Dynamics-function relationship of N-terminal acetyltransferases: the β6β7 loop modulates substrate accessibility to the catalytic site

Angèle ABBOUD[1,2], Pierre BÉDOUCHA[1,2], Jan BYŠKA[1], Thomas ARNESEN[3,4,5] and Nathalie REUTER[2,6]*

[1]Department of Informatics, University of Bergen, Bergen, Norway;

[2]Computational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway;

[3]Department of Biological Sciences, University of Bergen, Bergen, Norway;

[4]Department of Biomedicine, University of Bergen, Bergen, Norway;

[5]Department of Surgery, Haukeland University Hospital, Bergen, Norway;

[6]Department of Chemistry, University of Bergen, Bergen, Norway

*Corresponding authors: Nathalie Reuter, Phone: +47 55 58 40 40, Email : nathalie.reuter@uib.no

Keywords: Acetylation; N-terminal acetyltransferases; protein dynamics; normal modes analysis; ligand specificity

Abstract

N-terminal acetyltransferases (NATs) are enzymes catalysing the transfer of the acetyl from Ac-CoA to the N-terminus of proteins, one of the most common protein modifications. Unlike NATs, lysine acetyltransferases (KATs) transfer an acetyl onto the amine group of internal lysines. To date, not much is known on the exclusive substrate specificity of NATs towards protein N-termini. All the NATs and some KATs share a common fold called GNAT. The main difference between NATs and KATs is an extra hairpin loop found only in NATs called β6β7 loop. It covers the active site as a lid. The hypothesized role of the loop is that of a barrier restricting the access to the catalytic site and preventing acetylation of internal lysines. We investigated the dynamics-function relationships of all available structures of NATs covering the three domains of life. Using elastic network models and normal mode analysis, we found a common dynamics pattern conserved through the GNAT fold; a rigid V-shaped groove, formed by the β4 and β5 strands and three relatively more dynamic loops α1α2, β3β4 and β6β7. We identified two independent dynamical domains in the GNAT fold, which is split at the β5 strand. We characterized the β6β7 hairpin loop slow dynamics and show that its movements are able to significantly widen the mouth of the ligand binding site thereby influencing its size and shape. Taken together our results show that NATs may have access to a broader ligand specificity range than anticipated.

Author summary

N-terminal acetylation concerns 80% of eukaryotic proteins and is achieved by enzymes called the N-terminal acetyltransferases (NATs). They belong to the large family of acetyltransferases and adopt the GNAT fold. Interestingly most lysine acetyltransferases (KATs), which acetylate specifically internal lysines, share the same fold. Rationale for the ligand recognition by the GNAT enzymes remains unclear. Proteins are dynamic entities that utilize their structural flexibility to carry out functions in living cells. By studying the dynamics throughout the entire NATs family, we found that the slow dynamics of the fold is strongly conserved. We also revealed the mobility of the active site lid, namely the β-hairpin loop β6β7, which is one of the main structural differences between the NATs and the KATs. The size and shape of the ligand binding site depend on movements of that β-

hairpin loop. We suggest that in attempts of mapping NATs specificity or ligand design the fold flexibility should be taken into consideration.

Introduction

Acetyltransferases are enzymes catalysing the transfer of an acetyl group from the co-factor acetyl-coenzyme A (Ac-CoA) to a substrate. Among them, lysine acetyltransferases (KATs) and Nα-terminal acetyltransferases (NATs) perform protein acetylation to either lysine side chains or N-termini of polypeptide chains, respectively. NATs acetylate 80 to 90% of the proteins of the human proteome [1] and N-terminal acetylation has been shown to play a role in various biological processes from protein folding to gene regulation [2]. Dysregulation or mutations of NATs have been linked to several diseases including tumour development [2–5] and initiatives are already undertaken to develop inhibitors targeting the relevant NATs [6].

Most acetyltransferases share the GNAT fold (Gcn5-related N-acetyltransferases) [7]. It consists of a three-layered αβα sandwich containing seven β-sheets and four α-helices (Fig. 1). The GNAT fold displays two features that are conserved through most of the KATs and NATs and are related to the transfer of an acetyl to an amino group. The first is a conserved sequence motif essential for Ac-CoA binding (Q/RxxGxG/A) and located on the turn between strand β4 and helix α3 [7–9]. The β4 strand together with β4-α3 and α4 form much of the Ac-CoA binding site. The second salient feature of the GNAT fold is the V-shaped configuration of the two parallel strands β4 and β5, forming a groove where the extremities of the Ac-CoA and of the substrate peptide meet, positioning the acetyl group and the amino group close enough for the catalytic reaction to occur [7].

Fig 1. GNAT fold and substrate binding sites in NAT catalytic domains. (A) Cartoon representation of the GNAT fold of human Naa50 (PDB ID: **3TFY**). It consists of 6 β-sheets (green) and 4 α-helices (salmon pink) organised in the following topology: B1-H1-H2-B2-B3-B4-H3-B5-H4-B6. The Ac-CoA represented in sticks with the backbone in grey sits between helix α4 and the β4α3 loop. (B) Human Naa50 (cartoons and blue solvent-accessible surface) bound to substrate Met-Leu-Gly (van der Waals spheres) (PDB ID:

3TFY). (C) Naa60 (grey solvent-accessible surface) bound to a bisubstrate CoA-Ac-Met-Lys-Ala-Val. Loops β6β7 and α1α2 are labelled in (B) and (C).

Substrate specificity varies drastically within the NATs, offering a palette of enzymes able to target a large spectrum of N-terminal sequences. As of now, eight NATs (NatA-NatH) have been identified in eukaryotes [2,10,11], three in prokaryotes (RimI, RimJ and RimL) [12,13] and one in archaea (referred here as ArNat) [14]. NATs are classified based on their composition and substrates. Comparison of structures of NATs resolved by X-ray crystallography and KATs revealed the presence of a seventh β-strand at the C-terminus as the main conserved characteristic of the NATs [15–17]. The extra β-hairpin β6β7, the α1α2 loop and the helix α2 contain amino acids forming the boundary of the binding pocket, which we hereafter refer to as the *mouth* of the binding pocket. NAT substrates usually position two to three residues in the peptide binding site [15–18]. Two conserved tyrosines located on the β6β7 loop and another one on the α1α2 loop have been shown to interact with the substrate backbone via hydrogen bonds [19]. Both loops cover the groove containing the catalytic site (Fig. 1B) to which substrate peptide and Ac-CoA bind from opposite sides. The loops have been proposed to prevent the access of internal lysine to the catalytic site and as a consequence prevent their acetylation by NATs [15,20]. However, there have been reports of lysine acetylations by NATs [18,21–27]. Moreover, NATs can be inhibited by so-called bisubstrate inhibitors consisting of a short polypeptide covalently bound to the Ac-CoA [6,17,28,29]. The X-ray structure of the human NatF bound to bisubstrate CoA-Ac-MKAV$_7$ shows that the inhibitor is placed in the Ac-CoA and substrate binding site with the β6β7 hairpin loop hanging over the top of it [28] (Fig. 1C). This structure raises the question of how the bisubstrate accesses both the ligand and Ac-CoA binding sites and as a consequence suggests the mouth of the active site needs to be able to open up.

Using relatively short 10 nanoseconds-long molecular dynamics simulations of the catalytic domains of two human NATs, we earlier showed for hNaa50 that (i) helix α2 undergoes flexibility changes upon ligand binding and (ii) the β6β7 hairpin loop was the most flexible region of the protein [19]. Likewise, we observed the β6β7 loop of the human Naa10 to be flexible [30] in 100 ns-long molecular simulations. The time-scale of these simulations did not allow us to investigate conformational changes happening on longer time-scales, such as displacements of loops. Enzyme dynamics is known to be important for their function.

Notably, catalytic residues are reported to be placed at rigid positions of the protein structure, while amino acids involved in substrate binding are likely to have a higher flexibility [31–33]. Furthermore, functionally relevant flexibility is conserved between enzymes sharing the same fold. Normal mode analysis (NMA) using elastic network models (ENM) is an efficient computational method that has proven reliable to characterize the flexibility properties intrinsic to protein structures [34–40]. It has also been successfully used to conduct comparative analyses of multiple protein structures [31,34].

In this study, we characterize the dynamics of all available structures of NATs and compare them to uncover dynamics patterns intrinsic to the GNAT fold and relevant for function.

Results

Our dataset consists of 34 structures of 15 distinct proteins listed in Table 1 and is representing ten types of NATs defined according to their composition and substrates (S1 and S2 Tables). The dataset spans the three domains of Life, where six out of eight NATs stem from eukaryotes (NatA, NatB, NatD, NatE, NatF, NatH), three from bacteria (RimI, RimJ and RimL) and one from archaea (ArNat). In what follows, the ten types of NATs found in our dataset are referred to by the name of the catalytic subunit of each NAT complex (Naa10, Naa20, etc… reported in Table 1 in the column titled "Group").

Table 1. N-terminal acetyltransferases (NATs) found in PROSITE, Uniprot and PDB databases.

Group	Super Kingdom /	E.C	Uniprot	PDB ID	Chain
ArNat	Archaea	2.3.1.-			
	Sulfolubus		Q980R9	**2x7b**[41], 4lx9[42], 4r3k[43], 4r3l[43], 5c88[44]	A, A, A, A, A
	Thermoplasma		Q97CT7	**4pv6**[45]	A
Naa10	Eukaryota	2.3.1.255			
	Schizosaccharomyces		Q9UTI3	**4kvm**[17], 4kvo[17], 4kvx[17]	E, E, B
	Saccharomyces		P07347	4xnh, 4xpd, 4y49, 4hnw, 4hnx, 4hny	B, B, B, B, B, B
Naa20	Eukaryota	2.3.1.254			
	Candida		C4YDZ9	5k04[46], **5k18**[46]	B, B
Naa40	Eukaryota	2.3.1.257			
	Schizosaccharomyces		Q9USH6	**4ua3**[47]	A

	Homo sapiens		Q86UY6	**4u9v**[47], 4u9w[47]	A, B
Naa50	Eukaryota	2.3.1.258			
	Homo sapiens		Q9GZZ1	2ob0, 2psw, **3tfy**[15], 4x5k	A, A, C, A
Naa60	Eukaryota	2.3.1.259			
	Homo sapiens		Q9H7X0	5hgz[18], **5hh0**[18], 5hh1[18], **5icv**[28], 5icw[28]	A, A, A, B, A
Naa80	Eukaryota	-			
	Drosophila		Q59DX8	**5wjd**[48], 5wje[48]	
RimI	Bacteria	2.3.1.128			
	Salmonella		Q8ZJW4	2cnm[29], **2cns**[29], 2cnt[29]	A, A, A
	Escherichia		P0A946	**5isv**	A
RimJ	Bacteria	2.3.1.128			
	Aliivibrio		Q5DZH6	**3igr**	A
RimL	Bacteria	2.3.1.128			
	Salmonella		Q8ZPC0	1s7f[49], 1s7k[49], **1s7l**[49], 1s7n[49], 1z9u	A, A, A, A, B
	Thermus		Q5SHD1	**2z0z**[50], 2zxv[50]	A, A
			Q72HN8	2z10[50], 2z11[50]	A, A

Representative structures chosen for each Uniprot code are highlighted with a bold PDB ID. Two representatives were selected for the Naa60 group as they have different topologies (5hh0 contains one extra helix at the C-terminus). PDB IDs written in light grey are available structures that are not included in our dataset due to poor quality (see Methods). One chain was selected for the calculations. For information on the dataset preparation the reader is referred to the experimental procedures.

(1) The NATs' functional diversity builds upon the GNAT fold and is fine-tuned by accessory structural elements

We aligned all structures in the dataset using the multiple structure alignment tool MUSTANG [51] (see Experimental procedures). The structure of Naa50 (PDB ID: **3TFY**) was used as reference. The alignment led to 119 C-alpha atoms' positions conserved (Fig. 2A). While sequence similarity between pairs of NATs is relatively low (23% identity on average), the secondary structure elements of the GNAT fold align well (Fig. 2B). The region between the end of strand β4 and helix α4 is the

most conserved sequence-wise. It contains several of the amino acids involved in catalysis and located on strands β4 and β5, as well as the Ac-CoA binding motif (R/QxxGxG/A on β4-α3) (Fig. 2B). Noticeably, position 213 of the alignment (Asn114 in Naa50) is an asparagine conserved through all the 14 NATs, except in Naa80 where it is replaced by an aspartate (Asp 127) (Fig. 2B). This residue sits close to the oxygen of the Acetyl-CoA pantothenic acid for all structures [43]. Structural differences between NATs are restricted primarily to the N- and C-terminal regions, before helix α2 and after strand β6, respectively. The differences stem either from the longer elements of the GNAT domain or from additional accessories. The latter are secondary structure elements that are not part of the GNAT fold such as the N-terminal helix in Naa40, referred to as helix α0 or the sixty-one amino acid-long extra C-terminal helical segment in Naa60. Only the first thirty amino acids are resolved in the X-ray structure (α5 on Fig. 2C) but two helices are predicted from sequence analysis [52].

Fig 2. Structural alignment of Naa representatives. (A) MUSTANG structural alignment of all NAT structures listed in Table 1. The backbone of each structure is represented with lines except for that of the reference structure Naa50 (PDB ID: **3TFY**), which is represented with cartoons. The red beads represent the 116 aligned C-alpha atoms. (B) Multiple sequence alignment resulting from the structural alignment. Naa50 sequence is written with blue fonts, Ac-CoA-binding motifs are highlighted with cyan boxes and residues involved in the catalytic activity with green boxes. Sequences are labelled with the PDB ID from which their secondary structure elements are retrieved. All are shown above the sequence alignment except for *S. cerevisiae* Naa10 due to poor secondary structure annotation (PDB ID: **4Y49**). The image results from the use of ESPript [53]. (C) Cartoon representation of the shared GNAT fold (in grey) and structural variations: helix α0 in Naa40, helix α5 in Naa60, long β6β7 loops in Naa40 and Naa60, long α1α2 loops in Naa40 and RimL, long β3β4 loops in ArNats and Naa60.

We quantified the structural similarity between Naas by calculating pairwise root mean square deviations (RMSD) between the thirty-four structures of the dataset. The clustered values presented on a heatmap reveal two main groups (Fig. 3). As expected, RMSD values between structures belonging to the

same group are below 1Å, which is in agreement with the fact that structures within a group are orthologues and/or structures of the same protein but in different forms (e.g. apo vs. holo) (S1 Table). To evaluate the influence of the redundancy present in the dataset, we performed the analysis on a representative dataset that only consists of one structure per Uniprot accession number (Table 1) and found no difference in the clustering (S2 Fig.).

Fig 3. Heatmap representation of the pairwise Root Mean Square Deviations (RMSD). The dendrogram reflects the hierarchical clustering based on the RMSD values. The heatmap color scale goes from red (0Å < RMSD < 1Å; structural similarity) to yellow (3Å < RMSD < 4Å). Names of enzymes from eukaryotes are highlighted in blue, those of bacteria in purple and green is used for archaeal NATs.

The first group consists of: ArNats, Naa10, Naa20, Naa50, Naa60, RimI. In this cluster, the closest structures are the archaeal Naas and the eukaryotic Naa10 with RMSD values of 1.2 Å. The RMSDs between ArNats (e.g. 4lx9) and Naa10, Naa50, Naa60 and the RimI are lower than 2Å. The structural proximity of the archaeal Naa with enzymes belonging to other groups is in agreement with what is known about its substrate specificity. The archaeal Naa from *Sulfolobus* uses two different catalytic strategies; like Naa10 enzymes it can acetylate serines or it can acetylate methionines like the Naa50 enzymes (S2 Table). Mutations of key residues from the α1α2 loop were shown to shift the substrate specificity from small amino acids to methionines [42]. In this study, Liszczak et al. suggested these mutations as part of a model of the evolution of a eukaryotic ancestor to a more diverse family with different substrate specificity. The second main group consists of three clusters Naa40, Naa80 and the bacterial RimJ and RimL, which appear to be the most structurally distant from other structures in the dataset with RMSD values between 2.5 and 3.9 Å. They are composed of longer elements in the GNAT fold that influence the orientation of the secondary structure moving them further away from the other NATs. As shown in Fig. 2B the entire region from α1 to α2 is longer in RimL than in other Naas (6, 4 and 7 additional residues for helices α1, α2 and α1α2 loop, respectively). Naa40 also has an extended α1 helix of eight amino acids and an extra N-terminal

helix α0 consisting of 17 amino acids. This α0 helix sits under the GNAT fold and changes the topology of the region β1-α2. The α1α2 loop and the longer α1 helix cover the active site and the β6β7 hairpin loop is flanked away from the active site (Fig. 2C). The structure of the β6-β7 region in Naa80 is different from that of typical NATs. It has a shorter β6-strand, which leads to a different orientation of the β6β7 loop and a ligand binding site opened more widely than in the other NATs [48].

(2) Measure of pairwise flexibility similarity between NATs yield a grouping that coincides with similarity in substrate selectivity

The Bhattacharyya score (BC score) quantifies the intrinsic dynamics (dis)similarity between each pair of aligned cores of proteins in a dataset [54]. To this goal we perform normal mode analyses using elastic network models for each of the structures present in our dataset. We then calculated and compared the BC scores of the representative dataset (Table 1). A heatmap representation of those values together with a dendogram representing the clustering is shown in Fig. 4A. We observe that the structures are sorted in three main groups containing (1) Naa10, Naa20, Naa50, Naa60, archaeal NATs, bacterial RimI; (2) bacterial RimJ and RimL, as well as Naa40 which is the only eukaryotic NAT in this group and (3) Naa80. Interestingly, the first group is sorted between the Naas acetylating methionine: ArNats, Naa50 and Naa60 and the one acetylating small residues or alanine exclusively: Naa10 and RimI, respectively (S2 Table). Naa20 acetylates methionines followed by acidic residues and is clustered with Naa10. The latter has also been shown to shift substrate specificity towards acidic residues in its uncomplexed form [55]. Naa40 is also one of the most selective Naas since it acetylates only the Serine of the N-terminal of histones H4 and H2A. The bacterial Rims also have a narrow specificity and acetylate only ribosomal proteins (S2 Table). Naa80 is sharing the lowest BC scores with the other NATs. In addition of having a substrate-binding site wider than the other NATs, it also has a restricted substrate specificity towards the N-terminus of actin.

The BC-clusters show that flexibility (dis)similarities are encrypted in the coarse-description of the structure (actually restricted to its backbone) and in the harmonic potential of the ENM model (directly related to the proteins packing density).

Fig 4. **Comparison of the dynamics of the NATs using the Bhattacharyya score (BC score) on a non-redundant dataset.** (A) Heatmap representation of the pairwise BC score between the representative structures (see Experimental procedures). The color scale of the BC score goes from red for high similarity in intrinsic dynamics to yellow for higher dissimilarity. The dendogram is the hierarchical clustering of the pairwise BC score. The names of the groups are written in boxes colored in blue for eukaryotes, purple for bacterial and green for archaeal NATs. (B) Cartoon representation of the structures aligned and used to calculate the pairwise BC score. The helices α1 and α2 are colored according to the cluster they belong to (see color of boxes on the axes of the dendrogram). The first cluster composed of the archaeal Naa, Naa10, Naa20, Naa50, the Naa60 and the RimI (colored in dark blue) shares a shorter helix α1 than that of the second cluster consisting of Naa40 (colored in cyan), RimJ and RimL (colored in orange). Naa80 (colored in magenta) separates from the others; it has the shortest α1α2 loop and the widest binding site of all NATs.

(3) ENM-NMA reveals the signature flexibility pattern of the GNAT fold

Similarity in fold or topology is generally associated with similarity in flexibility and dynamics [33,56]. We here intend to characterize the flexibility intrinsic to the topology of the GNAT fold. We then compare the results from ENM-NMA for each structure to reveal flexibility patterns intrinsic to the common fold. We first compute the normalised fluctuations for each amino acid in each structure, and the correlations between pairs of amino acids (Fig. 5A). The latter reveals how local motions are coupled across different regions of the fold.

Fig 5. Normalized fluctuations and correlations. (A) Aligned normalized fluctuations for all NAT structures. For clarity one color is used for structures belong to the same group. The fluctuations depicted in plain lines are for positions where all structures align, while the dotted lines reflect the fluctuations of positions where not all NATs align. The secondary structures aligned in than 80% of all NATs are shown using green bars for the B-strands and red bars for the helices. (B) The correlation map of Naa50 (PDB ID:

3TFY) shows the correlation patterns found in all maps (shown in Supporting Information Figure S3). Long-range correlations are found within two blocks highlighted within the green and pink frames. For all NATs the highest correlations are found within these blocks and not in-between. (C) Schematic representation of the main fluctuations and correlations on a cartoon representation of Naa50 (PDB ID: **3TFY**). The two blocks of correlations are distinguished by the color of their surface (green and pink). Regions with highest fluctuations are shown in red and those with lowest fluctuations in yellow.

Ac-CoA and the peptide substrate do not affect fluctuations significantly. X-ray structures in our dataset have been solved in different forms including with or without Ac-CoA and/or peptide substrate. For the sake of consistency, we should perform all computations using an ENM of the enzymes only. We thus first evaluate the effect of removing the Ac-CoA and peptide substrate from X-ray structure on the computed flexibility of the fold. We chose to compute the modes and normalized atomic fluctuations for the yeast Naa10 as crystal structures of different forms of the enzyme are available, notably two slightly different conformations, structures including Ac-CoA as well as structures including a bisubstrate inhibitor that mimics the presence of both Ac-CoA and peptide substrate [17]. Ac-CoA is included in the ENM as 11 beads chosen to represent an atom every 4 Å (see Experimental procedures) Normalized fluctuations are shown in S1 Fig. The positions of the minima and maxima are not affected by removing the Ac-CoA or substrate. When Ac-CoA or the bisubstrate are removed from the three structures (PDB IDs: **4KVO**, **4KVX** and **4KVM**) the mobility of β4α3 and of the N-terminal end of helix α4 increases. Further, removing the bisubstrate from 4kvm also influences the fluctuations of loop β6β7 since it lies close to it. Generally, fluctuations are affected locally (i.e. at the Ac-CoA and ligand-binding sites) but not to a significant extent compared to the differences observed between NATs (Fig. 5A).

The V-shaped β-strands characteristic of the GNAT fold form a rigid core of the enzyme, while loops surrounding the binding site are mobile. The normalized fluctuations are plotted on Fig. 5A. We observe that the β-strands are the most rigid elements in all structures with β4 and β5 strands having the lowest

fluctuations. Interestingly these two strands carry most of the catalytic amino acids (Fig. 2B) and our observation matches earlier reports of catalytic residues being positioned at particularly rigid points in protein structures [32,33,57–59]. Furthermore, β4 and β5 are not assembling into a sheet along their whole length despite their proximity. Instead they form a "V shape" splitting the seven-stranded beta-sheet and creating a crevice where Ac-CoA and peptide substrate meet (Fig. 1).

In the region of the helices α1 and α2, we notice a similar pattern of flexibility between all the structures where the loop α1α2 and the helix α2 fluctuate more. Molecular dynamics simulations of the human Naa50 and Naa10 have shown that the flexibility of helix α2 is decreased in the presence of a substrate [19]. This region is also involved in the complex formation with the subunit Naa15 [30]. We observed high fluctuations for long unfolded N- and C-terminal ends. Besides those regions where fluctuations cannot be calculated reliably, the highest fluctuations are observed for the β3β4 and β6β7 loops. Two tyrosines located at positions 239 and 240 of the alignment (Y138 and Y139 in Naa50), and conserved across several groups (not in RimL, Naa40 and Naa80), are located on loop β6β7. They are known to be involved in substrate binding and form hydrogen bonds with the substrate [19].

The GNAT fold is divided in two dynamical domains on either sides of the β5 strand. We calculated the correlations for each of the representative structures. The representative correlation map is shown as a heatmap in Fig. 5B. All NATs share the same pattern consisting of two blocks with relatively little correlations between them (S3 Fig.) indicating that the proteins contains two dynamical domains [60]. The boundary between the two coincides with the V-shape split between β4 and β5; the first domain starts at sheet β1 and ends before strands β5, and the second domain starts at strand β5 and ends with strand β7. Within each domain, pairs of neighbouring β-strands are strongly correlated, as expected for beta-strands involved in the same sheet [61], but to a lesser extent for β4-β5. This is explained by the distance between β4 and β5 and the split in the fold at the β4-β5 interface. Correlations between strands and helices are weaker in general and happen through extremities of helices only. This is in agreement with what we observed for enzymes with the TIM barrel fold [33]. β5 shows strong correlations with strands on either side; β4 and β7. Correlations between domains are weak in comparison. There is an anti-correlation of such

type between loops α1α2 and β6β7, which are in contact with each other forming a lid above the catalytic site. The potential flexibility of the loop β6β7 to open and move independently of the other block reinforces the idea that it could create more room for the substrate to enter. Apart from the loop, it seems that the movement of the pink block (Fig. 5C) independently from the first block will influence the size of the binding pocket.

(4) Structural differences in the N-terminal-α2 region contribute to difference in dynamics between NATs

Differences in intrinsic dynamics are encoded in the structure. The GNAT fold has a region of high variability from the N-terminal to the α2 helix [7] (Fig. 2C and 4B). As noted earlier the helices α1 and α2 are longer in Naa40, RimJ and RimL than in the other NATs from group 1. RimJ and L have 1.5 additional turns in each of the two helices compared to structures clustered in group 1, while the two extra turns of helix α1 in Naa40 brings its C-terminal end over the active site at a location overlapping with that of the β6β7 hairpin loop in the other NATs. As a result β6β7 is protruding further away from the protein core than in the other NATs and shows very large fluctuations as calculated from the modes (Fig. 2C and Fig. 5A). Furthermore Naa40 has an extra N-terminal helix α0, the movements of which are correlated with strands β3 and β4 as well as with loop α3β5 (Fig. 6). On the contrary Naa80 has a shorter α1-α2 loop as well as a different orientation of its β6β7 loop.

Fig 6. Correlations maps for the representative of the Naa40 group. Helix α0, which is not present in other NATs, has high correlations with sheets β3 and β4, and with the α3β5 loop.

The fact that clustering structures according to the pairwise BC score sorts them into groups that differ by their N-terminal end to α2 region indicates that the dynamics of this region is significantly different between the three groups.

(5) The mobility of the β6β7 loop influences peptide substrate accessibility to the catalytic site

The low frequency modes yield deformations at the lowest energetic cost and have been shown to be functionally relevant [62,63]. Kurkcuoglu et al. showed that global motions of functional loops were important for fitting and binding of the substrate [64]. Given the high mobility of the β6β7 loop and its position with respect to Ac-CoA and the substrate binding site, it is legitimate to wonder how its movements would influence the fairly narrow tunnel in which substrate and Ac-CoA meet. In order to determine the contributions of loop motions to the accessibility of the active site, we evaluated the changes to the surface area of the tunnel entrance when the structure conformation is modified following the lowest energy normal modes.

Further, we calculated how the surface area changes when the structure is modified following the individual lowest energy normal modes (see Experimental procedures). A reconstitution of the protein side chains from our C-alpha model was necessary to calculate the surface area of the mouth of the tunnel using CAVER Analyst [65] (see Experimental procedures). To this goal we first defined a static clipping plane using three residues (Tyr31, 73 and 138) lining the entrance (see Experimental procedures). We generated a trajectory along the first six lowest-frequency normal modes and held this static clipping plane for each frame of the trajectory. Note that the orientation of the clipping plane was extracted from the conformation of the X-ray structure and the same orientation was used in each frame. We then calculated the value of the surface area throughout the trajectory. We also calculated the surface area at a clipping plane 2Å away from the *mouth*, where the second and third residues of the substrate sit, in order to verify that no amino acid closes the access upwards of the active site.

The lowest energy modes show concerted motions of the three loops β3β4, β6β7 and α1α2. They surround the active site and their movements modify the shape of its *mouth* and its surface area (Fig. 7). However, the motions of each loop seem to have a different effect on the surface area. As illustrated in Fig. 7B slight movements of Tyr138 in Naa50, found on the β6β7 loop, drastically change the opening of the entrance (from 50.5Å to 71.1Å). The β6β7 loop provides a simple steric regulation of the entrance of the binding site due to its position as a lid above the active site.

Fig 7. Influence of the flexibility of loop β6β7 on the surface area of the cavity mouth. Amino acids Y138, Y70 and Y26 line the *mouth* of the cavity in the X-ray structure of Naa50 (PDB ID: **3TFY**). (A) Representation of the low-energy modes affecting the surface area of the *mouth* the most. Naa50 is in cartoon representation with arrows indicating the direction of the movements corresponding. Only the vectors with a minimum length of 4Å are represented, with in red colour the positive factors and in blue the negative ones. The trajectory is 9 frames long, with the X-ray structure coordinates found in the middle (frame 5). Four frames are used to represent each direction of the vectors (from frame 1 to 4 for the positive vectors and from frame 6 to 9, for the negative vectors). The contours in cyan shades represent the evolution of the mouth surface area throughout the trajectory. The bold contour is obtained from the X-ray structure. The surface areas are also calculated two angstroms before the mouth of the cavity (contour in pink shades) to ensure that no amino acid comes closing the mouth. (B) Close up of the red square found on the cartoon representation of the mode 10 in (A). Clip plane of the protein at the mouth area defined by the three tyrosines (shown in pink spheres). The red surface represents the rest of the tunnel calculated by CAVER Analyst where the N-terminal and the Ac-COA meets (see Experimental procedures). The contours are the surface areas of the mouth depicted in cyan shades in (A). The scale indicates the range change in size of the surface throughout the trajectory. The bold contour is the contour in the X-ray structure and has a size of 50,5Å. Small movements of the Y138 in the low-energy mode 10 impact greatly the surface area of the entrance of the binding site (from 2.5 to 70Å).

Discussion

The variety of NATs enables the acetylation of a wide variety of N-termini of proteins at different localizations in the cell. The GNAT fold shared among all NATs and with other acetyltransferases such as KATs offers a common scaffold to perform the catalytic activity. One particular feature of the GNAT fold in the NATs is a tight entrance of the binding site, where the β6β7 hairpin loop shields the active site and

forms a tunnel together with the α1-α2 region (except in Naa80). The loop seems to enable the entrance of only N-terminal amino acids (Fig. 1C). However, it remains unclear how substrates are able to enter through the tight *mouth* of the tunnel. Furthermore the path through which bisubstrate inhibitors bind remains uncertain. Using ENM-NMA we have compared the intrinsic dynamics of the GNAT fold in all available structures of NATs covering all domains of Life. We have shown that i) there is a dynamic pattern in the fold common to all NATs, ii) the β6β7 loop is highly mobile and iii) intrinsic dynamics of the loop regulates the accessibility of the binding site.

We identified a common pattern in the intrinsic dynamics of the GNAT fold, where the β-sheet core is more rigid than the rest of the structure. The two strands, β4 and β5, are the least flexible elements of the β-sheets. Interestingly, the β4 and β5 strands carry residues involved in the proton wire essential for the transfer of the acetyl. The rigidity of the catalytic core in the GNAT fold is in agreement with case-studies of enzyme dynamics where residues involved in catalysis are found to be placed at rigid conserved positions of the fold thus stabilizing amino acids essential for the function [33,56,58].

The catalytic core is at the crossroads of two dynamic domains that move independently from each other. The region between the C-terminal end of α3 and the N-terminal end of β5 is found to have hinge residues that are important pivots in larger motions of the protein. The pathological Naa10 p.V107F mutation at the N-terminal end of β5 causes a 95% reduction of the catalytic activity compared to Naa10 WT [66]. In this study, Popp et al. built a homology model of the mutant V107F and observed a disruption of hydrophobic contacts with the Met98 found on helix α3 of the human Naa10. This correlates with our results and shows that perturbation of the packing density of hinges region affect function through a modification of the intrinsic structural flexibility [67].

The most mobile regions in NATs are the functional loops α1α2, β3β4 and β6β7. Substrate-binding residues have a tendency to be in more flexible regions than catalytic residues [56,68,69] and the α1α2 and β6β7 loops carry residues involved in substrate binding [19]. The C-terminal end of helix α2 has been identified as carrying a mutation (p. S37P) in the human Naa10 that is causative of the lethal Ogden syndrome [4]. Using molecular dynamics simulations on the model of the human NatA complex, we have

earlier shown that this mutation decreased the fluctuations of the α1α2 loop and of the α1 helix. Moreover the mutation impairs the catalytic activity and the formation of the NatA complex, inducing a reduction of NatA-mediated N-terminal acetylation and affecting cell proliferation [4]. The α1 and α2 helices are also part of the binding interface with Naa15, the auxiliary subunit of the NatA complex [17] (S2 Table). The loop β3β4 is 10 to 15 residues longer in Naa60 than in the other NATs and mutations on key residues disrupt interactions with the β5, β6, and β7 strands, leading to altered catalytic efficiency and protein stability [70]. Finally the β6β7 loop, one of the most flexible loops, contains two tyrosines conserved through most of the NATs, which make hydrogen bonds with the backbone of the first and second amino acids of the substrate [19].

The highly mobile α1α2 and β6β7 loops flank the catalytic core regulating the size of the binding site. Moreover the regions from the N-terminal to α2 and β6-β7 have structural differences among the NATs. The BC score, which quantifies similarities of intrinsic dynamics between the aligned core regions of proteins, highlighted dissimilarities between Naa40, the bacterial RimL and RimJ in one hand, and the rest of the NATs in the other hand; while Naa80 clusters on its own. Interestingly, Naa40, RimL and RimJ have longer α1 and α2 helices and a longer α1α2 loop, which results in slight changes of the active site shape. As far as we are aware of, Naa40 is the only known NAT with a different position of the substrate in the active site. As shown in Magin et al., all the substrates in other structures have their 2nd and 3rd residues sitting close to the α1 and α2 helices, while the α2 helix in Naa40 obstructs this region shifting the substrate towards the β5 and β7 strands [47]. The bacterial RimL is only active as a homodimer unlike the other NATs, which tend to be active as monomers complexed with auxiliary subunits forming heterodimers or heterotrimers (S2 Table). The β6β7 loop is part of the RimL dimerization interface. Helix α2 is tilted away from β4 yielding a larger mouth of the cavity compared to the other NATs. The longer elements in Naa40, RimJ and RimL thus illustrate how secondary structure elements lining the binding site affect its size, shape, and accessibility. In the case of Naa80, the opening to the binding site is wider than in other NATs and this is thought to play a role in its specificity for the acidic actin N-termini [48].

The collective motions known to support function have been shown to be well described by the lowest energy modes [63,71]. Those modes in NATs contain large movements of the β6β7 hairpin loop with hinges

found at the extremities of the neighbouring strands. We have demonstrated the impact of this motion on the shape and size of the entrance of the substrate binding site using Naa50. It acetylates methionine residues followed by hydrophobic residues [72,73]. As shown in Fig. 1B, the entrance of the active site tightly surrounds its substrate. From the X-ray structure we can measure that the methionine is approximately 6Å long and observe that Val29, Tyr31, Tyr138 and Tyr139 tightly surround it. The movements of Tyr138 alone, which is located on the β6β7 loop, cause a widening of the mouth in the low energy modes. This tyrosine is actually not present in Naa40, and the bacterial RimL where it is replaced by a glycine in Naa40 and asparagine or alanine in RimL. These changes might affect the substrate specificity. In general it has been observed that regions playing a role in ligand-binding have a tendency to explore more of the deformation pattern available than the catalytic site without impairing enzyme function [74].

Taken together, we propose that the dynamics of the two domains and the high mobility of the β6β7 loop give the ligand binding site a flexibility that is important for its selectivity. For the first time, we suggest a possible conformational change that would explain how the long bisubstrates are able to bind both the Ac-CoA and peptide substrate binding sites. Our calculations show that hinges of these movements are located at V131 and A145 in Naa10 (PDB ID: **4KVM**). Further, this loop represents a striking structural difference between NATs and KATs and is hypothesized to prevent the entrance of internal lysines in the active site [15,47,75]. Our calculations show that fairly small displacements of the β6β7 loop as a rigid entity modify the accessibility to the active site and the Ac-CoA. β-hairpin loops are known to have a strong hydrophobic core and interstrand interactions keeping them compact and the turn stable [76]. The human Naa10 has been shown to acetylate internal lysines of various proteins [24,26,27,77] and the auto-acetylation on its K136 found on the β6β7 loop could be the reason of its shift of substrate specificity towards internal lysine [78]. Acetylation of unexpected substrates might be enabled by an increased mobility of loop β6β7 triggered by either a particular substrate or experimental conditions.

The importance of the dynamics-function relationship of the GNAT fold in the NATs is undeniable. Our work fills a gap in the understanding of the versatility and broad substrate specificity of the NATs enzymes [4,19]. Our results are relevant for those seeking to design inhibitors of NATs involved in cancer, Huntington's disease or other pathologies. In addition, we reveal the role played by the flexibility of the

β6β7 hairpin loop in the opening of the ligand binding site. Further investigations are needed to experimentally evaluate the extent of the influence of the loop mobility on NATs activity and substrate specificity. This could be done by mutagenesis experiments where selected amino acids in hinge regions could be replaced by glycine or proline to increase or reduce loop mobility. Such an approach would present the advantage of not affecting the structure and stability of the β-hairpin itself [76,79].

Materials and Methods

Dataset preparation

Scrutinizing protein structure and fold databases such as CATH or SCOP we could not find a specific grouping for all NATs. We used the annotation of GNAT domain from PROSITE [80] (PROSITE code: PS51186) and "N-terminal protein amino acid acetylation" as a biological process in PDBe [81]. We collected more than 160 structures and filtered down to 45 structures that had annotations as N-alpha acetyltransferases or N-terminal acetyltransferases (Table 1). From these we excluded eleven structures for which the X-ray structure had unresolved segments within the GNAT fold (S1 Table).

We formed 10 functional groups: Naa10, Naa20, Naa40, Naa50, Naa60, Naa80, archaeal NATs, RimI, RimJ and RimL. These groups have been constituted either by considering the Enzyme Commission (EC) number, their functional annotation in scientific literature when available or the kingdom of the organism the protein is found in. All structures files were prepared for the calculations by selecting one chain in the assembly and removing the Ac-CoA or peptide substrate if present. The reference set consists in one structure, called representative, for each Uniprot code in each functional group (see PDB IDs in bold in Table 1).

Structural alignment

The structural alignment was obtained using MUSTANG [51] which has been shown to be efficient on distant related proteins [33,34]. The algorithm performs a progressive pairwise alignment using the position of Cα atoms. It extends the pairwise structural alignments into multiple structure alignments by

recalculating a pairwise residue-residue score at each step of the extension and progresses using a guide tree. The pairwise RMSD between Naas structures cluster is displayed as a heatmap plotted with the R function pheatmap [82]. The alignment provided is primarily used for the comparison of the intrinsic dynamics as described in Refs. [33,83].

Elastic network model and Normal mode analysis (ENM-NMA)

The normal mode analysis has been performed using WEBnm@ [83]. The web-tool is using an Elastic Network Model (ENM) modelling protein structures as a network of nodes, the Cα atoms, connected together by Hookean springs. We used the Calpha force field [84,85], as implemented in the Molecular Modelling Toolkit [86]. It uses a pair potential to describe the interactions between two Cα atoms as:

$$V_{ij}(\boldsymbol{r}) = \frac{k_{ij}}{2}\left(\|\boldsymbol{r}_{ij}\| - \|\boldsymbol{r}_{ij}^0\|\right)^2$$

where r_{ij} is the distance vector between two Cα atoms i and j in the configuration r of the protein, r_{ij}^0 is then the same pairs of atoms i and j at the equilibrium conformation and k_{ij} is the non-uniform force constant defined by:

$$k_{ij} = \begin{cases} a\,\boldsymbol{r}_{ij}^0 - b, for\ \boldsymbol{r}_{ij}^0 < d \\ c\left(\boldsymbol{r}_{ij}^0\right)^{-6}, for\ \boldsymbol{r}_{ij}^0 \geq d \end{cases}$$

with $a = 8.6 \times 10^5 \text{kJ mol}^{-1}\text{nm}^{-3}$; $b = 2.39 \times 10^5 \text{kJ mol}^{-1}\text{nm}^2$; $c = 128\ \text{kJ mol}^{-1}\text{nm}^4$ and $d = 0.4$ nm.

The potential energy of the network model is then the sum of all the atomic configurations:

$$V(\boldsymbol{r}) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} V_{ij}(\boldsymbol{r})$$

where N is the number of nodes in the network.

For the normal modes calculation of the holo form of *Schizosaccharomyces pombe* Naa10 we represented CoA by 11 beads placed at the positions of atoms distant of 3 to 4Å, namely C, C3P, C6P, C9P, CCP, P1A, C4B, P3B, N9A, N6A and N3A.

Normalized fluctuations

The fluctuations F_i give the variances of each atom position and are given by:

$$F_i = \sum_{m=7}^{3N} \frac{\|[\boldsymbol{d}_m]_i\|^2}{\boldsymbol{\lambda}_m}$$

where $\boldsymbol{d}_m$is the displacement vector of the atom i in mode m. F_i is then the sum of the all the displacements of i for all the non-trivial modes that are weighted by their eigenvalues.

Correlations

The matrix of correlations is calculated from the normal modes [87] which quantifies the coupling between two atoms i and j as:

$$C_{ij} = \frac{\sum_{m=7}^{3N} \frac{1}{\boldsymbol{\lambda}_m}[\boldsymbol{v}_m]_i \cdot [\boldsymbol{v}_m]_j}{(F_i)^{\frac{1}{2}} (F_j)^{\frac{1}{2}}}$$

$C_{ij} = 1$ when the motions are completely correlated and $C_{ij} = -1$ when they are completely anti-correlated.

Bhattacharyya coefficient score

Finally we used the Bhattacharyya coefficient (BC) score to compare the effective covariances of the common aligned cores of two structures A and B as implemented in WEBnm@ and is based on Fuglebakk et al. [54] :

513 $$BC(p_a,p_b) = \frac{|\tilde{\boldsymbol{A}}|^{\frac{1}{4}}|\tilde{\boldsymbol{B}}|^{\frac{1}{4}}}{\left|\frac{1}{2}(\tilde{\boldsymbol{A}}+\tilde{\boldsymbol{B}})\right|^{\frac{1}{2}}}$$

514

515 **Trajectories and side chains reconstruction.**

516 We selected the 6 first non-trivial modes from the set of normal modes of the protein of interest. For each
517 mode, we generated a trajectory consisting of nine snapshots; we displaced the initial Cα positions
518 following the mode and using arbitrary amplitudes in either direction around the X-ray structure (the mode
519 vectors were multiplied by -12, -9, -6, -3, and 3, 6, 9, and 12).
520 We used the Molecular Modelling ToolKit (MMTK,[86]) to reconstruct the side chains in order to obtain
521 trajectories at all-atom resolution. We calculated the 3D transformation necessary to superimpose the initial
522 all-atom structure onto each of the snapshots of the Cα trajectories. This was done by minimizing the RMS
523 difference between the initial all-atom structure and the Cα trace snapshots. The 3D transformations were
524 not computed on the overall structure but locally using an iterative process. We used sliding windows that
525 were three amino acids long to compute the transformation, which is then applied only to the central amino
526 acid for which the side chain is reconstructed. The process is then iterated by sliding by one residue along
527 the protein sequence.

528

529 **Visualisation and calculation of the cavity**

530 The analysis was performed using CAVER Analyst [65]. We first selected three Cα atoms from amino acids
531 (Tyr31, 73, and 138) that are lining the mouth of the cavity in the X-ray structure. In 3D space, these three
532 atoms define a parallel plane to the tunnel mouth. Then a set of intersecting spheres, with a radius of 1Å, is
533 placed on a line perpendicular to this plane to fill up the length of 5Å into the cavity. Using this geometrical
534 structure as a base, we computed a cavity surface in each frame using the algorithm described in [65]. We
535 do not use the extension of the algorithm proposed by Jurcik et al. [88] as it was developed for the detection
536 of deeply buried voids inside proteins. While in our case the cavities in NATs are not fully surrounded by

amino acids and are closer to the protein surface. Finally, the contours of the surface area of the mouth are extracted using the MoleCollar algorithm part of CAVER Analyst and first described in Ref. [89]. The MoleCollar algorithm estimates the empty space area by triangulating a contour that is obtained by cutting the cavity surface by a plane and summing up the areas of all resulting triangles. The algorithm was modified to keep the same orientation of the cutting plane throughout the trajectory in order to follow the evolution of the surface area along the trajectories.

Acknowledgements

We would like to thank Konrad Hinsen (Centre de Biophysique Moléculaire in Orléans, France) for fruitful discussions on trajectory creation; Edvin Fuglebakk (Institute of Marine Research, Bergen, Norway), Sandhya Tiwari (Computational Structural Biology Research Unit, RIKEN, Japan), Bojan Krtenic (Department of Biological Sciences, University of Bergen) and Simon Mitternacht (University of Bergen Library) for their valuable comments on the manuscript.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Author contributions:

Conceived and designed the experiments: AA, NR, JB, PB and TA. Performed the experiments: AA. Analysed the data: AA, NR. Code development for CAVER Analyst: JB. Code development to recreate all-atoms trajectories: PB. AA and NR wrote the paper with contributions from all authors.

References

1. Arnesen T, Van Damme P, Polevoda B, Helsens K, Evjenth R, Colaert N, et al. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. Proc Natl Acad Sci U S A. 2009;106: 8157–62. doi:10.1073/pnas.0901931106
2. Drazic A, Myklebust LM, Ree R, Arnesen T. The world of protein acetylation. Biochim Biophys Acta - Proteins Proteomics. 2016;1864: 1372–1401. doi:10.1016/j.bbapap.2016.06.007
3. Kalvik T V., Arnesen T. Protein N-terminal acetyltransferases in cancer. Oncogene. 2013. pp. 269–276. doi:10.1038/onc.2012.82
4. Myklebust LM, Van Damme P, Støve SI, Dörfel MJ, Abboud A, Kalvik T V., et al. Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. Hum Mol Genet. 2014;24: 1956–1976. doi:10.1093/hmg/ddu611
5. Dörfel MJ, Lyon GJ. The biological functions of Naa10 - From amino-terminal acetylation to human disease. Gene. 2015. pp. 103–131. doi:10.1016/j.gene.2015.04.085
6. Foyn H, Jones JE, Lewallen D, Narawane R, Varhaug JE, Thompson PR, et al. Design, synthesis, and kinetic characterization of protein N-terminal acetyltransferase inhibitors. ACS Chem Biol. 2013;8: 1121–1127. doi:10.1021/cb400136s
7. Vetting MW, Luiz LP, Yu M, Hegde SS, Magnet S, Roderick SL, et al. Structure and functions of the GNAT superfamily of acetyltransferases. Arch Biochem Biophys. 2005;433: 212–226. doi:10.1016/j.abb.2004.09.003
8. Lu L, Berkey KA, Casero RA. RGFGIGS Is an Amino Acid Sequence Required for Acetyl Coenzyme A Binding and Activity of Human Spermidine / Spermine N 1 Acetyltransferase *. J Biol Chem. 1996;271: 18920–18924. doi:10.1074/jbc.271.31.18920
9. Marmorstein R. Structure of histone acetyltransferases. J Mol Biol. 2001;311: 433–444. doi:10.1006/jmbi.2001.4859
10. Dinh T V., Bienvenut W V., Linster E, Feldman-Salit A, Jung VA, Meinnel T, et al. Molecular identification and functional characterization of the first Nα-acetyltransferase in plastids by global

acetylome profiling. Proteomics. 2015;15: 2426–2435. doi:10.1002/pmic.201500025

11. Drazic A, Aksnes H, Marie M, Boczkowska M, Varland S, Timmerman E, et al. NAA80 is actin's N-terminal acetyltransferase and regulates cytoskeleton assembly and cell motility. Proc Natl Acad Sci. 2018; 201718336. doi:10.1073/pnas.1718336115

12. Yoshikawa A, Isono S, Sheback A, Isono K. Cloning and nucleotide sequencing of the genes rimI and rimJ which encode enzymes acetylating ribosomal proteins S18 and S5 of Escherichia coli K12. MGG Mol Gen Genet. 1987;209: 481–488. doi:10.1007/BF00331153

13. Tanaka S, Matsushita Y, Yoshikawa A, Isono K. Cloning and molecular characterization of the gene rimL which encodes an enzyme acetylating ribosomal protein L12 of Escherichia coli K12. Mol Gen Genet. 1989;217: 289–93. doi:10.1007/BF02464895

14. Mackay DT, Botting CH, Taylor GL, White MF. An acetylase with relaxed specificity catalyses protein N-terminal acetylation in Sulfolobus solfataricus. Mol Microbiol. 2007;64: 1540–1548. doi:10.1111/j.1365-2958.2007.05752.x

15. Liszczak G, Arnesen T, Marmorsteins R. Structure of a ternary Naa50p (NAT5/SAN) N-terminal acetyltransferase complex reveals the molecular basis for substrate-specific acetylation. J Biol Chem. 2011;286: 37002–37010. doi:10.1074/jbc.M111.282863

16. Magin RS, Liszczak GP, Marmorstein R. The molecular basis for Histone H4- and H2A-specific amino-terminal acetylation by NatD. Structure. Elsevier; 2015;23: 332–341. doi:10.1016/j.str.2014.10.025

17. Liszczak G, Goldberg JM, Foyn H, Petersson EJ, Arnesen T, Marmorstein R. Molecular basis for N-terminal acetylation by the heterodimeric NatA complex. Nat Struct Mol Biol. 2013;20: 1098–1105. doi:10.1038/nsmb.2636

18. Chen JY, Liu L, Cao CL, Li MJ, Tan K, Yang X, et al. Structure and function of human Naa60 (NatF), a Golgi-localized bi-functional acetyltransferase. Sci Rep. Nature Publishing Group; 2016;6: 1–12. doi:10.1038/srep31425

19. Grauffel C, Abboud A, Liszczak G, Marmorstein R, Arnesen T, Reuter N. Specificity and Versatility of Substrate Binding Sites in Four Catalytic Domains of Human N-Terminal Acetyltransferases.

PLoS One. 2012;7. doi:10.1371/journal.pone.0052642

20. Magin RS, Liszczak GP, Marmorstein R. The molecular basis for Histone H4- and H2A-specific amino-terminal acetylation by NatD. Structure. 2015;23: 332–341. doi:10.1016/j.str.2014.10.025

21. Montgomery DC, Sorum AW, Meier JL. Chemoproteomic profiling of lysine acetyltransferases highlights an expanded landscape of catalytic acetylation. J Am Chem Soc. 2014;136: 8669–8676. doi:10.1021/ja502372j

22. Liu Z, Liu Y, Wang H, Ge X, Jin Q, Ding G, et al. Patt1, a novel protein acetyltransferase that is highly expressed in liver and downregulated in hepatocellular carcinoma, enhances apoptosis of hepatoma cells. Int J Biochem Cell Biol. 2009;41: 2528–2537. doi:10.1016/j.biocel.2009.08.009

23. Yang X, Yu W, Shi L, Sun L, Liang J, Yi X, et al. HAT4, a Golgi Apparatus-Anchored B-Type Histone Acetyltransferase, Acetylates Free Histone H4 and Facilitates Chromatin Assembly. Mol Cell. Elsevier Inc.; 2011;44: 39–50. doi:10.1016/j.molcel.2011.07.032

24. Yoon H, Kim HL, Chun YS, Shin DH, Lee KH, Shin CS, et al. NAA10 controls osteoblast differentiation and bone formation as a feedback regulator of Runx2. Nat Commun. Nature Publishing Group; 2014;5: 1–14. doi:10.1038/ncomms6176

25. Ohkawa N, Sugisaki S, Tokunaga E, Fujitani K, Hayasaka T, Setou M, et al. N-acetyltransferase ARD1-NAT1 regulates neuronal dendritic development. Genes to Cells. 2008;13: 1171–1183. doi:10.1111/j.1365-2443.2008.01235.x

26. Wang ZLW. Inactivation of androgen-induced regulator ARD1 inhibits androgen receptor acetylation and prostate tumorigenesis. Proc Natl Acad Sci U S A. 2012;188: 2015–2016. doi:10.1016/j.juro.2012.07.065

27. Seo JH, Park JH, Lee EJ, Vo TTL, Choi H, Kim JY, et al. ARD1-mediated Hsp70 acetylation balances stress-induced protein refolding and degradation. Nat Commun. Nature Publishing Group; 2016;7: 1–14. doi:10.1038/ncomms12882

28. Støve SII, Magin RSS, Foyn H, Haug BEE, Marmorstein R, Arnesen T. Crystal Structure of the Golgi-Associated Human Nα-Acetyltransferase 60 Reveals the Molecular Determinants for Substrate-Specific Acetylation. Structure. 2016;24: 1044–1056. doi:10.1016/j.str.2016.04.020

639 29. Vetting MW, Bareich DC, Yu M, Blanchard JS. Crystal structure of RimI from Salmonella
640 typhimurium LT2, the GNAT responsible for N(alpha)-acetylation of ribosomal protein S18. Protein
641 Sci. 2008;17: 1781–90. doi:10.1110/ps.035899.108
642 30. Myklebust LM, Van Damme P, Stove SI, Dorfel MJ, Abboud A, Kalvik TV, et al. Biochemical and
643 cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. Hum Mol Genet.
644 2014;24. doi:10.1093/hmg/ddu611
645 31. Micheletti C. Comparing proteins by their internal dynamics: Exploring structure-function
646 relationships beyond static structural alignments. Phys Life Rev. Elsevier B.V.; 2013;10: 1–26.
647 doi:10.1016/j.plrev.2012.10.009
648 32. Carnevale V, Raugei S, Micheletti C, Carloni P. Convergent dynamics in the protease enzymatic
649 superfamily. J Am Chem Soc. 2006;128: 9766–9772. doi:10.1021/ja060896t
650 33. Tiwari SP, Reuter N. Similarity in Shape Dictates Signature Intrinsic Dynamics Despite No
651 Functional Conservation in TIM Barrel Enzymes. PLoS Comput Biol. 2016;12: 1–26.
652 doi:10.1371/journal.pcbi.1004834
653 34. Fuglebakk E, Tiwari SP, Reuter N. Comparing the intrinsic dynamics of multiple protein structures
654 using elastic network models. Biochim Biophys Acta. The Authors; 2015;1850: 911–922.
655 doi:10.1016/j.bbagen.2014.09.021
656 35. Rueda M, Chacón P, Orozco M. Thorough Validation of Protein Normal Mode Analysis: A
657 Comparative Study with Essential Dynamics. Structure. 2007;15: 565–575.
658 doi:10.1016/j.str.2007.03.013
659 36. Berendsen HJ, Hayward S. Collective protein dynamics in relation to function. Curr Opin Struct
660 Biol. 2000;10: 165–169. doi:10.1016/S0959-440X(00)00061-0
661 37. Sathiyamoorthy K, Vijayalakshmi J, Tirupati B, Fan L, Saper MA. Structural analyses of the
662 Haemophilus influenzae peptidoglycan synthase activator LpoA suggest multiple conformations in
663 solution. J Biol Chem. 2017;292: 17626–17642. doi:10.1074/jbc.M117.804997
664 38. Chirasani VR, Revanasiddappa PD, Senapati S. Structural plasticity of cholesteryl ester transfer
665 protein assists the lipid transfer activity. J Biol Chem. 2016;291: 19462–19473.

doi:10.1074/jbc.M116.744623

39. Schushan M, Rimon A, Haliloglu T, Forrest LR, Padan E, Ben-Tal N. A model-structure of a periplasm-facing state of the NhaA antiporter suggests the molecular underpinnings of pH-induced conformational changes. J Biol Chem. 2012;287: 18249–18261. doi:10.1074/jbc.M111.336446

40. Valadié H, Lacapčre JJ, Sanejouand YH, Etchebest C. Dynamical properties of the MscL of Escherichia coli: A normal mode analysis. J Mol Biol. 2003;332: 657–674. doi:10.1016/S0022-2836(03)00851-9

41. Oke M, Carter LG, Johnson KA, Liu H, McMahon SA, Yan X, et al. The scottish structural proteomics facility: Targets, methods and outputs. J Struct Funct Genomics. 2010;11: 167–180. doi:10.1007/s10969-010-9090-y

42. Liszczak G, Marmorstein R. Implications for the evolution of eukaryotic amino-terminal acetyltransferase (NAT) enzymes from the structure of an archaeal ortholog. Proc Natl Acad Sci U S A. 2013;110: 14652–7. doi:10.1073/pnas.1310365110

43. Chang Y-Y, Hsu C-H. Structural Basis for Substrate-specific Acetylation of Nα-acetyltransferase Ard1 from Sulfolobus solfataricus. Sci Rep. 2015;5: 8673. doi:10.1038/srep08673

44. Chang YY, Hsu CH. Multiple Conformations of the Loop Region Confers Heat-Resistance on SsArd1, a Thermophilic NatA. ChemBioChem. 2016;17: 214–217. doi:10.1002/cbic.201500568

45. Ma C, Pathak C, Jang S, Lee SJ, Nam M, Kim SJ, et al. Structure of Thermoplasma volcanium Ard1 belongs to N-acetyltransferase family member suggesting multiple ligand binding modes with acetyl coenzyme A and coenzyme A. Biochim Biophys Acta - Proteins Proteomics. 2014;1844: 1790–1797. doi:10.1016/j.bbapap.2014.07.011

46. Hong H, Cai Y, Zhang S, Ding H, Wang H, Han A. Molecular Basis of Substrate Specific Acetylation by N-Terminal Acetyltransferase NatB. Structure. Elsevier Ltd.; 2017;25: 641–649.e3. doi:10.1016/j.str.2017.03.003

47. Magin RS, Liszczak GP, Marmorstein R. The molecular basis for Histone H4- and H2A-specific amino-terminal acetylation by NatD. Structure. Elsevier; 2015;23: 332–341. doi:10.1016/j.str.2014.10.025

48. Goris M, Magin RS, Foyn H, Myklebust LM, Varland S, Ree R, et al. Structural determinants and cellular environment define processed actin as the sole substrate of the N-terminal acetyltransferase NAA80. Proc Natl Acad Sci. 2018; 201719251. doi:10.1073/pnas.1719251115

49. Vetting MW, De Carvalho LPS, Roderick SL, Blanchard JS. A novel dimeric structure of the RimL Na-acetyltransferase from Salmonella typhimurium. J Biol Chem. 2005;280: 22108–22114. doi:10.1074/jbc.M502401200

50. Sakamoto K, Murayama K, Oki K, Iraha F, Kato-Murayama M, Takahashi M, et al. Genetic Encoding of 3-Iodo-l-Tyrosine in Escherichia coli for Single-Wavelength Anomalous Dispersion Phasing in Protein Crystallography. Structure. Elsevier Ltd; 2009;17: 335–344. doi:10.1016/j.str.2009.01.008

51. Konagurthu AS, Whisstock JC, Stuckey PJ, Lesk AM. MUSTANG: a multiple structural alignment algorithm. Proteins. 2006;64: 559–574. doi:10.1002/prot.20921

52. Aksnes H, Goris M, Strømland Ø, Drazic A, Waheed Q, Reuter N, et al. Molecular determinants of the N-Terminal acetyltransferase Naa60 anchoring to the Golgi membrane. J Biol Chem. 2017;292: 6821–6837. doi:10.1074/jbc.M116.770362

53. Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 2014;42. doi:10.1093/nar/gku316

54. Fuglebakk E, Echave J, Reuter N. Measuring and comparing structural fluctuation patterns in large protein datasets. Bioinformatics. 2012;28: 2431–2440. doi:10.1093/bioinformatics/bts445

55. Van Damme P, Evjenth R, Foyn H, Demeyer K, De Bock P-J, Lillehaug JR, et al. Proteome-derived Peptide Libraries Allow Detailed Analysis of the Substrate Specificities of N $^{\alpha}$ -acetyltransferases and Point to hNaa10p as the Post-translational Actin N $^{\alpha}$ -acetyltransferase. Mol Cell Proteomics. 2011;10: M110.004580. doi:10.1074/mcp.M110.004580

56. Tiwari SP, Reuter N. Conservation of intrinsic dynamics in proteins — what have computational models taught us? Curr Opin Struct Biol. Elsevier Ltd; 2018;50: 75–81. doi:10.1016/j.sbi.2017.12.001

57. Zen A, Carnevale V, Lesk AM, Micheletti C. Correspondences between low-energy modes in

720 enzymes: dynamics-based alignment of enzymatic functional families. Protein Sci. 2008;17: 918–
721 929. doi:10.1110/ps.073390208
722 58. Yang LW, Bahar I. Coupling between catalytic site and collective dynamics: A requirement for
723 mechanochemical activity of enzymes. Structure. 2005;13: 893–904. doi:10.1016/j.str.2005.03.015
724 59. Zheng W, Tekpinar M. Large-scale evaluation of dynamically important residues in proteins
725 predicted by the perturbation analysis of a coarse-grained elastic model. BMC Struct Biol. 2009;9:
726 45. doi:10.1186/1472-6807-9-45
727 60. Reuter N, Hinsen K, Lacapère JJ. Transconformations of the SERCA1 Ca-ATPase: A normal mode
728 study. Biophys J. 2003;85: 2186–2197. doi:10.1016/S0006-3495(03)74644-X
729 61. Fenwick RB, Orellana L, Esteban-Martín S, Orozco M, Salvatella X. Correlated motions are a
730 fundamental property of β-sheets. Nat Commun. 2014;5. doi:10.1038/ncomms5070
731 62. Ahmed A, Villinger S, Gohlke H. Large-scale comparison of protein essential dynamics from
732 molecular dynamics simulations and coarse-grained normal mode analyses. Proteins Struct Funct
733 Bioinforma. 2010;78: 3341–3352. doi:10.1002/prot.22841
734 63. Echave J. Why are the low-energy protein normal modes evolutionarily conserved ?*. Pure Appl
735 Chem. 2012;84: 1931–1937. doi:http://dx.doi.org/10.1351/PAC-CON-12-02-15
736 64. Kurkcuoglu Z, Bakan A, Kocaman D, Bahar I, Doruker P. Coupling between Catalytic Loop
737 Motions and Enzyme Global Dynamics. PLoS Comput Biol. 2012;8: 1–11.
738 doi:10.1371/journal.pcbi.1002705
739 65. Jurcik A, Bednar D, Byska J, Marques SM, Furmanova K, Daniel L, et al. CAVER Analyst 2 .0 :
740 Analysis and Visualization of Channels and Tunnels in Protein Structures and Molecular Dynamics
741 Trajectories. Bioinformatics. 2018; 9–10. doi:10.1093/bioinformatics/bty386/4993945
742 66. Popp B, Støve SI, Endele S, Myklebust LM, Hoyer J, Sticht H, et al. De novo missense mutations in
743 the NAA10 gene cause severe non-syndromic developmental delay in males and females. Eur J Hum
744 Genet. 2015;23: 602–609. doi:10.1038/ejhg.2014.150
745 67. Shahmoradi A, Sydykova DK, Spielman SJ, Jackson EL, Dawson ET, Meyer AG, et al. Predicting
746 Evolutionary Site Variability from Structure in Viral Proteins: Buriedness, Packing, Flexibility, and

Design. J Mol Evol. 2014;79: 130–142. doi:10.1007/s00239-014-9644-x

68. Bahar I, Cheng MH, Lee JY, Kaya C, Zhang S. Structure-Encoded Global Motions and Their Role in Mediating Protein-Substrate Interactions. Biophysical Journal. 2015. pp. 1101–1109. doi:10.1016/j.bpj.2015.06.004

69. Lovera S, Morando M, Pucheta-Martinez E, Martinez-Torrecuadrada JL, Saladino G, Gervasio FL. Towards a Molecular Understanding of the Link between Imatinib Resistance and Kinase Conformational Dynamics. PLoS Comput Biol. 2015;11. doi:10.1371/journal.pcbi.1004578

70. Støve SII, Magin RSS, Foyn H, Haug BEE, Marmorstein R, Arnesen T. Crystal Structure of the Golgi-Associated Human Nα-Acetyltransferase 60 Reveals the Molecular Determinants for Substrate-Specific Acetylation. Structure. 2016;24: 1044–1056. doi:10.1016/j.str.2016.04.020

71. Tama F, Sanejouand Y-H. Conformational change of proteins arising from normal mode calculations. Protein Eng Des Sel. 2001;14: 1–6. doi:10.1093/protein/14.1.1

72. Evjenth RH, Brenner AK, Thompson PR, Arnesen T, Frøystein NÅ, Lillehaug JR. Human protein N-terminal acetyltransferase hNaa50p (hNAT5/hSAN) follows ordered sequential catalytic mechanism: Combined kinetic and NMR study. J Biol Chem. 2012;287: 10081–10088. doi:10.1074/jbc.M111.326587

73. Evjenth R, Hole K, Karlsen OA, Ziegler M, Amesen T, Lillehaug JR. Human Naa50p (Nat5/San) displays both protein Nα- and Nε-acetyltransferase activity. J Biol Chem. 2009;284: 31122–31129. doi:10.1074/jbc.M109.001347

74. Velázquez-Muriel JA, Rueda M, Cuesta I, Pascual-Montano A, Orozco M, Carazo JM. Comparison of molecular dynamics and superfamily spaces of protein domain deformation. BMC Struct Biol. 2009;9: 1–14. doi:10.1186/1472-6807-9-6

75. Magin RS, March ZM, Marmorstein R. The N-terminal acetyltransferase Naa10/ARD1 does not acetylate lysine residues. J Biol Chem. 2016;291: 5270–5277. doi:10.1074/jbc.M115.709428

76. Jager M, Deechongkit S, Koepf EK, Nguyen H, Gao J, Powers ET, et al. Understanding the mechanism of beta-sheet folding from a chemical and biological perspective. Biopolymers. 2008;90: 751–758. doi:10.1002/bip.21101

774 77. Lim JH, Park JW, Chun YS. Human arrest defective 1 acetylates and activates beta-catenin,
775 promoting lung cancer cell proliferation. Cancer Res. 2006;66: 10677–10682. doi:10.1158/0008-
776 5472.CAN-06-3171
777 78. Seo JH, Cha JH, Park JH, Jeong CH, Park ZY, Lee HS, et al. Arrest defective 1 autoacetylation is a
778 critical step in its ability to stimulate cancer cell proliferation. Cancer Res. 2010;70: 4422–4432.
779 doi:10.1158/0008-5472.CAN-09-3258
780 79. Reinke T. Muller, Timothy Travers, Hi-jea Cha, Joshua L. Phillips, S. Gnanakaran KMP. Switch
781 loop flexibility affects substrate transport of the AcrB efflux pump. J Mol Biol. Elsevier Ltd; 2017;
782 22. doi:10.1016/j.jmb.2017.09.018
783 80. Sigrist CJA, De Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, et al. New and continuing
784 developments at PROSITE. Nucleic Acids Res. 2013;41: 344–347. doi:10.1093/nar/gks1067
785 81. Velankar S, Van Ginkel G, Alhroub Y, Battle GM, Berrisford JM, Conroy MJ, et al. PDBe:
786 Improved accessibility of macromolecular structure data from PDB and EMDB. Nucleic Acids Res.
787 2016;44: D385–D395. doi:10.1093/nar/gkv1047
788 82. Kolde R. Package `pheatmap'. Bioconductor. 2012; 1–6.
789 83. Tiwari SP, Fuglebakk E, Hollup SM, Skjærven L, Cragnolini T, Grindhaug SH, et al. WEBnm@
790 v2.0: Web server and services for comparing protein flexibility. BMC Bioinformatics. 2014;15: 427.
791 doi:10.1186/s12859-014-0427-6
792 84. Hinsen K. Analysis of domain motions by approximate normal mode calculations. Proteins Struct
793 Funct Genet. 1998;33: 417–429. doi:10.1002/(SICI)1097-0134(19981115)33:3<417::AID-
794 PROT10>3.0.CO;2-8
795 85. Hinsen K, Petrescu AJ, Dellerue S, Bellissent-Funel MC, Kneller GR. Harmonicity in slow protein
796 dynamics. Chem Phys. 2000;261: 25–37. doi:10.1016/S0301-0104(00)00222-6
797 86. Hinsen K. The molecular modeling toolkit: A new approach to molecular simulations. J Comput
798 Chem. 2000;21: 79–85. doi:10.1002/(SICI)1096-987X(20000130)21:2<79::AID-JCC1>3.0.CO;2-B
799 87. Ichiye T, Karplus M. Collective motions in proteins: A covariance analysis of atomic fluctuations in
800 molecular dynamics and normal mode simulations. Proteins Struct Funct Bioinforma. 1991;11: 205–

217. doi:10.1002/prot.340110305

88. Jurčík A, Byška J, Sochor J, Kozlíková B. Visibility-based approach to surface detection of tunnels in proteins. Proc 31st Spring Conf Comput Graph - SCCG '15. 2015; 65–72. doi:10.1145/2788539.2788548

89. Byška J, Jurčík A, Gröller ME, Viola I, Kozlíková B. MoleCollar and Tunnel Heat Map Visualizations for Conveying Spatio-Temporo-Chemical Properties Across and Along Protein Voids. Comput Graph Forum. 2015;34: 1–10. doi:10.1111/cgf.12612

90. Arnesen T, Anderson D, Baldersheim C, Lanotte M, Varhaug JE, Lillehaug JR. Identification and characterization of the human ARD1-NATH protein acetyltransferase complex. Biochem J. 2005;386: 433–443. doi:10.1042/BJ20041071

91. Polevoda B, Cardillo TS, Doyle TC, Bedi GS, Sherman F. Nat3p and Mdm20p are required for function of yeast NatB Nα-terminal acetyltransferase and of actin and tropomyosin. J Biol Chem. 2003;278: 30686–30697. doi:10.1074/jbc.M304690200

92. Starheim KK, Arnesen T, Gromyko D, Ryningen A, Varhaug JE, Lillehaug JR. Identification of the human N α -acetyltransferase complex B (hNatB): a complex important for cell-cycle progression. Biochem J. 2008;415: 325–331. doi:10.1042/BJ20080658

93. Van Damme P, Lasa M, Polevoda B, Gazquez C, Elosegui-Artola A, Kim DS, et al. N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. Proc Natl Acad Sci. 2012;109: 12449–12454. doi:10.1073/pnas.1210303109

94. Song OK, Wang X, Waterborg JH, Sternglanz R. An Nalpha-acetyltransferase responsible for acetylation of the N-terminal residues of histones H4 and H2A. J Biol Chem. 2003;278: 38109–38112. doi:10.1074/jbc.C300355200

95. Arnesen T, Anderson D, Torsvik J, Halseth HB, Varhaug JE, Lillehaug JR. Cloning and characterization of hNAT5/hSAN: An evolutionarily conserved component of the NatA protein N-α-acetyltransferase complex. Gene. 2006;371: 291–295. doi:10.1016/j.gene.2005.12.008

96. Van Damme P, Hole K, Gevaert K, Arnesen T. N-terminal acetylome analysis reveals the specificity of Naa50 (Nat5) and suggests a kinetic competition between N-terminal acetyltransferases and

methionine aminopeptidases. Proteomics. 2015;15: 2436–2446. doi:10.1002/pmic.201400575

97. Aksnes H, Van Damme P, Goris M, Starheim KK, Marie M, Støve SI, et al. An organellar nα-acetyltransferase, naa60, acetylates cytosolic n termini of transmembrane proteins and maintains golgi integrity. Cell Rep. 2015;10: 1362–1374. doi:10.1016/j.celrep.2015.01.053

98. Roy-Chaudhuri B, Kirthi N, Kelley T, Culver GM. Suppression of a cold-sensitive mutation in ribosomal protein S5 reveals a role for RimJ in ribosome biogenesis. Mol Microbiol. 2008;68: 1547–1559. doi:10.1111/j.1365-2958.2008.06252.x

Supporting Information

S1 Fig. Effect of the presence of a cofactor or a bisubstrate on the normalized fluctuations of the Naa10 from *Saccharomyces pombe*. Naa10 from *S. pombe* has been crystallized with Naa15 (PDB id: 4kvm and 4kvo) and without Naa15 (PDB id: 4kvx). The main difference between the structures of the two states is a rearrangement of helix α2 and of the α1α2 loop. The experimental structures of the complexed Naa10 form contain either a bisubstrate inhibitor (4kvm) or only a cofactor (4kvo). The structure of the uncomplexed Naa10 form has only the cofactor (4kvx). The computations were performed on the X-ray structures with their ligand bound, and on the same protein structures after removal of the ligand.

S2 Fig. Heatmap representation of the pairwise Root Mean Square Deviations (RMSD) for representatives structure. The dendrogram reflects the hierarchical clustering based on the RMSD values. The heatmap color scale goes from red (0Å < RMSD < 1Å; structural similarity) to yellow (3Å < RMSD < 4Å). Names of enzymes from eukaryotes are highlighted in blue, those of bacteria in purple and green is used for archaeal NATs.

S2 Fig. Correlations maps for the representative of each NAT group. Dark red and blue show correlated motions between pairs of residues (close to 1 or -1, respectively), when lighter colours refer to correlations close to 0. Long-range correlations are found within two blocks highlighted by the green and pink frames. For all the NATs the highest correlations are found within these blocks and not in-between.

S1 Table. Structure dataset.

S2 Table. Ligand specificity of NAT enzymes included in the dataset. The information in this table is collected from the following references: ArNat [14,42], NatA [1,17,55,90], NatB [91–93] , NatD [47,94], NatE [73,95,96], NatF [52,97], NatH [11,48], RimI [12,29], RimJ [12,98], RimL [13,49].

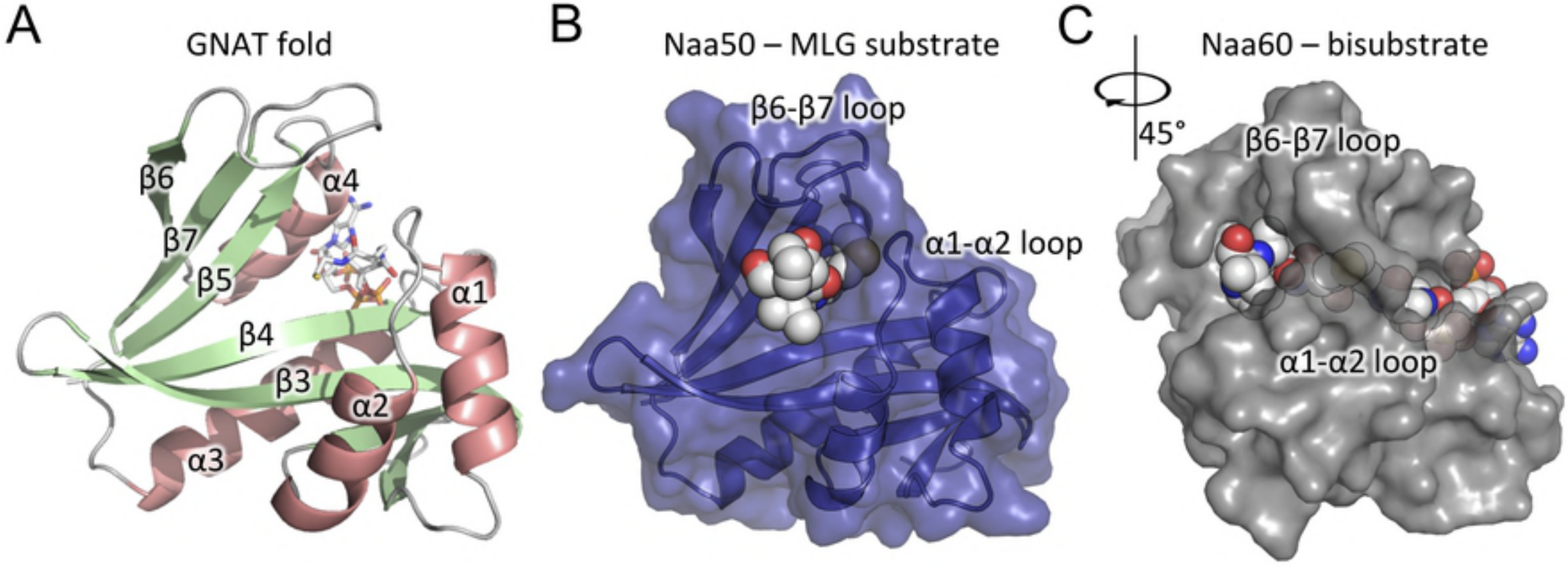

A
GNAT fold
β6
α4
β7
β5
α1
β4
β3
α2
α3
B
Naa50 – MLG substrate
β6-β7 loop
α1-α2 loop
C
Naa60 – bisubstrate
45°
β6-β7 loop
α1-α2 loop

Fig1

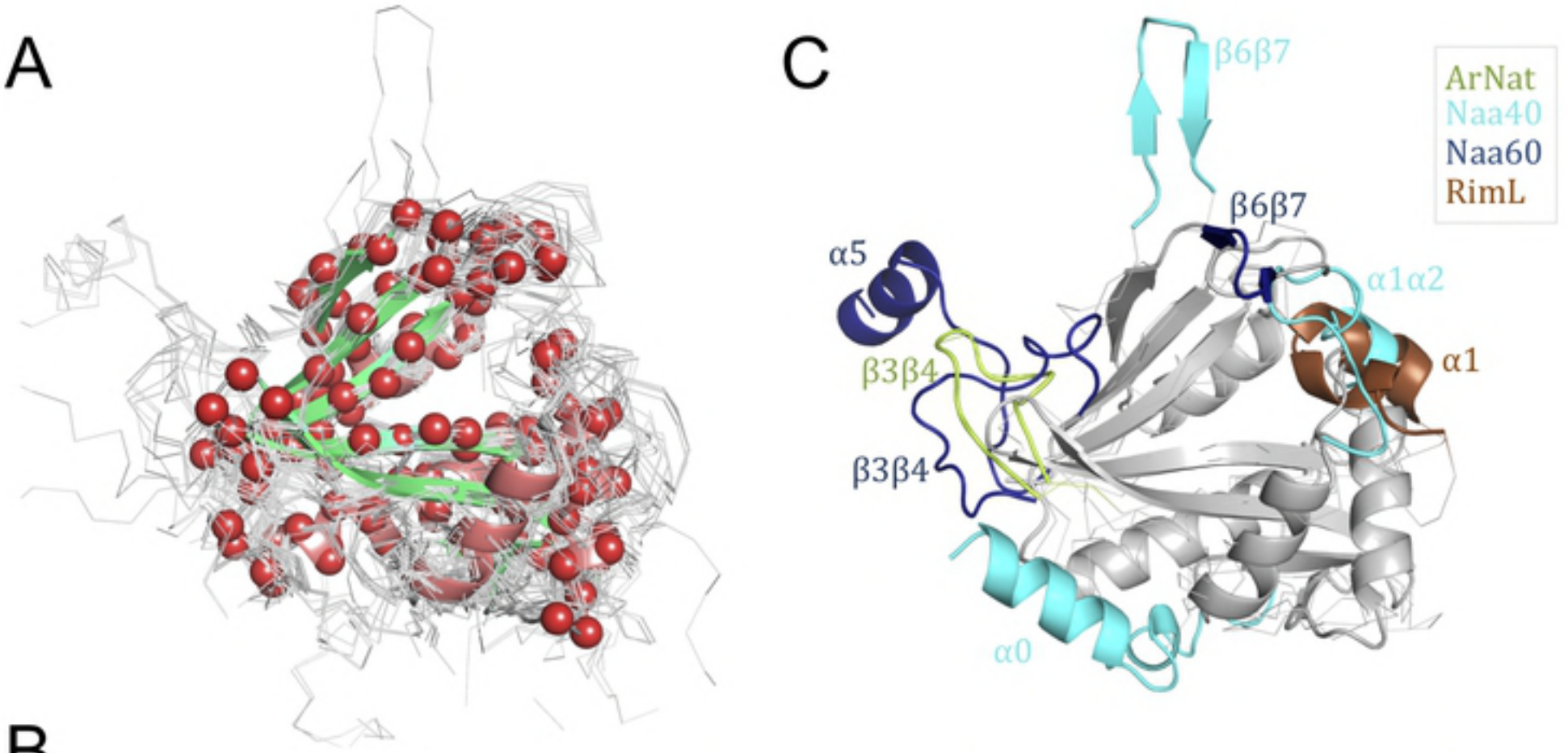


Fig2

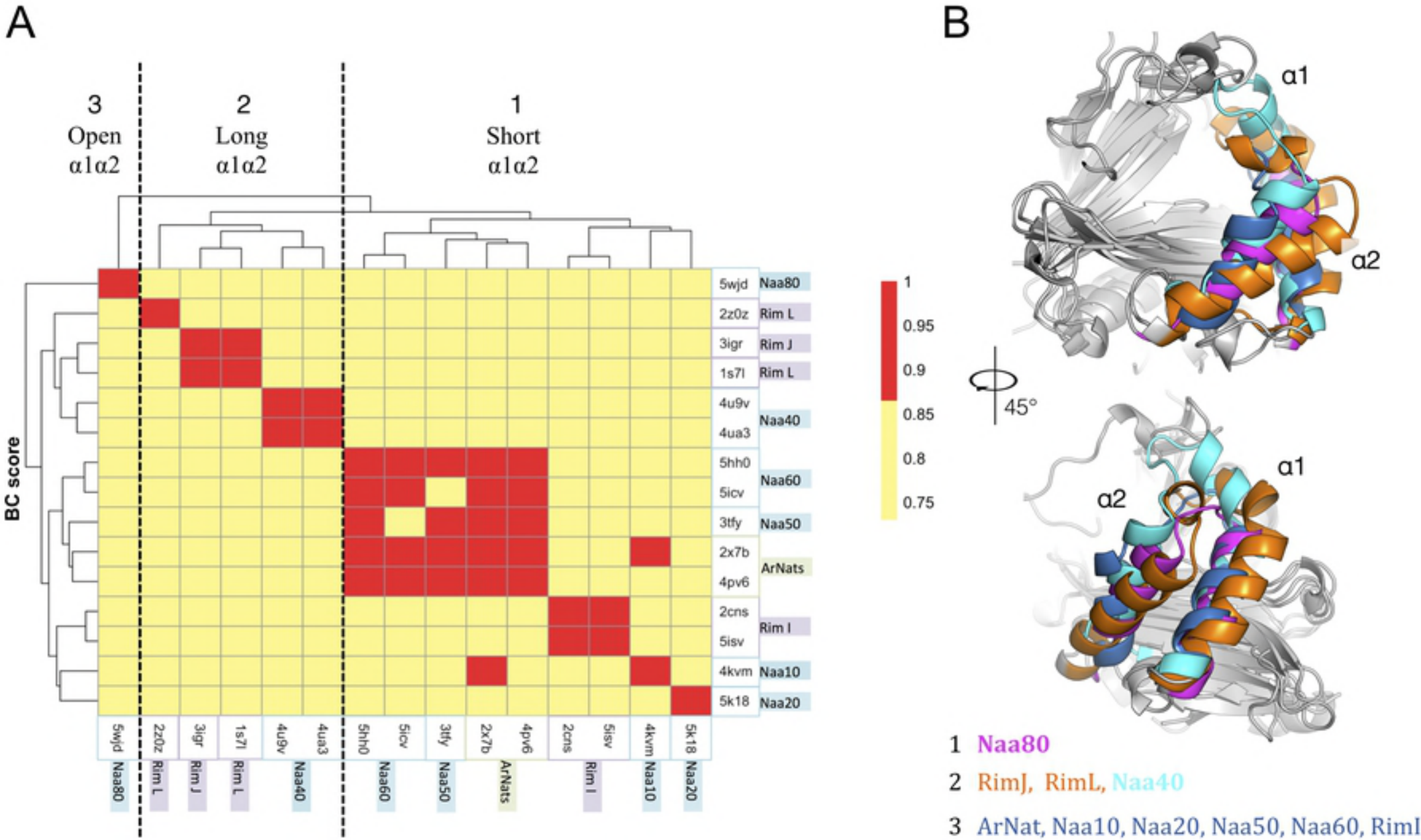
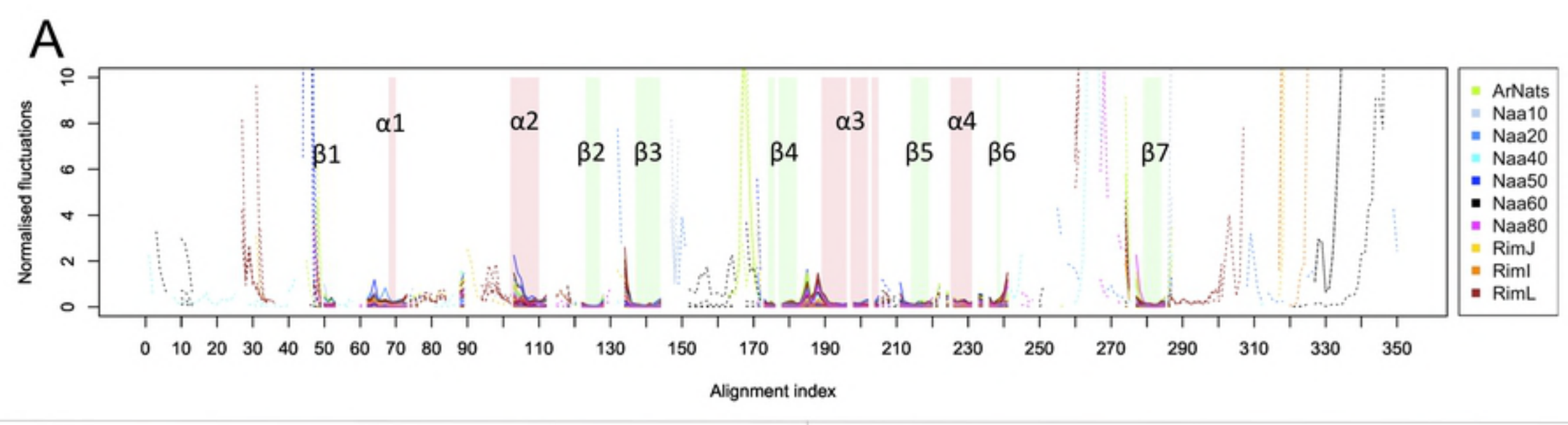
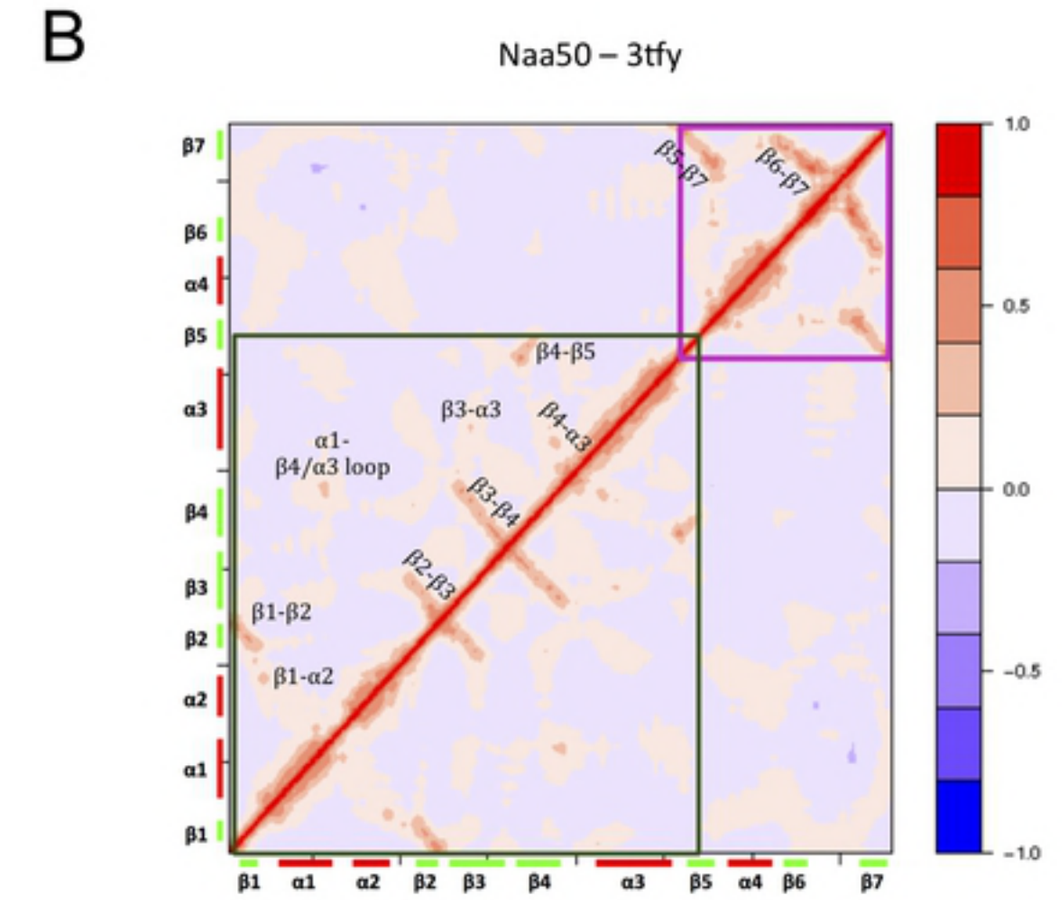

A
3
Open
α1α2
2
Long
α1α2
1
Short
α1α2
BC score
1
0.95
0.9
0.85
0.8
0.75
5wjd Naa80
2z0z Rim L
3igr Rim J
1s7l Rim L
4u9v
4ua3 Naa40
5hh0
5icv Naa60
3tfy Naa50
2x7b
4pv6 ArNats
2cns
5isv Rim I
4kvm Naa10
5k18 Naa20
B
α1
α2
45°
α1
α2
1 Naa80
2 RimJ, RimL, Naa40
3 ArNat, Naa10, Naa20, Naa50, Naa60, RimI

Fig4


A
Normalised fluctuations
0 2 4 6 8 10
β1 α1 α2 β2 β3 β4 α3 β5 α4 β6 β7
0 10 20 30 40 50 60 70 80 90 110 130 150 170 190 210 230 250 270 290 310 330 350
Alignment index
ArNats
Naa10
Naa20
Naa40
Naa50
Naa60
Naa80
RimJ
RimI
RimL


B
Naa50 – 3tfy
β7 β6 α4 β5 α3 β4 β3 β2 α2 α1 β1
β5-β7
β6-β7
β4-β5
β3-α3
β4-α3
α1-
β4/α3 loop
β3-β4
β2-β3
β1-β2
β1-α2
1.0
0.5
0.0
−0.5
−1.0
β1 α1 α2 β2 β3 β4 α3 β5 α4 β6 β7

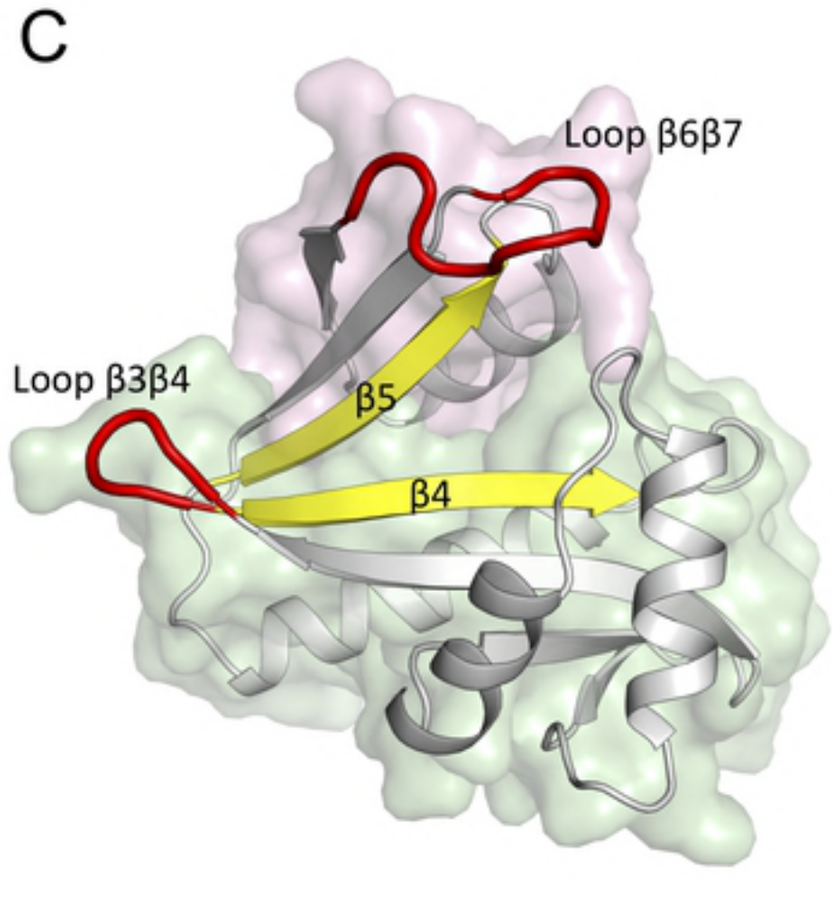

C
Loop β6β7
Loop β3β4
β5
β4

Fig5

Naa40 – 4u9v

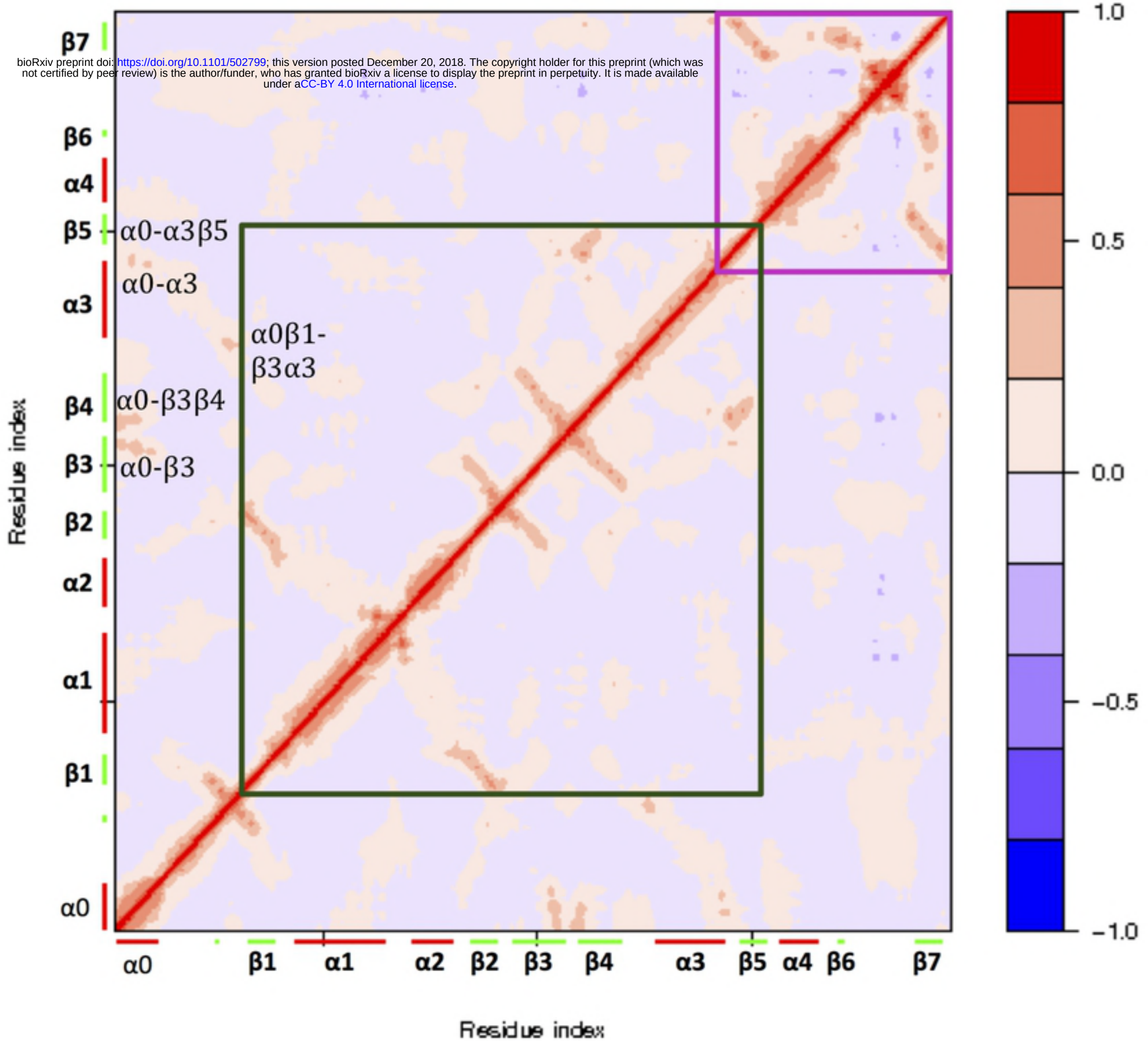


Fig6

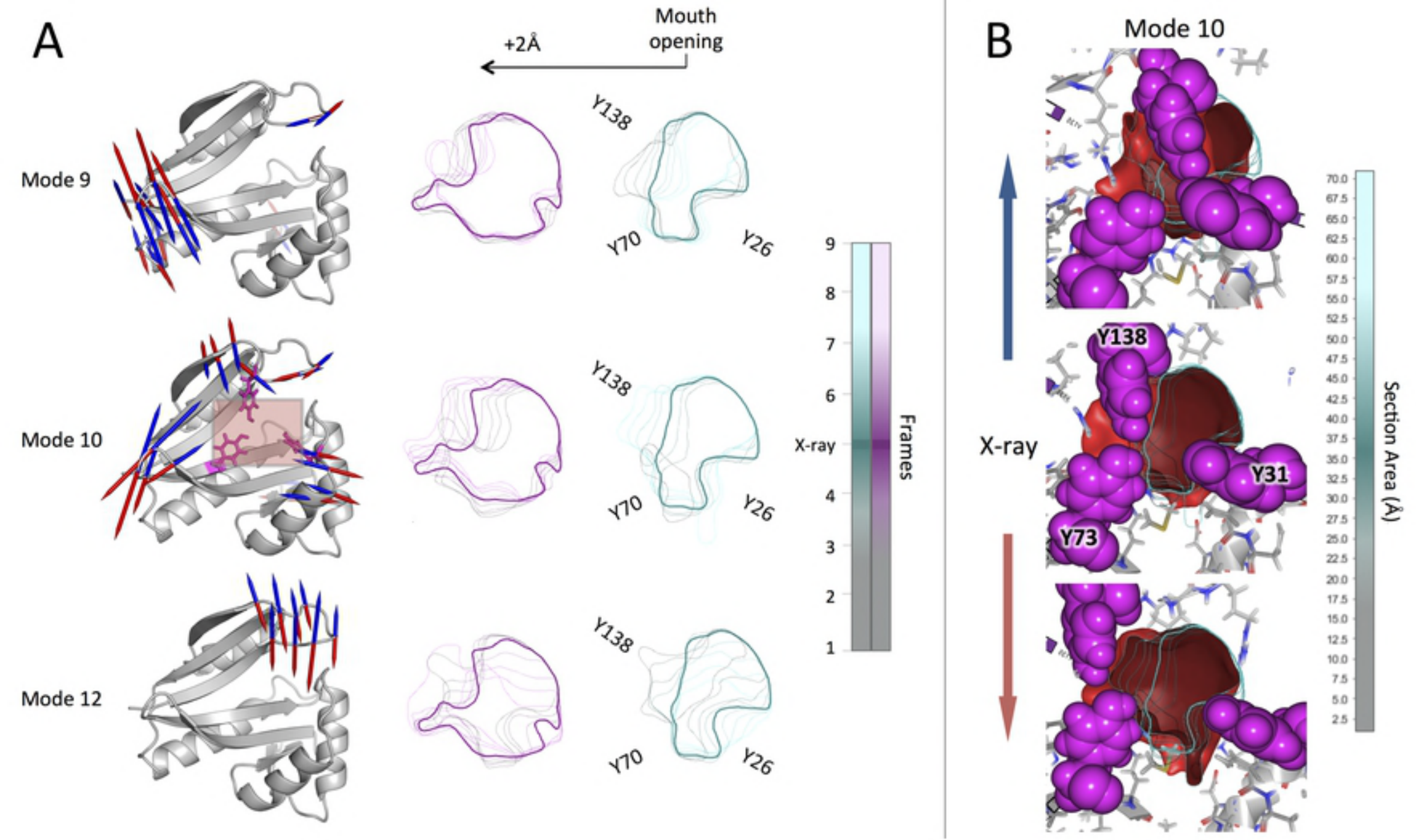

A
Mouth opening
+2Å
Mode 9
Mode 10
Mode 12
Y138
Y70
Y26
9
8
7
6
X-ray
4
3
2
1
Frames
B
Mode 10
X-ray
Y138
Y31
Y73
70.0
67.5
65.0
62.5
60.0
57.5
55.0
52.5
50.0
47.5
45.0
42.5
40.0
37.5
35.0
32.5
30.0
27.5
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
Section Area (Å)

Fig7

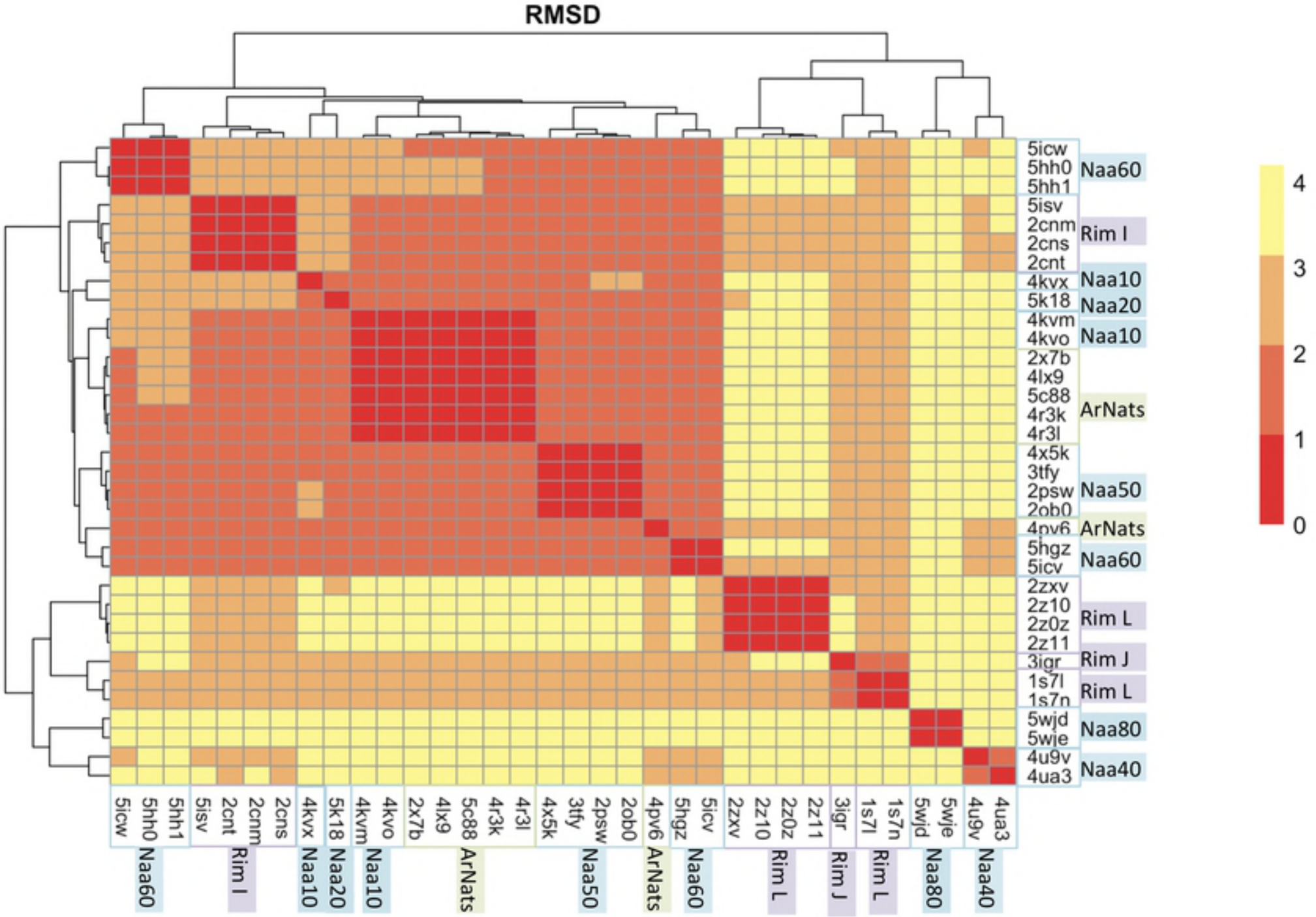

RMSD
5icw
5hh0
5hh1
Naa60
5isv
2cnm
2cns
2cnt
Rim I
4kvx
Naa10
5k18
Naa20
4kvm
4kvo
Naa10
2x7b
4lx9
5c88
4r3k
4r3l
ArNats
4x5k
3tfy
2psw
2ob0
Naa50
4pv6
ArNats
5hgz
5icv
Naa60
2zxv
2z10
2z0z
2z11
Rim L
3igr
Rim J
1s7l
1s7n
Rim L
5wjd
5wje
Naa80
4u9v
4ua3
Naa40
4
3
2
1
0

Fig3