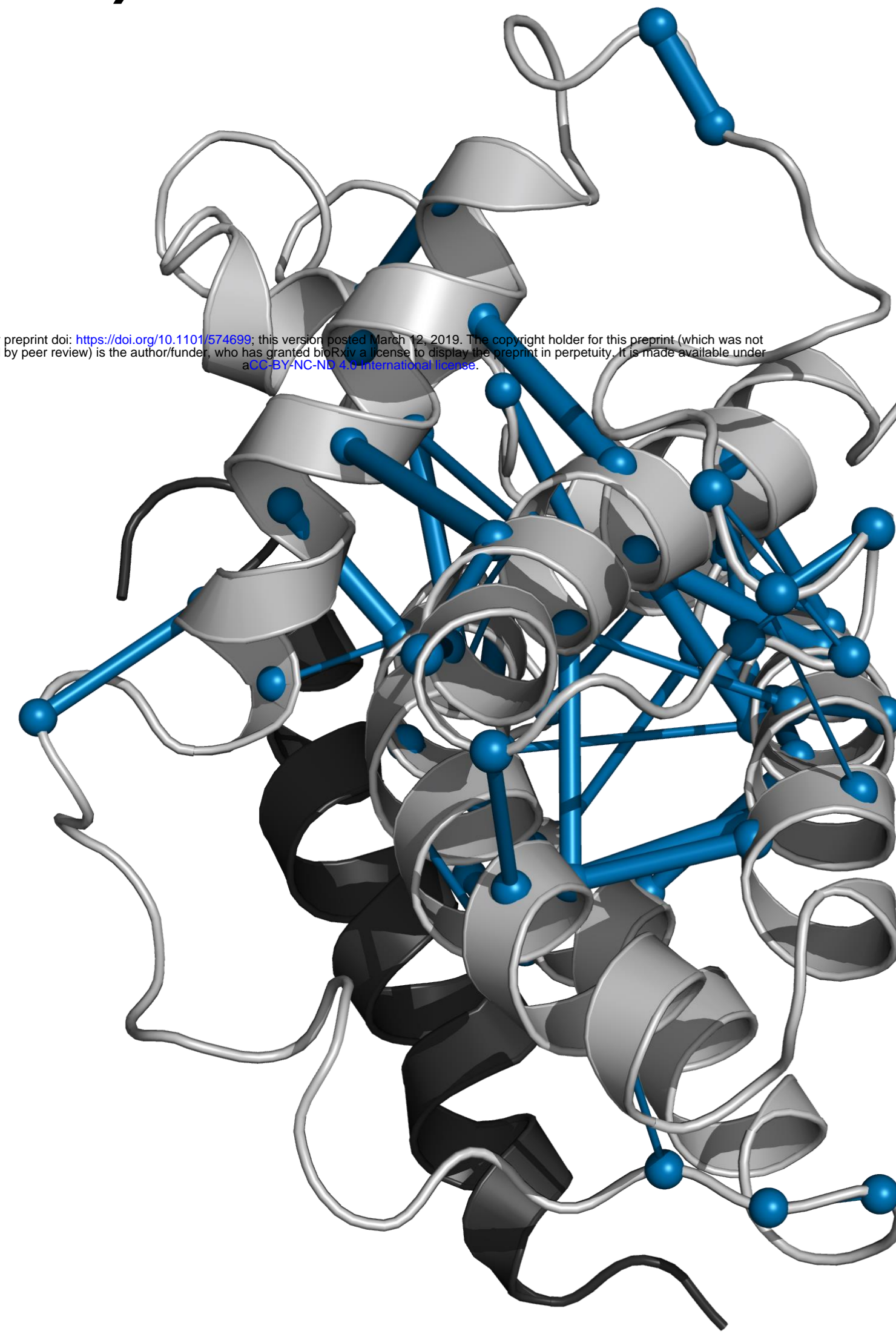
p53 binding site



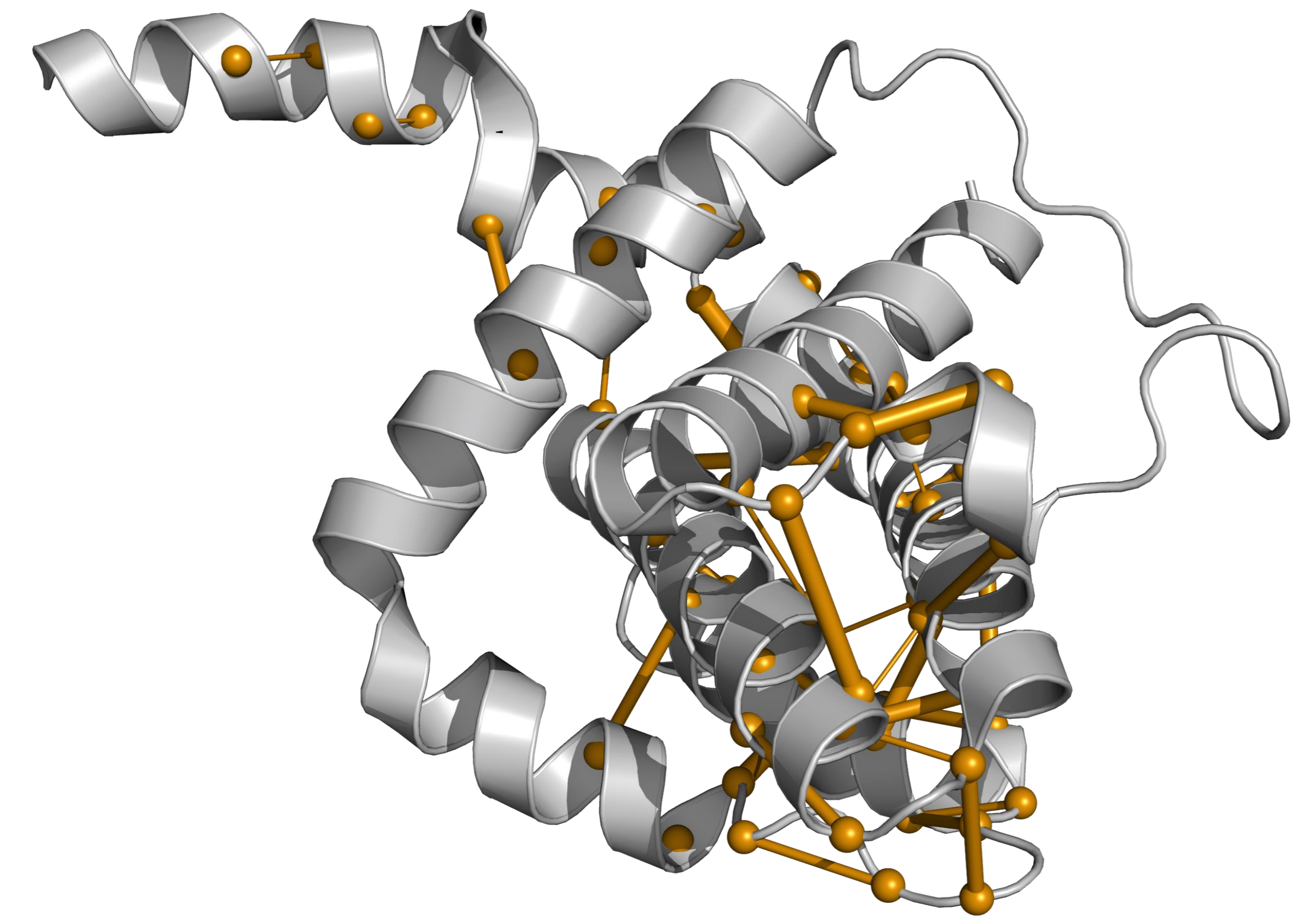
BH3-only proteins  
binding site

p53 binding site

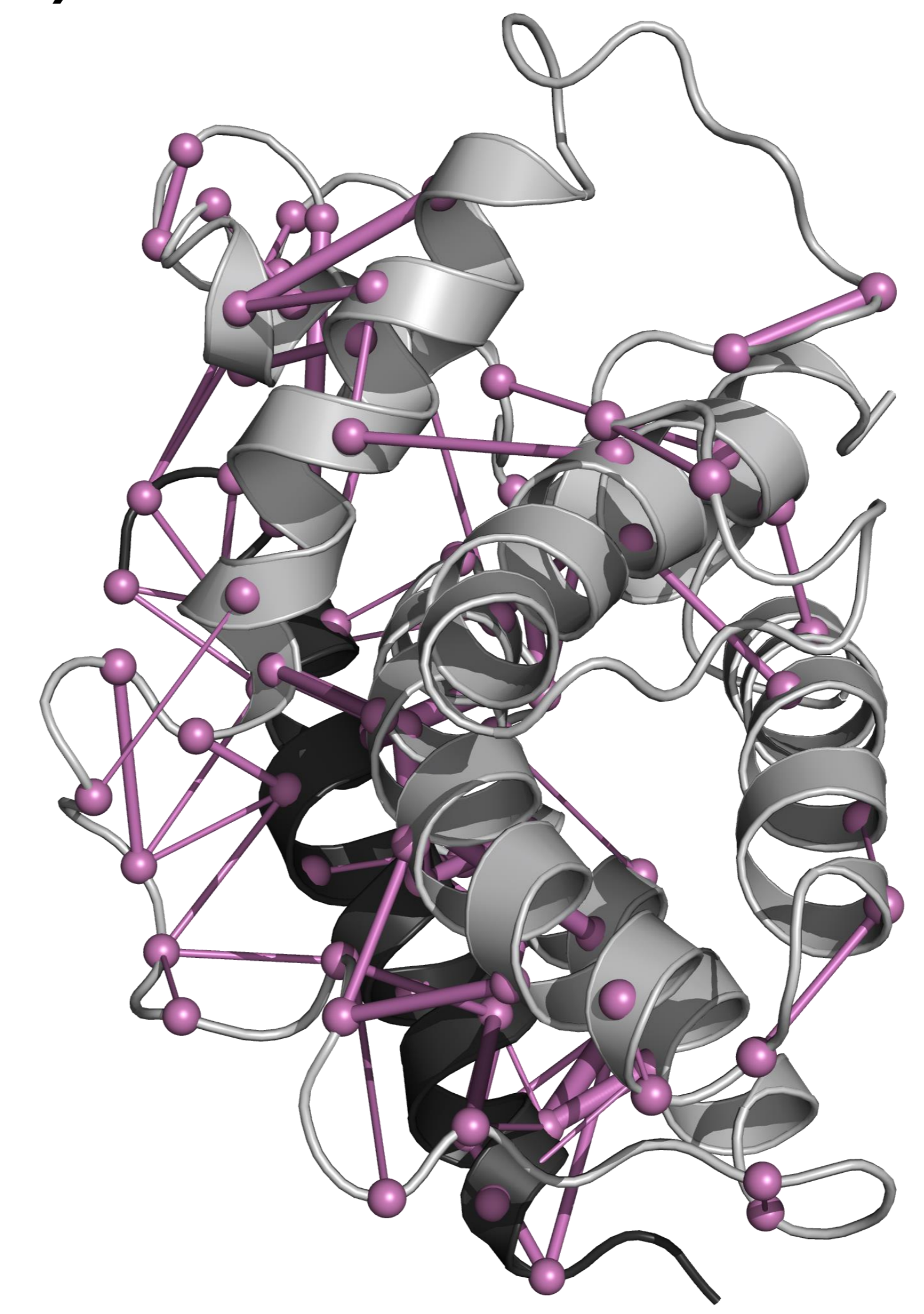
a)



b)



c)



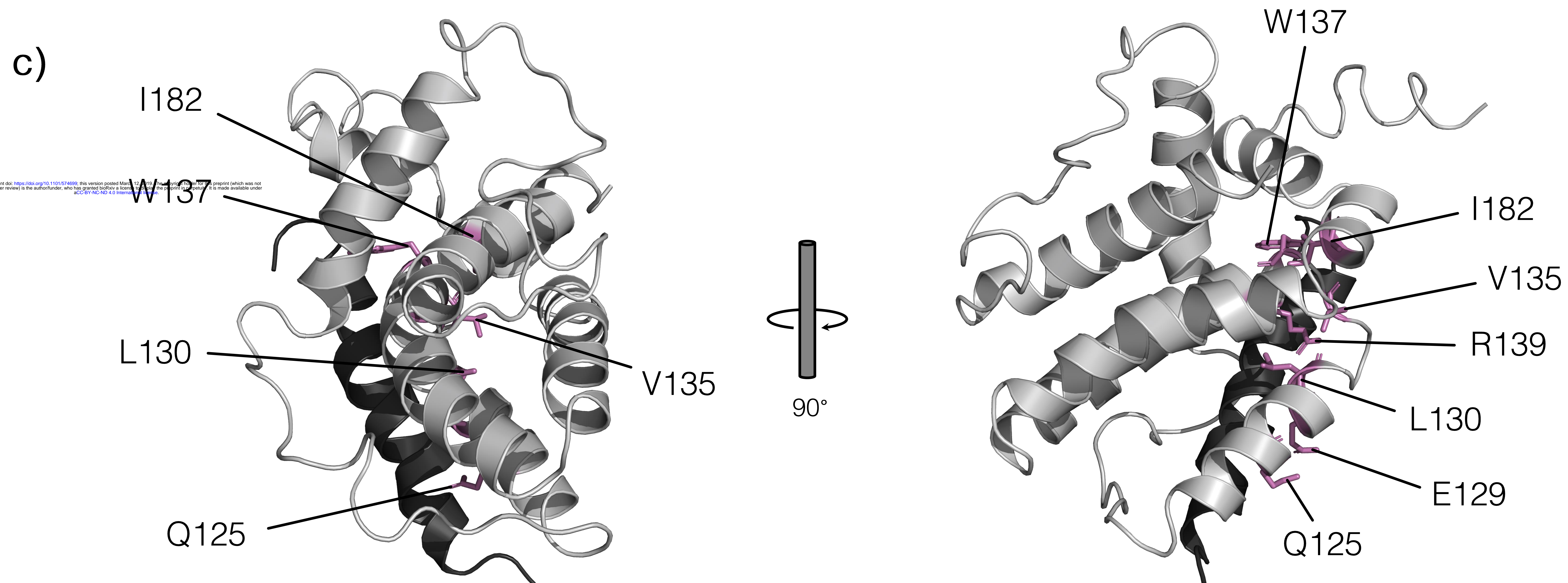
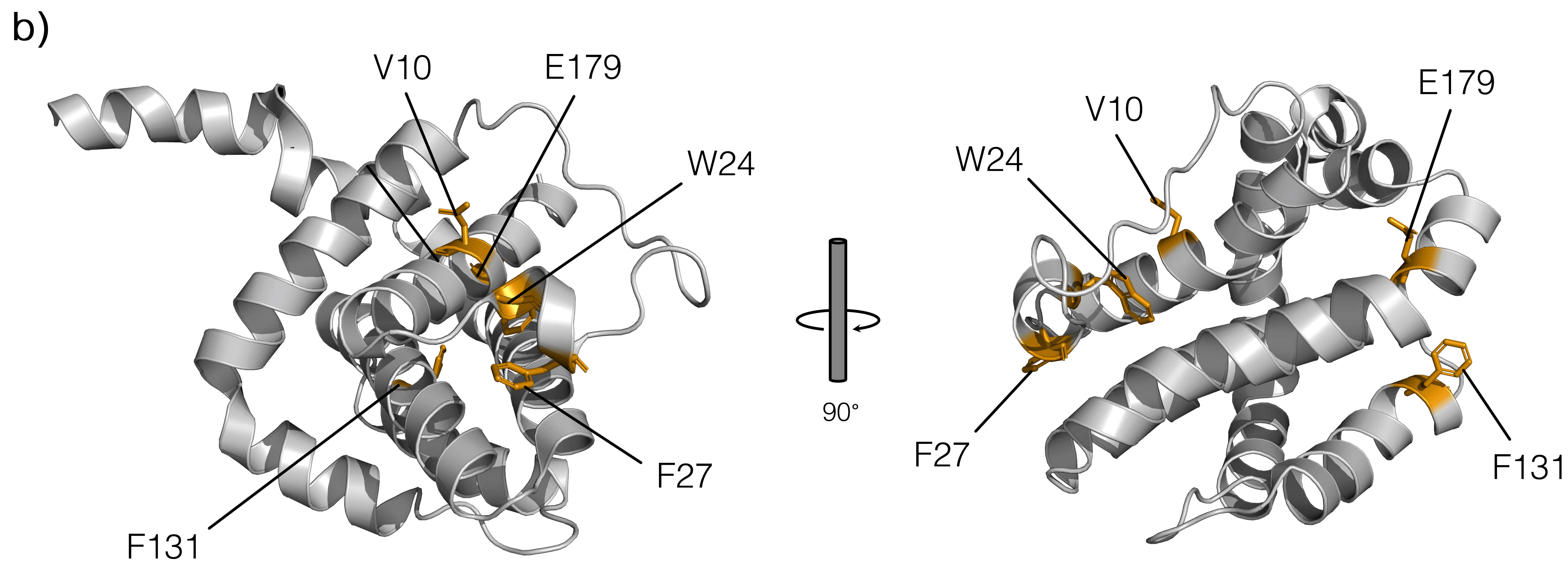
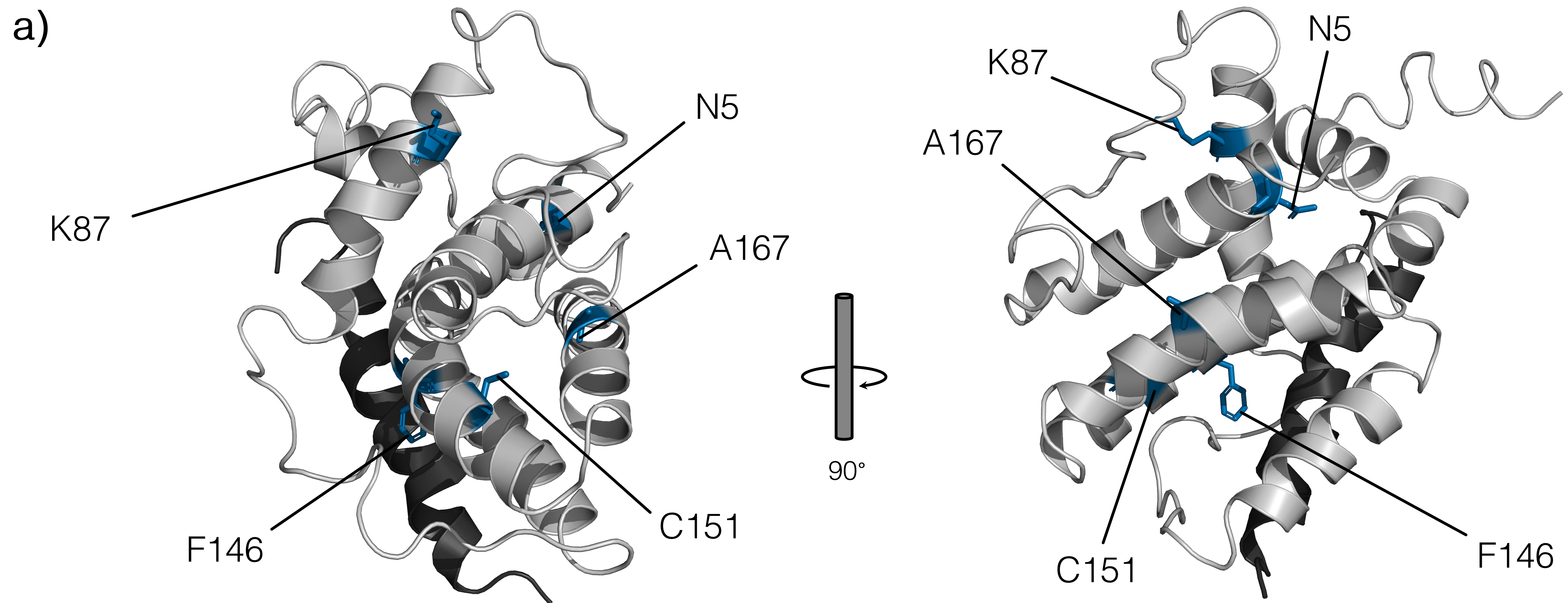
edges common to all PSNs



edges common to all PSNs of Bcl-x<sub>L</sub> free

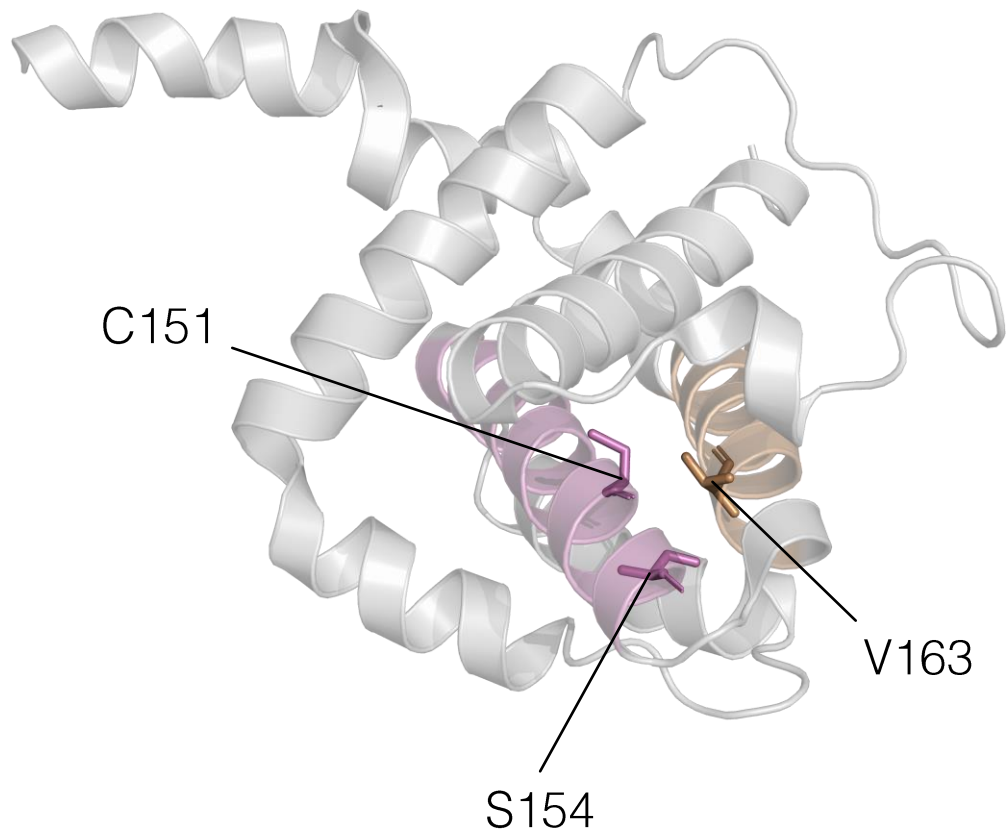


edges common to all PSNs of Bcl-x<sub>L</sub>-PUMA

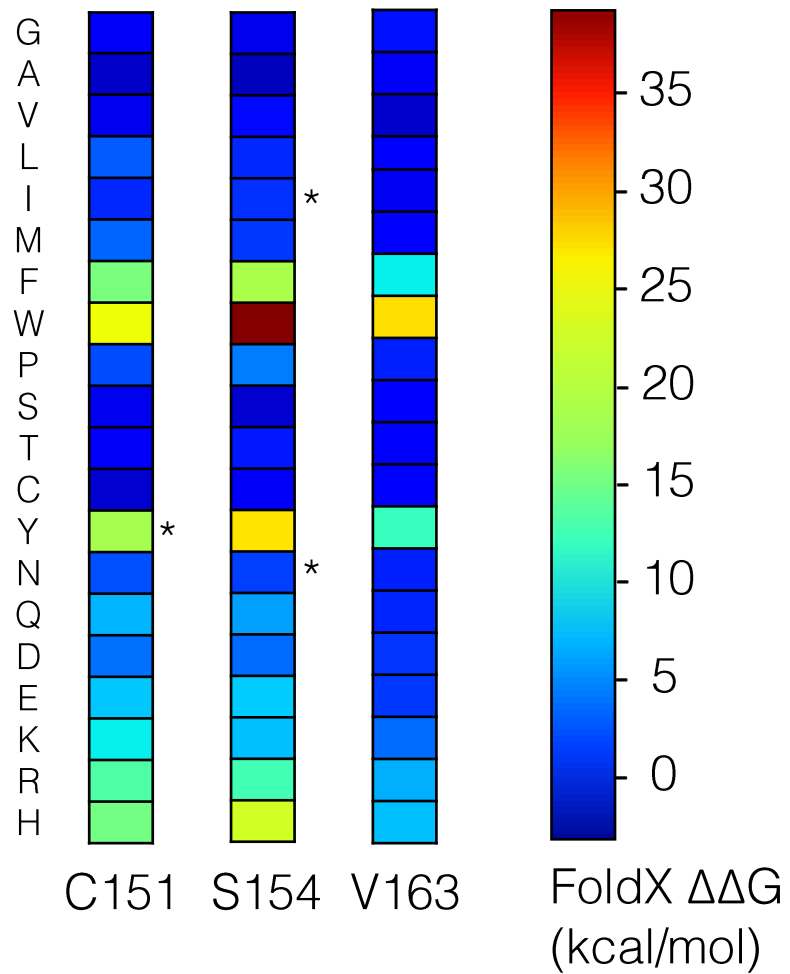


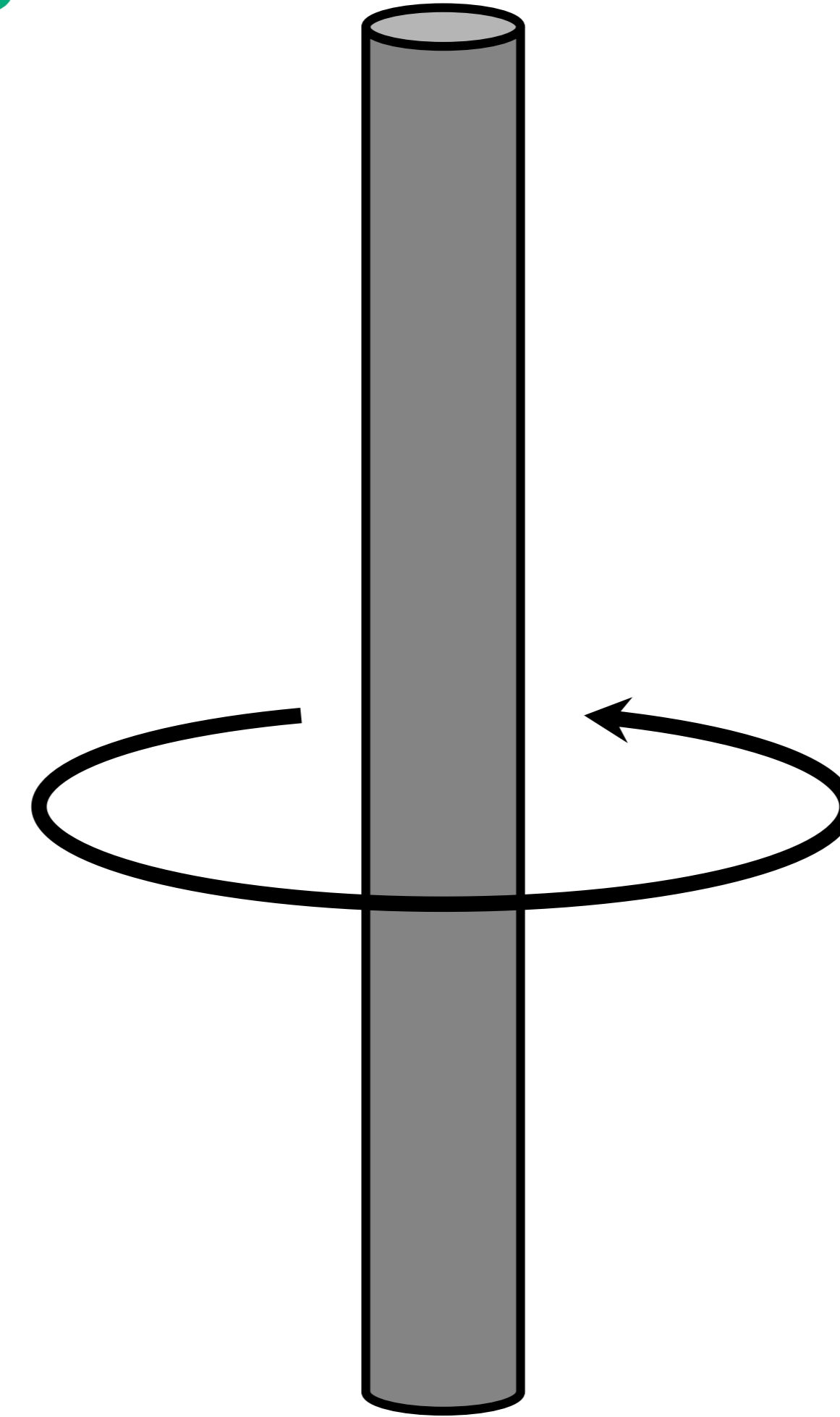
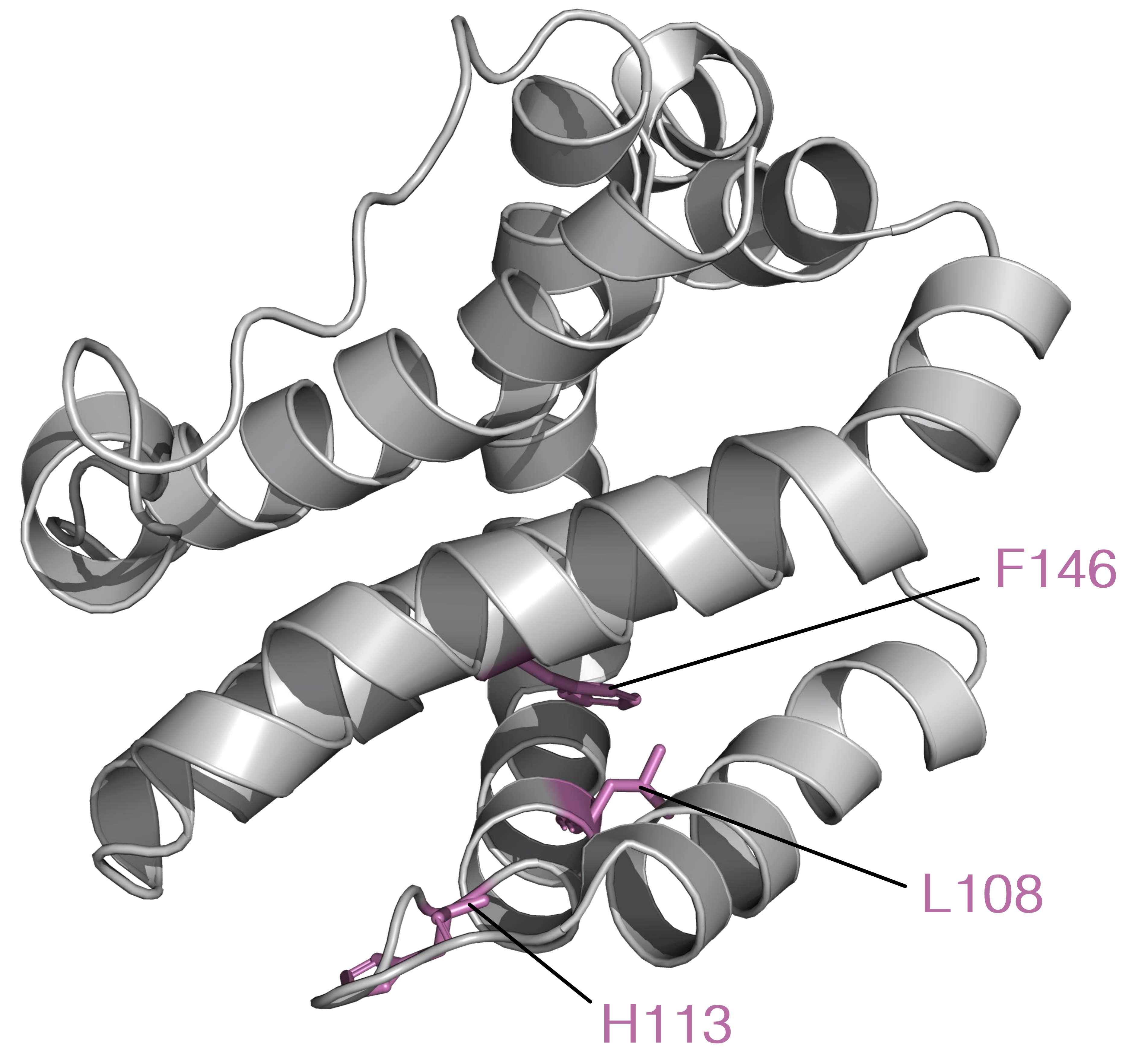
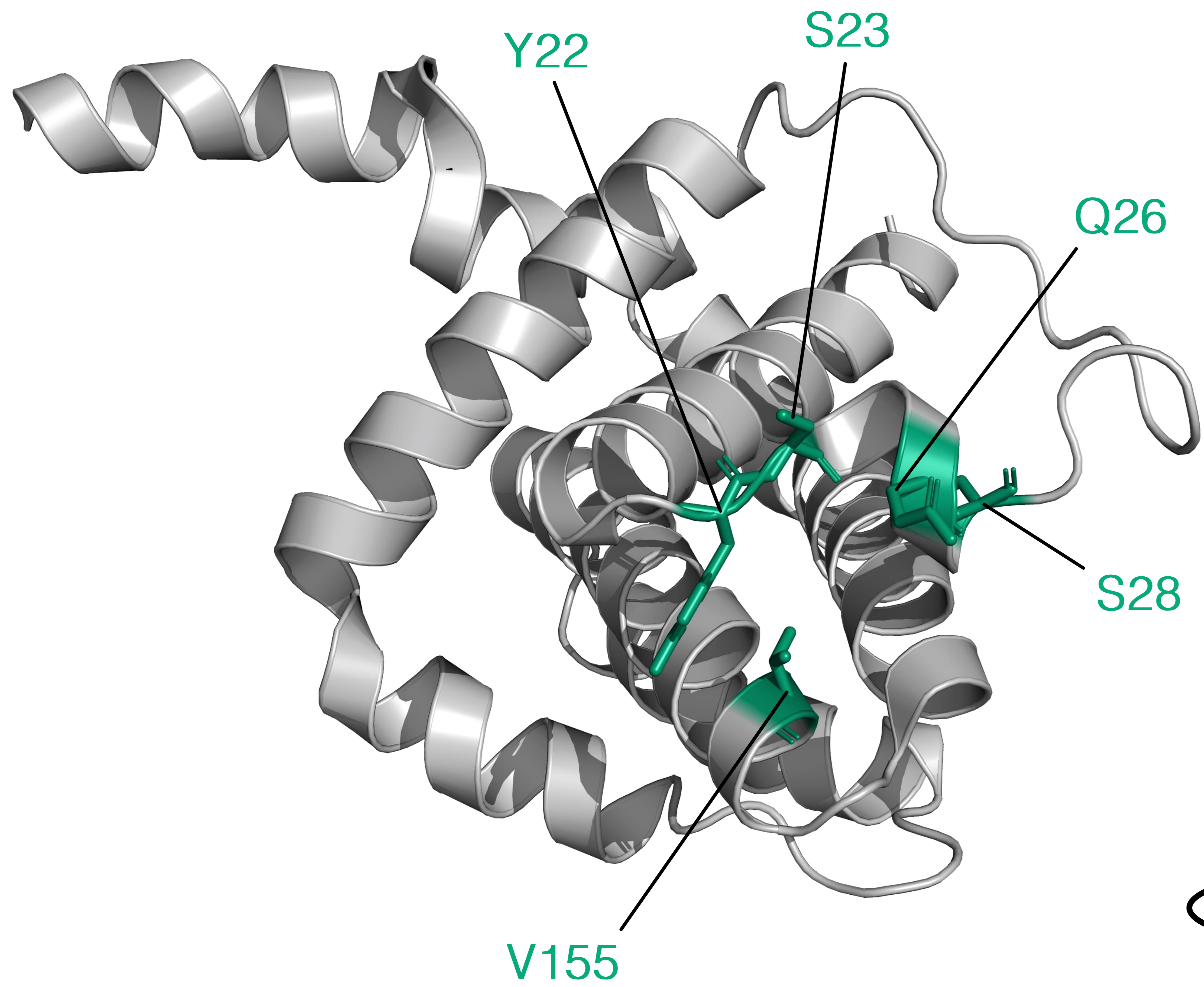
- hubs common to all PSNs
- hubs common to all PSNs of Bcl-x<sub>L</sub> free
- hubs common to all PSNs of Bcl-x<sub>L</sub> - PUMA

a)

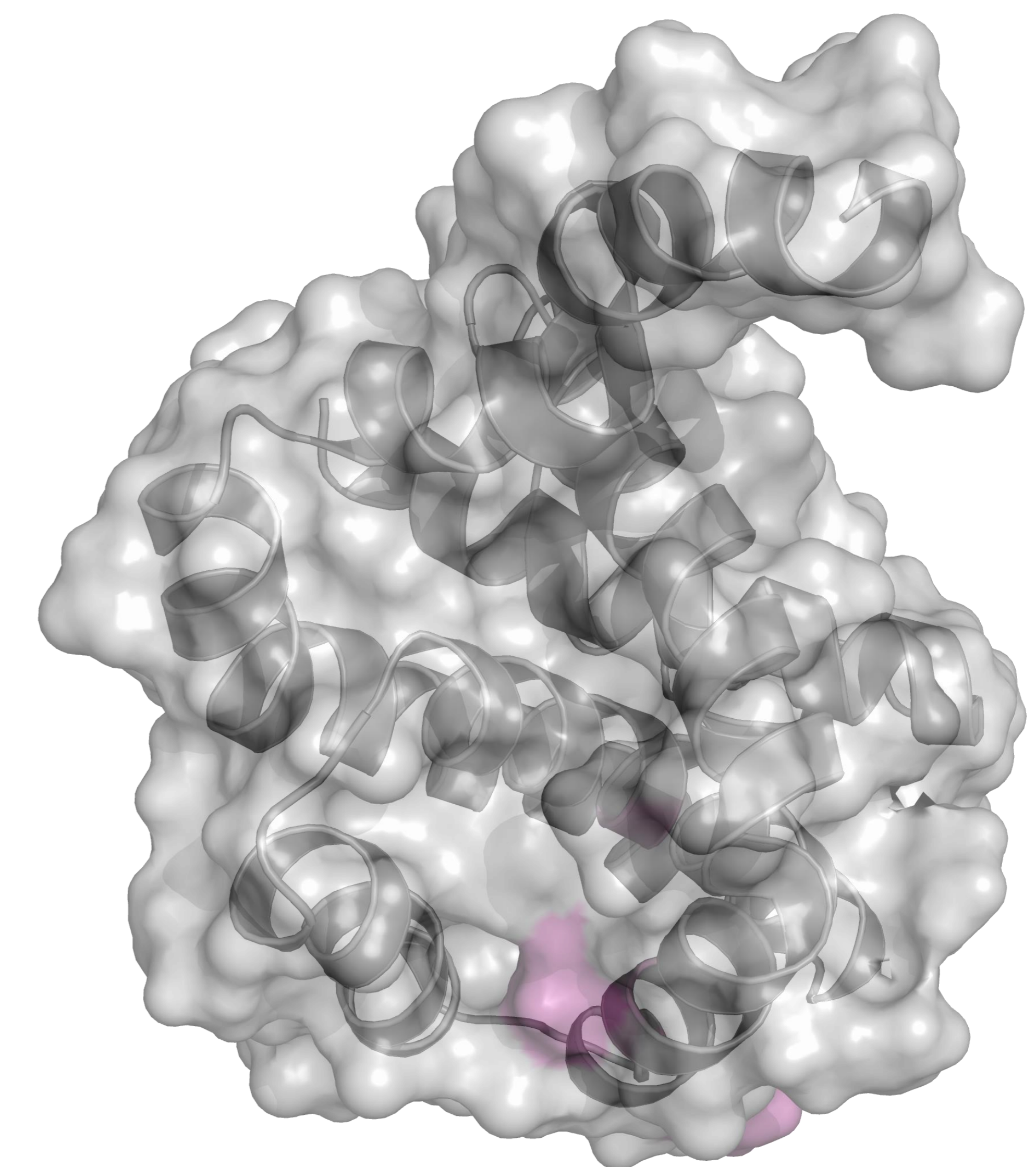
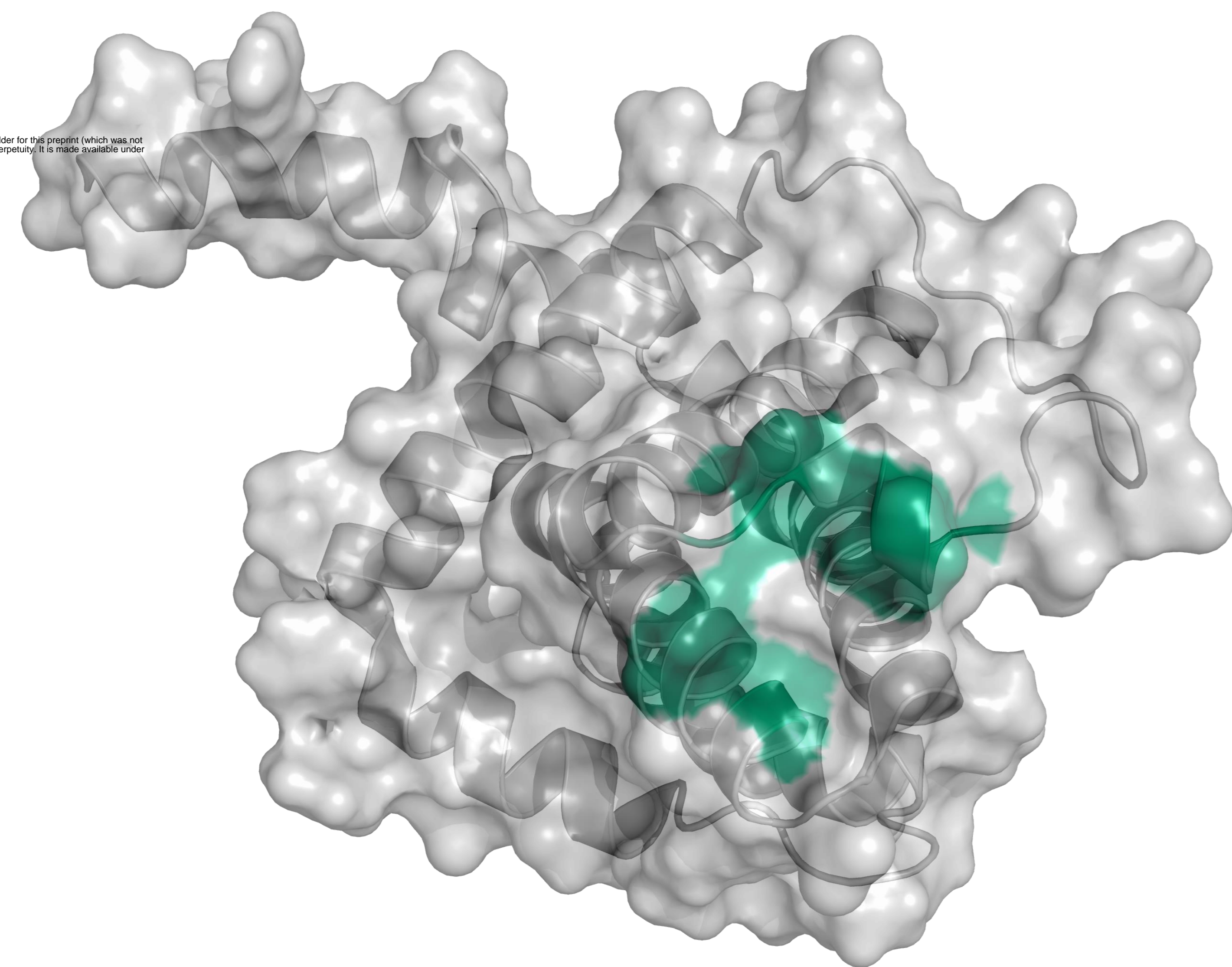


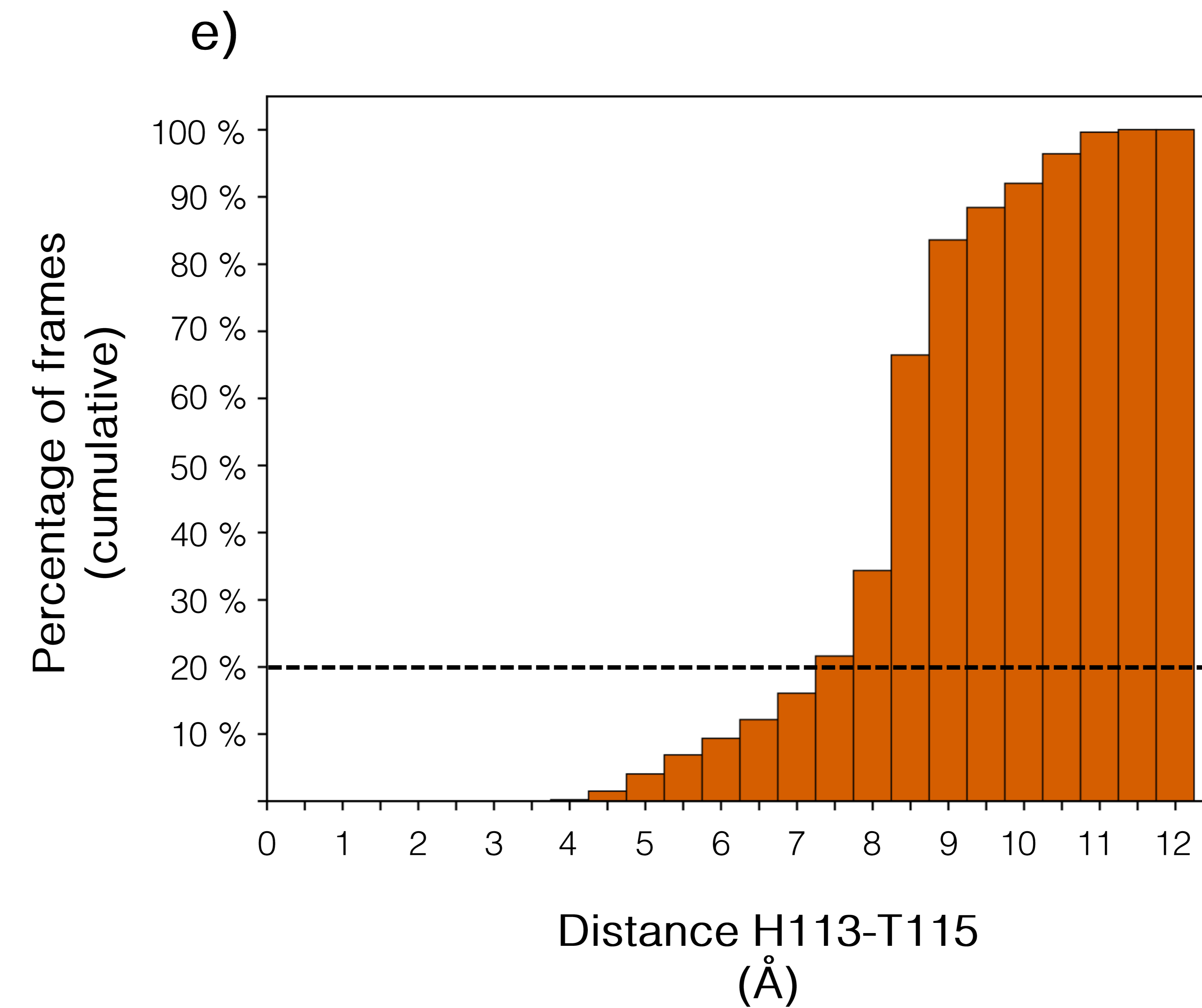
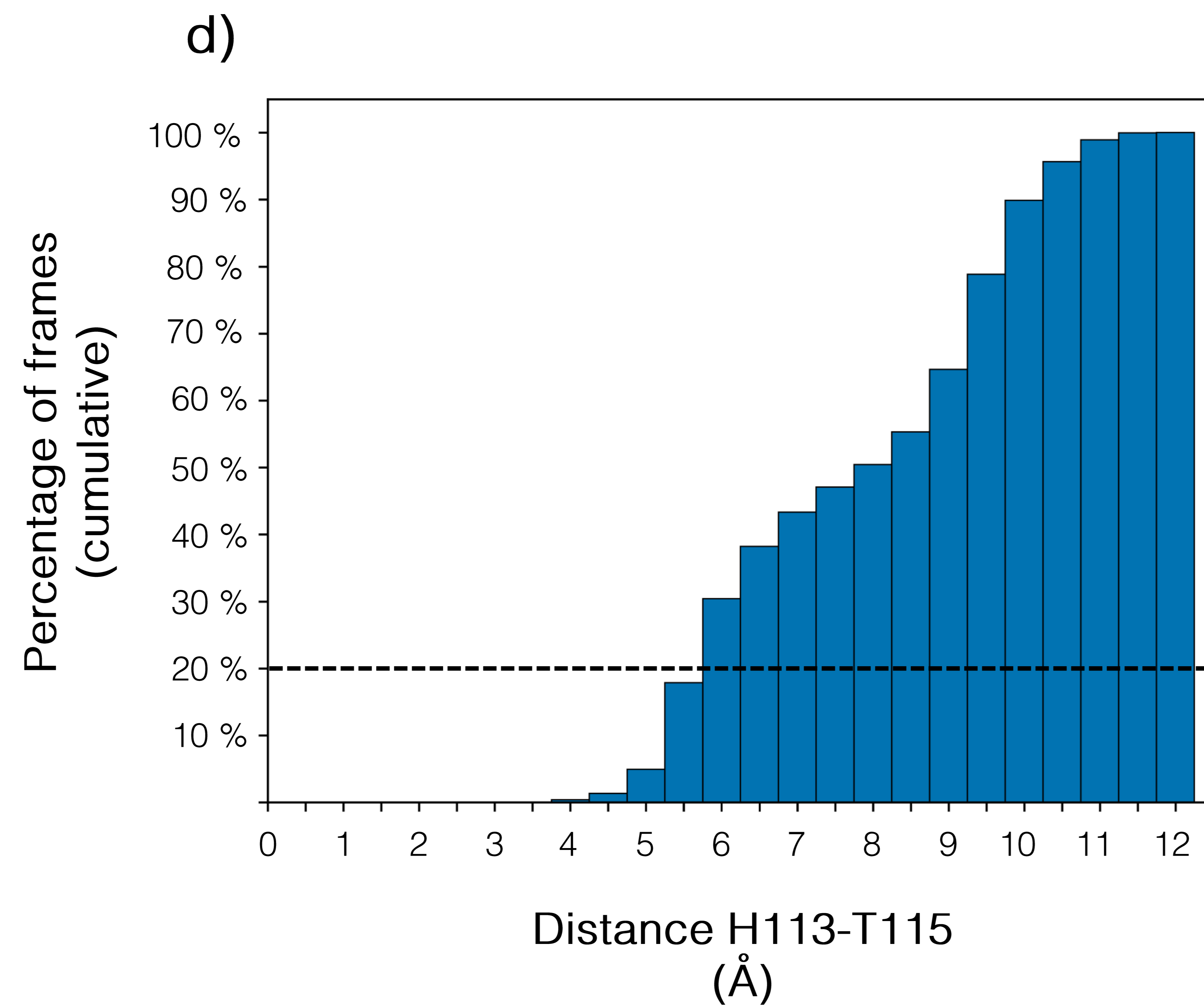
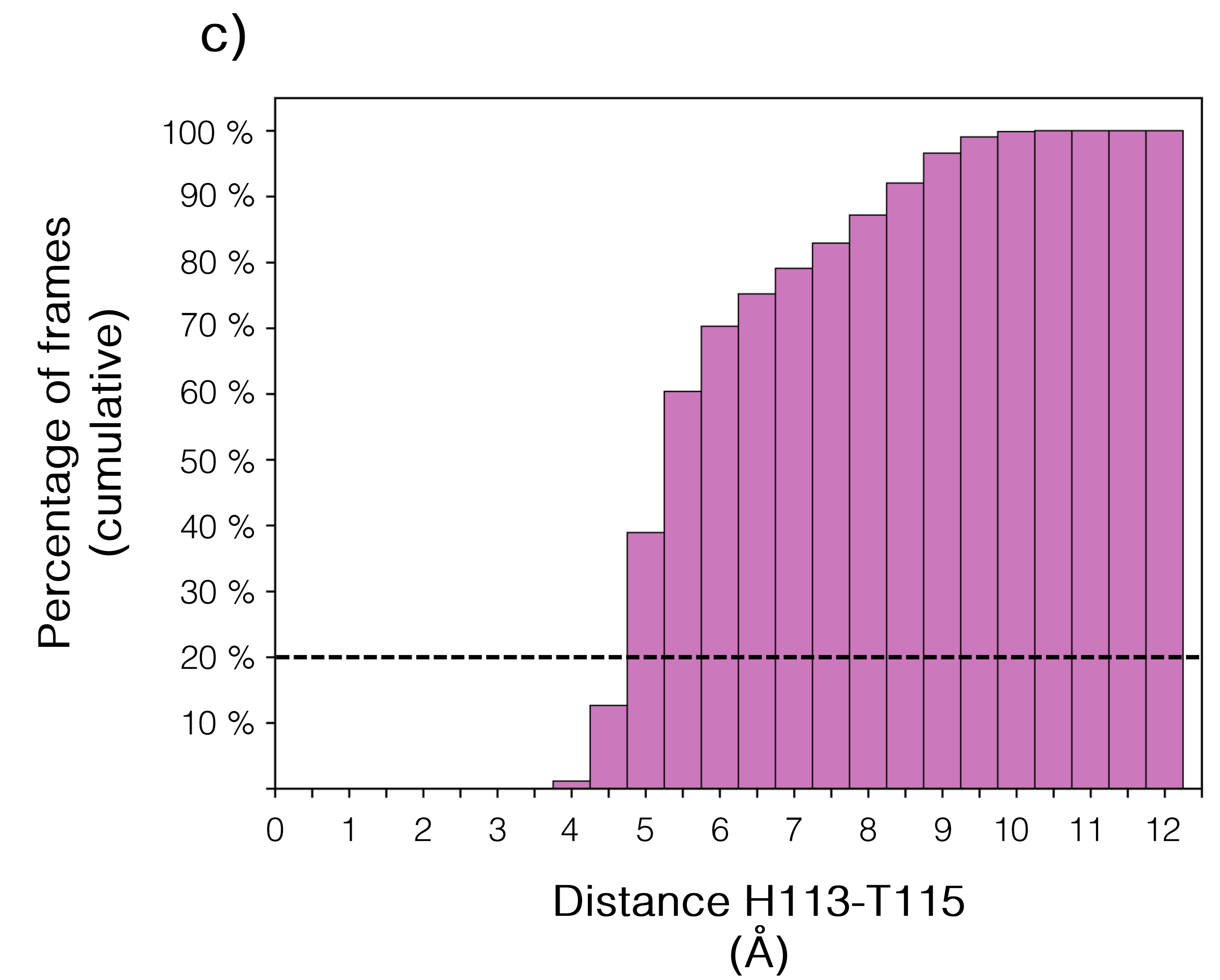
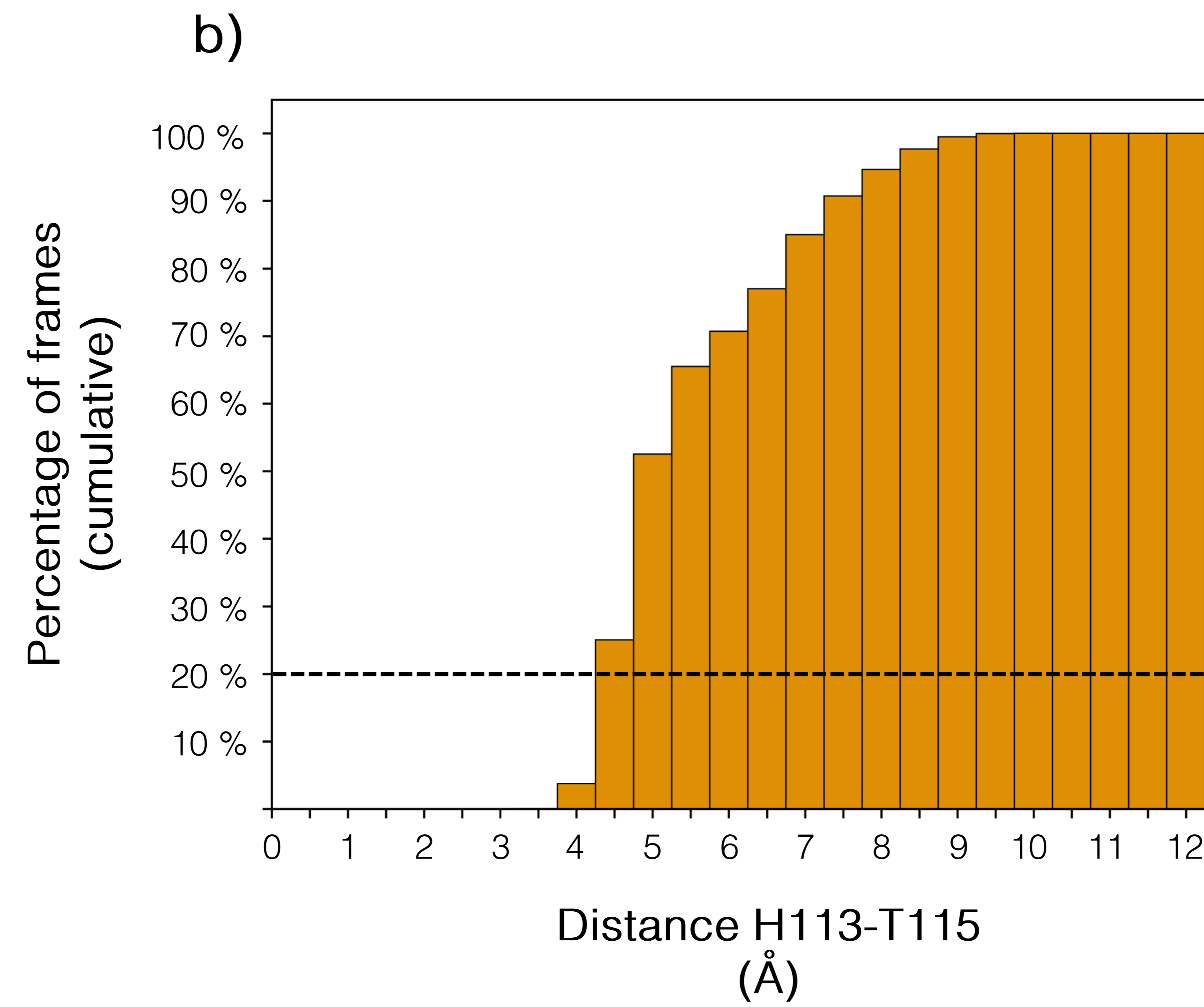
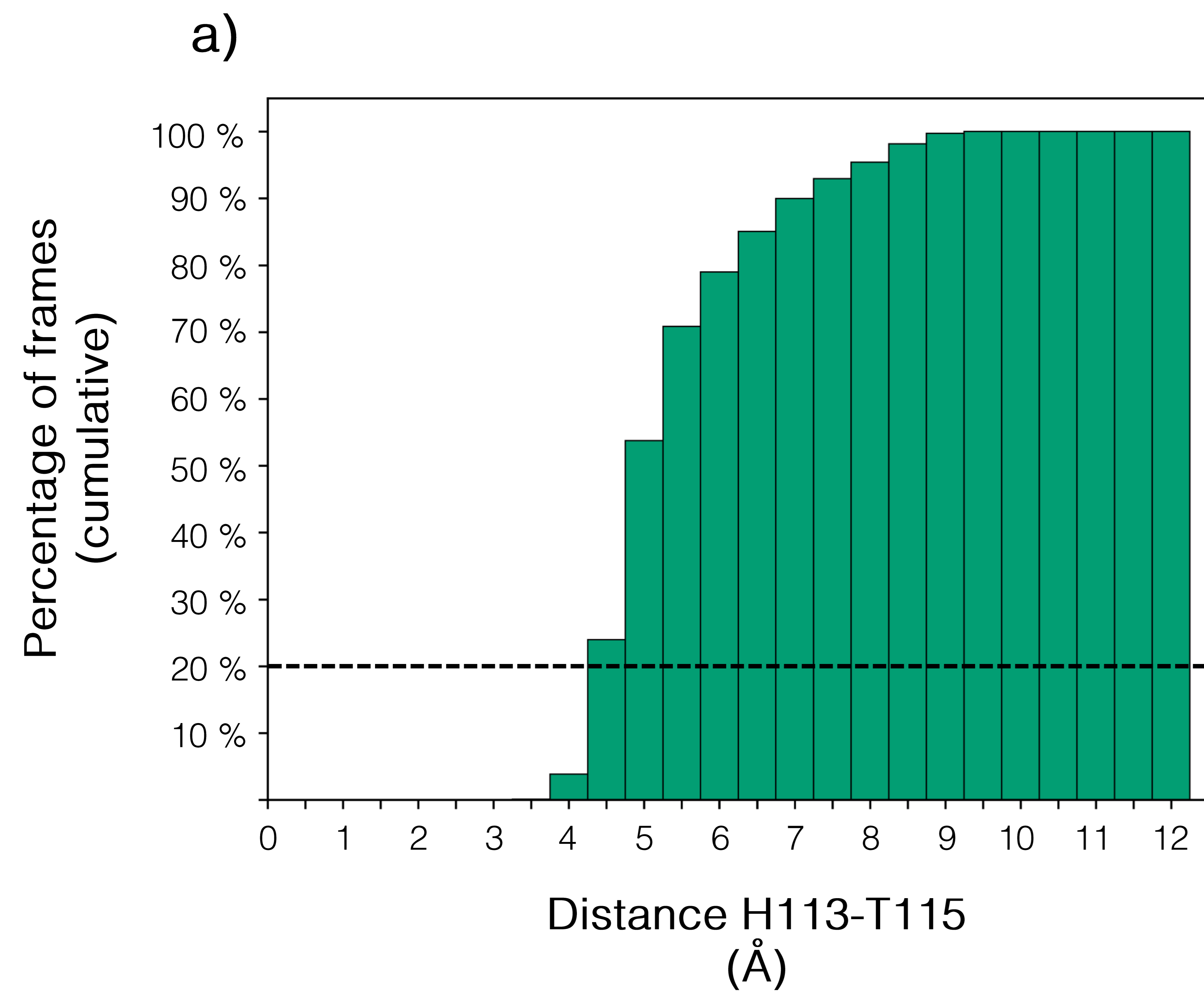
b)



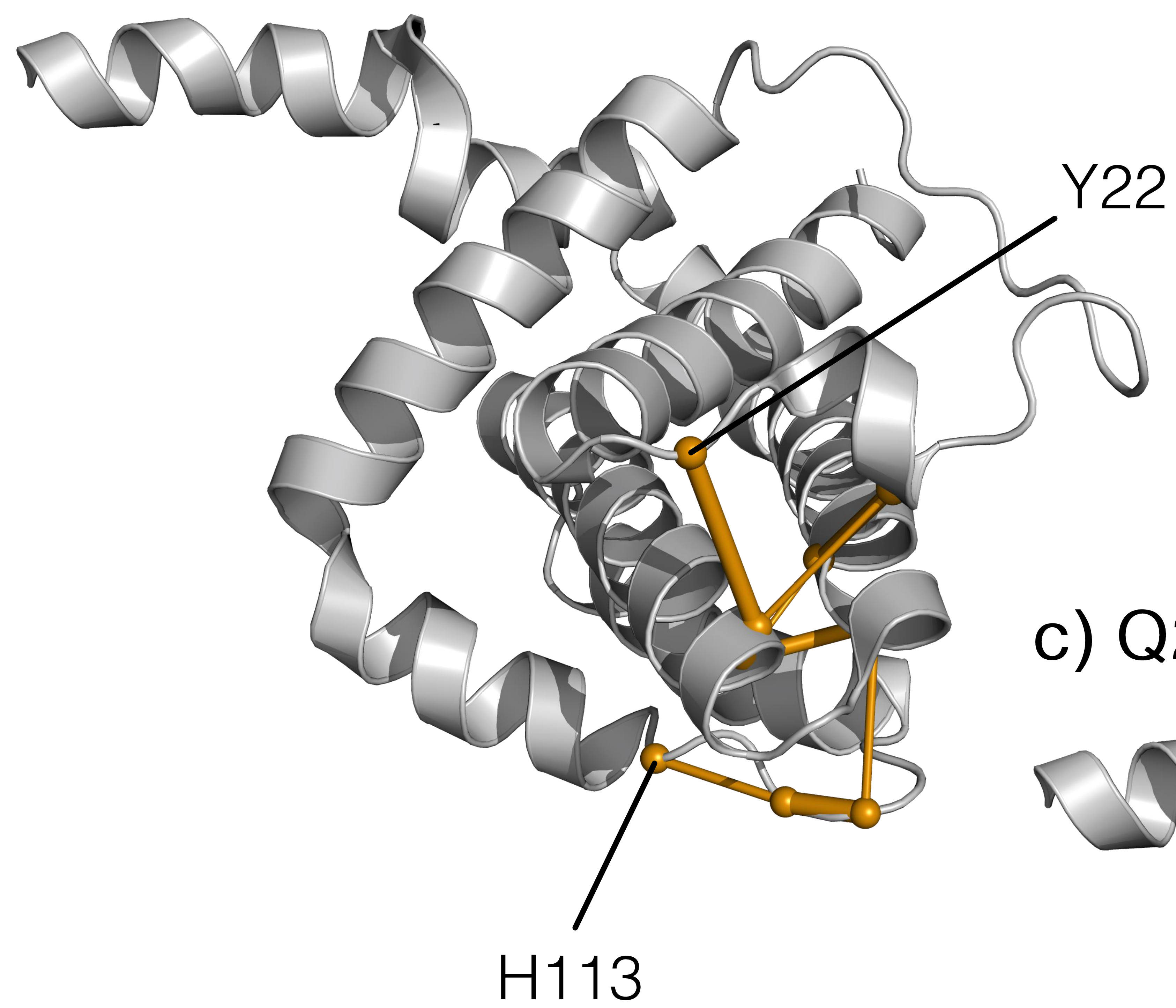


120°

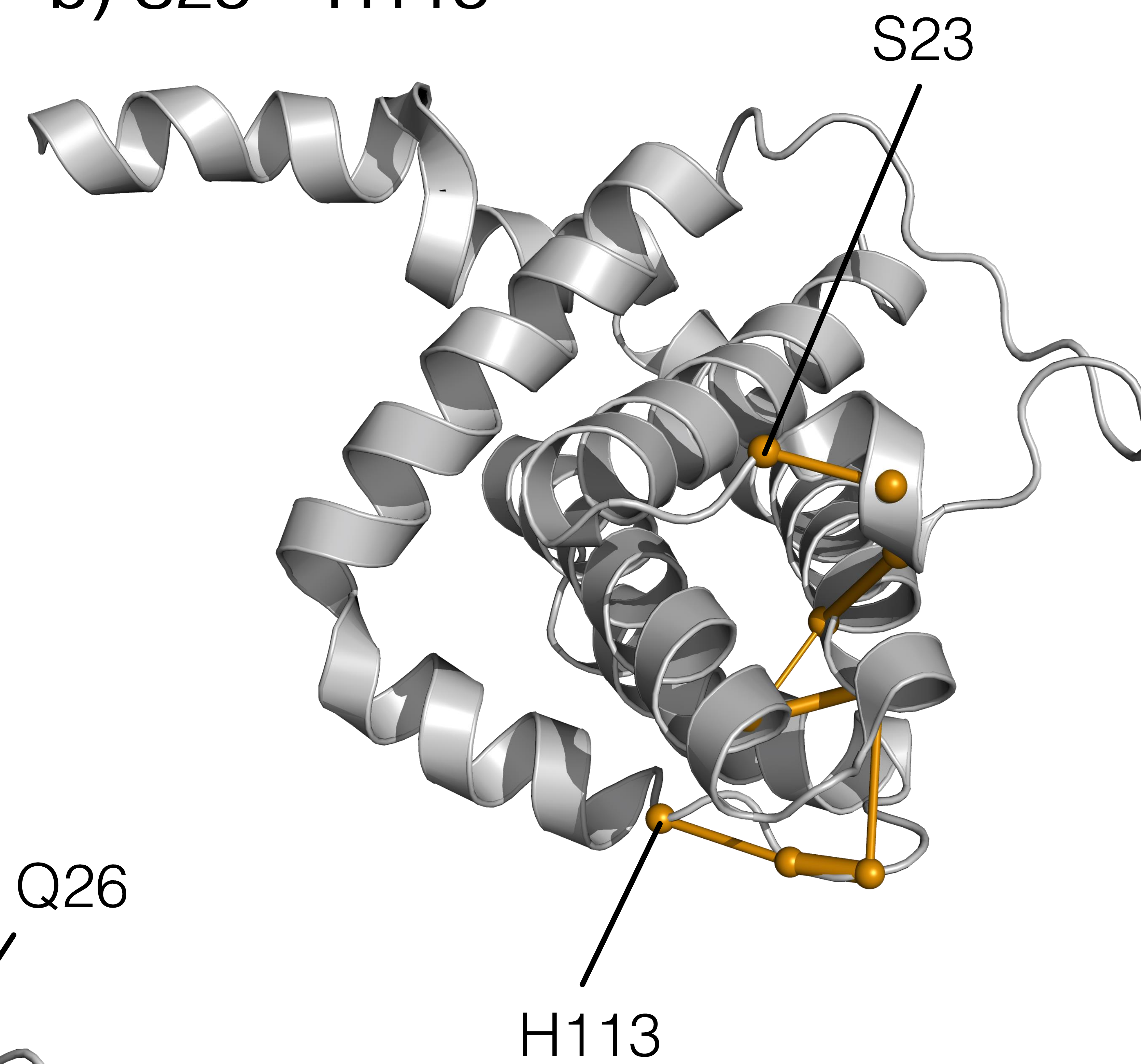




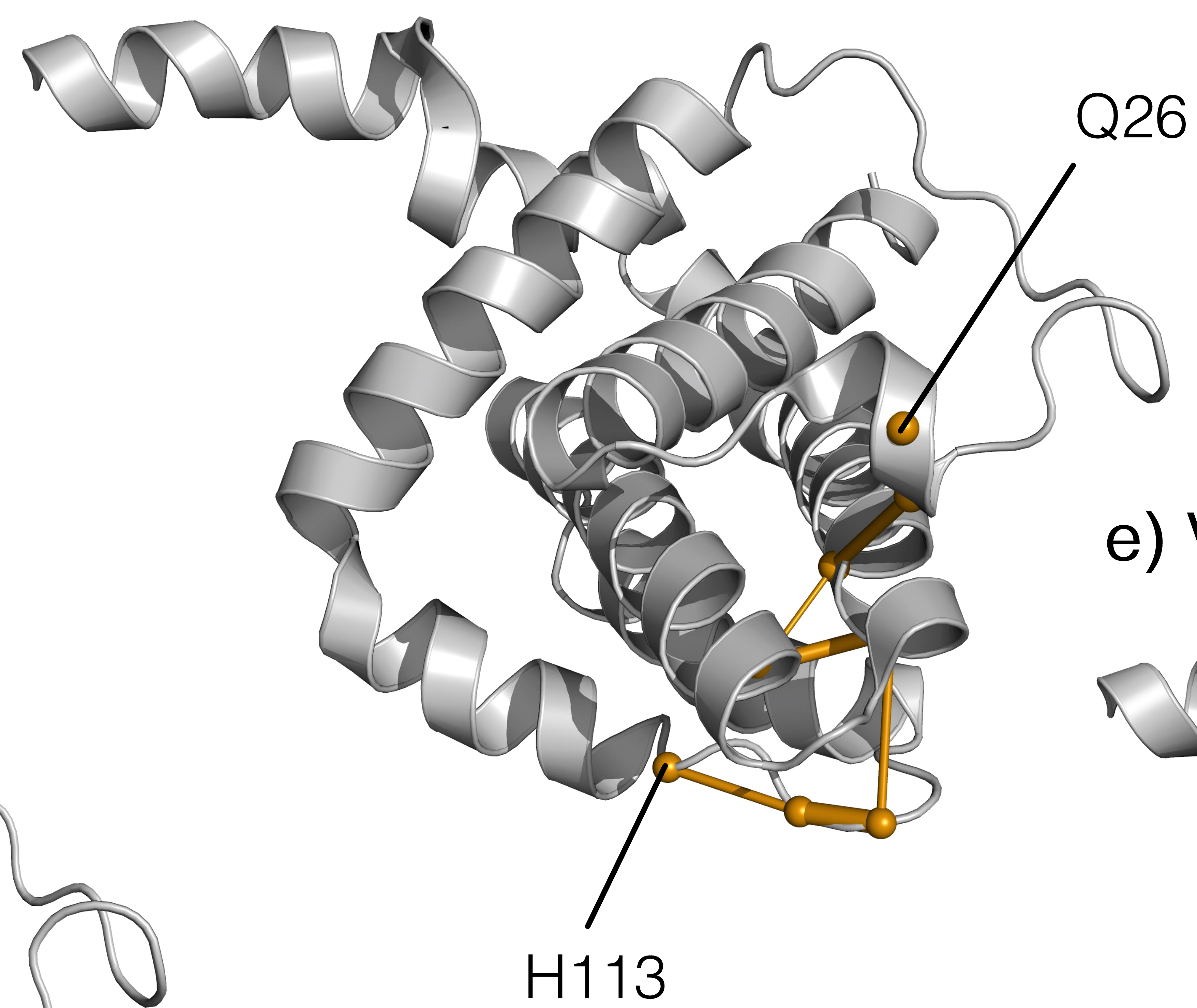
a) Y22 – H113



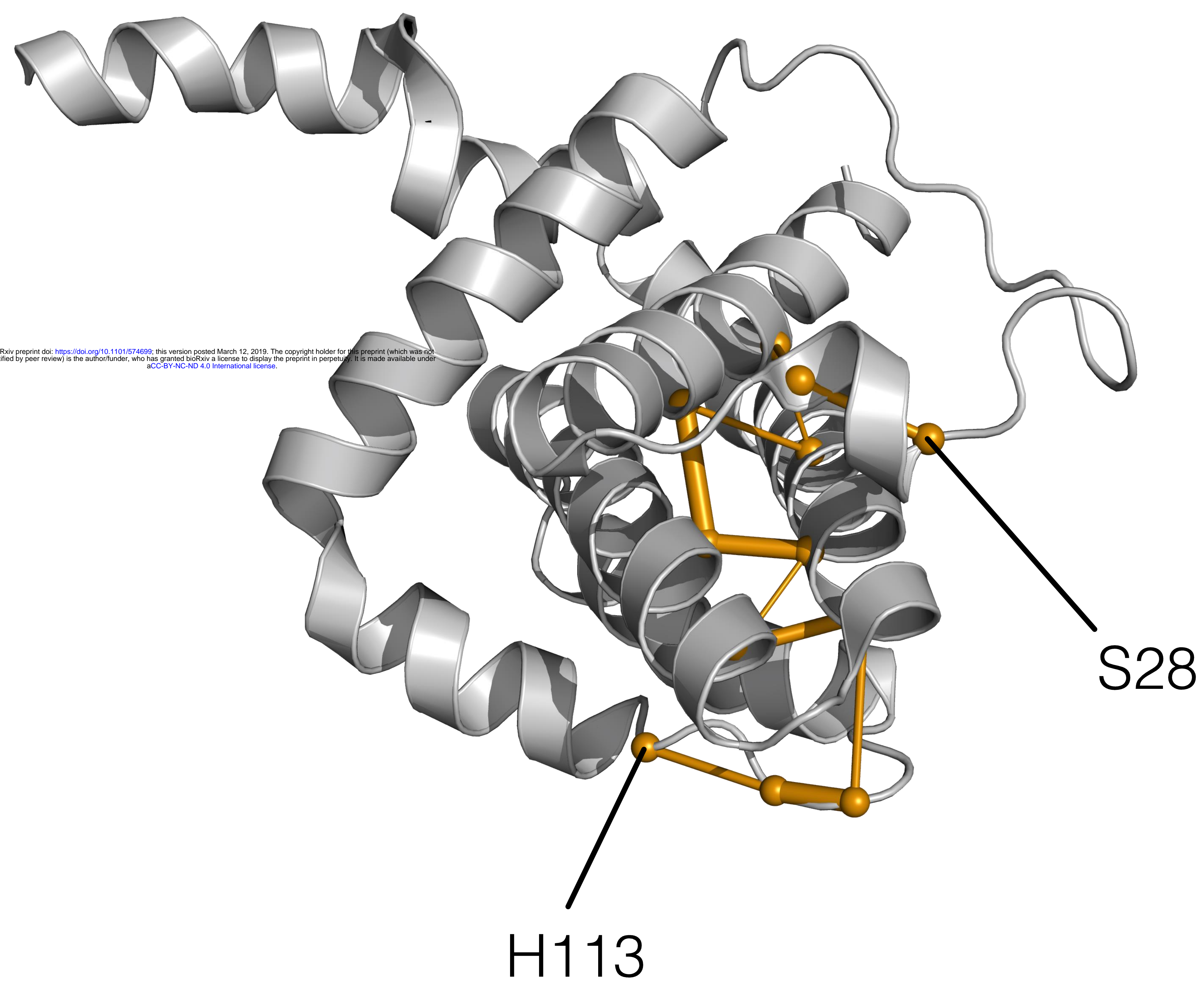
b) S23 – H113



c) Q26 – H113



d) S28 – H113



e) V155 – H113

