Biophysics and Physicobiology

Regular Article

Case study on the evolution of hetero-oligomer interfaces based on the differences in paralogous proteins

Saki Aoto[1] and Kei Yura[1,2,3]

[1]*Graduate School of Humanities and Sciences, Ochanomizu University, Bunkyo, Tokyo 112-8610, Japan*
[2]*Centre for Informational Biology, Ochanomizu University, Bunkyo, Tokyo 112-8610, Japan*
[3]*National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan*

We addressed the evolutionary trace of hetero-oligomer interfaces by comparing the structures of paralogous proteins; one of them is a monomer or homo-oligomer and the other is a hetero-oligomer. We found different trends in amino acid conservation pattern and hydrophobicity between homo-oligomer and hetero-oligomer. The degree of amino acid conservation in the interface of homo-oligomer has no obvious difference from that in the surface, whereas the degree of conservation is much higher in the interface of hetero-oligomer. The interface of homo-oligomer has a few very conserved residue positions, whereas the residue conservation in the interface of hetero-oligomer tends to be higher. In addition, the interface of hetero-oligomer has a tendency of being more hydrophobic compared with the one in homo-oligomer. We conjecture that these differences are related to the inherent symmetry in homo-oligomers that cannot exist in hetero-oligomers. Paucity of the structural data precludes statistical tests of these tendencies, yet the trend can be applied to the prediction of the interface of hetero-oligomer. We obtained putative interfaces of the subunits in CPSF (cleavage and polyadenylation specificity factor), one of the human pre-mRNA 3'-processing complexes. The locations of predicted interface residues were consistent with the known experimental data.

Corresponding author: Kei Yura, Centre for Informational Biology, Ochanomizu University, 2-1-1 Otsuka, Bunkyo, Tokyo 112-8610, Japan.
e-mail: yura.kei@ocha.ac.jp

The majority of the proteins in the cell form macromolecular assemblies. About 80% of proteins in *Escherichia coli* are in the form more than dimer [1]. The situation is assumed to be more or less the same in eukaryotes, including humans. These macromolecular assemblies are known to be a central player in many cellular functions, and their structural analyses are the interest of structural biologists [2]. For the determination of the structure of macromolecular assembly, single-particle cryo-electronmicroscopy (cryo-EM) is widely employed [3]. The technique has applied to the determination of the whole structures of ribosome [4] as well as spliceosome [5,6]. Structures obtained by cryo-EM are, however, in low resolution compared with those obtained by X-ray crystallography. A widely accepted strategy for the improvement in the resolution of the structure is to build a hybrid structural data by combining the cryo-EM low-resolution data with atomic data of the subunit structures and interfaces [7]. A number of methods have been developed to fit the high-resolution subunit structures into the contour of cryo-EM data [7]. This procedure cannot be completed only with structural data of the subunits and the macromolecular assembly, because there remain high degrees of freedom in positioning the subunit into the contour. Molecular dynamic simulation of the subunits, pre-docking calculation of two or more subunits, homology modelling and other data from small-angle X-ray scattering measurement and NMR for example are required for obtaining the structural data [8]. Hence for the hybrid method, data obtained by measuring or predicting the structures of the subunits and their interactions are valuable [9].

In the discipline of computational biology, interactions between proteins have been studied for a long time and the characteristics of the interface for the protein-protein

interactions have been compiled [10–12]. The characteristics were found to be different between obligate (always oligomer) and non-obligate complexes in such a way that the interface is generally larger and more hydrophobic in the former. The large interface seems to undergo conformational change upon complex formation. Janin *et al.* [13] summarized that the majority of the interface size was between 1,200 and 2,000Å², that the residues were hydrophobic and that the residue conservation was generally high. In addition, Hashimoto *et al.* [14] and Nishi *et al.* [15] reported that the residue insertion in the interface should have played an important role in oligomer formation. These characteristics can play important roles in predicting interfaces for protein-protein interactions, and hence may help structure fitting in cryo-EM data.

Most of the macromolecular assemblies are homo-oligomers [1], and hence most of the computational studies have been performed on homo-oligomer structures. However, there are a non-negligible number of hetero-oligomers in macromolecular assemblies and the knowledge of characteristics in the hetero-oligomer interfaces should also be required to grasp the whole picture of macromolecular assemblies. The paucity in the study of the interface of hetero-oligomers has been due to the fact that the number of hetero-oligomers with known three-dimensional structures was far smaller than that of homo-oligomers. However the increase in size of Protein Data Bank (PDB) [16] may allow us to study the interface of hetero-oligomers computationally.

We, therefore, computationally addressed the question on hetero-oligomer interfaces. Specifically we traced the presumptive evolutionary pathway from monomer to oligomer or small to big oligomers on both homo- and hetero-oligomers. This procedure can be carried out by identifying paralogous proteins in different oligomeric states in PDB. We then extracted information relatively unique to hetero-oligomer interface formation. We further tried applying the obtained data to hetero-oligomer interface prediction. The result of the prediction may be verified in the independently on-going structure determination study.

Methods

Retrieving paralogous proteins with different subunit composition from PDB

The process to find paralogues with different oligomer states is summarized in Figure 1. Initially, we ran BLASTClust [17] against the whole amino acid sequences of proteins in PDB [16] with 25% identity threshold and 90% coverage cut-off, to identify potential homologous sets. Amino acid sequences in each set were then linked to an entry in UniProt [18] using DBREF record in PDB entry and its biological oligomer state was retrieved from SUBUNIT record. When DBREF record was missing in PDB, we ran BLAST [17] for searching uniprot_sprot.fasta for an entry of a high score (sequence identity >95% and coverage >95%) derived from
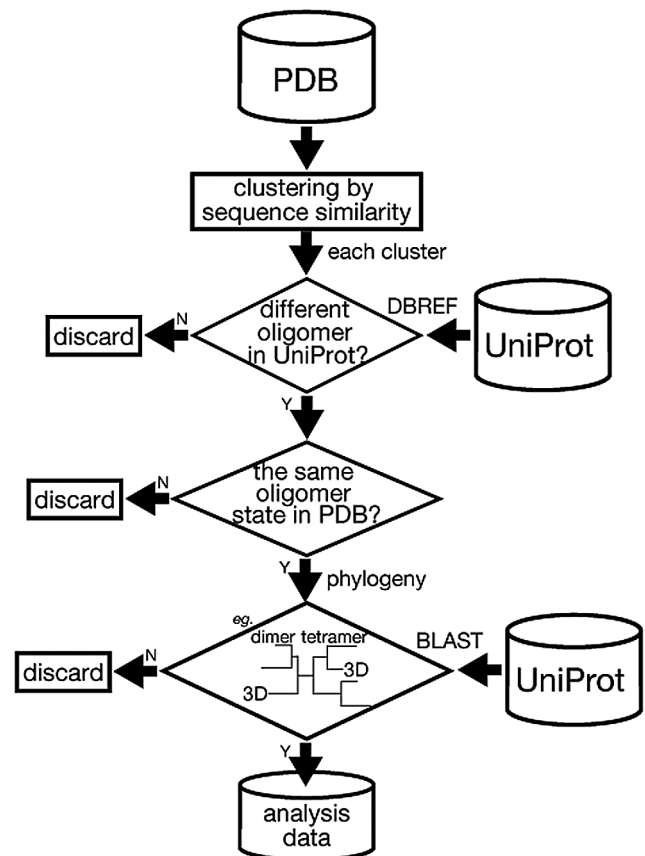

Figure 1   Procedure for data selection from the Protein Data Bank. The detail is described in Methods section.

the same species as the one in PDB. Finally a potential homologous set with entries having different SUBUNIT record in UniProt was selected as a candidate set of paralogous proteins with different oligomer state. The definition of paralogue here is not as strict as the one used in the study of molecular evolution. We used the term paralogue in such a case that two or more groups of proteins with different biological functions are apparently derived from a common ancestor.

The SUBUNIT record of UniProt in the entry did not always correspond to the oligomer state of the corresponding entry in PDB. We checked all the possible oligomer state of each PDB entry in the candidates of paralogous proteins by scrutinizing the oligomer states of author definition, of the asymmetric unit and of the biological unit, and if none of the oligomer states in PDB met the description in SUBUNIT record of UniProt, then we discarded the candidate set. The description of oligomer state in UniProt was sometimes inconsistent in a way that evident orthologues were annotated as being in different oligomer state. The contradiction may suggest that the protein should be actually in equilibrium of different subunit states or the subunit interactions should change in different conditions. In this case, we scru-

tinized the literatures and tried to set a consensus of the oligomer state.

A phylogenetic tree of each candidate set was drawn by the neighbour-joining method [19], based on the multiple sequence alignment of homologous sequences retrieved from UniProt by BLAST with an E-value threshold of 10^{-4}. The alignment was performed on ALAdeGAP [20]. If two or more distinct clusters were found in the tree with, at least, one sequence derived from PDB in each cluster, then the candidate set was selected as an analysis target.

Identification of subunit interface and surface residues

Residues at the subunit interface were detected by differences in accessible surface areas of the protein chain when the area was calculated in an oligomer state and a monomer state, as we previously employed in detecting protein-ligand interface [21]. We calculated the accessible surface area by the in-house program that is based on the method of Shrake and Rupley [22]. The accessibility of each residue was calculated based on the method described by Go and Miyazawa [23]. If the difference between the accessibility of a residue calculated on an oligomer state and that on the monomer state is no less than 0.03 (the detection limit of the in-house program), then the residue was defined as the one in the interface. If the difference is less than 0.03 and the accessibility on the oligomer state is more than 0.1, then the residue was defined as the one on the surface.

Assessing changes in the interface residues

Types of changes in the interfaces between the paralogues were classified as shown in Figure 2. In Figure 2, we assumed that one of the paralogues (paralogue B) is either homo- or hetero-dimer, hence this classification is not exhaustive, but in retrospect, the types shown in Figure 2 were sufficient for describing the data set we used in the present study.

Once the interface was identified, amino acid conservation on the interface could be measured. Conservation of amino acid residues at the interface was measured in two different scales. A first scale measured the variability of amino acid types amongst the orthologues and amongst paralogues by the following equation based on Shannon's entropy [24];

$$-\sum_{n=1}^{21} p_n^i \times \log_2 p_n^i - < -\sum_{n=1}^{21} p_n^s \times \log_2 p_n^s >_s ,$$

where p^i was a frequency of certain amino acid type in the multiple sequence alignment at a specific site of the interface and p^s was of the surface. The suffix n was defined from one to 21, because in this study, a gap was treated as the 21st amino acid type. The second term in the equation described the average evolutionary rate of the surface of the protein. The second term could level the measure amongst the proteins with different evolutionary rate. We called the scale ΔConservation.

A second scale measured the average chemical type of the residues amongst the orthologues and paralogues using the
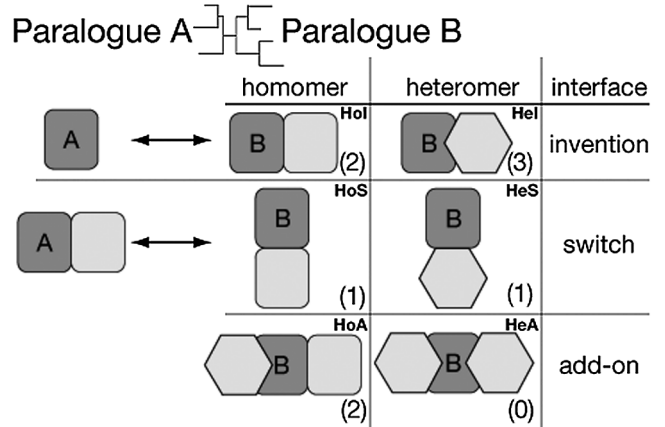

Figure 2 Classification of the differences in interface between paralogues. Number in parentheses is the count in Table 1. Each type is named based on its oligomer state and the putative evolutionary pathway of acquiring the interface. For example HoI stands for "homomer" and "Invention."

hydropathy index of Kyte and Doolittle [25]. The index ranged between –4.5 and 4.5. We assigned a value –5.0 to a "gap residue." A mean of the index was assigned to each site in the interface. We called the scale Hydropathy Index.

Results and Discussion

Potential paralogues with different subunit composition in PDB

We identified nine sets of entries (Table 1) following the procedures described in the methods section. The type of subunit changes can be classified based on the taxonomy depicted in Figure 2. Five out of nine sets were categorized to a transition from a monomer to multimer state. For the sake of clarity in the description, we will refer this type and other types using the term transition, but this does not indicate that we found any evidence of the evolution from the monomer to the multimer state. These five members were further divided into homomer-invention (HoI) and heteromer-invention (HeI) types. The remaining four cases were classified to homomer-add-on (HoA) type, homomer-switch (HoS) type and heteromer-switch (HeS) type (Table 2).

The dataset in Table 1 is apparently small compared with the ones in the previous similar studies such as the dataset by Hashimoto and Panchenko [14] and the one by Nishi *et al.* [15]. The difference seemingly derived from the difference in the aim of the study. The dataset of Hashimoto and Panchenko [14] was concentrated on homodimers. Nishi *et al.* [15] aimed to study the interface of homo-oligomers and the dataset was focused on homo-oligomers. Our dataset was rather focused on homo-/hetero-oligomers with paralogues, namely proteins likely derived from a common ancestor with different biological functions. In addition to the conceptual

Table 1   Paralogous proteins with different subunit composition

| set | Paralogue A | | | | | | Paralogue B | | | | |
---	Description*	identity (%)**	multimer	UniProtID	PDB	chain	Description*	identity (%)**	multimer	UniProtID	PDB	chain
1	deoxyribonucleoside kinase (5)	51±24	monomer	DNK_DROME	2JJ8	B	deoxycytidine kinase (8)	56±25	homodimer	DCK_HUMAN	2NO0	A
2	pitrilysin (11)	43±36	monomer	PTRA_ECOLI	1Q2L	A	insulin-degrading enzyme (15)	40±23	homodimer	IDE_HUMAN	2G47	A
3	HisA (535)	40±10	monomer	HIS4_THEMA	1QO2	A	HisF (529)	50±11	heterodimer	HIS6_THEMA	1GPW	A
4	L-aspartate oxidase (37)	36±12	monomer	NADB_ECOLI	1KNP	A	succinate dehydrogenase SdhA (30)	66±13	heterotetramer	SDHA_PIG	3AE9	A
5	myoglobin (104)	69±20	monomer	MYG_HUMAN	3RGK	A	hemoglobin beta chain (289)	62±16	heterotetramer	HBB_HUMAN	4HHB	B
6	threonine synthase (20)	36±13	homodimer	THRC_MYCTU	2D1F	A	cystathionine beta-synthase (33)	36±13	homotetramer	CBS_HUMAN	1M54	A
7	D-cysteine desulfhydrase (72)	56±23	homodimer	DCYD_SALTY	4D9F	D	tryptophan synthase beta chain (364)	52±15	heterotetramer	TRPB_SALTY	1K7F	B
8	MoeA (30)	28±18	homodimer	MOEA_ECOLI	2NQK	A	gephyrin (5)	57±36	homotrimer	GEPH_HUMAN	1JLJ	A
9	GatD (34)	47±12	heterodimer	GATD_METTH	2D6F	A	L-asparaginase I (25)	34±15	homotetramer	ASPG1_ECOLI	2HIM	A

*   The number in the parentheses is the number of homologous amino acid sequences retrieved from Uniprot.
**   A mean sequence identity with a standard deviation in each category.

Table 2   Comparison between the paralogues

set	Comparison			Separation Probability****	
	identity (%)**	Calpha RMSD	Interface***	ΔConservation	Hydropathy
1	32±4	1.63 Å (143 aa)	HoI	0.10	0.38
2	19±5	2.05 Å (428 aa)	HoI	0.20	0.06
3	24±3	2.22 Å (226 aa)	HeI	0.02	0.05
4	28±3	1.90 Å (499 aa)	HeI	0.22	0.01
5	25±3	1.47 Å (145 aa)	HeI	0.08	0.04
6	11±2	2.69 Å (290 aa)	HoA	0.06	0.20
7	14±2	2.73 Å (197 aa)	HoA	0.03	0.49
8	24±8	2.87 Å (135 aa)	HoS	0.11	0.40
9	24±3	1.96 Å (307 aa)	HeS	0.01	0.30

***   The classification of the interface is depicted in Figure 2.
****   Separation probability is calculated from the values of student-t along the X and Y axes in Figures 3 to 9.

reasons, we also encountered difficulty in making clear decision in the phylogeny step in Figure 1. Many phylogenetic trees have turned out hard to decipher and we inevitably discarded those data. The inclusion of those data will be our future work. Another note should be made on conformational difference between the protein in a monomer state and in an oligomer state. Conformation change in a protomer is expected when the structures of proteins in a monomer and an oligomer states are compared. In the current study, we checked the conformational change one by one and did not performed systematic study on it. The conformational change did not affect the size of the database in a retrospect. The systematic study will be our future work, too.

Detail of the interface patterns in each type
Homomer-invention (HoI) type

Two sets of enzymes are included in this category; a pair of deoxyribonucleoside kinase (dNK) and deoxycytidine kinase (dCK), and a pair of pitrilysin and insulin-degrading enzyme (Table 1).

dNK and dCK. dNK phosphorylates a variety of deoxyribonucleosides and demonstrates its activity in monomer state [26]. On the other hand, dCK is a deoxycytidine specific kinase and function in homodimer [27]. In PDB, the asymmetirc unit of dNK contains four subunits [28], but based on the experimental results, we singled out the longest chain (chain B) as a representative structure. The structure of dCK was determined in homodimer [29]. The superposition of the two structures depicted a facet of interface which contains 35 amino acid residues (Fig. 3A). The changes in properties of these residues were shown in Figure 3B. Most of the dots are distributed around 0.0 of ΔConservation, indicating that the degree of conservation of the residues in the interface of homodimer dCK is similar to the ones in the
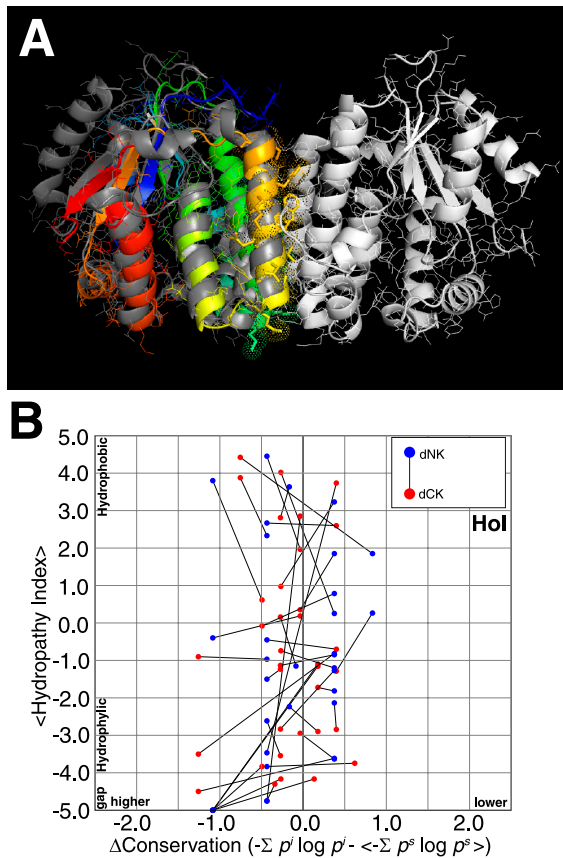
Figure 3 HoI (Homoer-Invention) type transition from dNK to dCK. A. Superimposition of the three-dimensional structures of dNK [28] and dCK [29]. dNK is a monomer enzyme and shown by a ribbon model in rainbow colour. dCK is a dimer and shown by a ribbon model, one subunit in grey and the other subunit in white. The subunit in grey was superimposed with dNK. The subunit interface residues corresponding in dNK were shown in dot surface and stick models. The superimposition calculation was performed with CE [30] and the result was visualized by PyMOL [31]. B. ΔConservation-Hydropathy Index plot of dNK and dCK. The horizontal axis is the value of ΔConservation. Zero means no difference between the surface and the interface, plus means that the residue in the interface is less conserved than the surface, and minus means that the residue in the interface is more conserved that the surface. The vertical axis is Hydropathy Index. Each dot represents an amino acid position in the interface of dCK. A dot in blue derives from dNK (paralogue A in Table 1) and a dot in red from dCK (paralogue B in Table 1). The Corresponding residues from both enzymes in the multiple sequence alignment are connected by an edge.

surface and the ones in the monomer dNK. However, there are ten exceptions that protrude in the left side of the graph. Of the ten, three red dots are connected to the blue dots in the right side, indicating that the residues are uniquely conserved in dimer interface of dCK. Out of the remaining seven blue dots, five dots are "gap residues" and the dCK has insertion residues against dNK. The remaining two dots are connected to the red dots in the right side.

The conservation of interface residues has been pointed out by many studies and reviewed by Janin *et al.* [13]. But in

this case study, the corresponding residues in dNK and in dCK are conserved in a similar manner and the distribution of dots in Figure 3B overlaps. We found three residues that improved ΔConservation by oligomerisation. These may be the ones called "hot spot," residues important for the stability of oligomer. Janin *et al.* [13] also summarized the hydrophobic tendency of the interface residues. If this is the case, then red dots tend to appear in the top of the graph and blue dots tend to appear on the bottom. But both red and blue dots are scattered along the vertical axis in Figure 3B. The importance of insertion residues in oligomerisation was pointed out by Hashimoto and Panchenko [14] and Nishi *et al.* [15]. In our case study on dNK/dCK, the insertion evidently played important role in interface invention, but the residues in the insertion are relatively less conserved.

Pitrilysin and insulin-degrading enzyme. Pitrilysin is a zinc metalloendopeptidase specific to a small peptide and active as a monomer, while a mammalian homologue, insulin-degrading enzyme, is a homodimer [32]. The structures of these enzymes are formed by two domains, namely N-terminal and C-terminal domains, and the structure of pitrilysin was determined in a relatively open conformation compared to the one of insulin-degrading enzyme. The dimer interface of insulin-degrading enzyme was found in the C-terminal domain, hence the structures were superimposed using C-terminal domain (Fig. 4A). Note that the ΔConservation-Hydropathy Index plot in Figure 4B shows four clear characteristics in interface. A first point is that there exist two putative "hot spot" residues protruding to the left in the graph. This characteristic is the same as the previous pairs up above. A second intriguing point is that many of the interface residues in insulin-degrading enzyme are insertion residues against pitrilysin. This is depicted by the localization of blue dots on the horizontal axis in Figure 4B. The characteristics suggested by Hashimoto *et al.* [14] and Nishi *et al.* [15] was well observed in this enzyme pair, too. A third point is that the interface residues had no clear tendency of change to hydrophobic amino acids. If it were, then the edge should have a tendency to go up in the plot with a blue dot on the bottom and a red dot on the top. On the contrary, the tendency shown in Figure 4B is the opposite; that is the edges tend to go down with blue dots on the top and red dots on the bottom, except the dots involving gaps in the alignment. And a fourth point is that many red dots are found on the right side (positive side) of the graph, indicating that the conservation is low compared with the residues on the surface.

Heteromer-invention (HeI) type

Three sets of proteins are classified in this category; a pair of HisA and HisF, a pair of L-aspartate oxidase and succinate dehydrogenase, and a pair of myoglobin and hemoglobin beta (Table 1).

HisA and HisF. HisA is phosphoribosyl isomerase A and located in the fourth step of histidine biosynthesis pathway. HisA was shown to have $(\beta\alpha)_8$-barrel fold [34] and to
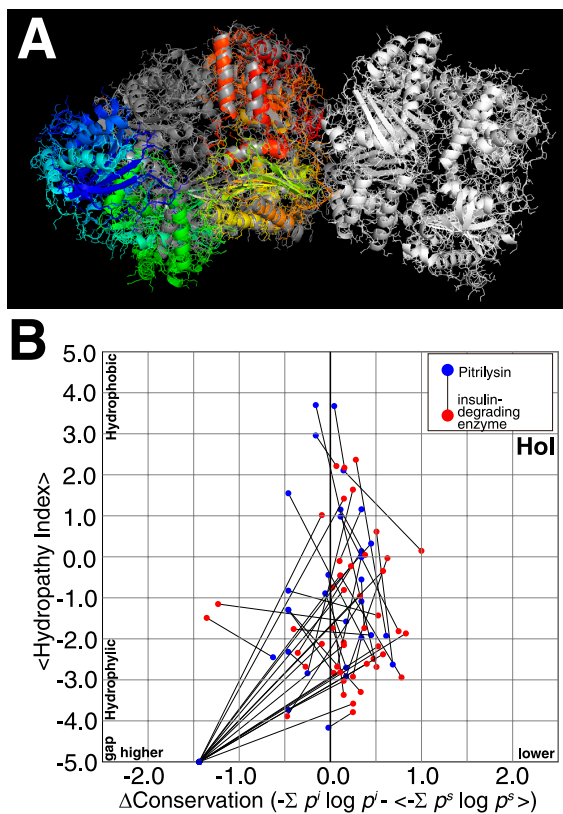
Figure 4 HoI (Homoer-Invention) type transition from pitrylysin and insulin-degrading enzyme. A. Superimposition of the three-dimensional structures of pitrylysin and insulin-degrading enzyme [33]. See the caption of Figure 3A for the colour. B. ΔConservation-Hydropathy Index plot of pitrylysin and insulin-degrading enzyme.

Figure 5 HeI (heteromer-invention) type transition from HisA to HisF. A. Superimposition of the three-dimensional structures of HisA [34] and HisF [37]. See the caption of Figure 3A for the colour. B. ΔConservation-Hydropathy Index plot of HisA and HisF.

function as a monomer in *Escherichia coli* and *Thermotoga maritima* [35]. HisF is located in the fifth step of histidine biosynthesis pathway and catalyzes the reaction of an imid-azole glycerol phosphate synthesis. The enzyme was shown to form a heterodimer with HisH, another subunit for the reaction catalysis [36] and the three-dimensional structure was determined in the complex [37]. The overall three dimensional structure of HisA and HisF was superimposed and HisH was found on the top of the barrel structure without any overlap with HisA (Fig. 5A). ΔConservation-Hydropathy Index plot (Fig. 5B) showed some peculiar differences from the ones in HoI types. There were 36 residue positions involved in the subunit interface and most of them (red dots) tend to locate in the upper area of the plot compared with the corresponding residues in HisA (blue dots) connected by edges. The separation in the distribution of the blue and red dots along the vertical axis is statistically significant (Wilcoxon test, $p < 10^{-3}$) (Table 2). Another tendency in the plot is that there are a number of red dots in the lower left side. The number of the red dots protruding to the left is apparently greater than the ones in HoI. These results demonstrate that interface residues tend to be hydrophobic and con-
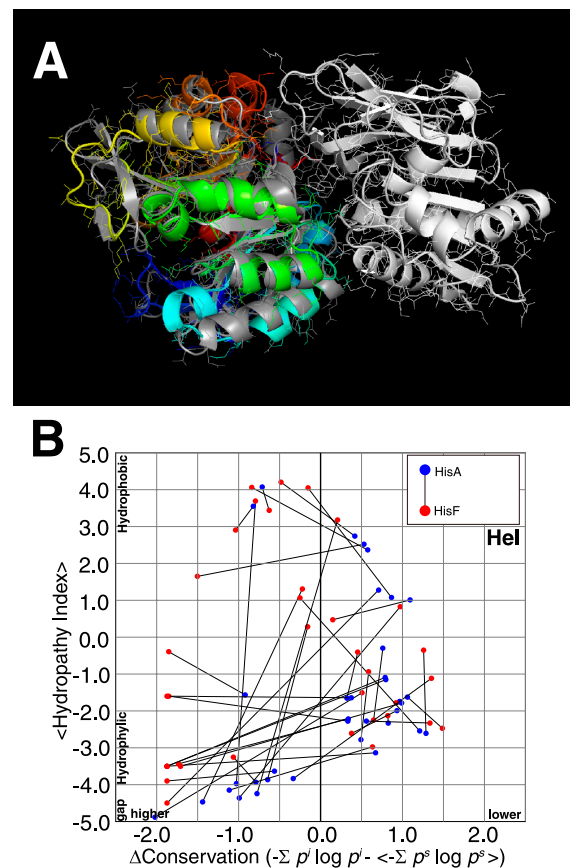
served against the corresponding non-interface residues. The interface characteristics are different from the ones of homo-oligomer.

L-aspartate oxidase and succinate dehydrogenase. L-aspartate oxidase is a flavoprotein involved in cofactor biosynthesis pathway and its activity was shown to perform in a monomer state in *E. coli* [38] and the three-dimensional structure was also determined as a monomer [39]. Succinate dehydrogenase is a huge enzyme complex involved in Krebs cycle. The protomer consists of four subunits, namely succinate dehydrogenase flavoprotein subunit (SdhA), succinate dehydrogenase iron-sulfur subunit (SdhB), and cytochrome b_{560} (SdhC, SdhD) in a cell [40]. The protomer further forms a trimer and the entire structure was already determined [41]. SdhA and L-aspartate oxidase are homologous and the three-dimensional structures are similar (Fig. 6A). SdhA forms a direct interaction with SdhB, hence the analyses here was performed over the heterodimer of SdhA and SdhB. On the ΔConservation-Hydropathy Index plot in Figure 6B, there are 78 residue positions and most of the blue dots derived from the residues of L-aspartate oxidase are located
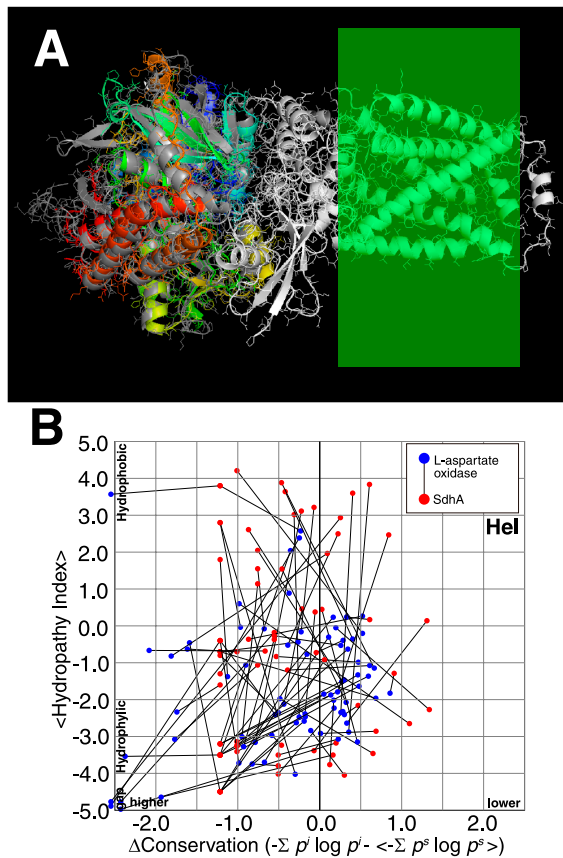
Figure 6 HeI (heteromer-invention) type transition from L-aspartate oxidase to SdhA. A. Superimposition of the three-dimensional structures of L-aspartate oxidase [39] and SdhA [41]. See the caption of Figure 3A for the colour. The green translucent cover indicates a cell membrane. B. ΔConservation-Hydropathy Index plot of L-aspartate oxidase and SdhA.
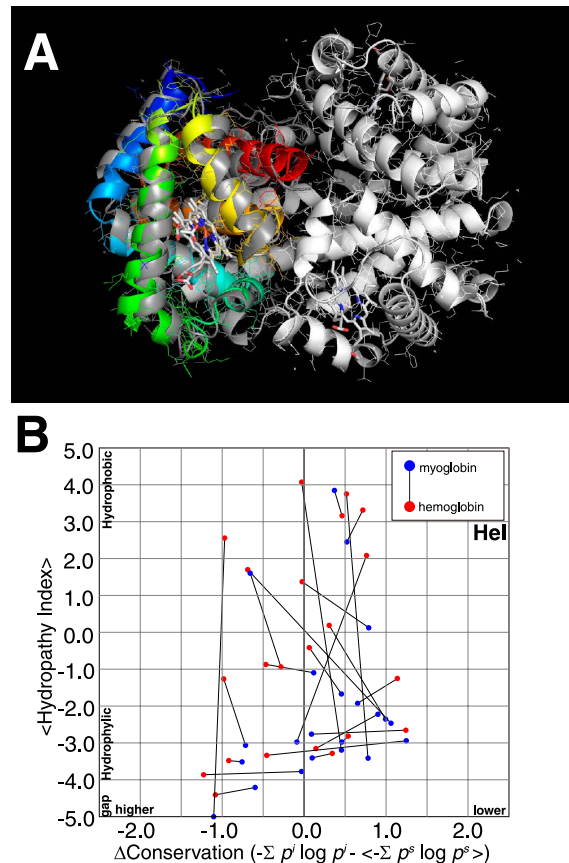

Figure 7 HeI (heteromer-invention) type transition from myoglobin to hemoglobin beta. A. Superimposition of the three-dimensional structures of human myoglobin [43] and human hemoglobin beta subunit [44]. See the caption of Figure 3A for the colour. B. ΔConservation-Hydropathy Index plot of myoglobin and hemoglobin beta subunit.

in the centre area surrounded by red dots. Especially in the upper area of the graph, the red dots are evident compared with the blue dots, indicating that the interface residues are hydrophobic.

Myoglobin and hemoglobin beta. Subunit interface of globin proteins has been studied extensively. Shionyu *et al.* extensively analyzed the interface of different globin proteins and found that a subtle difference of amino acid residues should have brought a drastic change in oligomer state in globin family [42]. We analysed here the interface difference between myoglobin, the monomer globin protein and the hemoglobin beta subunit, the closest homologue to myoglobin (Fig. 7). There are 22 residues in the interface of beta subunit to the other subunits and they also showed tendency similar as the other ones, namely increase in hydrophobicity, but the tendency was not statistically supported (Table 2).

Comparison between HoI and HeI.

The trends found in HeI type were different from the ones

in HoI type. The general difference in types of amino acid found in the interface has been reported a number of times [13] and interfaces of homodimers was reported to be less hydrophobic than the ones in heterodimer [45]. In this case study, by comparing the result of HoI and HeI, we suggest that the changes in amino acid types of interface for homomer should be within the range of the counterpart in the monomer state except for a few "hot spot" residues and that the changes for heteromer should be drastic especially in its hydrophobicity. The statistical significance of the differences was not clear, but the trends can be clearly found by calculating the separation probability of the dot distribution on the ΔConservation-Hydropathy Index plot along the ΔConservation axis and along the Hydropathy axis (Table 2). The probabilities of the separation in ΔConservation of HoI are generally higher than those of HeI, and the probabilities of the separation in Hydropathy Index are less than 5% in HeI. The difference can also be found in the role of insertion in the interfaces. In the homomer, insertions seemed to play important roles, but in heteromer, the contribution of insertion was less apparent.

Homomer add-on (HoA) type

Two sets of proteins are classified in this category; a pair of threonine synthase and cystathionine beta-synthase, and a pair of D-cysteine desulfhydrase and tryptophan synthase beta (Table 1). These four enzymes belong to the same super-family, the family of PLP-dependent enzyme. We decomposed the family into two groups based on the similarity of the amino acid sequences.

Threonine synthase and cystathionine beta-synthase. Threonine synthase derived from a bacterium *Corynebacterium glutamicum* was characterized as a monomer enzyme [46], but the one from another bacterium *Mycobacterium tuberculosis* as a dimer [47], and the one from *Arabidopsis thaliana* as a dimer [48]. This discrepancy in the oligomer state may be authentic, because the oligomer state could be found different in the orthologous proteins [13]. Cystathionine beta-synthase was characterized as tetramer [49]. Hence with this set of proteins, we can at least analyse the transition from homo-dimer to homotetramer (Table 1). When one subunit of the dimer of threonine synthase was superimposed to one sub-unit of the tetramer of cystathionine beta-synthase, the other subunit in the dimer of threonine synthase was automatically superimposed to another subunit of cystathionine beta-synthase, indicating that the tetramer interactions include the dimer interface (Fig. 8A). From the perspective of threonine synthase, cystathionine beta-synthase gained a new interface from the surface of threonine synthase to transform to tet-ramer. The changes in amino acid residues are shown in ΔConservation-Hydropathy Index plot in Figure 8B. A pair of two dots connected by an edge corresponds to a homologous pair of residue sites in the multiple sequence alignment of both enzymes, and the residue is on the surface of thre-onine synthase (blue dot) and in the interface of cystathi-onine beta-synthase (red dot). To consider a change specific for this transition, trends in interface can be a reference in the same proteins and shown in Figure 8C. In this figure, trends in threonine synthase subunit interface, which is also a subunit interface in cystathionine beta-synthase are shown. As expected from the trends in HoI, there are no clear tendencies in the direction of the edge and the location of dots except for a few red dots protruding to the left of the graph.

D-cysteine desulfhydrase and tryptophan synthase beta. Contrary to the result of the pair of threonine synthase and cystathionine beta-synthase, a pair of D-cysteine desulfhy-drase and tryptophan synthase beta had much clearer tendency in the interface. D-cysteine desulfhydrase catalyses a degradation reaction of D-cysteine with PLP in some specific bacteria [51], and considered to be a defence mechanism against D-cysteine. The enzyme appears to form a homo-dimer [51]. Crystal structure of the enzyme also forms a homodimer [52]. Tryptophan synthase beta subunit, together with an alpha subunit, is located at the terminal of trypto-phan biosynthesis pathway and catalyses the last reaction in the pathway, that is the synthesis of L-tryptophan from indole and L-serine. Alpha and beta subunit forms a hetero-
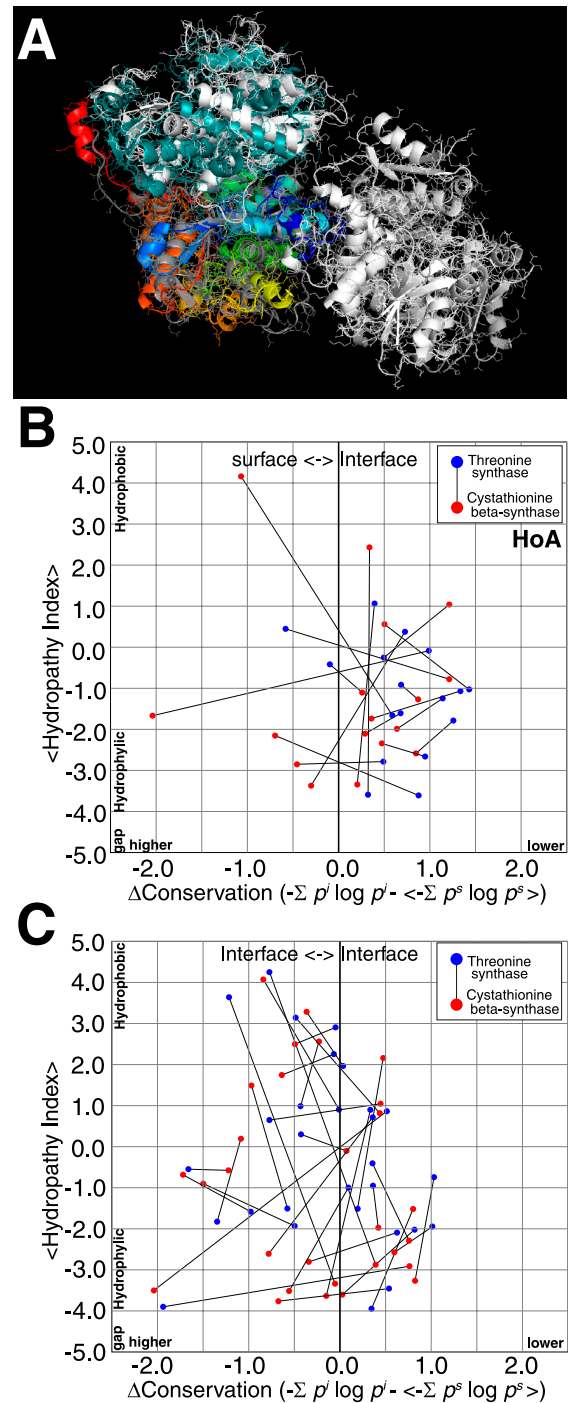

Figure 8 HoA (homomer-add-on) type transition from threonine synthase to cystathionine beta-synthase. A. Superimposition of the three-dimensional structures of threonine synthase homodimer [47] and cys-tathionine beta-synthase homotetramer [50]. All the subunits are shown by ribbon model. One subunit of threonine synthase is coloured in rain-bow, and the other subunit in deep cyan. The subunit of cystathionine beta-synthase superimposed to the subunit of threonine synthase is coloured in grey and all other subunits of the tetramer were coloured in white. B. ΔConservation-Hydropathy Index plot of threonine synthase and cystathionine beta-synthase. C. ΔConservation-Hydropathy Index plot for an amino acid residue site at all the interface in cystathionine beta-synthase. This is considered to be background behaviour of amino acid residue site to evaluate the behaviour of dots in B.
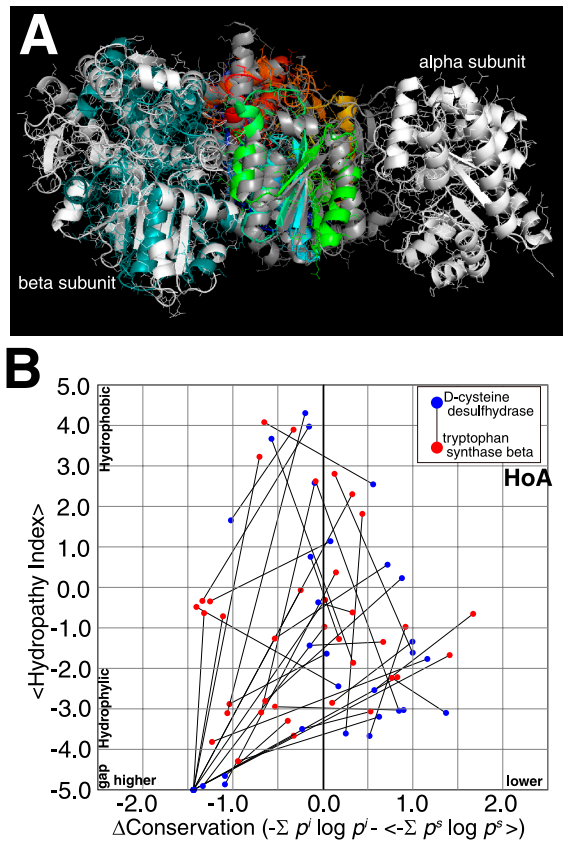
Figure 9 HoA (homomer-add-on) type transition from D-cysteine desulfhydrase to tryptophoan synthase beta. A. Superimposition of the three-dimensional structures of D-cysteine desulfhydrase homodimer [52] and tryptophan synthase alpha and beta heterotetramer [54]. For clarity, one of the alpha subunits of tryptophan synthase is deleted from the figure. When the beta subunit in grey is superimposed to one of the subunits in D-cysteine desulfhydrase, the other beta subunit in tryptophan synthase locates on the other subunit of desulfhydrase (deep cyan). See the caption of Figure 8A for the colour. B. ΔConservation-Hydropathy Index plot of D-cysteine desulfhydrase and tryptophan synthase beta subunit.

Figure 10 HoS (homomer-switch) and HeS (Heteromer-switch) types transition. A. Superimposition of the three-dimensional structures of MoeA [55] in rainbow colour and deep cyan, and gephyrin [56] in grey and white. B. Superimposition of the three-dimensional structures of GatD heterodimer [60] in rainbow colour and deep cyan, and L-asparaginase I [61] in grey and in white. C. Superimposition of the three-dimensional structures of GatD heterotetramer [60] in rainbow colour and deep cyan, and L-asparaginase I [61] in grey and white. Note that one of the subunits of GatD in deep cyan is in complete fit to a subunit of L-asparaginase I in white.

dimer and can complete the reaction, but alpha-beta-beta-alpha tetramer increases the reaction rates by one to two orders of magnitude and this tetramer is considered as a biological unit [53]. When the structures of both enzymes were compared, the dimer form of D-cysteine desulfhydrase is almost identical to the beta dimer in tryptophan synthase tetramer (Fig. 9A). Therefore, the configuration of tryptophan synthase beta subunit dimer is effectively the same as the one of D-cysteine desulfhydrase and tryptophan synthase beta subunit seems to acquire a new interface for alpha subunit at the time of evolving tryptophan biosynthesis pathway (Fig. 9A). Hence the interface is a heteromer interface and is expected to share trends found in HeI. As expected, hydrophobicity is in the tendency of increase and there are quite a number of red dots on the left side of the graph. The role of insertion residues is noticeable in this case, which was rather peculiar in HoI than HeI in the observation above.
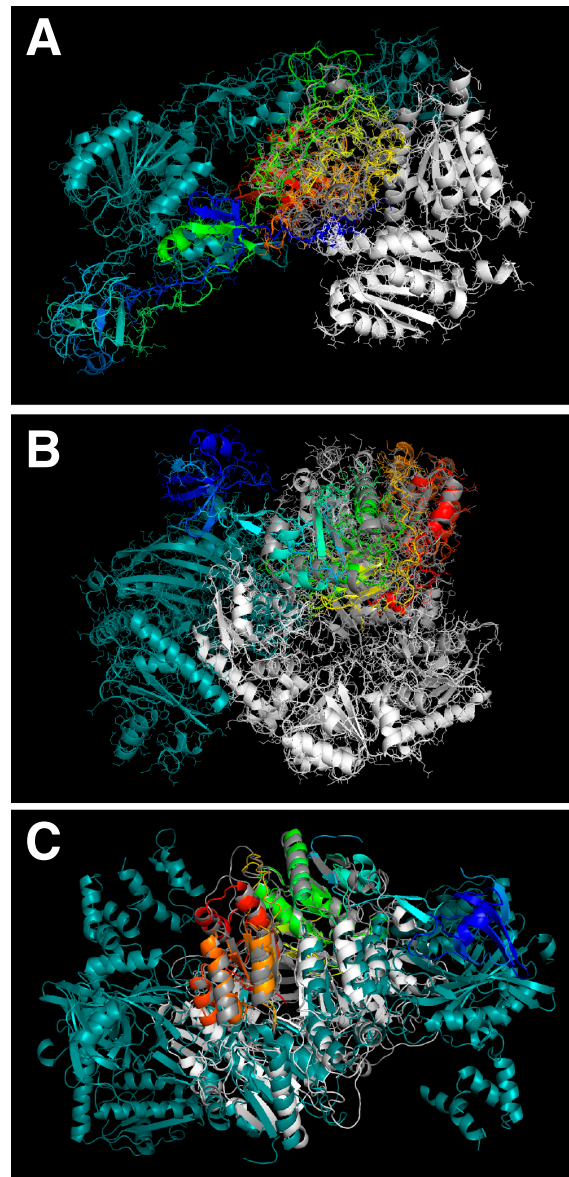
Homomer switch(HoS) type

MoeA and gephyrin. In a pair of MoeA and gephyrin, we found a complete switch of subunit interface (Fig. 10A). MoeA (rainbow and deep cyan in Figure 10A) catalyses the insertion of molybdate into molybdopterin and function as a homodimer [55]. Gephyrin is a multifunction protein initially considered to associate with a function of cytoskeleton,

but later found to have inhibitory function for glycine receptors [57]. The structure was solved both in its whole sequence (homodimer) [58] and in the C-terminal domain named G domain (homotrimer) only [56], and both forms seem to have biological implication. The posture of the homodimer is the same as the one of MoeA, but those of homotrimer is completely different (Fig. 10A). G domain of gephyrin is involved in both dimer interaction and trimer interaction, but the interface is completely different.

Heteromer switch(HeS) type

Archaeal glutamyl-tRNA amidotransferase D chain (GatD) that catalyses a conversion reaction from Glu to Gln on tRNA(Gln), is homologous to L-asparaginase I/II that catalyses a conversion reaction of Asp to Asn. GatD forms a heterodimer with GatE [59] and forms a complex with tRNA(Gln) [60]. On the other hand, L-asparaginase I forms a homotetramer, a dimer of dimers [61]. When the two structures were superimposed, a slight overlap between GatE and another subunit of L-asparaginase I was observed (Fig. 10B), but the interfaces are mostly on a different area, hence a different protomer interacts with different posture. An intriguing finding in this combination was that in the crystal, GatDE formed a dimer of dimers and the dimer interface of GatD is quite similar to the ones for L-asparaginase I (Fig. 10C). When a protomer of GatD and L-asparaginase I was superimposed, then the other GatD in the crystal and another L-asparaginase I subunit automatically superimposed themselves.

General trends in hetero oligomer interface

Some of the characteristics in oligomer interface as a whole were analyzed by Janin *et al.* [13]. The characteristics specific for homo-oligomers have also been noted in the previous studies by Hashimoto *et al.* [14] and Nishi *et al.* [15]. Briefly, the characteristics are higher conservation of amino acid residues, higher hydrophobicity and a significant role of insertion. In the present study, we extended the study of the characteristics to hetero-oligomer and elucidated the applicability of those trends to the heteromer interfaces. We applied the ΔConservation-Hydropathy Index plot to the study of interface and found that heteromer interface (HeI) should be peculiar in its hydrophobicity and higher conservation. The interface of homomer (HoI) has somewhat similar tendency in hydrophobicity to the one on the surface, but should be peculiar in the existence of a few "hot spot" residues. The significance of insertion residues in the interface seems different in homomer and heteromer, too. In the homomer, insertions seem to play important roles [14,15], but in heteromer, the contribution of insertion is less apparent. These results further suggests that, for instance, the strategy of interface prediction for homo-oligomer and hetero-oligomer should be different and that stress should be put to the different tendency.

We conjecture that the differences in the interface characteristics between the homomer and the heteromer stems from the existence of symmetry in homomer and non-existence of the one in heteromer. Homomer inevitably has symmetric axes in its oligomer structure and changes in a single amino acid residue have twice as much effect as the one in heteromer. The effect of a small number of changes in amino acid residues in homomer could add up to the effect equivalent to a large number of changes in heteromer. This conjecture may tell that the stability of homomers can be achieved by conservation of a very few hot spot residues, and hence the conservation in the interface of homomer is lower than that of heteromer.

Due to the difficulty of selecting hetero-oligomer interfaces and to the paucity in data of hetero-oligomers, this report turned out to be case studies for a couple of structures and should be improved with further study, yet this study evidently found different trends in heteromer interface compared with homomer interface.

Preliminary application of the trends to the prediction of heteromer interfaces - case study in cleavage and polyadenylation specificity factor, CPSF

The mechanisms of mRNA processing in eukaryotic cell is one of the central interests in molecular biology because of its importance in deciphering the information in genome sequence and of its time and space complexity [62]. The three-dimensional structures of the macromolecular assemblies involved in mRNA processing are being measured by many research groups to elucidate the mechanisms of transcription in eukaryotic cells. Cleavage and polyadenylation specificity factor (CPSF), one of the macromolecular assemblies for mRNA processing, involved in the initiation of polyadenylation step as its name suggested, has long been a target for the study and its structure is being analysed [63]. However, the complexity of the molecules still precludes the visualization of the assembly, hence the assembly can be a good target for computational prediction.

The subunit composition of CPSF has been studied and CPSF is found to be an assembly of at least five subunits in humans, namely CPSF160, CPSF100, CPSF73, CPSF30 and FIP1L1 [64]. FIP1L1 is rather specific to humans and species close to human. CPSF30 is a small protein compared with other three subunits. Hence the main architecture of CPSF is likely formed by CPSF160, CPSF100 and CPSF73. The active site for the cleavage of mRNA resides in CPSF73 and a part of its three-dimensional structure was already solved by X-ray crystallography [65]. The three dimensional structures of CPSF160 and CPSF100 have not been solved. The subunits of CPSF have paralogues in completely different macromolecular assemblies (Fig. 11) and the three-dimensional structures of the homologs of CPSF160 and CPSF100 were already registered in PDB. In fact CPSF100 and CPSF73 are homologous with each other. Based on these data, we approached the elucidation of heteromeric CPSF structure using the empirical rules we obtained in this study.
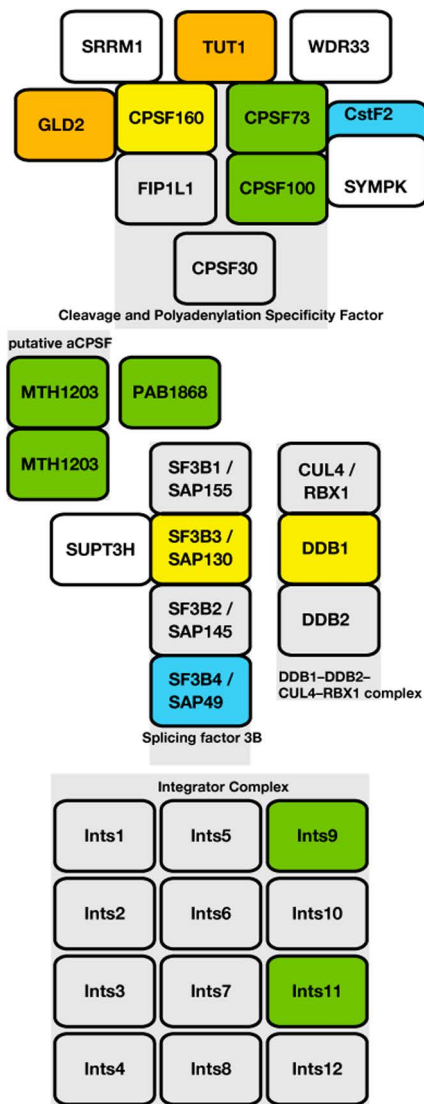
Figure 11 Subunit composition of cleavage and plyadenylation specificity factor (CPSF), DDB1-DDB2-CUL4-RBX1 complex, Splicing factor 3B (SF3B), Integrator Complex (Inr) and a putative archaeal CPSF (aCPSF). Many subunits are shared amongst the complexes. Subunits in the same colour share high sequence identity and quite likely to be derived from a common ancestor. CPSF73 and CPSF100 are homologue with each other and their homologues exist as a monomer in bacterium (PAB1868) and homodimer in archea (MTH1203). Homologue of CPSF160 exists in splicing factor 3B as SAP130 and DNA repair complex as DDB1. The three-dimensional structure of DDB1 has been solved [66].

Figure 12 shows the three-dimensional structure of CPSF160, CPSF100 and CPSF73 of humans. The structures of CPSF160 and CPSF100 were obtained by homology modelling. For CPSF160, the structure of HSDDB1 (PDB ID: 3ei1(A)) [66] was used as a template and alignment was performed by ALAdeGAP[20]. For CPSF100, the structure of CPSF73 (PDB ID: 2i7v(A)) [65] was used. The model structures were built by MODELLER [67] and the structure with the best DOPE energy was selected. For the prediction
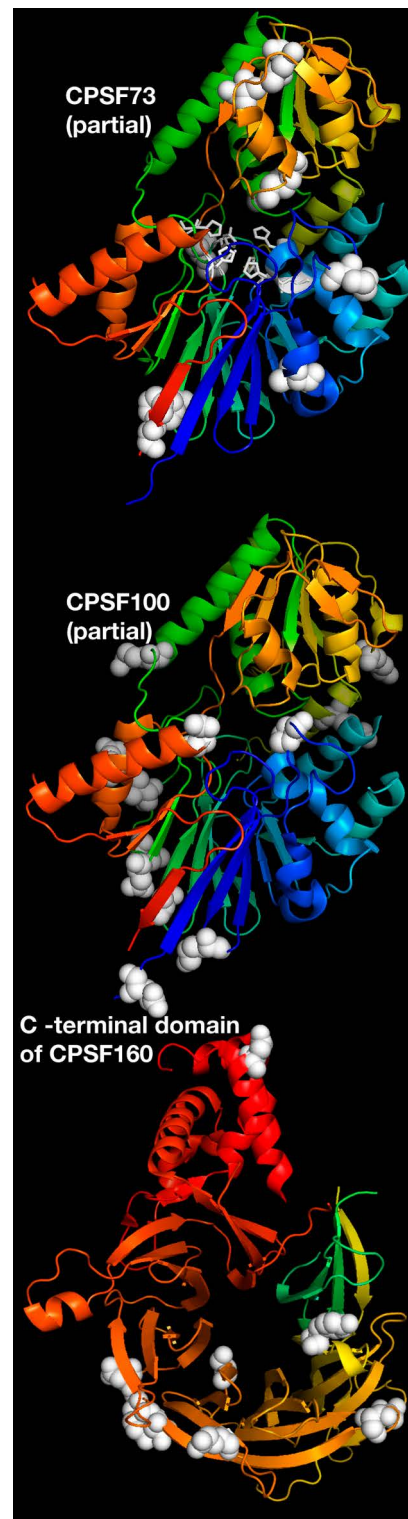

Figure 12 Structures of each subunit of human CPSF and their predicted interface residues for assembly. The structure of CPSF73 was already determined by X-ray crystallography and those of CPSF100 and CPSF160 were the prediction by homology modelling. The active sites on CPSF73 are shown in white stick model. The predicted interface residues are shown in white space-filling model. The interface residues were narrowed down to these residues based on the tendency of hetero-oligomer interfaces found in this study.

of interface residues, we aligned the paralogues of each CPSF subunit, identified the ΔConservation and Hydropathy Index of the surface residues in their monomer states and selected residue positions that met the trends we found above, namely high hydrophobicity and conservation change. The method in detail will be explained elsewhere. The predicted interface residues are shown in space-filling model. As described above, CPSF100 and CPSF73 are homologous, but only CPSF73 functions as an mRNA 3'-processing endonuclease [68] and CPSF100 does not. We presume that CPSF100 lost the endonuclease activity and started playing a role of scaffold in CPSF complex. The locations of predicted interface residues are consistent with this conjecture. The putative active site of CPSF73 is located on the C-terminal end of the β-sheet of the upper domain in Figure 12, where no predicted interface residue exists and the putative active site is accessible. The corresponding region in CPSF100 has predicted interface residues and should be used as an interface to CPSF160. The other cluster of predicted interface residues on CPSF100 (in the lower domain in Figure 12) may function as an interface to CPSF73, because the same surface is used for the subunit interface in the product of Mm0695 gene from *Methanosarcina mazei*, the putative archaeal homolog of CPSF [69]. The information obtained here will help build the structure of the whole CPSF and will be verified by cryo-EM analysis, which is on the way [63].

Acknowledgement

Conflicts of Interest

S. A. and K. Y. declare that they have no conflict of interest.

Author Contribution

S. A. and K. Y. designed the work. S. A. performed the calculations. S. A. and K. Y. wrote the manuscript.

References

[1] Goodsell, D. S. & Olson, A. J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 105–153 (2000).

[2] Harapin, J., Eibauer, M. & Medalia, O. Structural analysis of supramolecular assemblies by cryo-electron tomography. *Structure* **21**, 1522–1530 (2013).

[3] Bai, X-C., McMullan, G. & Scheres, S. H. W. How cryo-EM is revolutionizing structural biology. *Trends Biochem. Sci.* **40**, 49–57 (2015).

[4] Brown, A., Shao, S., Murray, J., Hegde, R. S. & Ramakrishnan, V. Structural basis for stop codon recognition in eukaryotes. *Nature* **524**, 493–496 (2015).

[5] Yan, C., Hang, J., Wan, R., Huang, M., Wong, C. C. L. & Shi Y. Structure of a yeast spliceosome at 3.6-angstrom resolution. *Science* **349**, 1182–1191 (2015).

[6] Hang, J., Wan, R., Yan, C. & Shi, Y. Structural basis of pre-mRNA splicing. *Science* **349**, 1191–1198 (2015).

[7] Lindert, S., Stewart, P. L. & Meiler, J. Hybrid approaches: Applying computational methods in cryo-electron microscopy. *Curr. Opin. Struct. Biol.* **19**, 218–225 (2009).

[8] Lopez-Blanco, J. R. & Chacon, P. Structural modeling from electron microscopy data. *Wiley Interdiscip Rev. Comput. Mol. Sci.* **5**, 62–81 (2015).

[9] Alber, F., Dokudovskaya, S., Veenhoff, L. M., Zhang, W., Kipper, J., Devos, D., *et al.* Determining the architectures of macromolecular assemblies, *Nature* **450**, 683–694 (2007).

[10] Nooren, I. M. A. & Thornton, J. M. Diversity of protein-protein interactions. *EMBO J.* **22**, 3486–3492 (2003).

[11] Todd, A. E., Orengo, C. A. & Thornton, J. M. Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* **307**, 1113–1143 (2001).

[12] Nooren, I. M. A. & Thornton, J. M. Structural characterisation and functional significance of transient protein-protein interactions. *J. Mol. Biol.* **325**, 991–1018 (2003).

[13] Janin, J., Bahadur. R. P. & Chakrabarti, P. Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.* **41**, 133–180 (2008).

[14] Hashimoto, K. & Panchenko, A. R. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. USA* **107**, 20352–20357 (2010).

[15] Nishi, H., Koike, R. & Ota, M. Cover and spacer insertions: Small nonhydrophobic accessories that assist protein oligomerization. *Proteins* **79**, 2372–2379 (2011).

[16] Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **10**, 980 (2003).

[17] Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

[18] The UniProt Consortium. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.* **39**, D214–D219 (2011).

[19] Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

[20] Hijikata, A., Yura, K., Noguti, T. & Go, M. Revisiting gap locations in amino acid sequence alignments and a proposal for a method to improve them by introducing solvent accessibility. *Proteins: Structure, Function, and Bioinformatics* **79**, 1868–1877 (2011).

[21] Kobayashi, E., Yura, K. & Nagai, Y. Distinct conformation of ATP molecule in solution and on protein. *BIOPHYSICS* **9**, 1–12 (2013).

[22] Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79**, 351–364 (1973).

[23] Go, M. & Miyazawa, S. Relationship between mutability, polarity and exteriority of amino acid residues in protein evolution. *Int. J. Pept. Protein Res.* **15**, 211–224 (1980).

[24] Schneider, T. D. & Stephens, R. M. Sequence logos: A new way to display consensus sequences. *Nucleic. Acids Res.* **18**, 6097–6100 (1990).

[25] Kyte, J. & Doolittle, R. F. A simple method for displaying the

hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

[26] Munch-Petersen, B., Piskur, J. & Søndergaard, L. Four deoxy-nucleoside kinase activities from *Drosophila melanogaster* are contained within a single monomeric enzyme, a new multi-functional deoxynuceoside kinase. *J. Biol. Chem.* **273**, 3926–3931 (1998).

[27] Eriksson, S., Cederlund, E., Bergman, T., Jörnvall, H. & Bohman, C. Characterization of human deoxycytidine kinase correlation with cDNA sequences. *FEBS Lett.* **280**, 363–366 (1991).

[28] Mikkelsen, N. E., Munch-Petersen, B. & Eklund, H. Structural studies of nucleoside analog and feedback inhibitor binding to Drosophila melanogaster multisubstrate deoxyribonu-cleoside kinase. *FEBS J.* **275**, 2151–2160 (2008).

[29] Sabini, E., Hazra, S., Konrad, M. & Lavie, A. Nonenantio-selectivity property of human deoxycytidine kinase explained by structures of the enzyme in complex with L- and D-nucleosides. *J. Med. Chem.* **50**, 3004–3014 (2007).

[30] Shindyalov, I. N. & Bourne P. E. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739–747 (1998).

[31] DeLano, W. L. The PyMOL Molecular Graphics System. Palo Alto, CA, USA: DeLano Scientific LLC (2008).

[32] Ding, L., Becker, A. B., Suzuki, A. & Roth, R. A. Comparison of the enzymatic and biochmical properties of human insulin-degrading enzyme and *Escherichia coli* protease III. *J. Biol. Chem.* **267**, 2414–2420 (1992).

[33] Shen, Y., Joachimiak, A., Rosner, M. R. & Tang, W. J. Structures of human insulin-degrading enzyme reveal a new substrate recognition mechanism. *Nature* **443**, 870–874 (2006).

[34] Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. & Wilmanns, M. Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* **289**, 1546–1550 (2000).

[35] Henn-Sax, M., Thoma, R., Schmidt, S., Hennig, M., Kirschner, K. & Sterner, R. Two $(\beta\alpha)_8$-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry* **41**, 12032–12042 (2002).

[36] Klem, T. J. & Davisson, V. J. Imidazole glycerol phosphate synthase: the glutamine amidotransferase in histidine biosynthesis. *Biochemistry* **32**, 5177–5186 (1993).

[37] Douangamath, A., Walker, M., Beismann-Driemeyer, S., Vega-Fernandez, M. C., Sterner, R. & Wilmanns, M. Structural evidence for ammonia tunneling across the $(\beta\alpha)_8$ barrel of the imidazole glycerol phosphate synthase bienzyme complex. *Structure* **10**, 185–193 (2002).

[38] Mortarino, M., Negri, A., Tedeschi, G., Simonic, T., Duga, S., Gassen, H. G., *et al.* L-aspartate oxidase from *Escherichia coli* I. Characterization of coenzyme binding and product inhibition. *Eur. J. Biochem.* **239**, 418–426 (1996).

[39] Bossi, R. T., Negri, A., Tedeschi, G. & Mattevi, A. Structure of FAD-bound L-aspartate oxidase: insight into substrate specificity and catalysis. *Biochemistry* **41**, 3018–3024 (2002).

[40] Stenberg, F., Chovanec, P., Maslen, S. L., Robinson, C. V., Ilag, L. L., von Heijne, G., *et al.* Protein complexes of the *Escherichia coli* cell envelope. *J. Biol. Chem.* **280**, 34409–34419 (2005).

[41] Yankovskaya, V., Horsefield, R., Törnroth, S., Luna-Chavez, C., Miyoshi, H., Léger, C., *et al.* Architecture of succinate dehydrogenase and reactive oxygen species generation. *Science* **299**, 700–704 (2003).

[42] Shionyu, M., Takahashi, K. & Go, M. Variable subunit contact and cooperativity of hemoglobins. *J. Mol. Evol.* **53**, 416–429 (2001).

[43] Hubbard, S. R., Hendrickson, W. A., Lambright, D. G. & Boxer, S. G. X-ray crystal structure of a recombinant human myo-globin mutant at 2.8 Å resolution. *J. Mol. Biol.* **213**, 215–218 (1990).

[44] Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R. The crystal structure of human deoxyhaemoglobin at 1.74 Å resolution. *J. Mol. Biol.* **175**, 159–174 (1984).

[45] Larsen, T. A., Olson, A. J. & Goodsell, D. S. Morphology of protein-protein interfaces. *Structure* **6**, 421–427 (1998).

[46] Malumbres, M., Mateos, L. M., Lumbreras, M. A., Guerrero, C. & Martin J. F. Analysis and expression of the *thr*C gene of *Brevibacterium lactofermentum* and characterization of the encoded threonine synthase. *Appl. Environ. Microbiol.* **60**, 2209–2219 (1994).

[47] Covarrubias, A. S., Högbom M., Bergfors T., Carroll, P., Mannerstedt, K., Oscarson, S., *et al.* Structural, biochemical, and *in vivo* investigation of the threonine synthase from *Mycobacterium tuberculosis*. *J. Mol. Biol.* **381**, 622–633 (2008).

[48] Laber, B., Maurer, W., Hanke, C., Gräfe, S., Ehlert, S., Messerschmidt, A., *et al.* Characterization of recombinant Arabidopsis thaliana threonine synthase. *Eur. J. Biochem.* **263**, 212–221 (1999).

[49] Meier, M., Janosik, M., Kery, V., Kraus, J. P. & Burkhard, P. Structure of human cystathionine β-synthase: a unique pyri-doxal 5'-phosphate-dependent heme protein. *EMBO J.* **20**, 3910–3916 (2001).

[50] Taoka, S., Lepore, B. W., Kabil, O., Ojha, S., Ringe, D. & Banerjee, R. Human cystathionine beta-synthase is a heme sensor protein. Evidence that the redox sensor is heme and not the vicinal cysteines in the CXXC motif seen in the crystal structure of the truncated enzyme. *Biochemistry* **41**, 10454–10461 (2002).

[51] Nagasawa, T., Ishii, T., Kumagai, H. & Yamada, H. D-cysteine desulfhydrase of *Escherichia coli*: Purification and characterization. *Eur. J. Biochem.* **153**, 541–551 (1985).

[52] Bharath, S. R., Bisht, S., Harijan, R. K., Savithri, H. S. & Murthy, M. R. N. Structural and mutational studies on substrate specificity and catalysis of *Salmonella typhimurium* D-cysteine desulfhydrase. *Plos One* **7**, e36267–e36267 (2012).

[53] Hyde, C. C., Ahmed, S. A., Padlan, E. A. Miles, E. W. & Davies, D. R. Three-dimensional structure of the tryptophan synthase alpha$_2$beta$_2$ multienzyme complex from *Salmonella typhimurium*. *J. Biol. Chem.* **263**, 17857–17871 (1988).

[54] Weyand, M., Schlichting, I., Marabotti, A. & Mozzarelli, A. Crystal structures of a new class of allosteric effectors complexed to tryptophan synthase. *J. Biol. Chem.* **277**, 10647–10652 (2002).

[55] Xiang, S., Nichols, J., Rajagopalan, K. V. & Schindelin, H. The crystal structure of *Escherichia coli* MoeA and its relationship to the multifunctional protein gephyrin. *Structure* **9**, 299–310 (2001).

[56] Sola, M., Bavro, V. N., Timmins, J., Franz, T., Richard-Blum, S., Schoehn, G., *et al.* Structural basis of dynamic glycine receptor clustering by gephyrin. *EMBO J.* **23**, 2510–2519 (2004).

[57] Schwarz, G., Schrader, N., Mendel, R. R., Hecht, H. & Schindelin, H. Crystal structures of human gephyrin and plant cnx1 G domains: Comparative analysis and functional implications. *J. Mol. Biol.* **312**, 405–418 (2001).

[58] Kim, E. Y., Schrader, N., Smolinsky, B., Bedet, C., Vannier, C., Schwarz, G., *et al.* Deciphering the structural framework of glycine receptor anchoring by gephyrin. *EMBO J.* **25**, 1385–1395 (2006).

[59] Tumbula, D. L., Becker, H. D., Chang, W. & Söll, D. Domain-specific recruitement of amide amino acids for protein synthesis. *Nature* **407**, 106–110 (2000).

[60] Oshikane, H., Sheppard, K., Fukai, S., Nakamura, Y., Ishitani,

R., Numata, T., *et al.* Structural basis of RNA-dependent recruitment of glutamine to the genetic code. *Science* **312**, 1950–1954 (2006).

[61] Yun, M. K., Nourse, A., White, S. W., Rock, C. O. & Heath, R. J. Crystal structure and allosteric regulation of the cytoplasmic *Escherichia coli* L-asparaginase I. *J.Mol.Biol.* **369**, 794–811 (2007).

[62] Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nature Review Genetics* **15**, 163–175 (2014).

[63] Xiang, K., Tong, L. & Manley, J. L. Delineating the structural blueprint of the pre-mRNA 3'-end processing machinery. *Mol. Cell. Biol.* **34**, 1894–1910 (2014).

[64] Shi, Y., Giammartino, D. C. D., Taylor, D., Sarkeshik A., Rice, W. J., Yates, J. R. III., *et al.* Molecular architecture of the human pre-mRNA 3' processing complex. *Mol. Cell* **33**, 365–376 (2009).

[65] Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethan-tham, V., Manley, J. L., *et al.* Polyadenylation factor CPSF-73 is the pre-mRNA 3-end-processing endonuclease. *Nature* **444**, 953–956 (2006).

[66] Scrima, A., Koničkova, R., Czyzewski, B. K., Kawasaki, Y., Jeffrey, P. D., Groisman, R., *et al.* Structural basis of UV DNA-damage recognition by the DDB1–DDB2 Complex. *Cell* **135**, 1213–1223 (2008).

[67] Šali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993).

[68] Ryan, K., Calvo, O. & Manley, J. L. Evidence that polyade-nylation factor CPSF-73 is the mRNA 3' processing endonu-clease. *RNA* **10**, 565–573 (2004).

[69] Mir-Montazeiri, B., Ammelburg, M., Forouzan, D., Lupas, A. N. & Hartmann, M. D. Crystal structure of a dimeric archaeal cleavage and polyadenylation specificity factor. *J. Struct. Biol.* **173**, 191–195 (2011).