THE PENNSYLVANIA STATE UNIVERSITY
MILLENNIUM SCHOLARS PROGRAM

DEPARTMENT OF CHEMISTRY

ALLOSTERIC NETWORKS WITHIN AND BETWEEN TRYPTOPHAN BIOSYNTHETIC ENZYMES

WOUDASIE ADMASU

SPRING 2021

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree in Biochemistry and Molecular Biology

Reviewed and Approved by the following:

David Boehr
Associate Professor of Chemistry
Thesis Supervisor

Denise C. Okafor
Assistant Professor of Biochemistry and Molecular Biology, and of Chemistry
Reviewer

Table of Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

1.1 Allostery and historical models

Allostery is the process of biological macromolecules transmitting the effect of binding at one site to a distant functional site[1], which consists of conformational and functional transitions in individual proteins. This process is most often used for regulation of enzyme activity but has a variety of purposes in protein biochemistry, including in signal transduction and gene regulation. Allosteric properties in enzymes also serve as suitable targets for drug design and novel biosensors[2], due to the possible impact that conformational changes can have on cellular function and disease states[3].

Cooperativity and allostery for enzymes with multiple binding sites are intrinsically linked. Cooperativity is the phenomenon that the binding of a substrate to one active site on an enzyme impacts substrate binding to a second site and is described by two main models: Monod-Wyman-Changeux (MWC) and Koshland-Nemethy-Filmer (KNF) models (**Figure 1.1**). With MWC, it's hypothesized that the protein in question only has two states: Low activity T state and high activity R state[4]. As the amount of bound ligand increases, the protein increasingly inhabits the high activity state, while at lowered amounts of bound ligand, the low activity T state is preferred. The KNF model, however, posits that ligand binding at one site causes conformational changes in other nearby sites, which affects their ability to bind ligand. The key difference that contrasts the KNF model to the MWC model is that the KNF model suggests that these proteins have many slightly different conformational states, which correspond to many modes of ligand binding. This explanation allows the KNF model to explain negative cooperativity (when binding of a ligand

decreases affinity and makes binding of other ligand molecules less likely), while MWC is incapable of that.

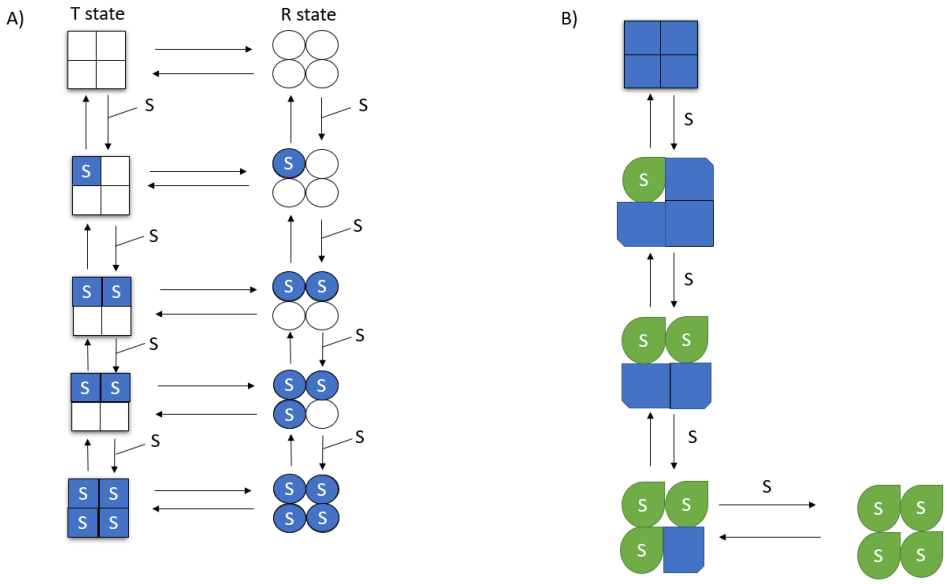

Figure 1.1 MWC vs KNF model. A) MWC model, showcases an equilibrium between the T and R states, and how transition is a concerted process, affecting all subunits simultaneously. B) KNF model, ligand binding at one site causing a conformational change and shifting the binding affinity in adjacent subunits only; a tense to relaxed transition is a sequential process.

The MWC model suggests only two different conformational states, while the KNF model suggests many different conformational states for proteins. To describe the conformational ensembles proposed by these models, the free energy landscape view can provide thermodynamic explanations of both models.

The free energy landscape of a system seeks to map out the possible states of that system. It is based on the idea that a protein has many thermodynamic configurations by which the free energy surface displays a number of local minima separated by barriers[11], which represent how fast or often a protein shifts between conformations (**Figure 1.2**). As globular proteins fold, they

undergo conformational changes that results in their free energy decreasing as they approach the native state. With the native state being the minima, the free energy landscape represents a funnel with the native state at the bottom[12]. For example, allosteric activation may operate as a bi-stable switch through a narrow window of allosteric ligand concentration[5], which shifts the protein population from an inactive to active state.

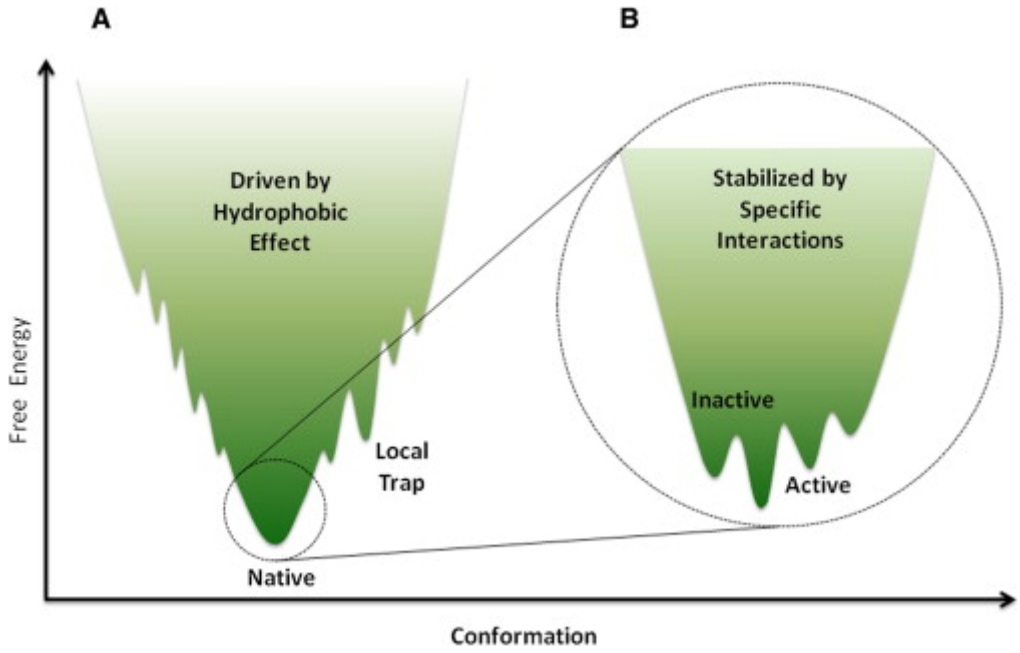

Figure 1.2 Free energy landscape of a protein. Schematic detailing protein folding as a function of conformations in a funnel shape due the hydrophobic driving force. This figure was adapted from (Ruth and Chung-Jung 2014).

1.2 Allosteric networks may mediate communication across a protein structure

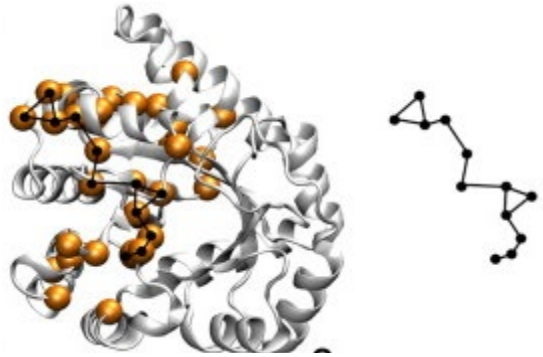

Figure 1.3 Partial amino acid interaction network of αTS. The nodes are amino acid residues, represented by circles, and the interactions between residues are indicated by lines joining the circles together. This figure was adapted from (O'Rourke et al. 2016).

While the free energy landscape provides a thermodynamic explanation of the conformational states of proteins, it does not explain the molecular details by which different conformations have different free energies. Free energy differences are ultimately related to the different interactions that these conformations have, both with their external environment and within the protein. These internal interactions may be part of allosteric amino acid networks, which are undirected networks consisting of amino acid residues and their interactions in a three-dimensional protein structure[6]. This is visually represented in **Figure 1.3**, which shows a partial amino acid network found through NMR-based methods. Understanding amino acid networks allows us to predict protein folding and identify functional residues within the protein. The network framework also allows us to analyze protein stability, which gives us a basic understanding of the thermodynamics involved in the protein folding process.

1.3 Tryptophan synthase as a model allosteric protein

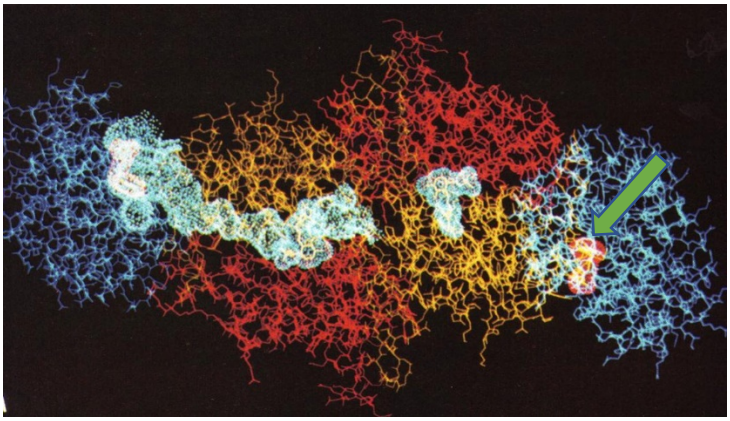

Figure 1.4 Three-dimensional structure of Tryptophan synthase from *Salmonella typhimurium*. The alpha subunits are blue, the beta subunits' N-terminal residues 1-204 and C-terminal residues 205-397 are yellow and red respectively, the green arrow indicates the position of bound indole propanol phosphate in the active site of the alpha subunit. This figure was adapted from (Miles 2013).

One model protein that we have chosen to use to understand allostery, cooperativity, free energy landscapes, and amino acid networks is tryptophan synthase (**Figure 1.4**). Tryptophan synthase (TS) is an allosteric α2β2 tetramer found in Eubacteria, Archaebacteria, Plantae, Protista, and Fungi. It is entirely absent in Animalia. Animals can consume tryptophan through protein sources, but bacteria are not capable of this. Thus, bacteria need to produce it for themselves through tryptophan biosynthesis, making this pathway a proposed target to generate novel antibiotics.

1.3.1 The alpha subunit

The alpha (α) subunit of TS is responsible for the conversion of indole-3-glycerol phosphate into D-glyceraldehyde-3-phosphate and indole. The alpha subunit belongs to the $(\beta/\alpha)_8$-

barrel family of enzymes, possessing an outer wheel of eight alpha helices and eight inner beta strands. The alpha subunit also includes three extra helices; α0, α2', and α8', which impact the positioning of Asp60, a catalytically important residue[13]. Asp60 is important for stabilizing the charge that develops on the indole ring nitrogen, while Glu49, another catalytically important residue, acts as an acid-base in 'push-pull' catalysis[13].

1.3.2 The beta subunit

The beta (β) subunit of TS reacts with L-serine in a pyridoxal 5'-phosphate-dependent reaction to give L-tryptophan and a water molecule. This happens through a series of chemical reactions. Pyridoxal 5'-phosphate is essential for these reactions to occur, as it allows for the coupling of indole (product from the alpha subunit reaction) to L-serine[14]. The beta subunits are much larger than the alpha subunits, folded into helix/sheet/helix structures[15].

1.3.3 Substrate channeling and allosteric communication between the alpha and beta subunits

As stated, TS is a heterotetramer protein complex, with the alpha and beta subunits arranged linearly. This works very well for the functionality of the enzyme, as interactions between the alpha and beta subunits greatly enhance their catalytic functions[8]. One of the most important interactions between them is the hydrophobic tunnel used to funnel indole to the beta subunit for its integration into L-tryptophan (with L-serine)[7]. This channeling also requires synchronization of the chemical reactions occurring at the alpha and beta active sites. This is coordinated by the binding of ligands in the alpha site and chemical steps taking place in the beta subunit[13], which may be connected to a possible allosteric network between the sites.

1.4 Other enzymes in the tryptophan biosynthetic pathway

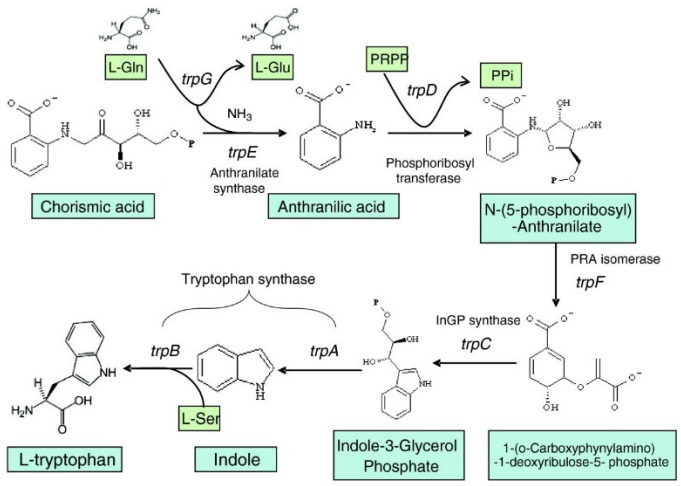

Figure 1.5 Schematic of tryptophan biosynthesis. Shows details of each step, including the gene that each enzyme is encoded in. This figure was adapted from (Kagan et al. 2008).

Our studies have focused on understanding allosteric networks within the alpha subunit and between the alpha and beta subunits. However, I also explore in this thesis the possibility of allosteric networks between other tryptophan biosynthetic enzymes, which warrants an explanation of the tryptophan biosynthetic pathway (see **Figure 1.5**). Briefly, the pathway for tryptophan biosynthesis starts with chorismate being converted into anthranilic acid by anthranilate synthase. Anthranilic acid has a ribosyl group transferred onto it and goes through an isomerization reaction by phosphoribosyl anthranilate isomerase (PRAI) to form 1-(o-carboxyphenylamino)-1-deoxyribulose-5-phosphate, which is used to form indole-3-glycerol phosphate by indole-3-phosphate synthase (InGP synthase or simply IGPS). TS catalyzes the last two steps of this process from here, removing the glycerol-3-phosphate group to form indole (by α subunits), and forming L-tryptophan in the final step (by β subunits).

1.5 Overview of thesis

The previous work in the Boehr lab identified allosteric amino acid networks in the alpha subunit using NMR methods. Here, I have taken a bioinformatics approach to elucidate allosteric networks using "frustration" and amino acid coevolution, both within the alpha subunit and between the alpha subunit and other enzymes of the tryptophan biosynthetic pathway. My results indicate that it is possible that allosteric networks may exist at sites on the alpha subunit outside of the alpha-beta interface, specifically on the outer sites of the alpha subunit and near the active site.

1.6 References

[1] Motlagh, H. N., Wrabl, J. O., Li, J., & Hilser, V. J. (2014). The ensemble nature of allostery. *Nature*, *508*(7496), 331–339. doi: 10.1038/nature13001

[2] Swain, J., & Gierasch, L. (2006). The changing landscape of protein allostery. *Current Opinion in Structural Biology*, *16*(1), 102–108. doi: 10.1016/j.sbi.2006.01.003

[3] Nussinov, R., & Tsai, C.-J. (2013). Allostery in Disease and in Drug Discovery. *Cell*, *153*(2), 293–305. doi: 10.1016/j.cell.2013.03.034

[4] Changeux, J.-P. (2012). Allostery and the Monod-Wyman-Changeux Model After 50 Years. *Annual Review of Biophysics*, *41*(1), 103–133. doi: 10.1146/annurev-biophys-050511-102222

[5] Hess, V. L., & Szabo, A. (1979). Ligand binding to macromolecules: Allosteric and sequential models of cooperativity. *Journal of Chemical Education*, *56*(5), 289. doi: 10.1021/ed056p289

[6] Yan, W., Zhou, J., Sun, M., Chen, J., Hu, G., & Shen, B. (2014). The construction of an amino acid network for understanding protein structure and function. *Amino Acids*, *46*(6), 1419–1439. doi: 10.1007/s00726-014-1710-6

[7]Dunn et al.; Tryptophan synthase: The workings of a channeling nanomachine *Trends Biochem. Sci, v*ol 33, no. 6, 2008, pp. 254-264

[8]Axe, J.M.; Boehr, D.D; Long-Range Interactions in the Alpha Subunit of Tryptophan Synthase Help to Coordinate Ligand Binding, Catalysis, and Substrate Channeling *J. Mol. Biol.* **2013**, *425*

[9] Martins, J. T., Bourbon, A. I., Pinheiro, A. C., Fasolin, L. H., & Vicente, A. A. (2018). Protein-Based Structures for Food Applications: From Macro to Nanoscale. *Frontiers in Sustainable Food Systems*, *2*. doi: 10.3389/fsufs.2018.00077

[10] Abrahams KA, Cox JAG, Fütterer K, et al. Inhibiting mycobacterial tryptophan synthase by targeting the inter-subunit interface. Sci Rep. 2017;7(1):9430. Published 2017 Aug 25. doi:10.1038/s41598-017-09642-y

Chapter 2

2.1 Introduction

2.1.1 Previous NMR results mapped out allosteric networks in the alpha subunit of tryptophan synthase

In this Chapter, I will be identifying potential allosteric networks in the alpha subunit of tryptophan synthase (TS) using two bioinformatic approaches, *frustration* and *amino acid covariation*, and then comparing against previous allosteric networks using NMR-based approaches. The previous studies in the Boehr lab used an NMR method called CHESCA (CHEmical Shift Covariance Analysis). In this method, a series of perturbations are made to a protein (e.g. adding different ligands) and then chemical shift changes are recorded. Statistical analyses are then used to identify resonances (which belong to the backbone amide groups of different amino acid residues) that have correlated chemical shift changes amongst this perturbation series. Those residues that have correlated chemical shift changes are then proposed to be part of the allosteric network. For the alpha subunit, the Boehr lab used a series of Ala-to-Gly perturbations, in which those amino acid substitutions appeared to shift the conformational state of TS towards a bound-like state[2]. The Boehr lab evaluated both the *resting* state in the absence of any substrate/products and the *working* state, in which NMR measurements were conducted under a dynamic chemical equilibrium between substrate (i.e. indole-3-glycerol phosphate) and products (indole and glyceraldehyde-3-phosphate). Importantly, the allosteric networks were somewhat different between the resting and working states[7]. More recently, the Boehr lab used a similar analysis to map out allosteric networks across each state in the catalytic cycle, and found that many network connections strengthened or weakened across catalytic cycle, suggesting a higher level of network organization[3].

2.1.2 Frustration and amino acid covariation within the alpha subunit of tryptophan synthase

The NMR results represent a great benchmark for comparing other methods to identify allosteric networks in proteins. Two other methods include protein frustration and amino acid covariation, which use the known structure and sequences of a protein, respectively. Protein frustration is a concept that compares the energy distributions with respect to structural decoys[1]. The structural decoys, in this case, occur from amino acids in contact pairs being systematically changed, which perturbs the structure with each change and changes the total energy of the protein. Parra et al. (2016) designed a so-called protein "frustratometer" that identifies different parts of the protein as neutral, minimally, and highly frustrated through the measurement of how favorable an interaction is. Sites of high local frustration, which I will focus on in this Chapter, often correlate with functional regions and regions involved in allosteric transitions[1].

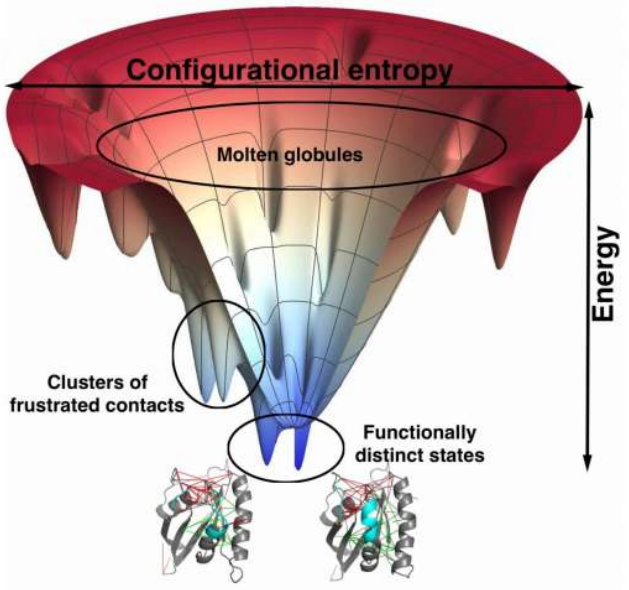

Figure 2.1. Protein free energy landscape detailing how the energy of frustrated contacts differ from molten globule and functionally distinct states. Figure adapted from (Chen et al. 2020).

In amino acid covariation analysis, large multiple sequence alignments are used to identify the correlation between amino acid changes within a protein. That is, when a residue at one site in a protein changes, does another residue change in a specific way? Within proteins, the covariation of residues at different sites may suggest coordination between them within a network, which could be important for biological events such as catalysis or conformational changes[4]. Protein frustration relates to the covariation of amino acids within the structure because highly frustrated residues that also covary with certain other residues provide evidence for the possibility that those residues are part of an allosteric network within the protein.

Figure 2.2. Example of covarying residues in sequence alignment. Sequences 1-5 show a covarying pair in the second and last positions in red and blue, respectively. Green indicates positions that are highly conserved. Black indicates positions that are not conserved and do not covary.

Using these methods, I find that covarying amino acids in proposed networks include those at the active site, the interface where the beta subunit would bind, and the outer region outside of those areas.

2.2 Methods

The RaptorX-Complex Contact webserver was used to predict covarying residues within two proteins (http://raptorx.uchicago.edu/ComplexContact/). This was done through an ultra-deep learning model trained from single-chain proteins that learns contact occurrence patterns from solved protein structures[5].

The Frustratometer webserver (http://frustratometer.qb.fcen.uba.ar/) we utilized used an algorithm that systematically perturbed the given protein structure through the systematic changing of amino acids in residue pairs (each change is referred to as a decoy). The total energy of the protein would change with each perturbation, which would allow for the energy of the decoys to be compared to the native energy distribution. The frustration index used to compare the calculations of the total energy is a Z-score of the energy of the native pair compared to N amount of decoys[6]. A contact is considered highly frustrated by the server if the frustration index is lower than -1.

2.3 Results and Discussion

2.3.1 Potential allosteric networks in αTS identified via the frustratometer

The main objective of this study was to investigate the frustration indices of TS bound with different ligands. Here, we were mostly concerned with those residues with high frustration indexes, as these have been proposed to be the mostly likely involved in functionally-relevant dynamics and/or allosteric networks[1], and our studies focused only on residues in the alpha subunit of tryptophan synthase (αTS). As different ligands (bound to either αTS or βTS) may lead to changes in noncovalent interactions, both at the binding site and more distant from the binding site, we thought it was worthwhile to evaluate frustration in a number of complexes, including TS bound with the αTS substrate mimic N-[1H-indol-3-yl-acetyl] aspartic acid (PDB 1K3U; **Figure**

2.2, Table 2.1), TS bound with the αTS transition state analog 4-(2-hydroxy-4-fluorophenylthio)-butylphosphonic acid (PDB 1C9D; **Figure 2.3, Table 2.2**), TS bound with αTS inhibitor F9 and L-tryptophan in the beta site (PDB 5CGQ; **Figure 2.4, Table 2.3**) and TS bound with the βTS quinoid intermediate (PDB 3CEP; **Figure 2.5, Table 2.4**). These structures resemble αTS with substrate bound (PDB 1K3U), during catalysis (PDB 1C9D), just before substrate channeling (PDB 5CGQ) and after substrate channeling (PDB 3CEP). We note that different allosteric networks were identified by NMR in αTS without ligands bound (i.e. the *resting* state) and during active catalysis (i.e. the *working* state).

Based on the Frustratometer webserver, I identified the top 15 most frustrated residues in αTS and mapped these onto the TS protein structure (using PDB 3CEP in all cases), so that similarities and differences between the most frustrated residues could be more easily compared. As there are only small structural changes in αTS (root mean square deviation ~ 3-5 angstroms) as it binds different ligands, it was not surprising that many of the most frustrated residues were the same among all four structures, including Glu2, Pro28, Gln32, Asp46, Asp56, Asp130, Ser135 and Ala263, and/or near similar regions of the protein structure (N.B. the amino acid types are based-off of what is found in *E. coli* αTS). Many of these highly frustrated residues are in structurally/functionally important regions of αTS, including at/near the active site (e.g. Pro96), and at/near the αTS/βTS binding interface (e.g. Pro155, Asp124, Glu135). Other residues were distant from both of these regions (e.g. Glu2). While some of these amino acid residues were clustered in space, other residues were more isolated.

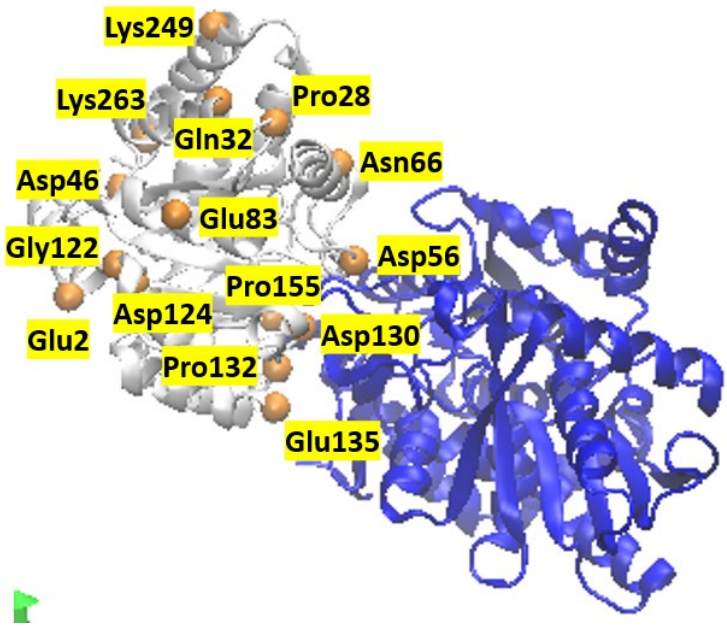

Figure 2.3. The most highly frustrated residues in αTS using PDB 1K3U (bound with αTS inhibitor N-[1H-indol-3-yl-acetyl] aspartic acid), according to the Frustratometer webserver plotted onto TS structure as colored spheres (using PDB 3CEP for comparisons). The alpha subunit is white, and the beta subunit is blue.

Table 2.1 Most frustrated residues in αTS when TS bound by N-[1H-indol-3-yl-acetyl] aspartic acid.

Residue Number	Index
Glu2	-2.227
Lys263	-1.708
Asp130	-1.517
Lys249	-1.405
Asp56	-1.381
Asp46	-1.366
Pro28	-1.342
Glu135	-1.322
Gln32	-1.297
Pro155	-1.285
Asp124	-1.273
Gly122	-1.268
Asn66	-1.258
Glu83	-1.239

Pro132	-1.235

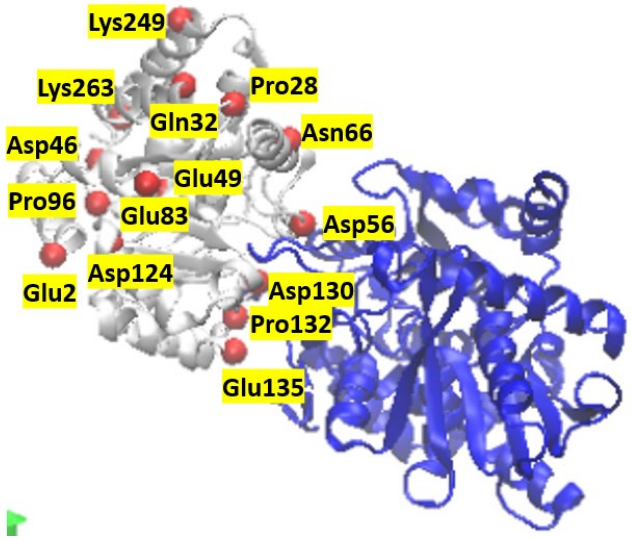

Figure 2.4. The most highly frustrated residues in αTS using PDB 1C9D (bound with transition state analogue inhibitor 4-(2-Hydroxy-4-Fluorophenylthio)-Butylphosphonic Acid), according to the Frustratometer webserver plotted onto TS structure as colored spheres (using PDB 3CEP for comparisons).

Table 2.2 Most frustrated residues in αTS when within TS bound by 4-(2-hydroxy-4-fluorophenylthio)-butylphosphonic acid.

Residue Number	Index
Glu2	-1.95
Asp130	-1.462
Lys249	-1.412
Asp46	-1.388
Lys263	-1.384
Asp124	-1.349

Asn66	-1.315
Gln32	-1.295
Glu135	-1.282
Pro132	-1.274
Pro28	-1.259
Glu83	-1.237
Asp56	-1.23
Glu49	-1.222
Pro96	-1.211

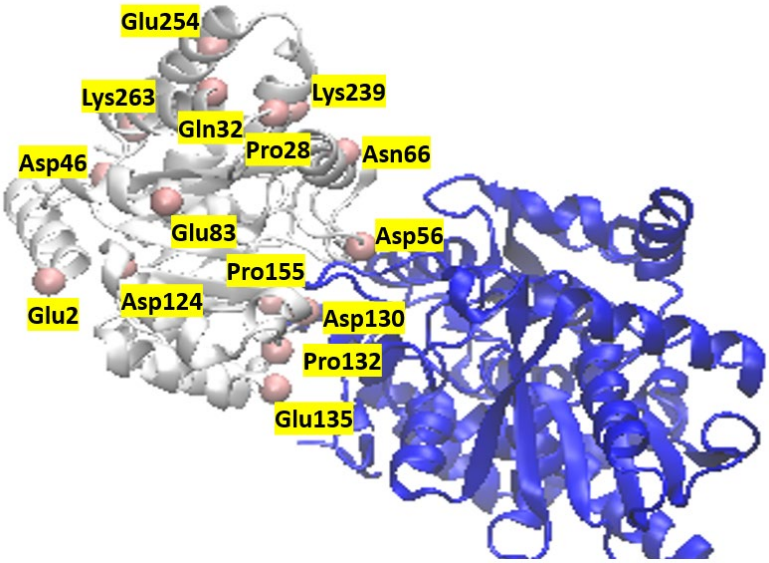

Figure 2.5. The most highly frustrated residues in αTS using PDB 5CGQ (bound with F9 ligand in α site and L-Tryptophan in β site), according to the Frustratometer webserver plotted onto TS structure as colored spheres (using PDB 3CEP for comparisons).

Table 2.3 Most frustrated residues in αTS when TS bound by F9 ligand and L-tryptophan.

Residue Number	Index
Glu2	-1.927
Lys263	-1.759
Asp130	-1.49
Asp46	-1.435

Asp56	-1.38
Lys239	-1.368
Asn66	-1.363
Pro28	-1.342
Gln32	-1.291
Glu254	-1.275
Pro132	-1.273
Glu135	-1.259
Asp124	-1.249
Glu83	-1.24
Pro155	-1.212

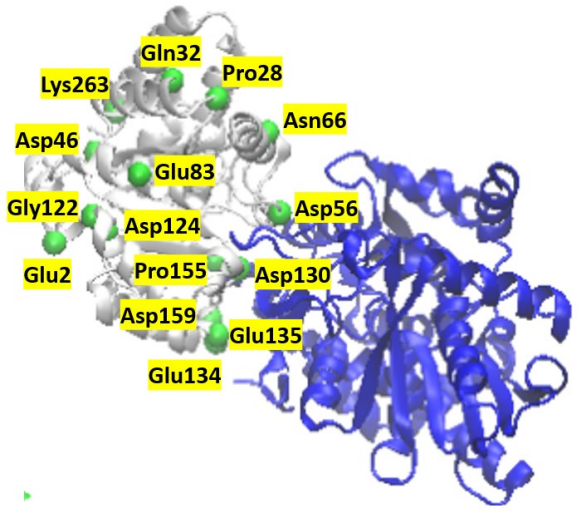

Figure 2.6. The most highly frustrated residues in αTS using PDB 3CEP (TS quinoid intermediate), according to the Frustratometer webserver plotted onto 3CEP as colored spheres.

Table 2.4 Most frustrated residues in αTS when TS is bound with the quinoid intermediate.

Residue Number	Index
Glu2	-1.941
Asp159	-1.671
Asp56	-1.57
Asp130	-1.517
Pro155	-1.402
Asp46	-1.4
Pro28	-1.338

Asn66	-1.326
Gln32	-1.303
Gly122	-1.289
Glu135	-1.261
Asp124	-1.221
Glu134	-1.219
Lys263	-1.211
Glu83	-1.204

2.3.2 Potential allosteric networks within αTS from covariation analysis

In order to assess the possibility of allosteric networks within αTS, covariation analysis from the RaptorX-Complex Contact webserver was used to determine the most probably covarying residues within the αTS subunit. The webserver generates results both for interpair 'contacts' (i.e. between two provided protein sequences) as well as intrapair 'contacts' (i.e. within one of the provided protein sequences), with this section focusing exclusively on the intrapair covarying residues. In Chapter 3, I cover results between αTS and βTS, and between other enzymes of the tryptophan biosynthetic pathway. The top intrapair residues for αTS are listed in **Table 2.5** and mapped in **Figure 2.6**.

Table 2.5 Intrapair covarying residues within αTS. Note that the residue type was derived from *E.coli* αTS.

αTS residue 1	αTS residue 2	k
Ala43	Val259	93.44
Ala47	Pro96	83.48
Ser125	Ala149	82.8
Gly211	Ile232	82.11
Pro21	Ala45	80.4
Gly98	Ser125	79.66
Gly51	Leu100	79.1
Pro217	Ala265	78.26
Val20	Ala47	77.63
Phe22	Glu49	76.37

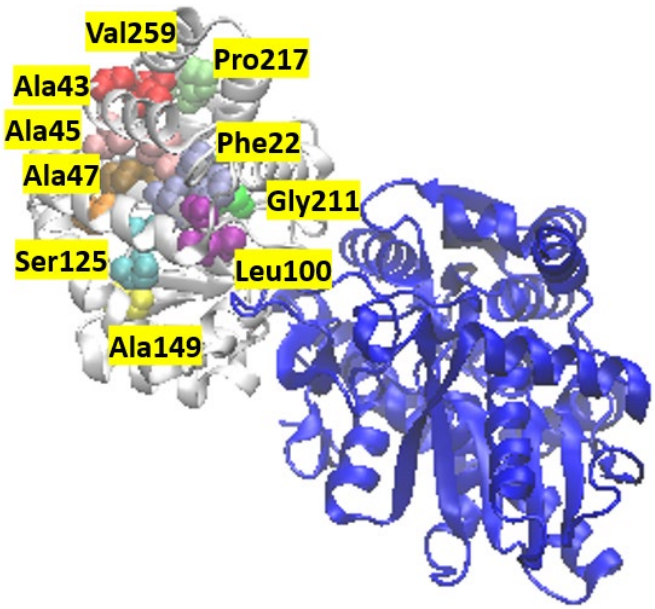

Figure 2.7. Intrapair covarying residues within αTS predicted by RaptorX were mapped onto αTS to assess their validity.

In Table 2.5, k refers to the contact prediction accuracy of the server. The higher the value, the more likely that two residues contact each other based on the sequence data. The intrapair covariation data generated a number of pairs that covary and are likely to contact each other. Among them, residues included those around the active site, including Ala43, Ala45, Ala47, Glu49 (directly involved in chemical catalysis) Gly51 and Pro96. Pro96 is in contact with Ala47, and Val259, respectively.

2.3.3 Comparisons between frustration, amino acid covariation and previous NMR analysis to identify αTS allosteric networks

Table 2.6 Previous NMR analyses identify potential allosteric networks in αTS in the *resting* and *working* states (Axe et al., 2014)

resting state			*working* state		
cluster 1	cluster 2		cluster 1	cluster 2	
Gly172	Asp46	Ala129	Val197	Asp124	Val126
Gln141	Thr39	Val259	Phe107	Ile41	Asn104
Tyr115	Leu11	Val128	Ile166	Ala47	Val106
Ser168	Glu31	Ile36	Leu48	Gln32	Val131
Phe152	Phe114	Asn104	Thr39	Ser125	Val128
Ala167	Leu100	Leu40	Leu162	Ile111	Ala129
Gly170	Leu25	Ile240	Ala167	Asp46	Val133
Leu191	Leu34	Asp124	Gly170	Ala73	Leu50
Leu176	Gln10	Phe72	Gly49	Ala45	Val121
Gln210	Asp27	Ile41	Ile151	Gly26	Gly51
Ala236	Lys35	Val126	Ala198	Lys35	Phe139
Ala198	Ile52	Leu50	Val257	Phe19	Ala116
Ala205	Gly75	Glu49	Ala254	Glu31	Phe72
	Ala71	Asp130	Thr266	Gly61	Leu127
	Ala73	Ala222	Ser168	Ile30	Gly211
	Val106	Ile111	Ala205	Leu99	Ala236
	Gly61	Phe139	Gly98	Leu100	Ala142
	Ala103	Ala45	Met101	Leu34	Ala229
				Ile97	Gln219
				Gly75	Ala190
				Ile36	Leu105

				Leu25	Phe114
				Asp27	Ile240
				Ala71	Ala149
				Ala103	

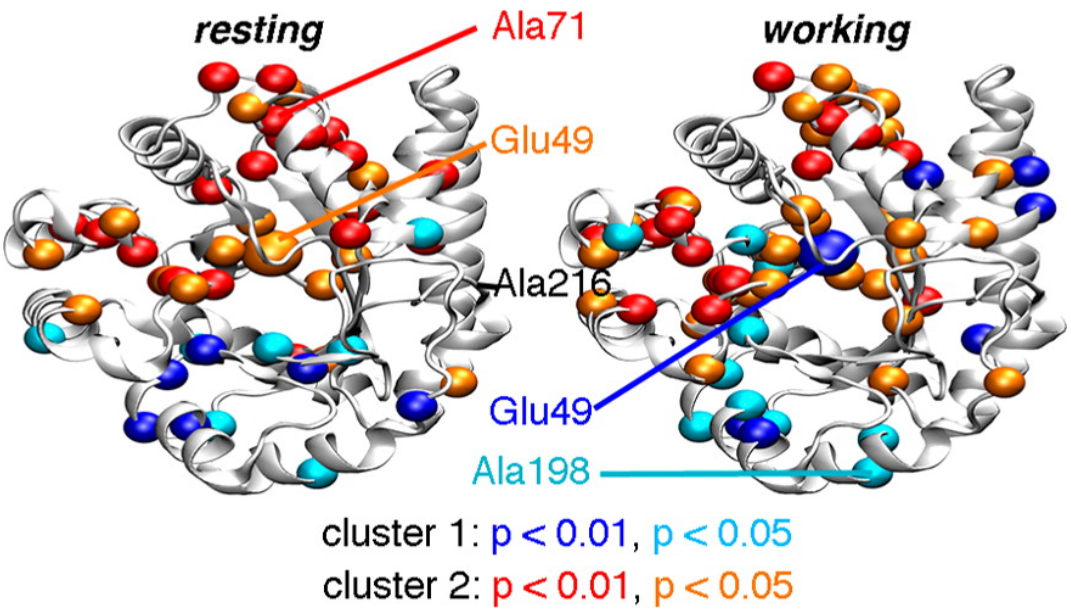

Figure 2.8. Allosteric networks are different between the *resting* and *working* states of αTS. Residues in *resting*/*working clusters 1* and *2* are plotted as blue/cyan and red/orange spheres onto the αTS structure (PDB 1K3U). This figure was reproduced from (Axe et al., 2014)

According to previous NMR results, through the use of CHESCA, allosteric networks change depending on the functional state (see **Figure 2.7, Table 2.6**). These functional states are differentiated as *resting* (ligand free) and *working* (turning over substrate or product). More specifically, active site residues directly involved in chemical catalysis change networks depending on the enzyme being in the *resting* or *working* state. Dynamic amino acid networks are critical for positioning and structural dynamics of catalytically relevant amino acids[7]. Between these NMR results, frustratometer results, and amino acid covariation results, Glu49, Pro96,

Leu100, and Asp46 are common results. There are also residues in similar regions that are shown in the results; Asp124, Pro21, Asn66, Leu100 and Lys263. In previous NMR results, Glu49 was found in different clusters depending on the state of the enzyme (*resting* or *working*)[7]. Furthermore, it was shown to have many correlations with residues in one of the *working* state clusters, those residues being absent in the *resting* state. This showed that Glu49 can occupy different positions, which help regulate enzyme activity. This is further supported by the amino acid intrapair covariation data, as Glu49 is shown to come into contact with Phe22 based on the *E. coli* sequence data, and both Glu49 and Asp46 nearby are both highly frustrated across different ligand-bound states. As has been stated before, areas of high local frustration are indicative of regions that are functionally important and/or regions involved in allosteric transitions, and with its repeated appearance, this appears to be true for Glu49 and Asp46, both in functional importance and their involvement in potential allosteric networks. Pro96 is shown to come into contact with Ala47, shown in the intrapair covariation data, and both residues have been identified as potential allosteric network points from previous NMR results. Pro96 is also highly frustrated, and residues near Ala47, like Asp46, are highly frustrated as well, which further supports their involvement in possible allosteric networks. Leu100 and Gly51 are shown to likely come into contact from the intrapair covariation data and Leu100 and residues near Gly51 (such as Leu50) have been identified by (Axe et al., 2014) as involved in potential allosteric networks. In addition, both Leu100 and residues near Gly51, like Asp56, are highly frustrated, suggesting that they are not only functionally important (near α/β interface), but providing further supporting evidence for their involvement in potential allosteric networks. There are also differences between the NMR, covariation, and frustratometer data. For example, Phe139, Glu2, and Ile232 are separately found

in previous NMR results, frustration results, and intrapair covariation, respectively, but not in the other methods.

2.4 References

[1]Parra, R. G., Schafer, N. P., Radusky, L. G., Tsai, M.-Y., Guzovsky, A. B., Wolynes, P. G., & Ferreiro, D. U. (2016). Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Research*, *44*(W1). https://doi.org/10.1093/nar/gkw304

[2]Axe, J. M., & Boehr, D. D. (2013). Long-Range Interactions in the Alpha Subunit of Tryptophan Synthase Help to Coordinate Ligand Binding, Catalysis, and Substrate Channeling. *Journal of Molecular Biology*, *425*(9), 1527–1545. https://doi.org/10.1016/j.jmb.2013.01.030

[3]O'Rourke, K. F., Sahu, D., Bosken, Y. K., D'Amico, R. N., Chang, C.-en A., & Boehr, D. D. (2019). Coordinated Network Changes across the Catalytic Cycle of Alpha Tryptophan Synthase. *Structure*, *27*(9). https://doi.org/10.1016/j.str.2019.05.013

[4] Niks, D., Hilario, E., Dierkers, A., Ngo, H., Borchardt, D., Neubauer, T. J., … Dunn, M. F. (2013). Allostery and Substrate Channeling in the Tryptophan Synthase Bienzyme Complex: Evidence for Two Subunit Conformations and Four Quaternary States. *Biochemistry*, *52*(37), 6396–6411. https://doi.org/10.1021/bi400795e

[5]Zeng, H., Wang, S., Zhou, T., Zhao, F., Li, X., Wu, Q., & Xu, J. (2018). ComplexContact: a web server for inter-protein contact prediction using deep learning. *Nucleic Acids Research*, *46*(W1). https://doi.org/10.1093/nar/gky420

[6]Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. R. Gonzalo Parra, Nicholas P. Schafer, Leandro G. Radusky, Min-Yeh Tsai, A.Brenda Guzovsky, Peter G. Wolynes and Diego U. Ferreiro. Nucleic Acids Research 2016; doi: 10.1093/nar/gkw304

[7] Axe, J. M., Yezdimer, E. M., O'Rourke, K. F., Kerstetter, N. E., You, W., Chang, C.-en A., & Boehr, D. D. (2014). Amino Acid Networks in a (β/α)8 Barrel Enzyme Change during Catalytic Turnover. Journal of the American Chemical Society, 136(19), 6818–6821. https://doi.org/10.1021/ja501602t

Chapter 3

3.1 Introduction

3.1.1 Allosteric networks between the alpha and beta subunits of tryptophan synthase

In the previous chapter, I identified potential allosteric networks within the alpha subunit of tryptophan synthase (αTS). These networks included residues at/near the active site and at/near the surface that interacts with the beta subunit (βTS). Allosteric networks which bridge the alpha-beta interface may be one way these subunits communicate to coordinate their functions. Here, the analysis will focus on amino acid covariation, similar to that used in Chapter 2, but the focus will be on covariation between the subunits, and between αTS and other tryptophan biosynthetic enzymes.

3.1.2 Networks between the alpha subunit and other tryptophan biosynthetic enzymes may suggest metabolon formation

Some of the allosteric network residues are near surfaces other than the alpha-beta interface, raising the possibility of allosteric networks between the alpha subunit of TS and other tryptophan biosynthetic enzymes, such as might be found in the formation of a so-called "metabolon". A metabolon is a protein complex of sequential metabolic enzymes and associated cellular structural elements commonly seen in metabolic pathways[3]. The purpose of a metabolon is to make a biological process more efficient by channeling the intermediate product of one enzyme to the next enzyme in the metabolic pathway. In addition, due to the high degree of intracellular organization, there are energetic advantages that allow for the increase in enzymatic activity[3]. In tryptophan biosynthesis, not only do the last two enzymes in the pathway form a complex (i.e. the alpha and beta subunits of TS), but the previous two enzymes (IGPS and PRAI)

can also be found together on the same polypeptide chain in some bacterial species (e.g. *E. coli*). TS and the bifunctional IGPS-PRAI might also associate, even weakly, to form a larger complex. (Refer to Figure 1.5 for more information on the tryptophan biosynthetic pathway).

The identification of covarying residues between the alpha subunit of TS and other TS enzymes could suggest direct contact between those proteins, which entertains the possibility of metabolon formation between them. It may also suggest new potential interactions between them that may be a form of positive control for tryptophan biosynthesis.

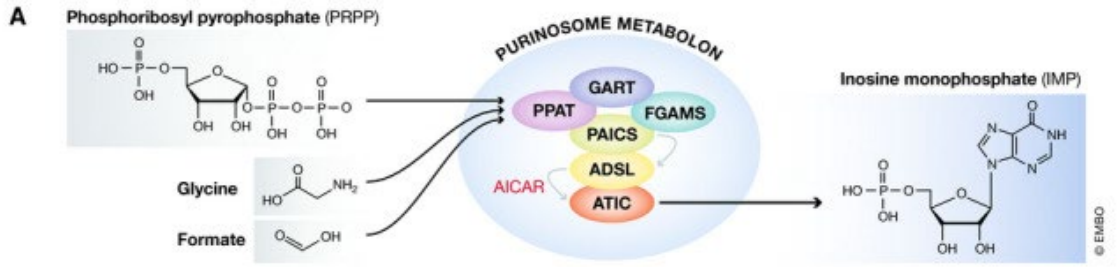

Figure 3.1 Purinosome metabolon. Schematic shows the metabolic pathway that the substrates (PRPP, glycine, and formate) are channeled through, involving all six enzymes, that produces inosine monophosphate (IMP). This figure was adapted from (Zhang and Fernie, 2020).

3.2 Methods

Similar to Chapter 2, the RaptorX-ComplexContact web server was utilized to predict the interfacial contacts between two potentially interacting protein sequences. The server used an ultra-deep learning model trained from single-chain proteins to predict contacts in a pair of proteins.

3.3 Results and Discussion

3.3.1 Amino acid covariation between the alpha and beta subunits of tryptophan synthase

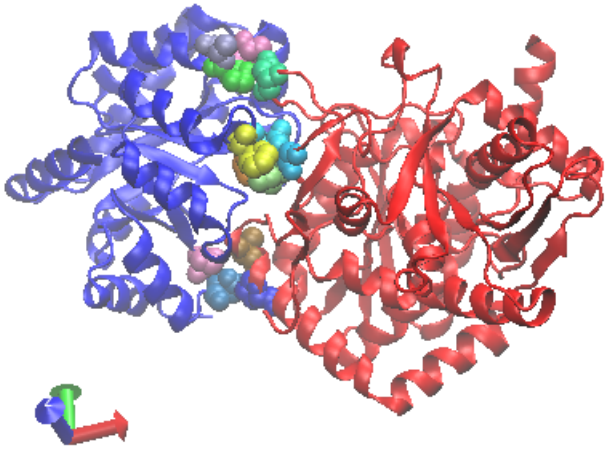

Figure 3.2 Covaring residues between αTS and βTS according to the RaptorX-ComplexContact web server. αTS is in red and βTS is in blue. Covarying residues are also listed in Table 3.1.

In order to assess the effectiveness of the server, contacts between αTS and βTS were tested as a positive control, as they are already known to form a complex. All the contacts were found along the α/β interface, which solidifies the method used by the web server. It is important to note that many of the residues within αTS interact with a single βTS residue in clusters, one example of this being Thr77, Pro78, and Ala79 of the αTS interact with Pro291 of the βTS. The interactions between these residues are non-covalent between the side chains of the polar or non-polar amino acids. Similarly, Ala103 and Phe107 interact with Leu278 of the βTS non-covalently through their side chains. In Chapter 2, intrapair covarying contacts were identified within αTS. However, none of the residues identified in that method were identified again here at the α/β interface.

Table 3.1 Interpairs of covarying residues in αTS and βTS according to the RaptorX-ComplexContact web server. Note that the residue type was derived from *E.coli* TS.

αTS residue	βTS residue
Thr77, Pro78, Ala79	Asp291
Ala103, Phe107	Ile278
Asn157, Pro155	Ile20
Phe107, Asn104	Gly277
Asn157	Pro23

3.3.2 Amino acid covariation between αTS and the previous enzyme in the pathway, indole-3-glycerol phosphate synthase

The RaptorX-ComplexContact webserver appeared to identify intermolecular contacts between αTS and βTS, and had identified potential allosteric network residues within αTS (see Chapter 2). Given these results, it was intriguing to consider if this method could also identify potential intermolecular contacts between other enzymes in the tryptophan biosynthetic pathway. Preliminary NMR analysis from Winston and Boehr (data not shown) indicated potential weak interactions between αTS and the upstream enzymes IGPS and PRAI. Such interactions might allow the formation of a metabolon for tryptophan biosynthesis. It is noted that the IGPS and PRAI enzymes are found on the same polypeptide chain in some bacterial species (e.g. *E. coli*), indicating that there is some evolutionary pressure to keep these enzymes together. A weak complex may

form between αTS-βTS and IGPS-PRAI, and the RaptorX-ComplexContact webserver could reveal interesting interactions.

It should be noted that most of the residues identified in the following analyses had more than 45% probability of covariation, which suggests that it is extremely likely that they covary. This is not as likely as the residues identified for αTS-βTS, which had a probability of 75% or more. Probabilites generated by the server for covariation indicate how likely it is that the residue pair covaries; the higher the probability, the more likely that pair covaries. The range of probabilities from the αTS-βTS comparison was 75-88%, and for the αTS-IGPS and αTS-PRAI comparisons it was 34-56%, indicating that there was substantially less confidence in the latter predictions.

Table 3.2 Covarying residues between αTS and IGPS

αTS Residue	IGPS Residue
Pro96	Phe40, Leu44
Val180	Lys13

One of the most prominent predictions for the covariation analysis between αTS and IGPS was located near the active site in the alpha subunit, namely Pro96 (N.B. residue types were identified according to the *E.coli* sequence). Half of the covarying residues on IGPS are on its outer region, which suggests that αTS and IGPS may associate. There is a possibility of new allosteric networks around the outer regions of IGPS. **Figure 3.3** showcases all the covarying residues mapped onto the PRAI-IGPS enzyme.

3.3.3 Amino acid covariation between αTS and phosphoribosylanthranilate isomerase

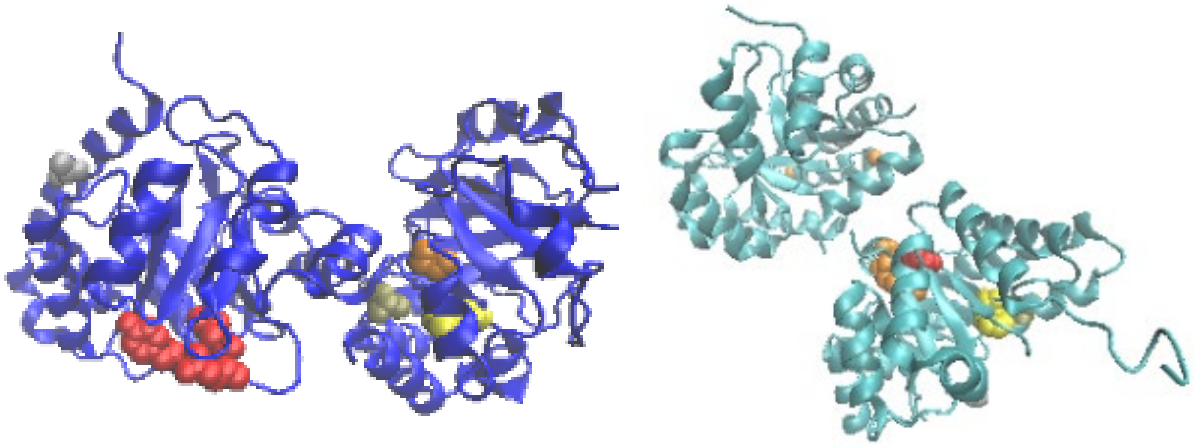

Figure 3.3 Covarying residues between PRAI-IGPS and αTS according to the RaptorX-ComplexContact web server. PRAI-IGPS is dark blue, while αTS is light blue. Covarying residues are color-coded and listed in Tables 3.2 and 3.3.

Table 3.3 Covarying residues between PRAI and αTS

αTS Residue	PRAI Residue
Tyr4, Leu7	Ala423
Phe22	Asp426
Ala236	Gly276

On the outer regions of αTS, residue Ala236 and Ala180 (**Table 3.2**) appear as top predictions between αTS and PRAI and IGPS, respectively. Ala180 is a part of loop 6 on αTS; conformational changes in loop 6 are important for substrate channeling between αTS and βTS[4]. Phe22 is also near the active site of αTS. The potential for an allosteric network involving these residues serves as an explanation for the covariation. Based on our results, we believe that it is possible that αTS, PRAI, and IGPS form a complex, based solely on the residues that had the

highest chance of covarying being functionally important. Experimental evidence will be needed to conclude that they do form a complex.

3.4 References

[1]Axe, J. M., & Boehr, D. D. (2013). Long-Range Interactions in the Alpha Subunit of Tryptophan Synthase Help to Coordinate Ligand Binding, Catalysis, and Substrate Channeling. *Journal of Molecular Biology*, *425*(9), 1527–1545. https://doi.org/10.1016/j.jmb.2013.01.030

[2]O'Rourke, K. F., Axe, J. M., D'Amico, R. N., Sahu, D., & Boehr, D. D. (2018). Millisecond Timescale Motions Connect Amino Acid Interaction Networks in Alpha Tryptophan Synthase. *Frontiers in Molecular Biosciences*, *5*. https://doi.org/10.3389/fmolb.2018.00092

[3]Ralston, L., & Yu, O. (2006). Metabolons involving plant cytochrome P450s. *Phytochemistry Reviews*, *5*(2-3), 459–472. https://doi.org/10.1007/s11101-006-9014-4

[4]O'Rourke, K. F., Sahu, D., Bosken, Y. K., D'Amico, R. N., Chang, C.-en A., & Boehr, D. D. (2019). Coordinated Network Changes across the Catalytic Cycle of Alpha Tryptophan Synthase. *Structure*, *27*(9). https://doi.org/10.1016/j.str.2019.05.013

Chapter 4

4.1 Conclusions and Future Directions

In summation, our findings reveal that it is possible that αTS and PRAI/IGPS form a complex. This is based upon the functional importance of the predicted covarying residues, how those residues are identified again as highly frustrated in different ligand-binding states, and how these identified residues coincided with residues that had been previously identified as part of potential allosteric networks in previous NMR studies involving the αTS.

However, more experimental studies based in structural analysis, such as NMR with the use of CHESCA, will be needed in order to confirm that allosteric networks exist between αTS and PRAI and IGPS, and that they form a complex. We have only investigated this possibility through the sequence data of these proteins, which does not capture the complete nuances of the structures and their interactions in space. However, we have found enough evidence to determine that the possibility is worth further investigation through experimental studies.