2024

INVESTIGATING DYNAMICS OF ELECTRON TRANSFER FLAVOPROTEIN AND ITS PARTNER FLAVODOXIN VIA SMALL ANGLE NEUTRON SCATTERING AND FLUORINE NMR

Sharique Ahmad Khan
University of Kentucky, skh249@uky.edu
Author ORCID Identifier:
https://orcid.org/0000-0002-1341-6975
Digital Object Identifier: https://doi.org/10.13023/etd.2024.120

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Sharique Ahmad Khan, Student

Anne-Frances Miller, Major Professor

Kenneth Graham, Director of Graduate Studies

INVESTIGATING DYNAMICS OF ELECTRON TRANSFER FLAVOPROTEIN AND ITS PARTNER FLAVODOXIN VIA SMALL ANGLE NEUTRON SCATTERING AND FLUORINE NMR

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Arts & Sciences
at the University of Kentucky

By
Sharique Ahmad Khan
Lexington, Kentucky
Director: Dr. Anne-Frances Miller, Professor of Chemistry
Lexington, Kentucky
[2024]

https://orcid.org/my-orcid?orcid=0000-0002-1341-6975

ABSTRACT OF DISSERTATION

INVESTIGATING DYNAMICS OF ELECTRON TRANSFER FLAVOPROTEIN AND ITS PARTNER FLAVODOXIN VIA SMALL ANGLE NEUTRON SCATTERING AND FLUORINE NMR

Anaerobic bacteria and archaea thrive in seemingly inhospitable environments because they are extremely energy efficient. Their capacity to fully utilize electrochemical energy is based in large part on their ability to conduct electron transfer bifurcation ('bifurcation') at strongly reducing potentials, thereby producing extremely potent reducing agents able to fix nitrogen and make molecular hydrogen. This chemistry is made possible by the use of a flavin as the site of bifurcation, supported by a specialized protein environment and mechanisms that control the flow of individual electrons.

Bifurcating electron transfer flavoproteins (Bf-ETFs) are versatile protein modules that provide the bifurcating capability associated with several metabolic functions. Bf-ETFs enable use of low-energy electron reserves such as NADH to charge the carriers ferredoxin and flavodoxin with high-energy electrons. Bf-ETFs possess two flavin adenine dinucleotide (FAD) cofactor. The bifurcating FAD (Bf-FAD) receives two electrons from NADH, subsequently distributing them through two distinct pathways. One pathway involves an exothermic electron transfer to a high-potential acceptor via the second FAD, designated the ET-FAD (electron transfer FAD). This provides the driving force needed to send the second electron to ferredoxin or flavodoxin, at lower (more reducing) potential.

Investigations centered on elucidating the crystal structure and internal dynamics of flavodoxin (Fld), a high-energy acceptor in the bifurcation process. ^{19}F NMR was used to examine conformational heterogeneity and dynamics of Fld free in solution, to characterize the flexibility of a 20-residue stretch of Fld's peptide chain that is believed to mediate the interaction between Fld and ETF. Detailed temperature-dependent NMR studies, alongside paramagnetic relaxation investigations comparing Fld in both its oxidized and semi-reduced forms, detailed internal dynamics pivotal to Fld's interactions with diverse partner proteins.

Complementary research explored the conformational dynamics of ETF, employing small-angle neutron scattering (SANS). This approach revealed a notable divergence from previously published structures, demonstrating presence of a more extended ETF conformation in solution. Significant reduction-triggered conformational alterations were also discerned via SANS by comparing the fully oxidized and reduced states of ETF. Molecular dynamics simulations-based data modeling suggest coexistence of multiple ETF conformations, ranging from extended to compact, in solution.

Finally, conformational consequences of complex formation between ETF and a partner protein were examined. We demonstrated isolation of a complex between ETF and is high-potential acceptor Butyryl CoA dehydrogenase (BCD). Innovative application of segmental deuteration of BCD in combination with SANS, enabled comprehensive insights into the conformational shifts of ETF upon complex formation. Contrast variation SANS,

utilizing 80% deuterated BCD, was used to identify the match point, paving the way for advanced analysis of the complex's structural dynamics.

This work enriches comprehension of the roles played by dynamics in bifurcation, and advances new technical approaches for future explorations of conformational changes within multidomain proteins.

KEYWORDS: Electron bifurcation, Electron transfer flavoprotein, Small angle neutron scattering, Conformation change, Extended & compact conformation, fluorine NMR

Sharique Ahmad Khan

(Name of Student)

04/17/2024

Date

INVESTIGATING DYNAMICS OF ELECTRON TRANSFER FLAVOPROTEIN AND ITS PARTNER FLAVODOXIN VIA SMALL ANGLE NEUTRON SCATTERING AND FLUORINE NMR

By
Sharique Ahmad Khan

Anne-Frances Miller
Director of Dissertation

Kenneth Graham
Director of Graduate Studies

04/17/2024
Date

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to Dr. Anne-Frances Miller, my advisor, and an inspiring scientist without who's mentorship and support this work would not have been possible. Thank you for your guidance, patience and unending optimism and enthusiasm that made me see the end of PhD. I owe all my success and learning to you.

I am truly grateful for the number of collaborations that shaped my PhD especially at ORNL. Dr. Hugh O'Neill provided me with great overall mentorship at ORNL and provided all the resources. Dr. Wellington Leite was instrumental in my learning and completion of experiments. Dr. Alan Hicks contributed significantly towards data analysis without which the work would be incomplete. I would also like to acknowledge other collaborators, Dr. James Brynes at Brookhaven National Lab, Dr. Biswajit Gorai at TU Berlin.

I would also like to thank my committee members Dr. David Rodgers, Dr. Christopher Richards and Dr. Chad Risko for giving me their valuable suggestions and feedback.

I am grateful to all the ex and current Miller lab members. Dr. Diessel Duan and Dr. Nishya Raseek were great help initially for technical training. I am also thankful to Debarati Das, Yutong Wan, Taylor Varner and Dallas Bell for their support and friendship.

Lastly, I would like to extend my deepest gratitude to my family and friends for their love and support. I could not have made it through PhD without the support and encouragement of my partner Anna-Bradtmueller. Thank you for always believing in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 Flavins and flavoproteins

Flavins are a group of yellow-colored organic compounds based on pteridine, responsible for a wide range of biological processes in all domains of life (Figure 1.1). They are present as cofactors in enzymes catalyzing diverse reactions [1, 2]. The two most common forms of flavins in biological systems are flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD)[3]. Riboflavin (vitamin B2) is the precursor for these co-factors. Humans must obtain this through their diet, because FAD and FMN are both electron carriers in vital redox reactions[4]. Plants, bacteria and yeast can synthesize riboflavin, and serve as our usual sources[5].

Flavins are characterized by their catalytically active isoalloxazine ring system, which is capable of undergoing oxidation-reduction reactions. The isoalloxazine ring system consists of three rings, the hydrophobic xylene ring, the hydrophilic dihydropyrazine ring and the pyrimidine ring[6]. Flavins can bind to flavoproteins via covalent and/or non-covalent interactions. Both FMN and FAD differ in their side chains; FMN is the mononucleotide form and has a phosphorylated ribose group, while FAD contains an additional adenine nucleotide (adenosine monophosphate) linking to the ribityl phosphate of FMN via its own phosphate group to make a dinucleotide. (Figure 1.1)

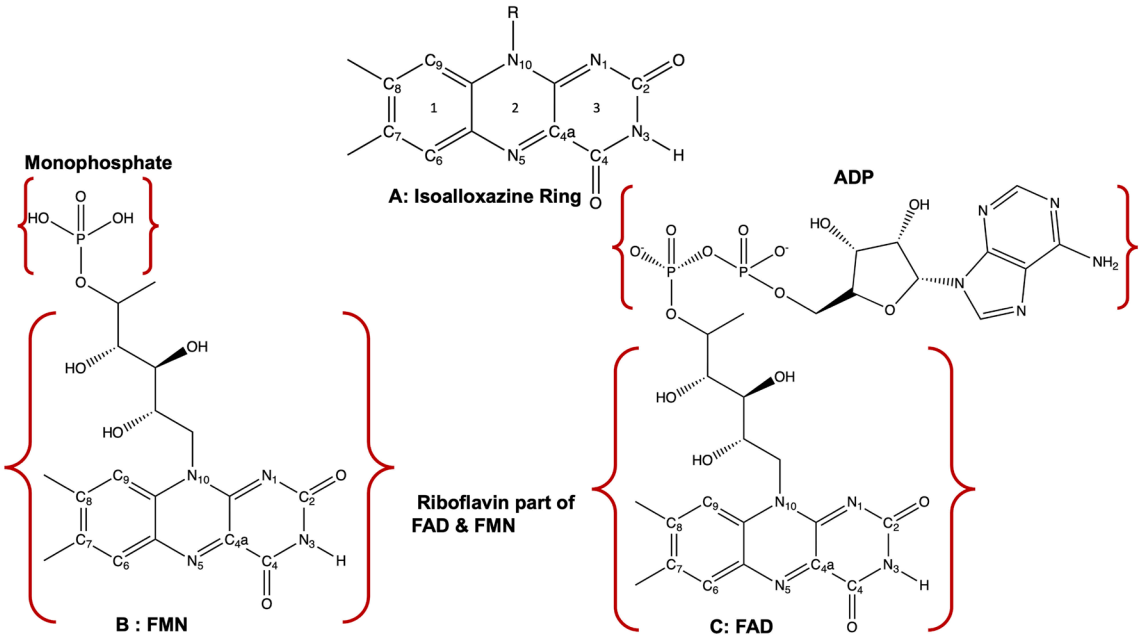

Figure 1:1 FAD&FMN.
A: Molecular labeling of the isoalloxazine ring system of Flavin molecule. The side chain R varies in FAD and FMN. B: The R side chain in FMN is ribose sugar attached to a phosphate group. C: FAD it contains ribose sugar and Adenine diphosphate (ADP).

All parts of FAD/FMN form interactions with protein residues that hold it in place in flavoproteins[7]. Majority of flavin-protein interactions are via the N10-ribityl side chain and pyrophosphate, that binds to conserved residues in flavoprotein family[3, 6]. However interactions involving the ring system, such as hydrogen bonding interactions via N1, O2, N3, O4 and N5, also tune the flavin's reactivity. Flavins' tendency to undergo reduction reactions is enhanced/suppressed by these hydrogen bonding interactions, as well as additional aspects of the binding site[8]. These tuning interactions allow flavins to serve diverse roles in electron transfer and metabolic processes.

Flavins can bind to proteins either tightly (as prosthetic groups) or loosely. The nature of the flavin-protein interaction can significantly affect the redox potential and reactivity of the flavin, thus influencing the enzyme's activity and specificity. Indeed, flavoproteins are involved in mitochondrial electron transport for ATP production, detoxification reactions in the liver, and biosynthesis and degradation of amino acids[9]. They are also involved in repair of DNA and regulation of cell growth and differentiation[10].

Flavin-dependent enzymes are associated with several diseases when mutated or improperly regulated. For instance, deficiencies in specific flavoproteins can lead to metabolic disorders, and mutations in genes encoding flavoprotein enzymes have been linked to cancer[11], neurodegenerative diseases[12], and other health conditions[13].

Flavins' numerous roles in protein chemistry and cell metabolism underscore their importance in the field of biochemistry and molecular biology, highlighting the intricate interplay between small molecule cofactors and protein function. Flavins are versatile in their redox properties, capable of undergoing one- or two-electron transfers. This flexibility allows flavin-dependent enzymes to mediate a wide range of reactions, including the oxidation of alcohols, aldehydes, and amino acids, as well as the reduction of cytochromes and quinones[14]. Thus, they are crucial for the metabolism of carbohydrates, fats, and proteins.

1.2 Redox properties of flavins

Flavins are notable for their distinct yellow color in their oxidized form. This is due to their extended conjugated π orbitals, which give flavins ability to absorb light in the visible spectrum. The extended conjugation system also allows flavins to accept and donate one or two electrons during catalysis[3]. The oxidized flavin (OX) forms a semiquinone (SQ) upon accepting one electron, or a hydroquinone (HQ) after accepting 2 electrons[15]. Each oxidation state has a distinct characteristic UV-Vis spectrum, making it easy to monitor redox reactions. The oxidized (OX) flavin has two major peaks, band I at around λ_{max} = 454 nm (can vary from 450-470 nm) and band II at 374 nm (can vary from 37—400 nm), Figure 2.1)[16]. Band 1 arises from a π-π* transition within the isoalloxazine ring system and band 2 arises from an n-π* transition, which tends to be more responsive to interactions with nearby amino acid residues or solvent molecules[17]. The maximum wavelength, λ_{max} at band I may vary slightly depending on the specific flavin molecule, solvent conditions, and other factors. Upon 1-electron reduction, The SQ formed can be neutral or anionic depending on whether a proton is acquired. The anionic SQ (ASQ) is characterized by increased absorbance at 374 nm and 525 nm whereas for neutral SQ (NSQ) the most prominent feature is a broad band near 650 nm. When forming ASQ from OX, band I intensity decreases and band II amplitude increases. Upon further reduction to the HQ state, the intensities of both band I and band II decrease. Flavoproteins such as ETF and Flavodoxin (Fld)are routinely studied via UV-Vis spectroscopy by performing stepwise titrations using reducing agents such as dithionite and titanium citrate, to form different redox states.

Figure 1.2 shows a progression of UV-Vis spectra revealing conversion among the flavin's oxidation states as the flavodoxin's bound FMN becomes increasingly reduced.

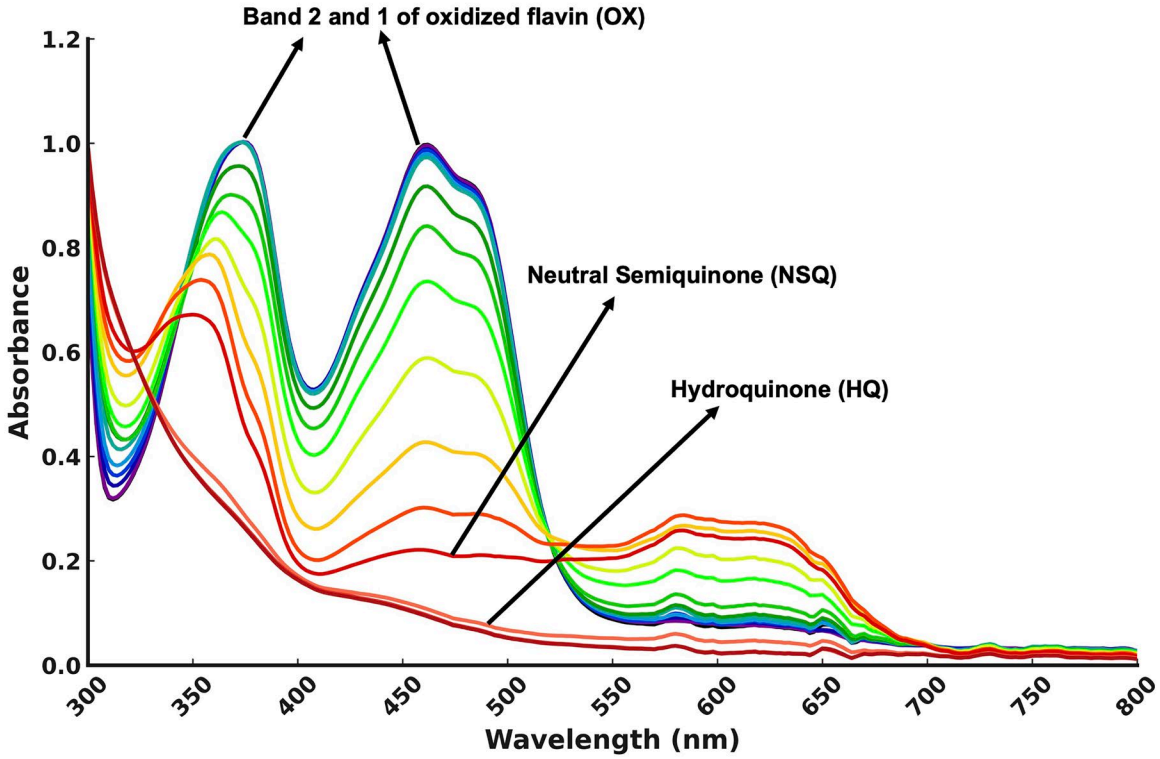

Figure 1:2 UV Vis spectra of Fld protein

Stepwise reduction of the Flavodoxin protein using titanium citrate. In the first few spectra when the protein is in oxidized state (OX), there are two maxima, band I around 454 nm and band II around 374 nm. The protein subsequently forms a neutral semiquinone (SQ) where band I and II absorbance decrease, and a distinct shoulder is formed between 550-700 nm. Fully reduced hydroquinone (HQ) state is formed towards the end.

1.3 Electron transfer flavoproteins (ETFs)

Electron transfer flavoproteins (ETFs) are FAD-containing heterodimeric proteins that are found in all kingdoms of life, from bacteria to humans[18]. They were first discovered in mitochondria and are located on the matrix face of inner mitochondrial membranes[19]. The mitochondrial ETFs (now also known as canonical ETFs) contain a single FAD that shuttles electrons from different kinds of fatty acyl CoA dehydrogenases (CADs) to the quinone pool, thus contributing electrons to the respiratory electron transfer chain[20]. ETF's sequential transfer of electrons from a variety of dehydrogenation reactions to the respiratory chain thus enables ATP production from diverse substrates. Indeed, mutations of some ETFs have been linked to metabolic disorders in humans [21].

A few decades after the discovery of canonical ETFs, other group of ETFs were reported, the so-called bifurcating ETFs (Bf-ETFs) which have two FADs instead of one[22] and participate in the process of electron bifurcation[23-27]. Both groups of ETFs are heterodimeric, composed of subunits referred to as Etf-A and Etf-B in this thesis. Based on their three-dimensional structures, the ETFs are described in terms of three domains. Domain I is composed of the N-terminal half

of the Etf-A subunit (also called L) and domain III is composed of most of the smaller Etf-B subunit (also sometimes called S). These two domains are tightly associated with the flavin head group of one of the FADs bound in the domain interface, in Bf-ETFs (Figure 1.3). We refer to this 'superdomain' uniting domains I and III as the 'base'. Domain II is formed mostly by the Etf-A subunit's C-terminal half and sits in a shallow bowl formed by the base. We refer to it as the 'head' because it turns.

Domain II contains the other FAD (called Et-FAD because it mediates electron transfer). This FAD is common to both the groups of ETFs. However canonical ETFs lack the FAD between domains I and III, retaining only an AMP molecule in Domain III. The AMP of canonical ETFs is superimposable on the AMP portion of the FAD of Bf-ETFs, in overlaid crystal structures of the two types of ETFs. This FAD unique to Bf-ETFs has been shown to be the site of bifurcation, so we refer to it as the bifurcating FAD (Bf-FAD).

The research discussed in this dissertation will address rotation of the head domain with respect to the base domain in Bf-ETFs. This movement is deemed critical to bifurcation because it is believed to regulate the number of electrons that pass from the Bf-FAD to the ET-FAD in each catalytic cycle (see section 1.6).

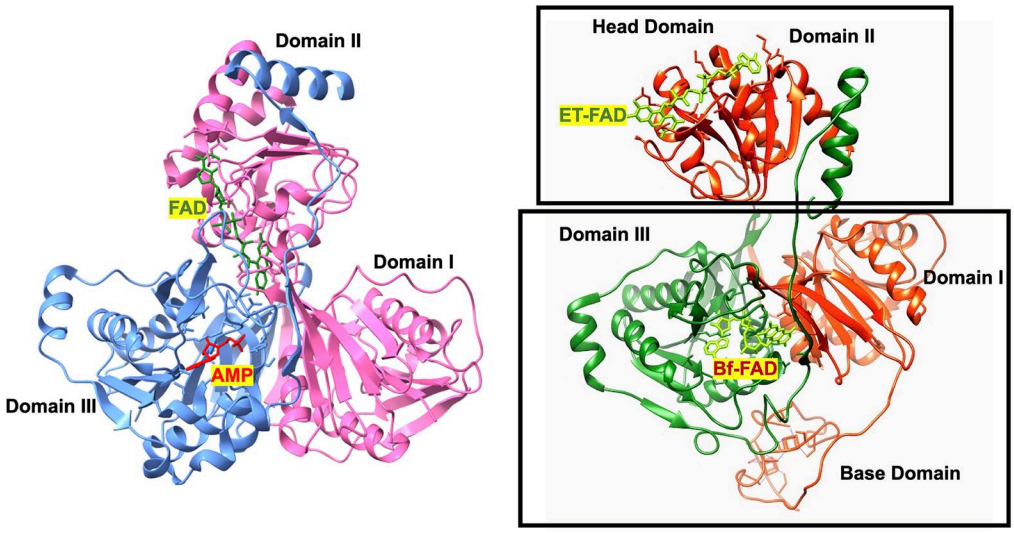

Figure 1:3 Ribbon diagram of canonical and bifurcating ETFs showing domains Canonical ETF[28] on the left with AMP (red) and FAD (green) as co-factors. Bifurcating ETF on the right[26] where the AMP is replaced by FAD. The canonical ETF is in close conformation with the two co-factors closed to each other. The bifurcating ETF is in open conformation with the two FADs separated from each other due to rotation of head domain.

1.4 Electron Bifurcation

Electron transfer bifurcation couples exergonic and endergonic electron transfers to conserve energy. In bifurcation, a thermodynamically unfavorable endergonic reaction is coupled to an exergonic reaction so that the overall reaction becomes spontaneous. It distributes energy among two electrons acquired as hydride, to one electron with a more positive reduction potential and one with a more negative reduction potential, compared to that of the hydride electron pair. Bifurcation augments the reducing power of one electron at the expense of the other. In biology, two versions of electron bifurcation exist, involving either quinones or flavins[29].

The best-known electron bifurcation was reported in the cytochrome bc1 complex of the respiratory chain and uses ubiquinol as the hydride donor (ubiquinone/ubiquinol; E_0 '= +90 mV), Rieske iron-sulfur protein (E_o'= +285 mV) as the high reduction potential acceptor and cytochrome b (E_o'= -90 mV) as the low reduction potential acceptor. This was first elucidated by Peter Mitchell in 1975. [30, 31]

1.5 Flavin based electron bifurcation (FBEB)

For almost four decades since it's discovery by Mitchell, electron bifurcation was linked only to the respiratory chain complex. In 2008 another form of bifurcation was reported in the context of anaerobic metabolism, the so-called Flavin-based electron bifurcation[23, 24, 29]. Bifurcation is now touted as a third mode of energy conservation, in addition to oxidative phosphorylation and substrate level phosphorylation.

Flavin-based bifurcation employs a hydride electron pair from NAD(P)H, coenzyme $F_{420}H2$, H_2, or formate[29, 32-34]. The low potential electrons generated are sufficiently reducing to reduce ferredoxin (or flavodoxins, when iron is scarce, Figure 1.4)[33, 35-37]. Ferredoxins (Fds) are low molecular weight, acidic iron sulfur proteins with either one [2Fe-2S] or one, two, or more [4Fe-4S] clusters. Ferredoxins characteristically transfer only one electron at a time with a reduction

potential, E_0‘= -420 mV, which is close to that of the hydrogen electrode at pH 7 (E_0‘= -414 mV).[29]

The first example of flavin based electron bifurcation was reported in 2008 *Clostridium kluyveri[23, 24]*. In *C. kluyveri* the butyryl-CoA dehydrogenase ETF complex (Bcd-EtfAB) was shown to couple the endergonic reduction of Fd (E_0’ = -420 mv) with NADH (E_0‘= -320 mV) to the exergonic reduction of crotonyl-CoA to butyryl-CoA (E_0 ‘= -10 mv) with NADH. The initial discovery was followed by reports of multiple ETFs in different organisms over the years. [23, 26, 38]

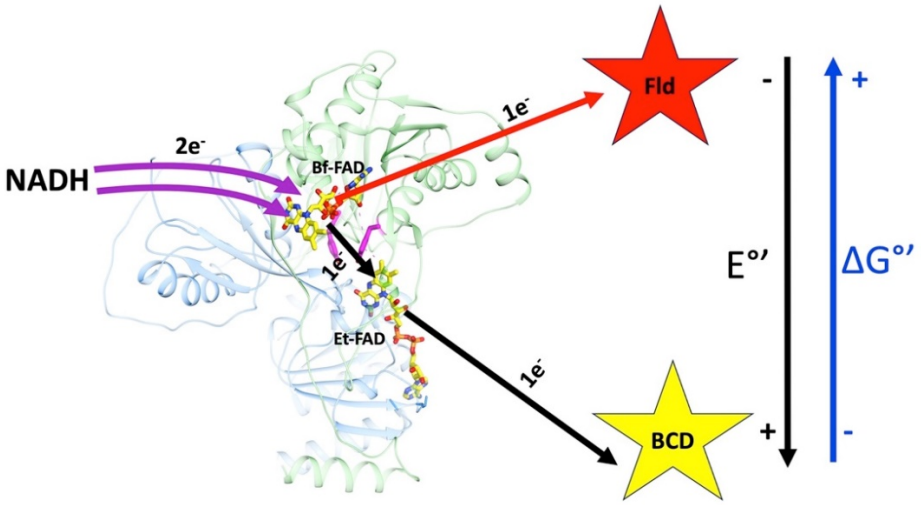

Figure 1:4 Process of flavin based electron bifurcation in a bifurcating ETF.
Two electrons from NADH are bifurcated by the Bf-FAD. One electron is transferred to a high potential and low energy acceptor such as butyryl CoA dehydrogenase (BCD) via Et-FAD and the other electron is transferred to low potential and high energy acceptor such as Fd or Fld. Fld is an acidic, highly soluble, low molecular weight protein with a tightly bound FMN co-factor.

1.6 Multiple conformations of ETF

Crystal structures of both groups of ETFs have captured two different conformations, 'open' and 'closed', described below. The closed conformation of canonical ETFs has been reported for ETFs from human (*huETF*), *Paracoccus dentitrificans* (*PdeETF*) and *Geobacter metallireducens[28, 39, 40]*. In the closed conformation (Figure 1.3), AMP and FAD molecules are separated by approximately 15 Å. It took a considerable effort to capture the open conformation via crystallography. The first effort to solve the structure of *Methylophilus methylotrophus* ETF (*MmeETF*) in complex with its partner trimethylamine dehydrogenase revealed that the head domain diffracted poorly. The authors concluded that binding to the partner reorients the head domain of the ETF. After performing mutagenic studies to find substitutions that destabilize the conformation observed in the absence of partner, the authors were able to crystallize the ETF-MCAD complex from human (*huETF*)which resulted in revealing the open conformation of the ETF[18]. It was clear that the head domain had undergone a rigid body reorientation relative to the base domain in the open conformation, when

compared to the closed conformation. In the open conformation, the bound AMP and FAD cofactors are approximately 38 Å apart.

The analogous conformations have also been reported in Bf-ETFs, although individual complexes are not identical. The closed conformation has been reported in the crystal structure of ETF from *Acidaminococccus fermentans[26]* (AfeETF, Figure 1.5). The open conformation was reported in ETF from *Clostridum difficile* (*CdiETF*)[41] (Figure 1.5) in complex with butyrylCAD and the *Acetobacterium woodi* Bf-ETF in complex with caffeoylCAD (*AwoETF*)[42]. In the closed conformation, Et-FAD (in the base) is 18 Å from the BF-Flavin. In the open conformation, the two FADs are 37 Å apart and the Et-FAD is exposed to the partner protein. Thus, rotation/movement of the head domain into the closed conformation moves the ET-flavin close to Bf-flavin, whereas the open conformation exposes the ET-FAD to partner proteins. Thus, the motion gates electron transfer (ET), allowing ET between the ET-flavin and partners when ETF is open, and approaching an interflavin distance that would allow electron tunelling between the ET- and Bf-flavins when the ETF is closed. An intermediate conformation was modeled based on the CryoEM data for the ETF from *Thermotiga mariitima[43]*. This intermediate conformation places the two FADs 25 Å apart. (Figure 1.5)

Because the open conformation isolates the low energy Et-FAD from the Bf-FAD, it may serve to prevent the high energy electron from using this exergonic path. In contrast, an 'ideal' bifurcating conformation in which the two FADs are close enough for electrons to tunnel between them has still not be observed. However

such a conformation need not be long-lived because of the high kinetic efficiency of electron tunneling. More other electron transfer mechanisms are also possible.

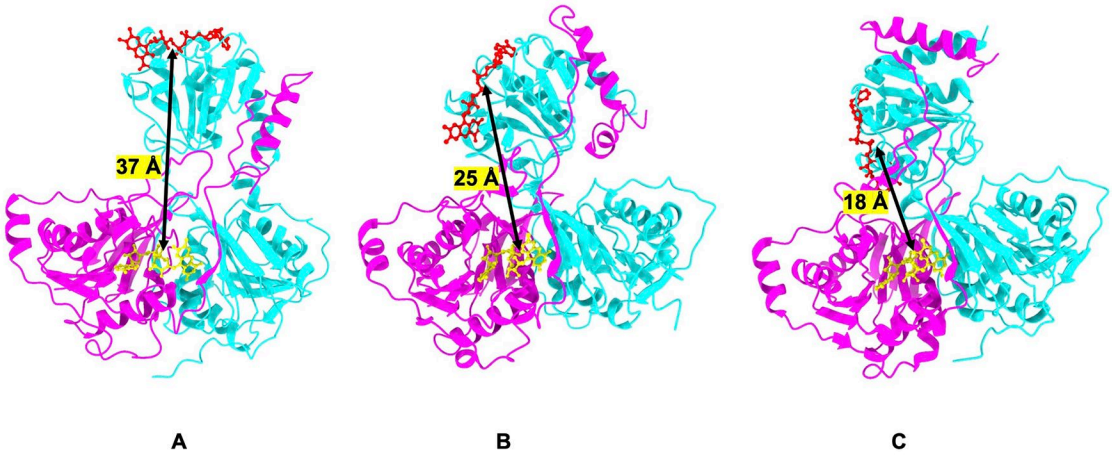

Figure 1:5 Reported conformations of bifurcating ETF
A) Open conformation from *C. difficile* ETF B) Intermediate conformation from CryoEM data from *T. mariitima* C) Closed conformation *A. fermentans* ETF.

To understand the process of electron bifurcation and how efficiency is achieved, we need more extensive sampling of the conformations adopted by ETFs under turnover conditions. The few structures available at present have all come from crystals, or freeze-trapped samples from whence a subset of images are selected and merged. However ETF is proposed to be sampling multiple different conformations during catalysis. We also need to identify each of the catalytic events that triggers conformation change, to understand the overall functioning of ETF.

This research is focused on exploring conformations of ETF as well as understanding the factors driving the conformational change in the Bf-ETFs.

1.7 Partner Proteins of ETF: Flavodoxin & Butyryl CoA dehydrogenase

A. fermentans and *M. elsdenii* both belonging to Negativicutes class of bacteria, in which the electron-bifurcating reduction of crotonyl-CoA is catalyzed by ETF and homotetrameric butyryl-CoA dehydrogenase (BCD)[44, 45]. BCD has a FAD in each subunit. Chowdhry *et al* proposed that in *A.fermentans[26],* NADH reduces the Bf-FAD of ETF into HQ which then delivers one high potential electron to Et-FAD of ETF which should be around 14 Å from Bf-FAD. As a result, a stable ASQ is formed at the Et-FAD. The remaining low potential and highly reactive SQ on the Bf-FAD donates the other electron to Fd or Fld. Since the Et-FAD is located on the mobile head domain, it is carried to the partner BCD by head rotation. It can then reduce the partner to form a SQ species on the BCD FAD. In a second round of catalysis, the BCD FAD becomes HQ and then reduces the crotonyl-CoA to butyryl-CoA.

Fixen *et al* [37]showed that under iron repleting conditions Fld can replace Fd as the low potential electron acceptor. Similar results were obtained in *A.fermentans*. We have exploited the absence of paramagnetic Fe in Flavodoxin, and used ^{19}F NMR to identify dynamic regions in Fld that may interact with the ETF. In addition to elucidating possible modes of interaction between Fld and ETF, it is also important to understand the significance of different conformations of ETF in their

interaction with the BCD partner proteins, since these affect the accessibility of the ET-FAD and therefore will modulate the efficiency of electron transfer.

1.8 Objectives & Summary

Domain scale motions have been documented for proteins such as the respiratory *bc1* [46]complex and Cytochrome P450 reductase (CPR)[47, 48]. Crystallography has provided valuable insights on multidomain proteins involved in electron transfer reactions and how electron transfer is coordinated with protein dynamics involving conformation changes. Side chain and cofactor interactions have also been studied extensively via crystallography. However, there are some disadvantages, such as enrichment of conformations that are rare in solution due to protein-protein interactions during crystallization, or crystal packing constraints that suppress conformational changes of interest.

Neutron scattering has been used to study the structure and dynamics of vast arrays of materials including biological systems[49]. Neutrons, being neutral in charge, are highly penetrating, yet their magnetic moment produces isotope-specific scattering that makes neutron scattering a very powerful technique[50]. Small angle neutron scattering (SANS) is one of the broadest applications of neutron scattering. As a structural probe, SANS is applicable to length scales ranging from one to hundreds of nanometers.

Biological systems are excellently suited for study by SANS. SANS can investigate homogenous, dilute solutions where the non-interacting particles are free to diffuse in their native structural state. This is the major advantage of SANS over other techniques, as proteins can be studied under catalytic conditions. SANS can be used to investigate the interactions between partner proteins, small ions or other molecules. Moreover, crystallography is not always a good choice for studying dynamics in biological macromolecular complexes. X-ray methods can also change the redox state of redox-active centers in proteins[51]. NMR is also limited to the study of smaller proteins and becomes extremely complicated to use to study large protein complexes. Small angle scattering techniques can provide structural insights in solution about the overall shape, size and association states of protein in solution. When combined with modern modeling tools, structural models can be designed which give important information about conformations of proteins in solution.

One of the primary objectives of this research was to establish new tools and techniques to address the dynamics in Bf-ETF and its partner flavodoxin. The majority of structural information available for Bf-ETFs has been obtained from X-ray crystallography and a couple of modeled structures based on Cryo-EM data. There is a critical need to investigate solution based structural methods as solution structures are physiologically more relevant. Proteins can have access to multiple conformations in solutions where they can roam freely and are less constrained

compared to crystals. Such freedom is critical in the study of ETFs because a conformation change is central to ETF function. Therefore, we have developed small angle scattering methods and ^{19}F NMR to study dynamics of Bf-ETF and its partners Fld and BCD, in solution. The benefits of using these techniques will be discussed in subsequent chapters.

We sought to address three major themes using SANS:

1. Establish the resting conformation of ETF in solution.
2. Characterize conformation changes associated with reduction of ETF.
3. Investigate conformation change elicited when ETF binds to its partner BCD.

Our work on the Bf-Etf from *A. Fermentans* using SANS demonstrates the existence of a conformational ensemble in the resting state (Chapter 3). The ETF is a very dynamic protein due to presence of flexible linkers connecting the head domain to the base domain. This flexibility can allow the head domain to move towards the base domain during catalysis when the Et-FAD accepts one electron from the Bf-FAD. The head domain can then again move to transfer that electron to the high potential acceptor i.e in our case the protein BCD. Our first important goal was to establish the dynamic nature of the resting conformation of ETF in solution on its own. We approached this by using SAXS first to confirm the stable, aggregation-free and monodispersed protein in solution, all the major requirements to model SANS data.

As SANS experiments are more extensive and expensive, we wanted to ensure data collected reflected only the best samples. Once we confirmed good sample behavior via SAXS we used SANS as our method of choice. One of the primary reasons for this was that flavins are prone to photoreduction by X-rays. By using neutrons, we obtain the ability to control the oxidation state of the ETF. Our major finding from performance of SANS on ETF alone was that the structure in solution cannot be described by any of the existing crystal structures. None of the existing conformations could explain the scattering data in terms of radius of gyration (R_g) and size distribution profiles characterized by pairwise distance probabilities (P(r)), and deviance between modeled and observed scattering envelopes (Chi^2). These are important parameters used to compare scattering data and simulated scattering data from existing structures. We then used molecular dynamics simulation-based methods to do flexible system modeling of ETF. We were able to show that in solution, ETF exists in equilibrium of at least two major populations of conformers, which differ with respect to their R_gs and therefore can be called 'extended' and 'compact'. These are present with almost equal distribution in resting ETF. In the extended conformation, the base domain is very far from the head domain whereas in the compact conformation, they are close to each other.

We investigated whether there is a change in this equilibrium when NADH reduces the protein, and indeed we saw a shift towards a more compact conformation when ETF is reduced. Thus, we were able to prove that the protein exists in more compact conformation after full reduction of ETF.

We also wanted to exploit a unique advantage of SANS, i.e the differential scattering of neutrons by hydrogen and deuterium that enables contrast variation methods. This allows SANS to distinguish among components of complexes when individual partners or parts of proteins are deuterated. The buffering solution is then optimized in its ratio of $^{1}H_2O$ to $^{2}H_2O$ to match the scattering density of the deuterated partner protein such that the latter has zero contrast from the medium and can be completely hidden. Such conditions enabled us to test for a possible conformation change in the non-deuterated ETF upon complexation.

BCD from *A. fermentans* has been used by Buckel's research group to demonstrate bifurcation activity, making it a better characterized system than the ETF-BCD systems from *C. difficile* and *M. elsdenii*. The existing work focuses on mechanistic aspects of bifurcation. To complement this, we investigated structural changes in ETF when it forms a complex with BCD. We first established that a complex forms between ETF and BCD under our conditions, using size exclusion chromatography. We partially deuterated the BCD and determined its match point (contrast = 0) using SANS. We then performed SANS on the ETF-BCD complex at the match point and investigated the structural changes to ETF using the same approach as ETF alone. Our main finding was that we still see an ensemble of conformations of ETFs, even in the complex. However the relative populations of extended and compact conformations were shifted, with the majority of ETF populating an extended conformation in complex with BCD, and a minority of ETF populating a compact ETF conformation in complex with BCD. Some free ETF was also present in the solution, suggesting relatively weak binding, consistent with presence of

multiple conformations in the complex. Thus, our work has opened new avenues by which to investigate the structures of other ETFs in solution, using SANS. Moreover we have demonstrated the feasibility of doing SANS as a function of redox state, and in comparing free and complexed ETF.

We also studied the other important partner of ETF, Fld. Flavodoxins are highly soluble and stable proteins. In conditions of iron scarcity, they can replace Fd as the low potential electron acceptor of ETFs. We installed ^{19}F NMR labels on the tyrosine residues of Fld to address dynamics in the long loop proposed to associate with partner proteins, and around the FMN (chapter 4). ^{19}F NMR can serve as a simple but very informative tool to understand conformational dynamics when ETF and Fld interact. For completeness, we also solved the crystal structure of Fld.

CHAPTER 2. BIOLOGICAL SMALL ANGLE NEUTRON SCATTERING (SANS) AND SMALL ANGLE X-RAY SCATTERING (SAXS)

Small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) are two popular methods for studying structural properties of proteins and other important bio-macromolecules in solution, in near physiological conditions. Synchrotron and neutron scattering facilities around the world now give researchers access to these techniques along with powerful computational tools and user-friendly packages to assist with data analysis. These scattering methods inform on the size, and shape of macromolecules, permitting determination of oligomeric state, low-resolution structure, and even flexibility and degree of folding. Moreover, these properties are determined for macromolecules in solution under near-natural conditions, so one can learn about structural changes in response to different conditions or additions to the solution.

SAXS involves scattering of X-rays and SANS employs neutron scattering. While both are interpreted based on the principles of small angle scattering (SAS, section 2.2) and share general common applications, each has its own advantages, so they are often used in different in applications. SAS is used to probe materials at length scales ranging from Å to nm small angle data can be used to test structural models of macromolecules constructed from crystallographic data sets, homology models or *ab initio* calculations. They can serve as a rapid assay for the presence of

aggregates and precipitates in the solution. SAXS experiments can be performed at synchrotron or in house sources, in a matter of minutes. SANS experiments require a beam of neutrons, and so can only be performed at nuclear reactors or spallation sources, at considerable expense. However, SANS provides a unique advantage in that scattering sign and efficiency varies by atomic nucleus. Thus, selective deuteration and contrast matching permit the study of individual components of large protein complexes, providing invaluable simplification and insight into the composition of complicated biological assemblies. This unique capability is exploited in studies discussed later in this chapter.

2.1 Physical principles of Small Angle Scattering

In a basic SAS experiment, the sample is exposed to neutrons or X-rays and the scattered radiation is detected by a detector. The SAS experiments are performed very close to the neutron or X-ray beam at small angles. The wavelength of X-rays used is typically around 0.15 nm[52]. The dual nature of matter is an important principle for SANS where neutrons are treated as waves. According to the de Broglie relation, the wavelength of a neutron can be calculated as:

$$\lambda = {}^{h}/_{mv} \qquad \text{[Eq. 2.1]}$$

Where λ is the wavelength, h is the Plank's constant, m is the mass of neutron and v is the neutron velocity. The range of wavelength is usually 2 to 20 Angstroms[49]

when the velocity is varied from hundred to a few thousands per second. The scattering results from the interreference of the secondary scattered waves from the atomic nuclei of the sample in SANS. In SAXS the scattering is due to the interference of secondary scattered waves from the electronic cloud[53].

During a scattering experiment, an incident beam of neutrons or X-rays is directed at a sample such as protein, nucleic acids and/or membranes and scattered radiation is recorded at a given angle relative to the beam[54]. (Figure 2.1)

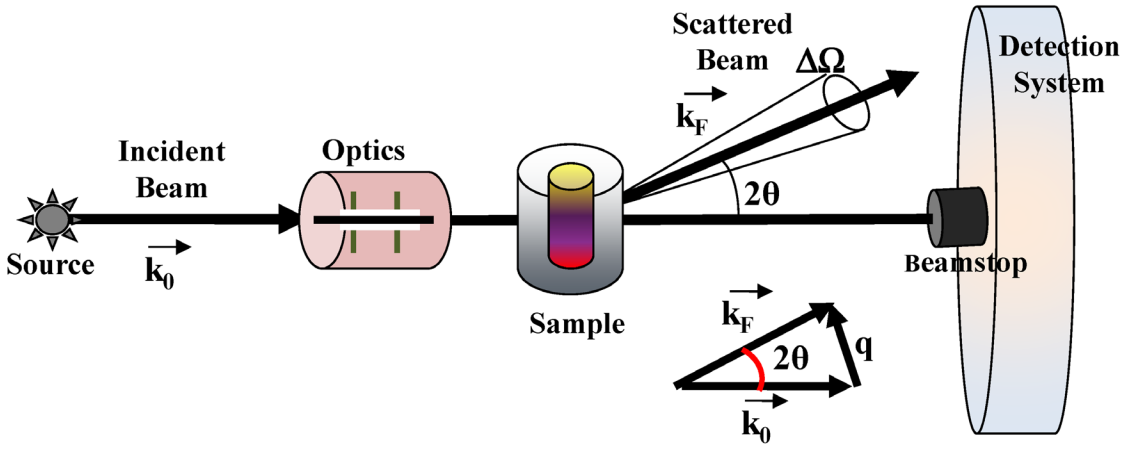

Figure 2:1 Typical scattering geometry for small angle scattering (SAS) experiments(Adapted from Lombardo et al,[54]).
An incident radiation beam from a neutrons or X-rays source impinging on the material system (sample) under investigation is scattered at a given scattering angle 2θ and is collected by a proper detection system. The difference between the scattered (k_F) and incident (k_0) wavevectors furnishes the scattering wavevector $q = |k_F - k_0| = (4\pi/\lambda)\sin(\theta)$, which has the dimensions of a reciprocal length (common units are $Å^{-1}$ or nm^{-1}).

The magnitude of the momentum transfer vector i.e., the scattering vector 'Q' is related to the wavelength lambda (λ) of the incident beam and the scattering angle 2θ and is defined by the following equation[54]:

$$Q = \frac{4\pi sin\theta}{\lambda} \qquad \text{[Eq. 2.2]}$$

The units for Q are usually inverse nanometers nm^{-1} or $Å^{-1}$. The intensity of the scattered X-rays or neutrons is recorded as a function of the scattering vector Q on a 2D detector by varying θ and denoted as the scattering intensity (I(Q)). The most general equation used to present the intensity of the scattering resulting from the particles present in the solution is, is given by[55]:

$$I(Q) = n\left\langle \left| \int_V (\rho(\vec{r}) - \rho_s)^2 e^{-i\,\vec{Q}\cdot\vec{r}} d^3 r \right|^2 \right\rangle \quad \text{[Eq. 2.3]}$$

Where *I*(Q) is the scattered intensity, n is the number of particles per unit volume, $\rho(\vec{r})$ is the scattering length density (SLD) of the particle at position r (discussed below), ρ_s is the scattering length density of the solvent and the integral is taken over the particle volume V. The integral is then averaged over time, all the orientations and the entire ensemble of structures in the solution.

The random positions and orientations of the particle result in an isotropic intensity. For monodispersed-non interacting particles, it is proportional to the scattering from a single particle averaged over all the orientations[56].

The intensity data are radially averaged, and the raw scattering data are converted to a one-dimensional scattering profile that contains average intensity I(Q) versus the scattering vector Q[57].

An important parameter that can be calculated from a good quality scattering data is the forward (or zero angle) scattered intensity, *I*(0). *I*(0) is the intensity of the scattered radiation through zero angle. In reality this value cannot be measured as it cannot be distinguished from the radiation that passes through the sample

unscattered. $I(0)$ is therefore calculated by extrapolation. It is related to the number of scattering particles per unit volume (N) and the particle volume (V) squared. It can also be expressed in terms of the mass per unit volume (C), molecular weight (MW), the contrast (Δρ) and the partial specific volume (v) of the particle[58]:

$$I(0) = N(\mathrm{Dr})^2 = \frac{C\mathrm{D}r^2v^2MW}{N_A} \text{ [Eq 2.4]}$$

Here N_A is the Avogadro's number.

2.1.1 Scattering length and Cross-section:

Both SANS and SAXS involve interference between the wavelets scattered by different elements in the system. In the case of neutron scattering, the incident wave from the neutron beam is represented as a plane wave and is described by the equation[59]:

$$\psi_0 = e^{ik_0z} \qquad \text{[Eq.2.4]}$$

The scattered wave is then a spherical wave and is described by the equation:

$$\psi_1 = \frac{b}{r}e^{ik_1.r} \qquad \text{[Eq. 2.5]}$$

Here r is the distance or position of the detector from the nucleus. *b* has dimensions of the length and measures the strength of scattering ability of the nucleus. *b* is constant for a given nucleus or isotope. In a regular SANS experiment, numbers of neutrons scattered are counted in a particular direction. If dΩ is the solid angle subtended by the detector, then the differential cross section is described by the equation[59]:

$$\frac{d\sigma}{d\Omega} = \frac{(neutrons\ s^{-1} scattered\ into\ d\Omega)}{\Phi d\Omega} \qquad \text{[Eq 2.6]}$$

Where Φ is the incident neutron flux (number of neutrons per unit area per second). The total scattering cross section is defined as the total number of neutrons scattered per second, normalized to the flux[59]:

$$\sigma_s = \int \left(\frac{d\sigma}{d\Omega}\right) d\Omega \qquad \text{[Eq 2.7]}$$

Where the integral is over all the directions. The total cross section can be related to b. If the incident neutrons have the velocity v, then the number of neutrons passing through an area dS s^{-1} is:

$$vdS|\psi^2| = vdS\frac{b^2}{r^2} = vb^2\, d\Omega \quad \text{[Eq. 2.8]}$$

From the definition of neutron cross section,

$$\frac{d\sigma}{d\Omega} = \frac{vb^2 d\Omega}{\Phi d\Omega} \qquad \text{[Eq 2.9]}$$

Where $\sigma_s = 4\Pi b^2$

σ_s has dimensions of area. The cross section is considered as the effective area which the target nucleus has available for the incident neutrons for elastic scattering. It is known as bound atom cross section as nucleus is considered fixed for the samples[60].

Neutrons are scattered isotopically from individual nuclei, whereas for X-ray scattering, the scattering originates in the electron cloud, which is very large compared to the X-ray wavelength. In the case of X-rays, the atomic form factors are Q dependent. For SAXS, the variation is small. The Thompson scattering

amplitude for electron is r_T=0.282 10^{-12} cm, so the X-ray scattering length of an atom, *f* is given by:

$$f = r_T Z \qquad \text{[Eq 2.10]}$$

Z is the atomic number. The scattering length increases with the number of electrons/atoms, while for neutrons the values of *b* vary from isotope to isotope[56]. If the nucleus has nonzero spin, it can interact with the neutron spin, and the total cross section contains both coherent and incoherent components[53, 61].

2.1.2 Coherent and Incoherent Cross-sections:

The neutron scattering cross section can be divided into coherent (*b*) and incoherent components. The coherent section contains information on interference effects arising from spatial correlations of the nuclei in the system, thus it gives the information on the structure of the sample. The incoherent cross section has no information on interference effects and forms an isotropic/flat background which must be subtracted off in SANS structural investigations. The coherent scattering length is analogous to the atomic form factor *f*, while there is no analogue for incoherent scattering length in X-rays.

When a neutron of spin 1/2 encounters a single isotope with nuclear spin I, the spin of the neutron nucleus system can assume two values, I±1/2 Therefore, the scattering lengths of the two system are denoted by b^+ or b^-. The number of spin states associated with each one of them are 2(I+1/2) +1=2I+2 and 2(I-1/2)+1= 2I respectively[59, 62]. The total number of states is 4I+2. For unpolarized neutrons and randomly oriented nuclear spins each spin state has the same probability. Thus,

the frequency of b^+ occurring is weighted by (I+1)/(2I+1), and for b^-, I/(2I+1). The coherent cross section for each isotope is given by $s_c = 4P\,\overline{b}^2$, where $\overline{b}$ represents the thermally averaged scattering length with + and – spin state populations. Similarly, the total scattering cross-section is given as $s_s = 4\pi\overline{b^2}$. The average coherent scattering length is given by[59]:

$$\overline{b} = \frac{1}{2I+1}[(I+1)b^+ + Ib^-] \quad \text{[Eq. 2.11]}$$

$$\overline{b^2} = \frac{1}{2I+1}[(I+1)(b^+)^2 + I(b^-)^2] \quad \text{[Eq 2.12]}$$

The difference between s_s and s_c is the incoherent cross section, σ_i

The isotopes with no spin such as ^{12}C, $b^2 = \overline{b} = \overline{b^2}$, therefore there is no incoherent scattering. Most of the atoms encountered in neutron scattering are mainly coherent scatterers (Carbon, Oxygen). The main exception is hydrogen (^1H), the spin up and spin down scattering lengths have opposite signs ($b^+ = 1.080*10^{-12}$ cm, $b^- = -4.737*10^{-12}$ cm)[59].

So, for hydrogen $\sigma_c = 1.76*10^{-24}$ cm^2 and $\sigma_i = 79.7 * 10^{-24}$ cm^2. Table 1 lists the coherent scattering lengths of the most common atoms in biology.

Table 2.1: Examples of the SANS and SAXS atomic coherent scattering lengths b (10−12 cm) of relevant elements of the biological systems[54].

Element	Z	SANS	SAXS
Hydrogen (1H)	1	−0.374	0.28

Deuterium (2D)	1	0.67	0.28
Carbon (12C)	6	0.66	1.69
Nitrogen (14N)	7	0.94	1.97
Oxygen (16O)	8	0.58	2.25
Aluminium (27N)	13	0.35	3.65
Phosphorous (31P)	15	0.51	4.23
Sulfur (32N)	16	0.28	4.50

Table 2.2: SANS and SAXS Scattering length density [SLD] (1010cm−2) for H_2O and D_2O molecules[54].

Molecule	SANS	SAXS
H2O	−0.56	9.39
D2O	6.73	9.38

2.2 Scattering length density Contrast variation

Scattering experiments can exploit or probe the composition of samples, as well as their particulate natures. Sample nature is represented substantially by the

scattering length density (SLD), which describes the sample material's interaction with the incident beam of neutrons or X-rays. SLD is defined as the sum of coherent scattering lengths over all atoms lying in a given volume V, divided by V. SLD describes how (on average) the component molecules that effectively scatters the incident radiation. The calculation of SLD consists of summation of the atomic scattering length of each atom over a volume that is representative of a macromolecule where scattering features do not vary. More specifically, the SLD from a (macro-)molecule containing x_i atoms i is given by:

$$\rho i = \sum \frac{xibi}{v_m} \qquad \text{[Eq 2.13]}$$

where the sum is extended to the scattering length contributions b_i from the different N atoms contained in the molecular volume v_m. [54]

In SAXS, SLD is always a positive number and is proportional to the atomic number. For SANS there is no direct proportionality and SLD is different for different atomic numbers/isotopes with positive or negative signs. Because the compositions of proteins, nucleic acids and lipids include abundant Carbon, Nitrogen, and diverse other atoms in smaller amounts, their SLDs differ from the Oxygen and Hydrogen-dominated aqueous solvents that are most commonly used[61, 63, 64]. This produces a contrast between the proteins and their solvent, with respect to scattering. Contrast is defined by the difference is scattering length density of the macromolecule (e.g protein) relative to its surrounding environment (buffer). The scattering intensity is dependent on the scattering contrast Δρ, and I (Q) can be written as the Fourier transform of spatial distribution and is given by the equation[55, 61]:

$$I(Q) = \frac{Dr}{V}\left|\int_V e^{-i\,Q\cdot r}\,dr\right|^2 \qquad \text{[Eq 2.14]}$$

Where r is the real space vector of atoms in the scattering particle and the distribution is integrated over the entire scattering volume V. Δρ is the scattering contrast between the solvent and the proteins.

The crucial feature of SANS that makes it very useful in biology is the different behavior of hydrogen and deuterium atoms under a neutron beam. Hydrogen and deuterium scatter with opposite signs. The coherent nuclear scattering length of hydrogen is -3.74 fm while that of the deuterium is 6.67 fm. In biomolecules hydrogen can be exchanged with deuterium which can have dramatic effects on scattering. The technique of contrast variation is based on differential scattering of hydrogen and deuterium. For SANS, samples can be dissolved in deuterated solvents so exchangeable hydrogens can be replaced by deuterium. The overall scattering of the biomolecule will change depending on the ratio of hydrogen to deuterium in the molecule. Biomolecules can also be produced to contain a desired ratio of H to D. Alternatively, for a biomolecule of fixed composition, buffer can be made using a ratio of H_2O to D_2O (called the match point) at which the SLD of the biomolecule is equal to that of the solvent. Table 2 above lists the SLD for H_2O and D_2O. There will be no contrast and the scattering due to the biomolecules will be removed along with that of the solvent by subtraction of scattering due to buffer (determined from measurement of a buffer-only control sample). Contrast variation enables structural characterization of two proteins forming a complex (one proteated and other deuterated). Figure 2.2 shows that in SANS experiment when such complex is formed, the scattering intensity of proteated protein can be contrast

matched with a buffer containing 42% D_2O, so that observed the scattering derives only from the deuterated protein. Similarly scattering from the deuterated protein can be contrast matched with 100% D_2O buffer, resulting in net scattering only from the proteated protein. In our work on ETF and BCD, we produced an 70 percent deuterated BCD protein. We calculated its match to be 74.5% D_2O: 25.5% H_2O buffer. Thus, at this condition, we were able to measure scattering from ETF without interference from that of BCD, in an ETF: BCD complex.

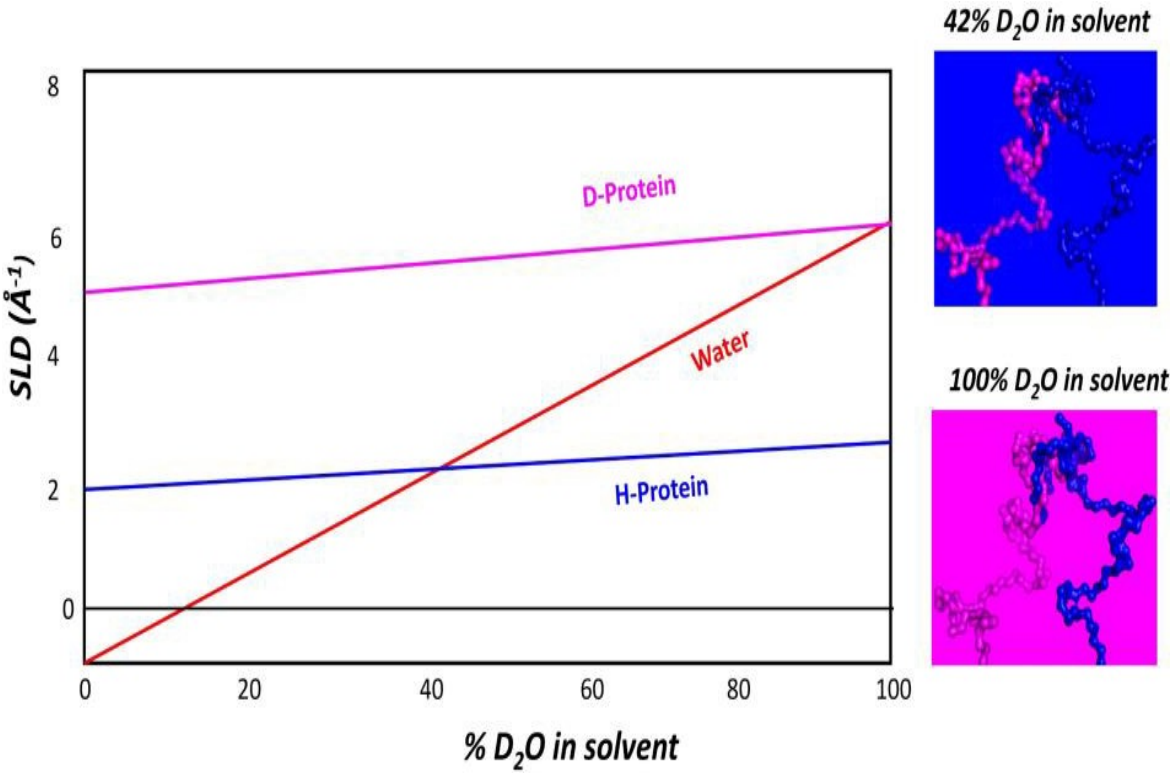

Figure 2:2 Contrast variation in SANS

The plot on the left shows how the scattering length densities of proteated or deuterated proteins change as a function of the D_2O/H_2O ratio in the solvent (due to equilibration of exchangeable Hs to the ratio of H:D in the solvent). It also depicts the variation in the scattering length density of the D_2O/H_2O mixtures themselves (red). Crossing points of either protein's like with that of solvent corresponds to a match point. In 42% D_2O in the solvent, proteated protein (blue line) is contrast matched to the solvent background and in 100% D_2O, deuterated protein (magenta) is contrast matched. Adapted from smith jc et al[63].

2.3 General Instrumentation:

A general schematic of SAS instrument is shown in figure 1 above.

1. Source

Neutrons are usually produced by nuclear reactors or spallation sources. Fission reactions are used at nuclear reactors to produce neutrons, usually from uranium (^{235} U). Alternatively, neutrons are generated upon bombardment of heavy metal targets with high-energy protons, at spallation sources[55]. The high flux isotope reactor (HFIR) at the Oak Ridge National Laboratory (ORNL) was used for SANS experiments for our research. ORNL also encompasses a spallation source called the spallation neutron source (SNS). X-rays are generally produced in laboratories using rotating anodes typically made of copper or molybdenum by thermionic emission. For more intense beams, X-rays are produced using synchrotron sources using particle accelerators, enabling rapid data collection and study of weakly scattering samples. X-ray Free Electron Lasers (XFELs) are also emerging as sources of X-rays[65].

2. Collimation System

The collimation systems for SANS are designed to produce a neutron beam of uniform intensity over its cross-section. The collimation system controls the beam

divergence, and thus determines the minimum scattering angle which ultimately decides the Q value. SANS instrument collimation system works like a pinhole camera, it involves the use of two apertures, round holes or rectangular slits that defines the size of the It consists of apertures and neutron guides for properly directing the beam[49, 55]. The aperture size is controlled and can be increased or decreased to obtain a beam with desired intensity. Guides are reflective tubes usually made with nickel or super mirror coatings to control the directionality of the beam without loss in intensity. The collimation system is very important for obtaining the best signal to noise ratio for the experiment, without loss of resolution. SANS instruments also incorporate a monochromator to select neutrons of a specific wavelength from the broader spectrum produced by the source. The monochromator is usually made up of single crystals of graphite or silicon. The neutron beam from the source is diffracted by the crystals. The wavelength can be selected based on Bragg's law of diffraction.

In SAXS, X-ray optics are used to shape and focus the X-ray beam onto the sample. The optics may include mirrors, lenses (such as compound refractive lenses), and monochromators[65]. The choice of optics is very important and depends on the source and the requirements of the experiment, such as the need for a monochromatic beam or a particular beam size.

3. Sample Environment

The sample environment is the chamber and environmental control system in which samples are staged and into which the beam is directed. The sample environment

varies depending on the type of experiment and conditions to be maintained. For SANS, it can involve temperature-controlled devices, pressure cells, magnetic fields, and other means of producing the conditions desired for the conduct of the experiments. It also incorporates means of positioning the sample in the path of the beam, and moreover of moving successive samples into position under remote control. Thus, sample environments can be designed to accommodate various sample types (solutions, gels, powders) and to control experimental conditions such as temperature, pressure, illumination, or shear.

4. Detector

In SANS the detector captures scattered neutrons after they interact with the sample. Modern SANS detectors are typically two-dimensional arrays which quantify scattered beam intensity as a function of position, so that scattering intensity is measured as a function of scattering angle. The main principle is that the detectors are wire detectors that collect charge that results from ionization events caused by neutrons impinging on a gas[55]. The most efficient SANS detectors are ^{3}He detectors[64]. Challenges include that these detectors must be sensitive to neutrons and must cover a large area with high spatial resolution to accurately map the scattering pattern. In SAXS, Charge-Coupled Devices (CCDs) are the most common detectors: These are sensitive to X-rays and capable of capturing high-resolution images of the scattering pattern.

5. Data Acquisition and Analysis System

The data acquisition system collects the signals from the detector and converts them into a format suitable for analysis. This system typically includes software to process and analyze the scattering data, extracting meaningful structural information about the sample. The analysis might involve correcting for background scattering, normalizing the data, and fitting it to models or using it to compute the size distribution and shape of the scatterers.

6. Beam Stop

A device placed between the sample and the detector to block the direct (unscattered) neutron or X-ray beam(Figure 1), preventing damage to the detector, and improving the dynamic range of the measurement.

2.4 Scattering Data Analysis

First, data are averaged, and the contribution of solvent (and matched components) are removed by subtraction of scattering from a sample containing the same medium but lacking the molecule or complex of interest.

2.5.1 Guinier Analysis

The first parameter that is usually calculated from the scattering data is the radius of gyration, R_g. It is defined as the root- mean squared distance of all elemental

scattering volumes from their center of mass weighted by their scattering densities. It is usually smaller than the actual radius, R. For a hollow spherical shell R_g= R but for a solid sphere R_g= $(3/5)^{1/2}$ R which is nearly equal to 0.77 R[58]. For elongated shapes of similar volume, the larger the axial ratio, larger is the R_g. R_g is thus an immediate and physical measure of size and shape of the protein and the oligomeric state. Proteins with different molecular weight and shape can still have similar R_g.[58] Guinier analysis is performed on the scattering data of macromolecular complexes to calculate R_g. Guinier analysis can provide quaternary structure information on macromolecules by informing about the size of the scattering particle. It can also be used to calculate molecular weight of the tumbling complex from the forward scattering density. The Guinier approximation for scattering at low scattering angles is[66]:

$$I(Q) = I(0)e^{\frac{-R_g^2 Q^2}{3}}$$ [Eq. 2.15]

Where I(0) and R_g are the forward scattering intensity and radius of gyration, respectively. For a monodispersed system, the linearization of the above equation (lnI(Q) vs Q^2) is known as Guinier plot and the I(0) and R_g values can be calculated from it. For globular and spherical particles, the Guinier approximation holds under only when Q_{min}*R_g≤0.65 and Q_{max}Rg≤1.3. Q_{min} and Q_{max} denote minimum and maximum Q values respectively.

The Guinier plot is also a crucial assessor of the viability of further SANS studies. If a linear Guinier plot is not obtained, more work will be required to identify better solution conditions, or possibly even a different variant of the protein that is better-

behaved. However, the nature of the deviation from linearity provides guidance for doing so[58]:

1. A plot that curves up as Q goes to zero suggests aggregation, radiation damage, inter-particle interactions or high polydispersity of the protein sample.

2. A plot that curved downward as Q goes to zero indicates inter-particle repulsion in the protein sample. This may be mitigated by higher ionic strength.

Figure 2.3 shows how a Guinier plot looks like when there is aggregation.

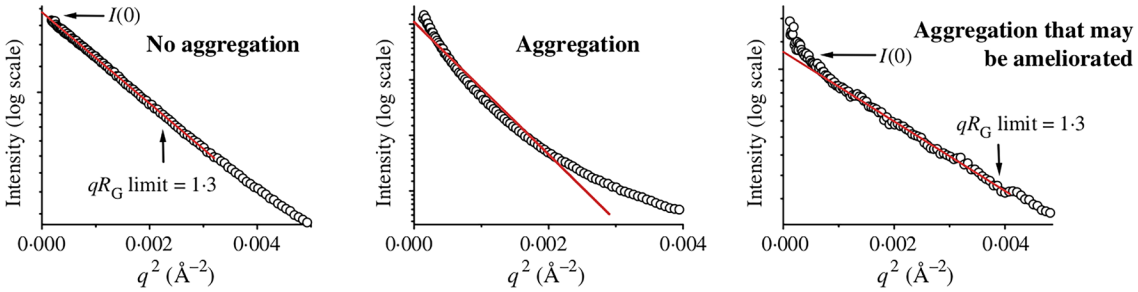

Figure 2:3 Guinier plots for different levels of aggregation.
Adapted from [67] showing no (left panel), large (middle panel) and small (right panel) amounts of aggregates seen in the Guinier fit.

2.5.2 Pairwise distribution function *P(r)*:

Although I(Q) vs Q plot is related to the shape of the protein in solution, it is not very informative about the structure of the protein. To get a better insight a Fourier transform is done on the scattering data to obtain the interatomic distance distribution function, *P(r)*(radial Patterson function). It is also referred to as pair distance or vector length distribution function. It is a histogram of distances between all possible pairs of atoms within a particle. The fourier transform that is used to evaluate *P(r)* is given by the equation:

$$P(r) = \frac{r}{2\pi^2}\sum_0^\infty I(q)qsin(qr)dq \qquad \text{[Eq. 2.16]}$$

In reality data cannot be collected from zero to infinity, therefore an indirect Fourier transform method is used with the assumptions that *P(r)* is zero at r=0 and at the maximum linear dimension D_{max}. D_{max} is an important parameter in the scattering data and high-quality scattering data is required to estimate proper values. In indirect Fourier transform the *P(r)* decays smoothly to zero without any noises like oscillations or deviations in the curve. The D_{max} value is increased or decreased until a suitable value is found.

Additionally, R_g and *I*(0) values can be calculated from this analysis. Comparison of the R_g and *I*(0) values obtained from Guinier analysis and from the *P(r)* function is a standard way to check the accuracy of scattering data analysis.

The scattering intensity is related to the *P(r)* function by the following equation:

$$I(Q) = \int_0^{D_{max}} P(r)\sin\frac{Qr}{Qr}dr \qquad \text{[Eq. 2.17]}$$

Finally, the *P(r)* profile provides the intermolecular (r) distances of the monodispersed scattering protein particles. The distance distribution data will appear as a proper Gaussian function distribution for a well folded globular protein. R_g is the second moment of the *P(r)* function and described by the equation:

$$R_g^2 = \frac{\int P(r)r^2dr}{2\int P(r)dr} \qquad \text{[Eq. 2.18]}$$

$I(0)$ is the zeroth moment, which corresponds to simply the area under P(r)

$$I(0) = 4\pi \int_0^{\mathrm{Dmax}} P(r)dr \quad \text{[Eq. 2.19]}$$

An important criterion for a good P(r) profile is that that the P(r) function falls gradually to zero at D_{max}. Figure 2.4 shows different profiles for the same Glucose isomerase protein. In the plot below, it is clearly visible for a D_{max} of 83, the P(r) function is forced abruptly down. For a D_{max} of 103, the function approaches smoothly to zero. For a D_{max} of 123 the function reaches zero but then oscillates about it. Therefore, 103 is a good value for D_{max}, whereas 83 is underestimated and 123 is overestimated.

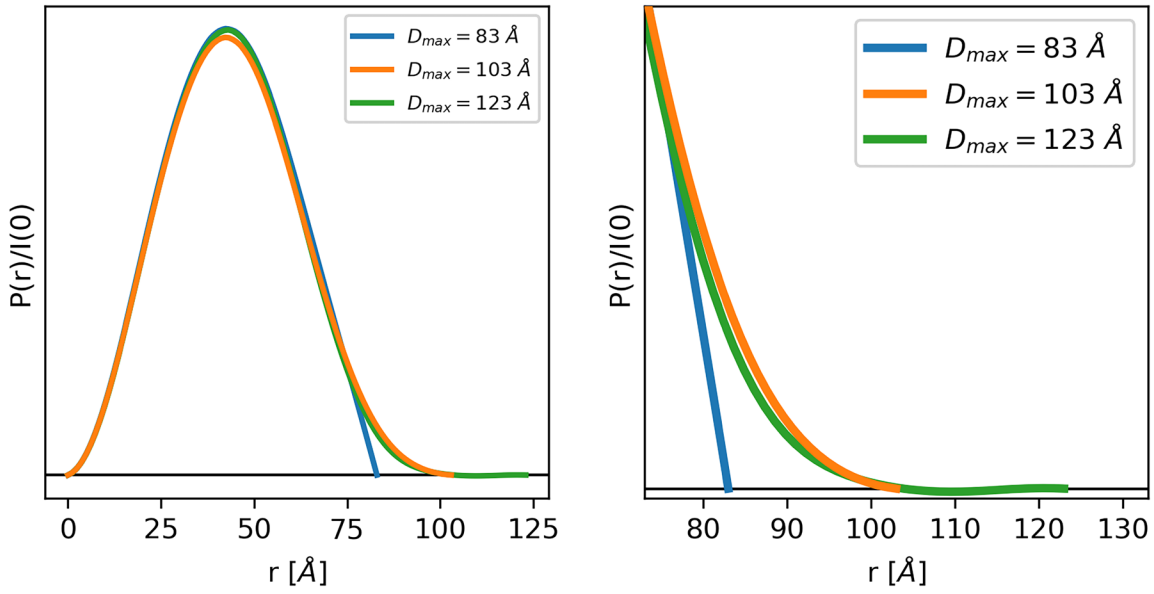

Figure 2:4 P(r) profiles of glucose isomerase
Adapted from BioXTS-RAW tutorial on Glucose isomerase. The left plot shows the full P(r) function. The different D_{max} values yield similar P(r) functions, so much so that they end up plotted on top of each other for most of their r values. The right plot is the same functions showing just the end, as P(r) approaches zero at D_{max}.

2.5.3 Kratky Analysis

The flexibility of proteins can result in specific effects on the small angle scattering pattern that can be used to characterize the flexibility of the protein. In a Kratky plot, the intensity is multiplied by Q^2 (plot of $I(Q)*Q^2$ *versus* Q) to scale down the very intense scattering at small Q. The Kratky plot is unique and distinct for a well folded (functional three-dimensional structure), flexible and denatured protein in the small angle regime (Figure 2.5). In a Kratky plot, a globular protein exhibits a Gaussian-shaped curve, characterized by values within the range $0 < Q*R_g < 0.3$ with its maximum occurring at the Q value that satisfies $Q*R_g \sim \sqrt{3}$. A partially

folded or a multidomain protein with long disordered regions can have a single maximum too, but it will not have a Gaussian bell shape and instead will tend to display an extended plateau with increasing Q. A completely unfolded protein will not have a single maximum but will display a sustained amplitude at higher Q values. Dimensionless Kratky plots are obtained by dividing I(Q) with I (0). Dimensionless Kratky plots can provide semi-quantitate analysis of flexibility and disorder. Figure 2.5 below shows Kratky plots for different types of proteins.

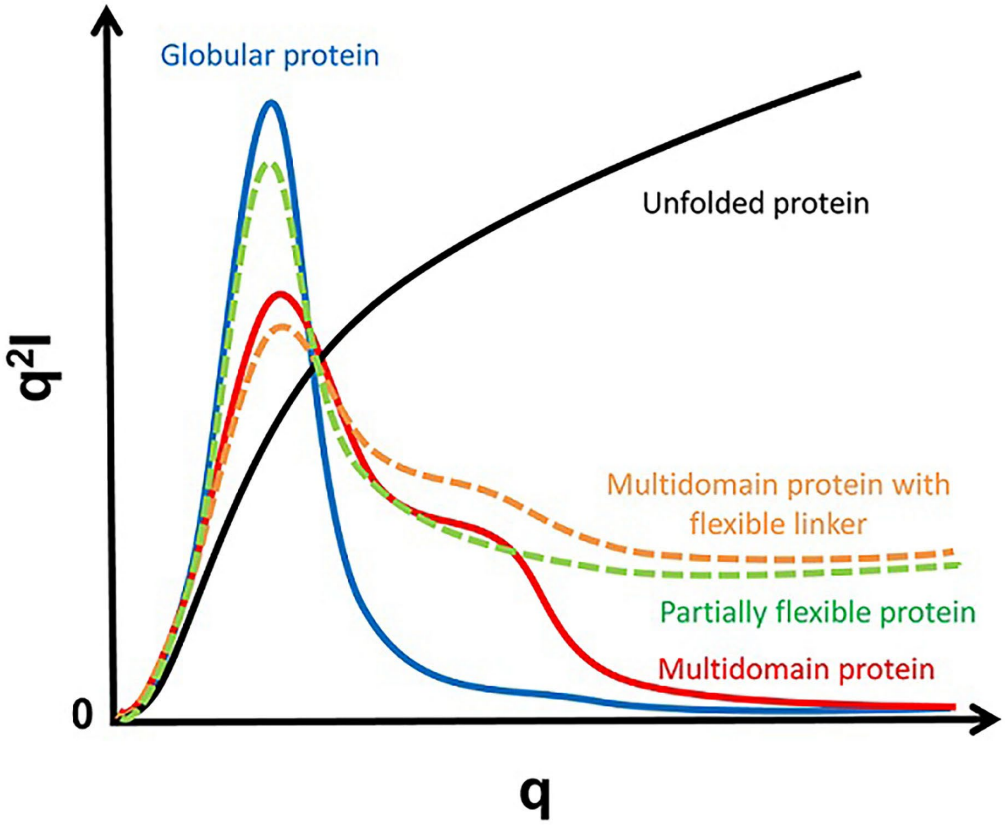

Figure 2:5 Kratky plots for different kinds of proteins
Adapted from Stanford SLAC SAXS tutorial. Kratky plot indicates different flexibility for different types of proteins.

2.5.4: χ2(Chi-squared)

χ2(Chi-squared) is a statistical measure used to assess the fit of a model to the experimental data. It quantifies the discrepancy between the observed scattering data and the scattering intensity predicted by a model. This model could represent the shape, size, and internal structure of macromolecules or complexes in solution. For the analysis of SAS data on proteins, a theoretical scattering profile is generated based on the crystal structure of the protein. The χ2 is then calculated by comparing the theoretical scattering and the experimental scattering to check the deviance.

The calculation of χ2in the context of biological SAS typically follows this formula:

$$\chi 2 = \frac{1}{N-P}\sum_{i=1}^{N}\left(\frac{I_{exp,i}-I_{model,i}}{\sigma_i}\right)^2 \quad \text{[Eq. 2.20]}$$

where:

- N is the number of data points in the experimental scattering curve.
- Pis the number of parameters in the model being fitted to the data.
- $I_{exp,i}$ is the experimentally measured scattering intensity at the *i*-th point.
- $I_{model,i}$ is the scattering intensity predicted by the model at the *i*-th point.
- σ_i is the experimental uncertainty (standard deviation) of the *i*-th data point.

Interpretation

- A low $\chi 2$ value indicates a good fit between the model and the experimental data, suggesting that the model accurately represents the structure that generates the scattering pattern.

- A high $\chi 2$ value suggests a poor fit, indicating discrepancies between the model and the experimental observations. This could mean that the model does not capture some aspects of the sample's structure or that there are systematic errors in the data.
- It's important to interpret $\chi 2$ values in the context of the specific system and data quality. Values close to 1 are generally considered indicative of a good fit, but this can vary depending on the specifics of the experimental setup and the data quality.

  Caveats
- The significance of $\chi 2$ depends on the accuracy of the estimated experimental uncertainties (σi). Underestimation or overestimation of these uncertainties can lead to misleading $\chi 2$ values.
- $\chi 2$ alone does not provide a complete picture of the fit quality. It should be used in conjunction with visual inspection of the fit and other statistical measures, as well as considering the physical plausibility of the model.

  Figure 2.6 below compares scattering profiles of a model lysozyme protein. The SAXS data for the same protein has been shown at good quality and when there is presence of aggregates or interparticle repulsion. The fits in green show how ideal scattering analysis should look like.

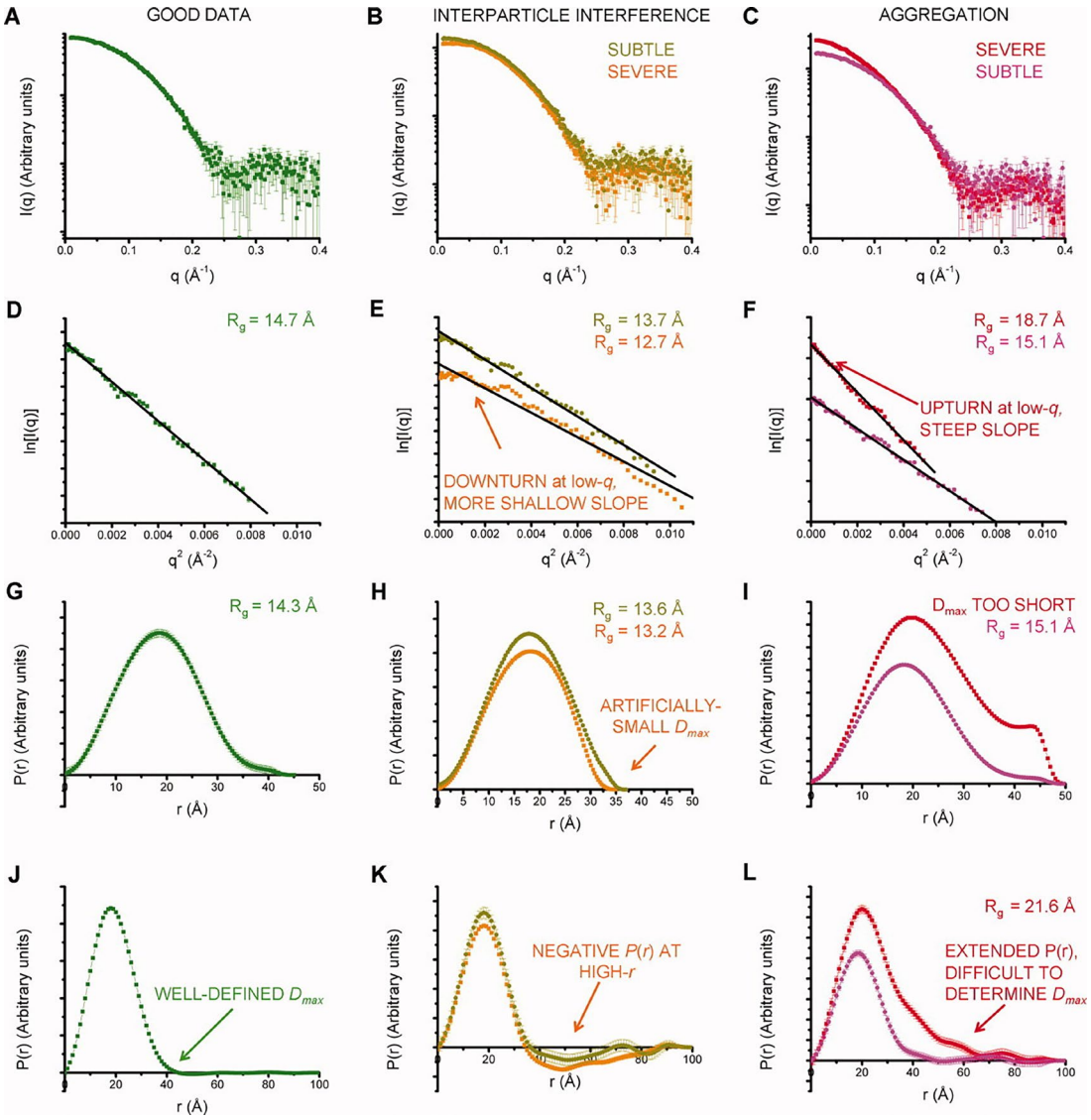
A
GOOD DATA
I(q) (Arbitrary units)
q (Å^{-1})
B
INTERPARTICLE INTERFERENCE
SUBTLE
SEVERE
C
AGGREGATION
SEVERE
SUBTLE
D
R_g = 14.7 Å
ln[I(q)]
q^2 (Å^{-2})
E
R_g = 13.7 Å
R_g = 12.7 Å
DOWNTURN at low-*q*,
MORE SHALLOW SLOPE
F
R_g = 18.7 Å
R_g = 15.1 Å
UPTURN at low-*q*,
STEEP SLOPE
G
R_g = 14.3 Å
P(r) (Arbitrary units)
r (Å)
H
R_g = 13.6 Å
R_g = 13.2 Å
ARTIFICIALLY-
SMALL D_{max}
I
D_{max} TOO SHORT
R_g = 15.1 Å
J
WELL-DEFINED D_{max}
K
NEGATIVE *P*(*r*) AT
HIGH-*r*
L
R_g = 21.6 Å
EXTENDED P(r),
DIFFICULT TO
DETERMINE D_{max}

Figure 2:6 SANS data quality analysis

Adapted from [58] The impact of particle interaction and clumping on data from small-angle scattering is challenging to discern merely from I(q) graphs (A–C). The use of Guinier charts (D–F) can be enlightening because deviations from a straight line distinctly signal particle interference (E, orange) or clumping (F, red). The presentation of P(r) with a selected singular D_{max} value (G–I) might mislead, often only flagging extreme clumping cases (I, red). Examining P(r) behavior with larger D_{max} figures (J–L) offers a more definitive insight into sample integrity. It's crucial to acknowledge that minor cases (K, dark yellow and L, magenta) might go unnoticed with these preliminary examination techniques, making it imperative to compare the findings against secondary benchmarks. When noticing differences around 1 Å in R_g, it's vital to understand that such discrepancies greatly exceed the accuracy with which R_g can usually be pinpointed (to within several tenths of an Ångström for small to medium proteins). This level of error can significantly skew any 3D reconstructions based on the scattering data.

2.6 Size Exclusion Chromatography

Size exclusion chromatography (SEC) is a chromatographic technique in which molecules in solution are separated on the basis of their sizes (and thus their molecular weights). It is routinely used to purify proteins before small angle scattering experiments because its operation on the basis of size also resolves different oligomers of the target protein, which would not be separable on the basis of affinity chromatography or ion exchange. Monodispersity as well as purity are pre-requisites for scattering experiments. The SEC column is packed with fine, porous beads. The beads can be composed of dextrose, agarose, or polyacrylamide, but they will contain pores that can designed to be of different sizes for different applications. Each size exclusion column is designed for a range of molecular weighs that it can separate. As the mobile phase (buffer solution) passes through the column, some particles are trapped in the pores. Larger particles are excluded

and therefore flow through the column more quickly than the smaller biomolecules. Therefore, the larger complexes of biomolecules elute first. The main requirement for SEC is that the analyte does not have any chemical interaction with the stationary phase and the separation is solely based on the size of the particles.

SEC is usually the final step of protein purification before scattering. It helps in removing any soluble aggregates and impurities from the protein sample. Figure 2.7 shows a SEC experiment being performed on AKTA FPLC system and a SEC elution profile. SEC can be coupled with SAXS or SANS to produce a hyphenated technique called SEC-SAXS or SEC-SANS. Proteins that aggregate fast and do not stay monodispersed for a long time may require such experiments, which enable immediate data collection upon elution of each fraction from the column.

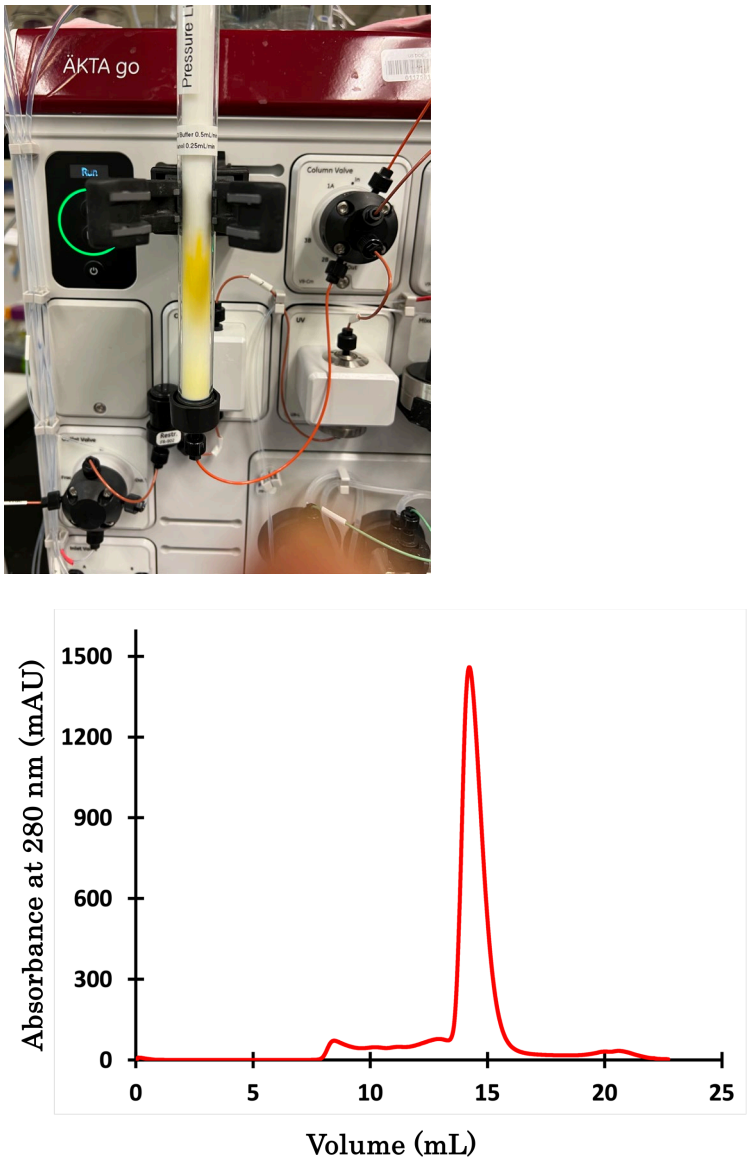

Figure 2:7 SEC experiment (left) performed on ETF protein using an AKTA FPLC system.
The SEC elution profile on the right shows a sample with relatively purity as indicated by the single dominant peak. Some aggregates and possible smaller impurities are seen, eluting before and after the main fraction.

CHAPTER 3. POPULATIONS OF EXTENDED CONFORMATIONS ARE SUBSTANTIAL AND RESPONSIVE TO OXIDATION STATE, BASED ON SMALL ANGLE NEUTRON SCATTERING OF BIFURCATING ELECTRON TRANSFER FLAVOPROTEIN.

I would like to acknowledge contribution from the following authors:

Alan Hicks: Computational SANS data modeling,

Wellington Leite: SANS

Biswajit Gorai: Metadynamics simulation

James Byrnes: SEC-SAXS

Maria-Andrea Mrogonski: Supervisor to Biswajit Gorai & collaborator

Hugh O'Neill: P.I at ORNL & supervisor

Anne-Frances Miller: P.I, supervisor, and corresponding author

Abstract

Electron transfer bifurcation enables biological systems to drive endergonic electron transfer by coupling it to exergonic transfer of a second electron. This enables production of strong reductants at the expense of more weakly reducing starting material, and in turn enables enzymes to drive difficult reactions based on modest but abundant fuel. In electron transfer flavoproteins (ETFs), a domain-scale conformational change is believed to sever the exergonic pathway after a single electron has used it, thereby preventing the energy dissipation that would accompany exergonic transfer of the second electron. To understand the

conformation change that participates in turnover, we have deployed small angle neutron scattering (SANS). We find that none of the extant ETF structures is able to explain the observed scattering, and nor is any individual conformation generated by two contrasting molecular dynamics protocols. To optimize a description in terms of an ensemble of conformations, we developed a genetic algorithm to identify ensembles best able to explain the SANS data. Ensembles containing a compact conformation comparable to one of the crystallographically documented conformations, accompanied by a much more extended one, succeeded. The extended conformations are different from any currently solved ETF structures. The extended conformations are also strongly populated at over 20%, and their population is responsive to reduction of the ETF's flavins, indicating that can contribute to turnover. We propose that the extended conformations provide an accessible efficient path for interconversion of the open and closed conformations believed essential at different points in catalysis. Their substantial population even at room temperature in solution indicates a very low kinetic barrier between the open and closed conformations, consistent with efficient rapid enzymatic turnover.

3.1. Introduction:

Electron transfer bifurcation (EB) enhances energy efficiency in both aerobic and anaerobic organisms, by coupling an endergonic (energy capturing) electron transfer to an exergonic (energy expending) transfer. Thus, an otherwise unfavorable transfer becomes spontaneous as part of the coupled process. In-effect,

the energy of one electron is augmented at the expense of the other one. The high-energy electron is transferred to a low reduction potential acceptor (low E°, strongly reducing), while the low-energy electron is acquired by a high-potential acceptor, ensuring the conservation of the system's total energy, despite its unequal redistribution among electron transfer paths. EB has been known since the 1970s, when it was elucidated by Mitchell et al in the context of the respiratory chain bc1 complex[30, 31, 68]. In this case, ubiquinol (E_0‘= +90 mV) is the source of a pair of electrons, the Rieske iron-sulfur cluster protein (E_0‘= +285 mV) acquires one at high E° thereby 'paying for' direction of the other to cytochrome *b* (E_0‘= -110) mV, as the low E° acceptor.

In 2008, Flavin-based EB (FBEB) was identified by Buckel *et al*. [24]in anaerobes. FBEB employs flavins instead of quinones as the site of bifurcation, with the lower E°s of flavins being consistent with the more reducing metabolism of anaerobes. Flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) serves as the site of bifurcation and yields reduced low-potential electron carriers such as ferredoxins or flavodoxins, based on abundant but less reducing two-electron donors, including NADH, reduced coenzyme F_{420}, H_2, or formate[29]. FBEB was first documented in *Clostridium kluyveri*, where the butyryl-CoA dehydrogenase (BCD) in combination with electron transfer flavoprotein (ETF) were demonstrated to couple endergonic reduction of ferredoxin (E_0‘= -420 mV) by NADH (E_0‘= -320 mV) to the exergonic reduction of crotonyl-CoA (E_0‘= -10 mV) to butyryl-CoA[24]. Since then, FBEB has been studied in numerous bifurcating ETFs (Bf-ETFs) from a growing diversity of species, including *Acidaminococcus fermentans*,

Acetobacterium woodii, *Rhodopseudomonas palustris*, and *Thermotoga maritima* [25, 26, 42, 43, 69].

ETFs are heterodimeric FAD containing proteins, that were originally discovered in mitochondria, where they associate with the inner membrane and contribute electrons to the electron transport chain[20]. Notably, mitochondrial ETFs, or 'canonical' ETFs, contain one FAD and one AMP as cofactors and shuttle electrons from fatty acyl CoA dehydrogenases to the quinone pool[70, 71], whereas bifurcating ETFs (Bf-ETFs) contain two FADs. The FAD that has no counterpart in canonical ETFs nevertheless binds such that its AMP fragment is superimposable with the AMP of canonical ETFs. This and conservation of residues nearby over both canonical and Bf-ETFs indicate that canonical ETFs evolved from bifurcating predecessors[72]. The FAD unique to Bf-ETFs is called the bifurcating FAD (Bf-FAD) and is the site of bifurcation[26, 72].

Biochemical and crystallographic investigations have informed our understanding of FBEB[25-27, 41, 42, 73, 74]. The two subunits of ETFs (EtfA and EtfB) fold into three domains, depicted in Figure 1. Domain I encompasses the N-terminal portion of the EtfA subunit, while domain III comprises most of EtfB, which is of a smaller size. The flavin moiety of one of the Bf-FAD is situated in the interface of domains I and III. The tightly associated domains I and III are termed the 'base' domain in what follows. Conversely, domain II, primarily formed by the C-terminal portion of EtfA, associates with the base via a hydrophilic interface and occupies a concave depression in one face of the base. Domain II is referred to as the 'head' domain below, or the 'shuttle' domain, and contains the FAD shared with canonical

ETFs. This FAD mediates electron-transfer to high-E° partner proteins, including butyryl-CoA dehydrogenase (BCD), so it is called the electron transfer FAD (ET-FAD).

Two distinct conformations have been captured in crystal structures of both canonical and Bf ETFs: 'open' and 'closed' [71] (also called B and D conformations[41]). The 'closed' conformation is exemplified by the crystal structure of ETF from *Acidaminococcus fermentans* (*Afe*ETF)[26], whereas the 'open' conformation is observed in the ETF from *Clostridium difficile* (*Cdi*ETF) in association with butyryl-CoA dehydrogenase [27] and the *Acetobacterium woodii* Bf-ETF in complex with caffeoyl-CoA dehydrogenase (*Awo*ETF)[42]. In the 'closed' state, the ET-flavin (within the head) is located 18 Å away from the Bf-flavin and partially occluded from solvent. In contrast, the 'open' conformation results from 80° rotation of the head domain that places the ET-FAD on an outward facing surface accessible to interacting protein partners, and separates the two flavins by 37 Å. Therefore, the two orientations of the head domain cause the ET-flavin to alternate between positions close to the Bf-flavin, *vs.* close to partners. This conformational conversion thus regulates the electron transfer process, enabling ET between the ET-flavin and its partners in the 'open' state, and possibly allowing for electron tunneling between the ET- and Bf-flavins in the 'closed' state. An intermediate state posited from cryo-electron microscopy (Cryo-EM) data of the ETF from *Thermotoga maritima[43]*, places the flavin moieties 25 Å apart, suggesting an intermediate in the motional mechanism. An 'ideal' bifurcating conformation, facilitating electron tunneling between the two flavins, remains to be

observed. However, such a conformation need not be abundant or persistent, due to the rapidity of electron tunneling. Regardless, the different captured conformations of the ETF suggest that large domain scale motion occurs in the course of turnover. It is important to characterize these motions in solution, free of crystal packing constraints or freeze-trapping. Moreover in order to be functional, rotation of the head domain must be coordinated with other elements of catalysis to maximize efficiency of EB[32].

Domain scale motion essential to catalysis has been documented in the respiratory bc1 complex [46, 75] and cytochrome P450 reductase (CPR)[76, 77]. These systems have revealed high resolution details of interactions between the protein domains and cofactors, and shown how protein conformations can modify the distance over which electron transfer occurs between cofactors[78-82]. However a complete picture requires insight from a variety of different methods. Crystallization can enrich certain conformations due to protein-protein interactions), and thus produce a biased picture or fail to capture a functional conformation. It is critical to also obtain information on proteins' structures in solution, under physiologically relevant conditions. Moreover use of X-rays in data collection can produce complications for flavoproteins, due to photoreduction of flavin by X-rays[83-85]. Forster resonance energy transfer (FRET) and other optical methods can provide detailed insights in solution [86] but suffer the same challenges, in addition to complications in interpreting the data, because flavins are fluorophores themselves, as well as quenchers.

Small-Angle Neutron Scattering (SANS) is exceptionally well-suited for probing the intricate internal architecture of protein assemblies, enabling the examination of domain-scale structures, internal reconfigurations and dynamics, in solution. It has been applied in the study of domain movements within CPR, NADPH-dependent sulfite reductases[87, 88], and more complex entities such as the 70s ribosome[89, 90]. In the case of CPR, Freeman et al [91]were able to link domain scale motion to different redox states of the flavins. In subsequent work [47] they also used SANS contrast matching with deuterated protein, to elucidate the solution structure of the electron transfer complex between CPR and its electron transfer partner cytochrome *c*.

Small angle X-ray scattering (SAXS) is also widely used to study protein conformation in solution[48, 92]. Like SAXS, SANS facilitates the acquisition of scattering profiles proteins offering insights into their overall shape. However, SANS distinguishes itself from SAXS through its non-intrusive nature that also does not perturb oxidation states. Neutrons, devoid of electrostatic charge and possessing relatively low kinetic energy, engage weakly with molecules, so there is no radiation damage or reduction of flavins—both concerns in SAXS and cryo-EM. These features render SANS particularly benign, making it suitable for elucidating structural and dynamic properties on distance scales ranging from nanometers to hundreds of nanometers[64]. SANS's utility extends to the analysis of dilute solutions and large complexes beyond the scope of Nuclear Magnetic Resonance (NMR) spectroscopy, perfectly aligning with the requirements for studying ETF systems. It accommodates conditions mimicking enzymatic turnover

without necessitating the use of contrast agents (employed in Electron Microscopy) or crystallization facilitators. Recent developments in the computational modeling of SANS data have opened new avenues for understanding the solution structures of multidomain proteins and complexes, including Bf-ETFs.

In our present investigation we have used *Afe*ETF as a model system to investigate domain scale motions in Bf-ETFs. We have delineated the resting conformation of Bf-ETF, proposing a structural ensemble in solution distinct and more nuanced than what has previously been reported. By contrasting SAXS and SANS data, we have demonstrated the susceptibility of ETF to populate different conformations under the influence of X-rays versus neutrons. Leveraging molecular dynamics-based modeling tools, we describe an equilibrium between extended and compact conformations of ETF, noting a shift in this equilibrium upon reduction of the bound flavins. We adapted the Bilbo-MD utility package traditionally employed for SAXS data modeling [93] and deployed it for analyzing the SANS data. We have also implemented a genetic algorithm (GA) to explore conformational ensembles and identify those that best adhere to the data. These are the first applications of these tools to SANS interpretation, to the best of our knowledge, and our development of these platforms also enabled us to explore the possible advantages of integrating structures obtained by metadynamics . This approach enabled a comprehensive investigation of the conformational landscape that may be explored by *Afe*ETF, via sampling thousands of conformations generated through Bilbo-MD and metadynamics simulations. Finally, we explored possible physiological triggers of conformational change in ETF, testing change in redox

state and partner protein binding. The latter was made possible by selective deuteration and contrast variation techniques that are unique capabilities of SANS. Thus, we have advanced tools available for studying domain-scale movements in proteins by SANS, but we also put for a new understanding of the dynamic rearrangements within ETF in terms of an ensemble of conformations wherein a significant population of extended conformations coexists with the familiar compact forms, and we propose that these extended conformations provide critical paths between the open and closed conformations, thereby enabling their interconversion as required for catalytic competence.

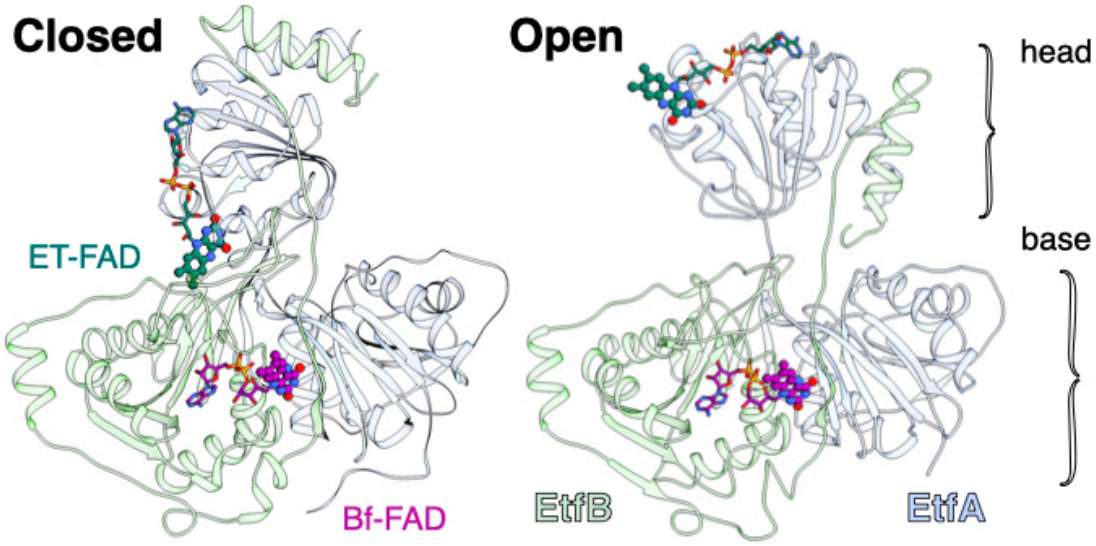

Figure 3:1 Comparison of two conformations captured in crystal structures.
Structures are based on 4KPU and 6FAH. FADs are shown in ball and stick with ET FAD in yellow and BF- FAD in green using the CPK convention for non-C atoms. The flavin headgroups are rendered with thicker bonds.

3.2. Results:

3.2.1 ETF Scattering cannot be explained by the conformations captured in crystal structures.

SANS of the oxidized (OX) state of *Afe*ETF immediately revealed complexity not evident from the crystal structures. Whereas pairwise distances approaching 100 ± 10 Å are evident from the pairwise distance distribution function profile (P(r)), the theoretical SANS calculated based on each of the crystal structures yielded distances no longer than 80 Å 4KPU (Figure 3.2A, methods section 3.4.6). Neither 4KPU (closed conformation) nor 6FAH (open conformation) was able to explain the experimental profile, yielding deviations (χ2 values) of 49 and >38, *vs*. target values of 1-5, even when we modeled the effects of the unstructured tags used to purify the protein. Nor did modelling based on the Cryo-EM structure improve agreement, as the χ2 exceeds 38 as well (7KOE, Table 3.1). The high χ2 values are consistent with the fact that both the open and closed conformations are compact, with radii of gyrations, R_gs, of 24.6 and 24.3 Å respectively (and only 24.0 Å for 7KOE). Thus, the longer scattering distances observed by SANS cannot be explained by any of the solid-state structural models, and additional more extended conformation(s) must be populated in solution.

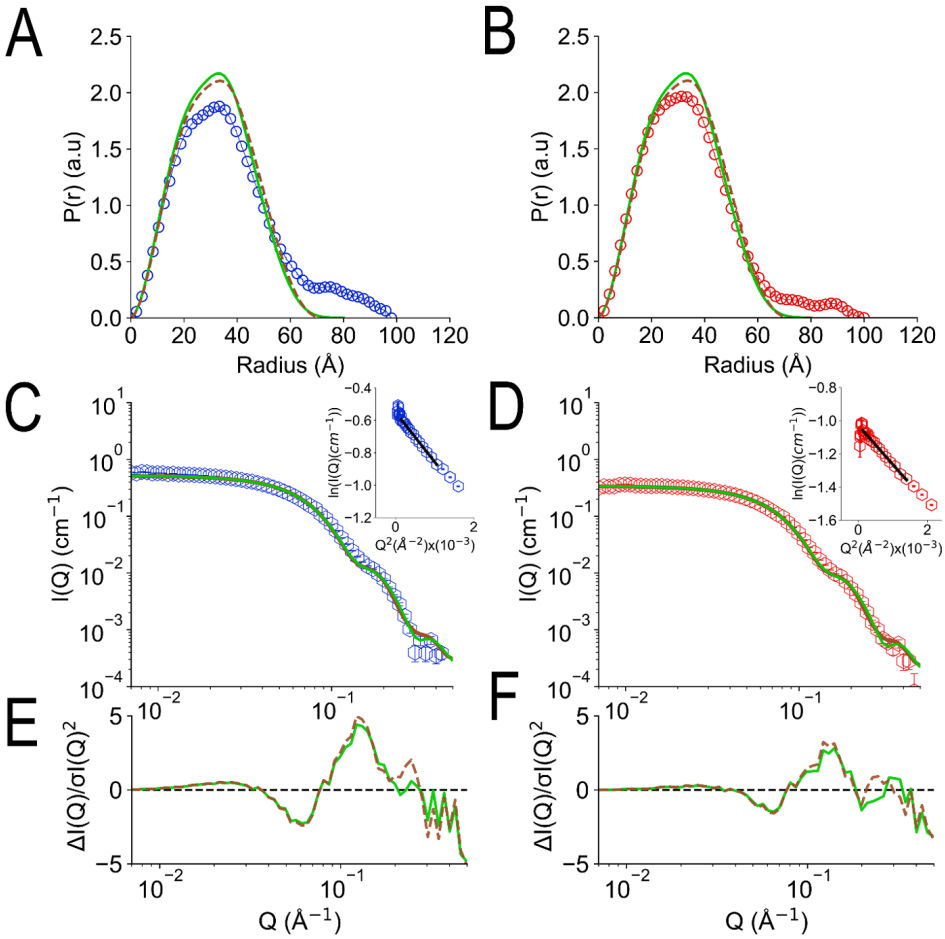

Figure 3:2 SANS demonstrates extended conformations in solution, whose population changes with oxidation state.

Color codes: Blue for oxidized and Red for reduced. Lime green 4KPU and orange 6FAH. Panels A and B: pair distance distribution function ,(P(r) profiles, and panels C and D: SANS profiles, with Error-normalized residuals plotted beneath (E and F). Insets to C and D present Guinier Fits results are compared for the OX state (A and C), and the RED state (B and D). The theoretical SANS and P(r) profiles calculated from AfeETF's crystal structure (4KPU) are overlaid with experimental data and shown as solid blue lines. Error-normalized residuals are shown below the SANS profiles. The inset in panels A and B are Guinier fits for the oxidized and reduced states, respectively. All data analyses were performed with Pepsi-SANS as described in Methods section 4.6

3.2.2 Oxidation state affects the conformation of bifurcating ETF

Both X-ray crystallography and Cryo-EM can result in flavin reduction [83-85] so we tested whether SANS of a reduced sample would better agree with structures from those methods. Figure 3.2 shows that SANS by reduced *Afe*ETF are no more compatible with the predictions of published structures, with $\chi 2$ values of 28.25 vs 4KPU, and larger values vs. the other structures. Regardless of oxidation state, none of the three solid-state structures explain the scattering observed at pairwise distances (P(r)) longer than 70 Å.

Because ETF's function involves alternation between oxidation states, we tested for conformational change coupled to the oxidation states of the flavins. The long (2 hr) data collection times involved in SANS and the relative instability of the fully reduced states require validation that the intended state persists over the duration of data collection. This is demonstrated by Figure 3.3. In particular, attainment and retention of full reduction of both flavins is demonstrated by four-fold diminution of absorbance at 454 nm. We note that even charge transfer features signaling retention of NAD^+ bound to reduced flavin persist at long wavelengths of 650 nm and beyond, documenting sustained anaerobicity of our samples and compatibility of reduced-state samples with SANS measurements.

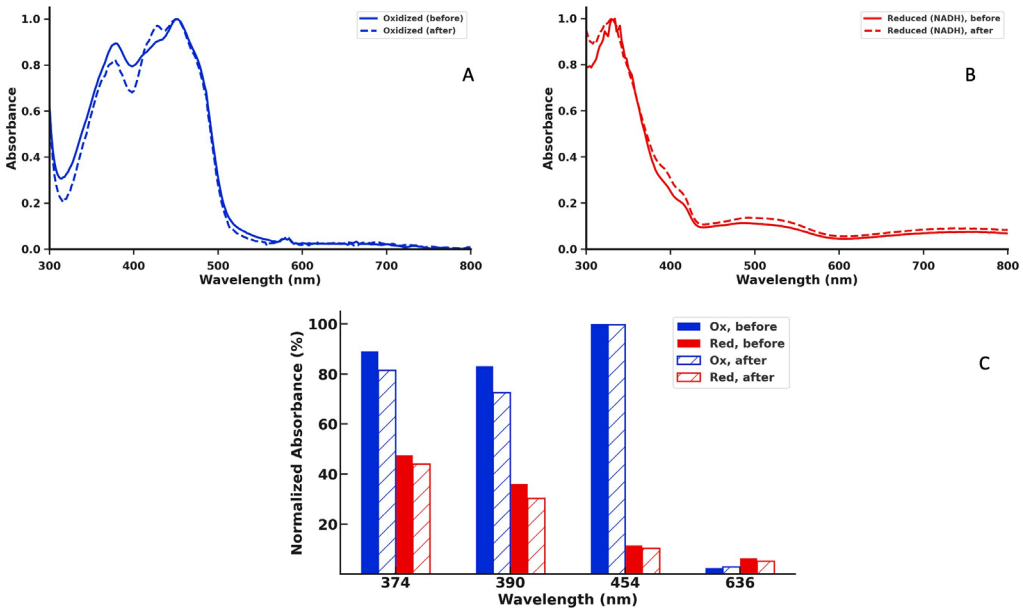

Figure 3:3 Optical spectra documenting oxidation states of samples. samples.
A. oxidized ETF before and after. B. Reduced ETF and after showing sample stayed reduced throughout the experiment. C. Bar graph showing scaled absorbances at 374, 390, 454 and 636 nm to compare spectral features before and after data collection.

Analysis of the SANS profile of RED AfeETF indicates an Rg of 25.8 +/- 0.1 Å, which is approximately 5 Å smaller than that of OX AfeETF, with an Rg of 30.3 ± 0.2 Å (Fig. 3.2). Thus, the local effects of adding a hydride (H-) to each flavin are coupled to long-range structural consequences. Because these are coupled to an element of catalytic turnover, the diminution in long-range distances may be related to events in turnover.

Table 3.1: Figures of Merit and Fitting Parameters Emerging from Analysis of SANS and SAXS, and Comparison with Theoretical Predictions

	ETF-Oxidized	**ETF-Reduced (NADH)**	**ETF -Reduced (Dithionite)**	**ETF Oxidized**
SAS technique	**SANS**	**SANS**	**SANS**	**SEC-SAXS**

employed

Guinier analysis				
I (0) (cm^{-1})	0.577±0.002	0.354±0.001	0.445±0.001	0.573±0.003
R_g (Å)	30.3 ± 0.2	25.8 ±0.1	25.9± 0.2	30.7 ± 0.3
QxR_g range	0.39 - 0.98	0.44 -1.36	0.18 - 1.28	0.21 - 1.29
***P*(r) analysis**				
I (0) (cm^{-1})	0.570±0.002	0.358±0.001	0.448±0.002	0.573±0.003
R_g (Å)	29.9±0.2	26.9±0.2	26.8±0.3	31.4±0.4
D_{max} (Å)	100 ± 10	100 ± 10	100 ± 10	120 ± 10
Q range ($Å^{-1}$)	0.01-0.42	0.01-0.49	0.01-0.39	0.01-0.51
***PDB fits*[a]**				
χ^2 (4KPU)	86.75	22.84	23.57	7.72
χ^2 (6FAH)	96.69	29.91	22.34	5.24
χ^2 (7KOE)	138.2	37.86	38.69	9.64

[a] Theoretical scattering was computed for each structure after modelling the full sequence of *Afe*ETF onto the structure in question. N- and C-terminal unstructured amino acids and tags were added and then predicted SANS profiles were generated using PEPSI SANS.

3.2.3 Sampling the conformational space of bifurcating ETFs using MD simulations

Although the SANS of RED *Afe*ETF documents smaller contributions from extended conformations, it is still not explainable by any of the known structures exemplified by 4KPU, 6FAH or 7KOE, alone or in combination. Therefore, we employed molecular dynamics (MD) to explore energetically accessible

conformations of *Afe*ETF and ensembles thereof to identify solutions compatible with the data. We performed two types of MD simulations: 1) We modeled rigid body motions of connected domains using Bilbo-MD's CHARMM simulation engine[94, 95], and 2) we employed metadynamic*s* algorithms to permit accelerated comprehensive exploration of atomistic dynamics. Both approaches employed a full-length model of *Afe*ETF, starting from the crystal structure 4KPU with unstructured residues including purification tags appended.

Bilbo-MD simulations explore a broad spectrum of conformations as displayed in the distance between the two FADs (R_FAD) plotted vs extension of the ETF R_g (Figure 3.4). Obtained conformations yield average $\chi 2$ values vs. the SANS data of 20.5 and 30.1 for the oxidized and reduced states. Figure 4 color-codes the quality of fit to SANS data, with $\chi 2$ values from 57 to 107 in blue to light blue for fits to OX-state data (panels A and B), and $\chi 2$ values from 20 to 50 in red to light orange for fits to RED-state data (panels C and D). Thus, intense colours represent conformations that better fit the data. The Bilbo conformers with R_g values similar to those of the crystal structures (4KPU and 4L2I), had $\chi 2$ values from 57 to 107 vs. OX data (blue to light blue, Fig. 3.4A) and 20 to 50 vs. RED data (red to light orange, Fig. 3.4C), respectively. Thus, the RED state is better described by conformations resembling the crystal structures. Conversely, the conformer with the minimum $\chi 2$ vs the OX state data (46.23) has an R_g of 29.9 Å and an R_{FAD} distance of 49.4 Å. The former is 5 Å larger than that of any solid-state structure and the latter is longer than yet observed in an ETF. In addition, despite the fact that solved structures fail to explain larger pairwise distances, individual

conformers that are more extended also fail, as conformers with R_g values greater than 29.9 Å (OX) and 26 Å (RED)display elevated $\chi 2$ values that increase with R_g, for both oxidation states. Thus, simply increasing the R_g of a single model did not yield an improved fit.

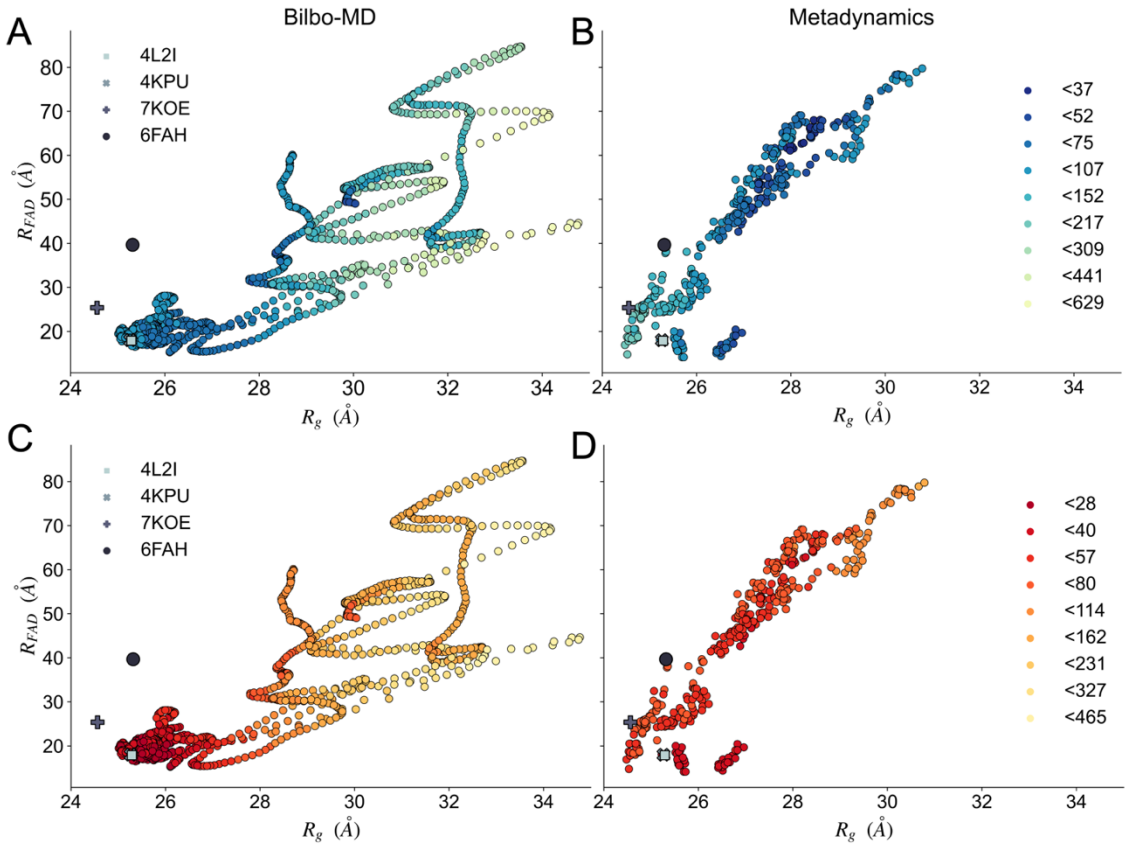

Figure 3:4 Correlation of conformational properties from the Bilbo-MD and Metadynamics ensembles to fits to scattering data.
The plots show the distance between C7M atoms of the two FADs (R_{FAD}) and the radius of gyration (R_g). The columns represent the simulation types (A,C) are Bilbo-MD, while (B,D) are metadynamics. The rows are the fits to oxidized data(A,B) and reduced data(C,D). The color keys for the fits to the oxidized(Blue to Green to Yellow) and reduced(Red to Orange to Yellow) data are the binned $\chi 2$ values with each color as the upper limit of the bin in order of low to high $\chi 2$. For reference, the R_{FAD} and R_g of the crystal and cryo-EM structures (4L2I, 4KPU, 7KOE and 6FAH) are displayed as well.

The conformers from metadynamics (Figure 3.4B,D) cover similar spans of R_{FAD} and R_g to the Bilbo-MD ensembles, but do not sample the most extended conformations. Metadynamics appears to enable models more in line with the crystal and cryo-EM structures than the Bilbo-MD conformers. The Metadynamics conformers share the same trend as the Bilbo-MD in terms of $\chi 2$, with smaller $\chi 2$ towards lower R_g and R_FAD for the RED state and intermediate R_g and R_FAD for the OX state. Models obtained via metadynamics attain lower $\chi 2$ values with minimum values of 26.64 and 23.25 for the OX and RED states, and Metadynamics yields individual conformers with better $\chi 2$ values to OX data at intermediate R_g values. Nevertheless, although both MD approaches improving the fits slightly to a single conformer, neither produced a conformation with a satisfactory $\chi 2$ to the SANS data, of either oxidation state.

We also quantified the ability of an ensemble of all conformers to account for the data. For the OX state, the average scattering profile from the entire ensemble of Bilbo-MD conformers (unweighted) fit the SANS data better ($\chi 2 = 20.5$) than did the prediction of any of the solved structures, or the average of the profiles of the ensemble of *Metadynamics* conformers ($\chi 2 = 42.97$). This was attributable to inclusion of conformers with higher R_g in the Bilbo-MD ensemble.

For the reduced state, the average of the conformational ensembles from Bilbo-MD and *Metadynamics* provide similar fits, with $\chi 2$ values of 30.15 and 33.76, respectively. However, the conformers close to the crystal structure (4KPU) gave lower $\chi 2$ values than the average scattering profile. These results suggest that the reduced state is more compact, like the crystal structures, whereas the oxidized state

of ETF is more extended and/or populates more extended conformations. Nevertheless a simple average of all conformers does not predict the experimental results. Thus, not all conformers are equally likely and/or conformations not sampled are required for a full description.

Since individual extended conformers had elevated $\chi 2$ values (but see intermediate R_g conformers in the OX state), we next addressed the possibility of explaining data assuming simultaneous population of more than one conformer, to form a conformational ensemble.

3.2.4 Identifying best-matching ensembles of conformers using a Genetic Algorithm

To determine the minimum number of conformations required to describe the SANS of *Afe*ETF, we employed a genetic algorithm (GA) ensemble optimization search (GA-SANS). We applied it to Bilbo-MD and *Metadynamics* ensembles separately, and also together. Our methodology was inspired by the Bilbo-MD framework, in which we have replaced the Minimal Ensemble Search (MES) with the GA. To the best of our knowledge, the current study represents the first application of a GA for refining protein ensembles against neutron scattering data. Ensemble optimization greatly improved the fit to data for both oxidation states. In the OX state, the best ensemble based on Bilbo-MD-derived conformers includes two conformations and fit the data with a $\chi 2$ value of 2.74 (Table 2). Ensembles with three and four conformations generally yielded larger $\chi 2$ values and were therefore not pursued . The optimal 2-conformer model includes a compact and an extended conformation with $R_g s$ of 25 Å and 33 Å, accounting for 48% and 52% of the population, respectively (Table 3.2).

Table 3.2: Best-Fitting Ensembles Identified by a Genetic Algorithm Approach, Relations to Crystal Structure of Participating Conformations and Ensemble's Quality of Fit to Data.

Sample	SAS technique	Source(s) of Conformers	Extended conformation[a]			Compact conformation[a]			$\chi 2$
			Population	R_g (Å)	RMSD[b] (Å)	Population	R_g (Å)	RMSD[b] (Å)	
Oxidized ETF	SANS	Bilbo-MD	0.52	32.2	18.31	0.48	25.1	4.96	2.74
		Metadynamics	0.61	30.8	15.89	0.39	24.5	7.54	8.27
		Combined	0.56	32.6	18.19	0.44	24.9	8.19	1.30
Reduced ETF (NADH)	SANS	Bilbo-MD	0.19	34.8	14.26	0.81	25.5	0.92	5.93
		Metadynamics	0.28	30.5	15.62	0.72	25.2	1.35	13.45
		Combined (One extended Two compacts)	0.20	34.8	14.26	Fully compact: 0.06	24.5	7.23	5.96
						Intermediate 0.74	25.5	0.93	
Reduced ETF (Dithionite)	SANS	Bilbo-MD	0.20	34.6	14.27	0.80	25.5	5.84	7.55
		Metadynamics	0.29	30.5	15.62	0.71	25.2	1.35	19.49
		Combined	0.20	34.6	14.27	0.80	25.5	5.84	7.55
Oxidized ETF	SEC-SAXS	Bilbo-MD	0.37	32.7	18.26	0.63	26.1	8.33	1.26
		Metadynamics	0.28	29.5	12.45	0.72	26.5	9.36	15.0
		Combined	0.38	32.6	18.20	0.62	26.1	8.33	1.26
ETF-dBCD complex									
ETF-dBCD complex	SANS	Bilbo-MD	0.65	65.8	13.99	0.20	56.3	4.21	0.80
Free ETF in ETF-dBCD complex			0	NA	NA	0.15	24.4	7.2	

[a] The extended and compact conformations obtained from each modeling method are listed.

[b] The RMSD is reported in Å with respect to 4KPU for *Afe*ETF and 5OL2 for the complex with BCD.

For the RED state, two conformations also sufficed to model the data, yielding a χ^2 value of 5.93 for an ensemble in which a compact conformation (R_g = 25.8 Å) is populated to 81% and an extended conformer (R_g = 34.3 Å) describes 19% of the population. The extended conformer identified for the OX state is different from than obtained for the RED state . However, the compact conformers identified for the OX and RED state are very similar (RMSD of 0.92 Å).

The GA also favored two-conformer models of conformers from Metadynamics, obtaining χ^2 values of 8.27 and 13.45 for the OX and RED states, respectively. Again, each ensemble comprised an extended and a compact conformation. The extended conformer accounted for 61 % of the population for the OX state and 28 % of the population for the reduced state. The larger population of an extended conformation from Metadynamics vs. Bilbo-MD could be partially due to these conformations being less extended (smaller R_gs of 30 and 28 Å) that the extended conformations identified from the Bilbo MD ensemble. Being less extended, larger amounts of them were required to achieve comparable fit.

When our GA protocol employed an ensemble combining conformations derived from both MD methods, the results for the OX state again supported a two-conformers model. Its significantly improved χ^2 value of 1.30 is a 2.0- and 6.4-fold improvement in χ^2 compared with models obtained from Bilbo-MD or

Metadynamics ensembles separately. The resulting model preserved the extended conformation obtained from Bilbo-MD alone and increased its weight by 4%. However, the compact conformer was selected from the cluster c7 of the *Metadynamics* simulations, from the same region of the $\chi 2$ landscape as the crystal structure (4KPU, RMSD of 8.19 Å).

The fit to the RED state data was not significantly improved by application of the GA algorithm to a combined ensemble of Bilbo-MD and *Metadynamics* conformations, as indicated by the obtained $\chi 2$ value of 5.89. This is in part because each of the individual ensembles supported better fits to RED than to OX ($\chi 2$ values of 5.93 and 13.45 for Bilbo-MD and *Metadynamics* -derived conformations). As in the other attempts, a three-conformer model yielded no better fit to RED data, with a $\chi 2$ value of 5.96. In the three-conformer model, the extended conformation remained a minor but essential contribution to the population (20%), while the compact conformer was divided into a previously identified form and a new one, with occupation of 74% and 6%, respectively. Although minor in terms of its population, this new compact conformer is interesting in that it stems from a different cluster (c2) in the *Metadynamics* simulation. It is more compact than the major conformer, and is structurally distinct from all other conformers in the RED ensemble and the combined OX ensemble.(RMSD 7.23). Thus, although the easier-to-fit RED state does not benefit statistically from either the merged ensemble or a three-conformer model, that larger simulation revealed additional structural variation of possible mechanistic significance (see below, Discussion).

The $\chi 2$ values relied upon by the GA to optimize models are based on agreement between theoretical and observed scattering profiles. However a more nuanced appreciation of the distinctions between the models is obtained by comparing theoretical predictions with different presentations of the data. Figure 3.5 shows the models' agreement with data, for the OX state and the RED state formed by reaction with NADH. Although the GA algorithm achieved slightly better fits to data when allowed to employ conformations from both Bilbo-MD and Metadynamics, all the models employing two conformations greatly outperformed the crystal structures in their agreement with the data. Thus we conclude that *Afe*ETF in solution is best described as an ensemble of at least two conformations of which one is comparably compact to the crystal structures or Cryo-EM model, but the other is considerably more extended. Although prior work has equated an extended conformation to the open conformation, we re-iterate that the structures of the open conformation have small R_gs comparable to those of the closed conformation. Based on their theoretical scattering profiles the open and closed conformations documented in crystal structures are not distinguishable by our SANS data or analogous SAXS. The extended conformation, shown by our GA to be essential to a complete description of *Afe*ETF in solution, is an additional conformation.

Our two MD protocols provided possible exemplars of what the extended conformation might look like. Although those arising from metadynamics appeal more to our sense of what a protein 'ought' to look like, the data argue that those produced by Bilbo-MD better describe the behaviour of *Afe*ETF. Specifically, predicted Kratky plots and P(r) profiles calculated from models employing Bilbo-

MD conformations (magenta**) agreed better with the data than did predictions from models based on metadynamics-derived conformations (**lime green). In the Bilbo-MD extended conformations, the head domain is essentially detached from the base, only tethered by linker polypeptides, whereas metadynamics conformations retain some interactions between the two domains. The better agreement provided by Bilbo-MD models indicates that the conformational ensemble of *Afe*ETF, especially when OX, includes conformations in which the head domain enjoys considerable orientational freedom from the bases, although its translation remains constrained by covalent attachment. Figure 3.6 shows our final models for the SANS data.

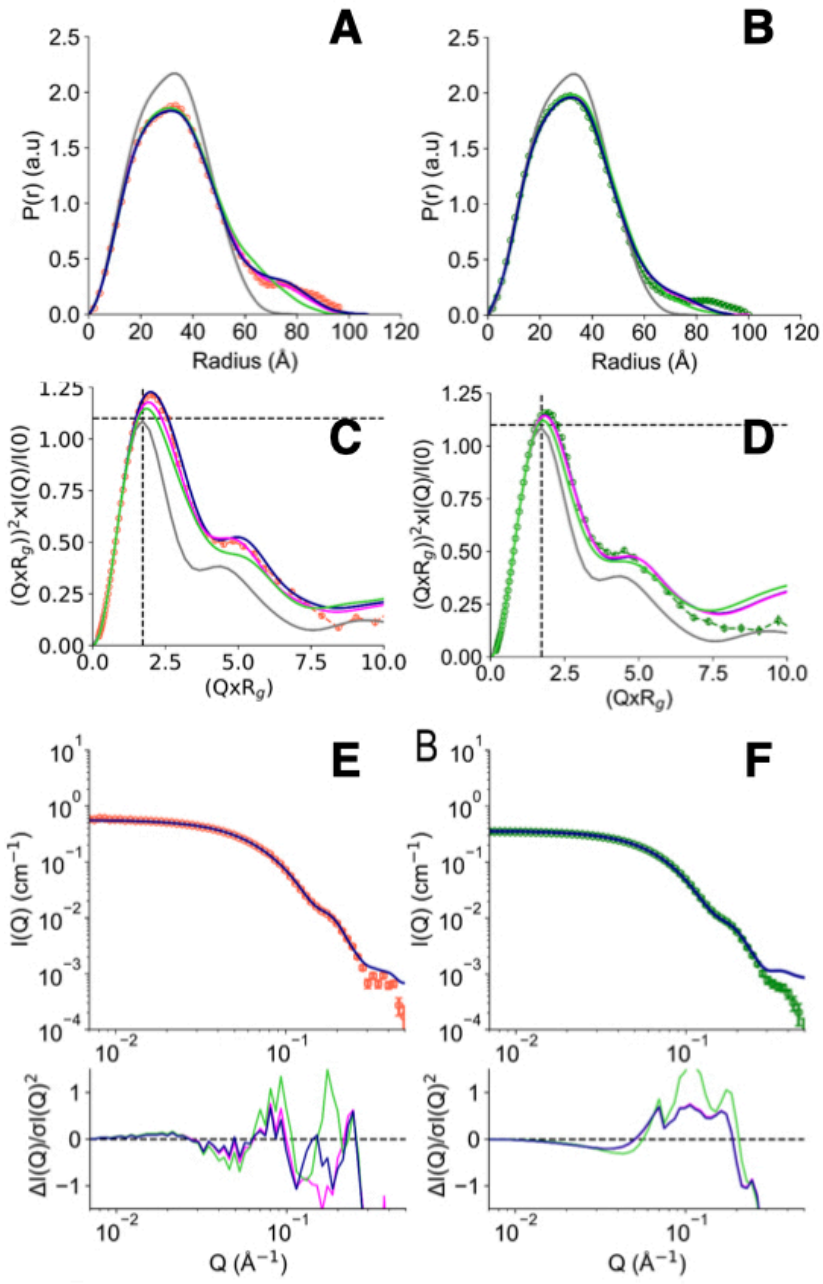
A
B
C
D
E
F
P(r) (a.u)
Radius (Å)
(QxR$_g$))^{2}xI(Q)/I(0)
(QxR$_g$)
I(Q) (cm^{-1})
ΔI(Q)/σI(Q)^2
Q (Å^{-1})

Figure 3:5 Agreement achieved with SANS data using a Genetic Algorithm and conformers from Bilbo-MD and Metadynamics.

Panels A and B compare the agreement with experimental P(r) profiles obtained by various fitting strategies. Fits to OX P(r) are in panel A and those to RED (NADH) ETF's P(r) are in Panel B, with the P(r) data shown as open circles in orange for OX and green for RED. Predictions of the best GA model for each state are the blue solid lines, and the models are those provided in Table 2 with $\chi 2$ =1.30 (OX) and 5.93 (RED). For comparison, the predictions obtained when the GA drew only on subsets of the conformations are also shown: solid magenta lines depict the optimized ensembles employing two Bilbo-MD-derived conformations ($\chi 2$ = 2.74 for OX, 5.93 for RED) and solid lime-green lines depict those based on two Metadynamics-derived conformations ($\chi 2$ = 8.27 for OX, and 13.45 for RED). The theoretical P(r) from the *Afe*ETF crystal structure 4KPU is also shown as a solid gray line. Panels C and D are the corrresponding normalized Kratky Plots and panels E and F are the corresponding SANS profiles, with their normalized residuals beneath them.

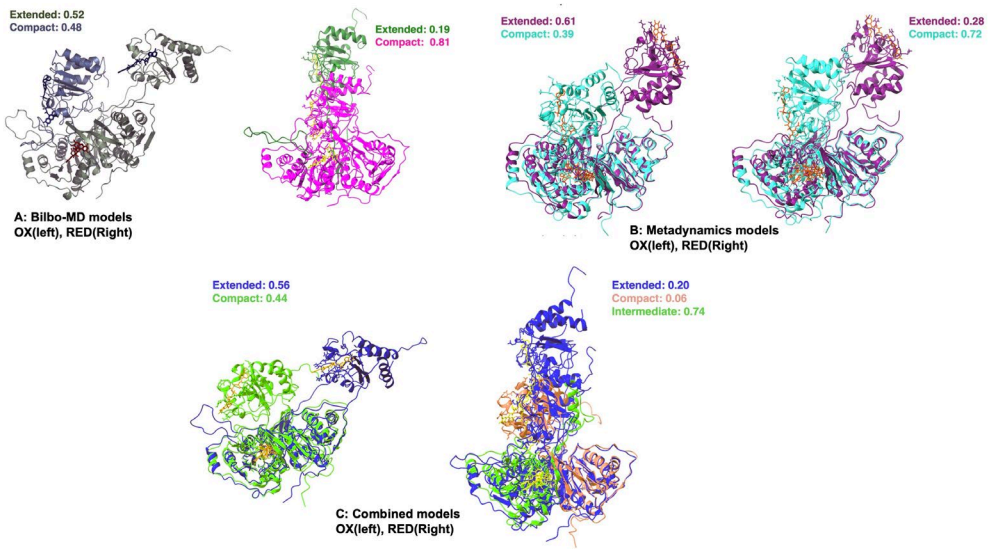

Figure 3:6 SANS data models.**A:** Bilbo MD models**. B:** Metadynamics models **C:** Models generated by the combined approach. In all the three approaches compact conformation dominates after the protein is reduced.

3.2.5 Interaction of bifurcating ETFs with binding partner BCD

The open conformation of ETF has only been observed when ETF is complexed with a partner, suggesting that partner binding alters the relative stabilities of accessible conformers. To investigate the effect of partner binding on the conformation of ETF in solution, we performed SANS experiments using the ETF in complex with the binding partner BCD. To differentiate scattering from ETF from scattering for BCD, we used partially deuterated BCD (dBCD) in combination with protiated ETF at a (1:1) molar ratio and concentrations of 26 μM. We performed SANS measurements in a 75% D_2O buffer to match the scattering from dBCD (minimal contrast between them). Under these conditions, scattering from protiated (^1H) ETF dominates the data after subtraction of buffer contributions (including dBCD). Thus, differences between SANS of ETF vs. SANS from ETF•dBCD are dominated by effects due to conformational changes in ETF alone. In the crystal structure of the ETF•BCD complex (5LOE.pdb), the BCD is a dimer. Each BCD binds an ETF and the two ETFs are spatially separated on opposite sides of the BCD dimer. The P(r) profile of ETF in the complex indicates an asymmetrical bimodal distribution of molecular densities with peak positions at 28 Å and 125 Å, and higher molecular densities for the first peak than for the second one (Figure 3.7). The first peak is consistent with the numerous shorter pairwise distances within the internal structure of ETF whereas the second peak demonstrates presence of longer correlation distances within the complex, best explained as distances between atoms in different ETFs. Note that the P(r) profile has a minimum at 75 Å, corresponding to half the center-to-center distance between

two separated ETFs. In addition, a third, smaller peak at ~ 200 Å, is likely indicative of intrinsic mobilities of the head domains of ETF, allowing the heads to sample larger head-to-head distances within the complex. Indeed, the SANS data indicate that ETF in the complex adopts an extended conformation, again as part of a multiconformational ensemble, as in the OX state (Figure 3.2A). However, such a conformation, or multiple angles between the head domain and the base, are not seen in the two ETFs present in the crystal structure of BCD_2•ETF_2 (5LOE.pdb[27]), and SANS predicted based on this structure is a poor fit to the SANS data of ETF in the complex.

A single conformation of ETF does not provide good agreement with the SANS data regardless of whether a compact or extended conformation is used. To find a minimal ensemble of conformations able to describe ETF complexed to BCD, we used our GA approach using Bilbo-MD -derived conformations to fit SANS. Our search considered an ensemble of conformations of physically separated ETF dimers. In addition, due to higher probability densities within the first peak in the P(r) profile we also included free ETFs in our ensemble. The best model fit the data as a mixture of 83.3% ± 0.9% ETFs bound on opposite sides of BCD (figure 3.8) in both an extended and a compact conformation, and 16.7% ± 1.2% of ETF free in a compact conformation. The best fit to the SANS is obtained when ETF that is bound in the complex populates an extended conformation to approximately 65% ETF whereas some 20 % of the population is compact. This modeling of the SANS complex data yielded a X^2 value of 0.80. The free ETF that was observed in the compact conformation was modeled with the same compact conformation as was

selected by GA in the OX state. For the ETF in complex with BCD, our data and analysis indicate an R_g of 57 ± 5 Å. This R_g *is 14* Å larger than the R_g calculated from the 5LOE starting conformation (R_g ^{GNOM} *of* 43 Å). Thus, complexation of ETF and BCD does not appear to trap better-defined conformations of each. This complex appears no better defined than its free participant ETF.

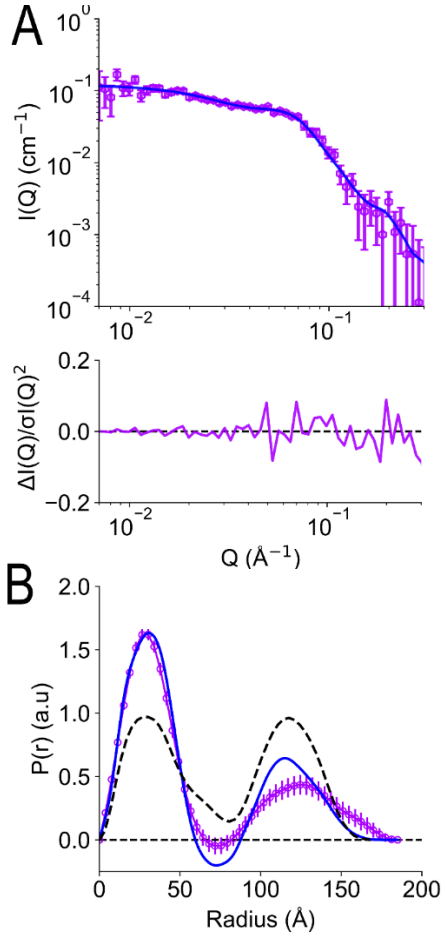

Figure 3:7 SANS in presence of partner protein BCD.
A. Scattering and residual profile of the ETF-dBCD complex. Blue is the theoretical scattering from the 5LOE PDB. Purple is the scattering from the ETF at BCD match point. B. P(r) profiles. The two peaks indicate physically separated units in space.

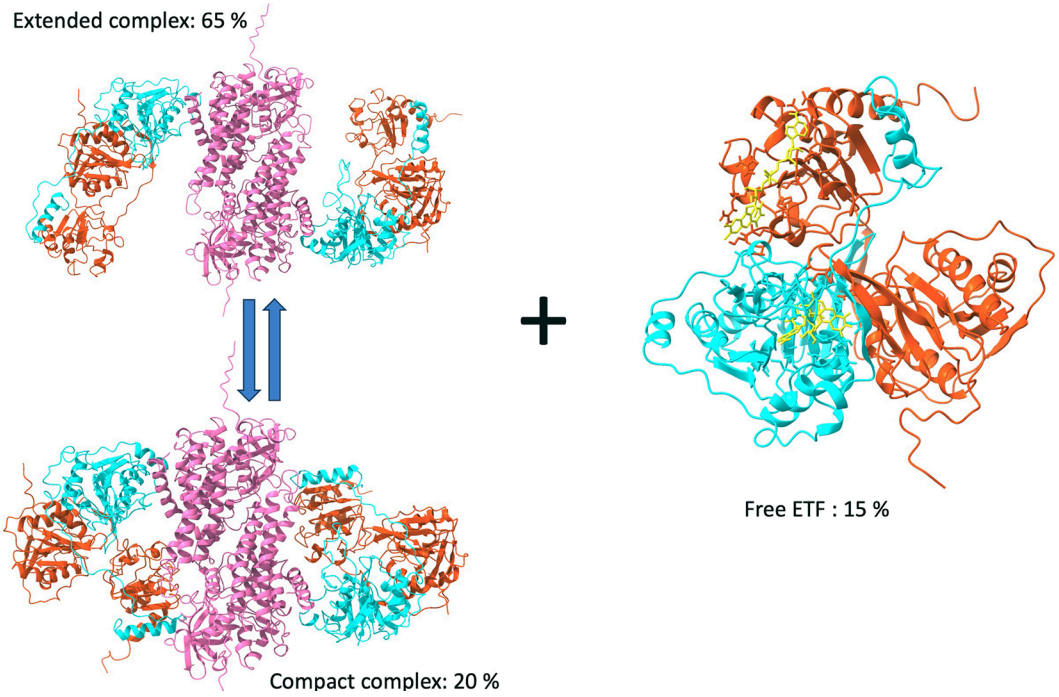

Figure 3:8 Modeling of SANS data on the ETF-dBCD complex
Within the complex ETF has both extended and compact conformations. The free ETF was in the compact conformation.

3.2.6 Comparison of SAXS and SANS and possible impact of SAXS on redox state of ETF

Because irradiation with X-rays can result in flavin reduction , and might therefore produce some RED ETF during SAXS data collection, we compared the scattering obtained from ETF using neutrons and X-rays as the irritation source. This was done using in-line size exclusion chromatography SAXS (SEC-SAXS). For these experiments all buffers were equilibrated with air so ETF began the experiments in the OX state and would be able to undergo reoxidation by dissolved O_2 in the event of photoreduction. Analysis of the SEC-SAXS profile of ETF indicates R_g^{SAXS}

=30.7 ± 0.3 Å and D_{max}^{SAXS} = 120 ± 10 Å. These are within error of the values obtained by SANS (R_g = 30.3 ± 0.2 Å, D_{max} = 110 ± 10 Å).

However, SAXS and SANS profiles of ETF display large differences in the high-Q region (Figure 3.7C). Specifically, the SANS profile displays a shoulder around 0.18 $Å^{-1}$, which is not present in the SAXS profile. In addition, the P(r) profile from the SANS data indicates a shoulder at 20 Å in addition to the main peak at 40 Å, whereas the P(r) profile from SAXS indicates a single peak at 35 Å, and the two differ significantly regarding population of long distances (and therefore the most extended conformations). Application of the GA to the SAXS data using conformation from the combined Bilbo-MD/*Metadynamics* ensembles indicated the presence of 62 % compact and 38% extended conformers. The compact conformation is the same one observed in the oxidized state from GA-SANS data. However, the extended conformer was distinct.

We do not claim that the best fitting ensembles obtained are unique solutions, and we expect that the conformations that suffice to model the data are indeed representing large populations that fluctuate about the average dimensions of the conformations chosen by the GA as best fitting. However data themselves indicate distinct conformational ensembles, so we are not surprised that the extended conformers identified as best-fitting deviate from one another, although additional work will be needed to learn about differences between the proteins in the samples. Despite being studied under the same oxidizing environment, samples exposed to X-ray rays have a different conformational ensembles from those studied using neutrons. The redox and photo-active flavins of ETF could be adopting different

oxidation states that are not produced when a neutron beam is used. Alternately, photoreduced flavins could be reacting with dissolved O_2 to release reactive oxygen species able to react with the protein.

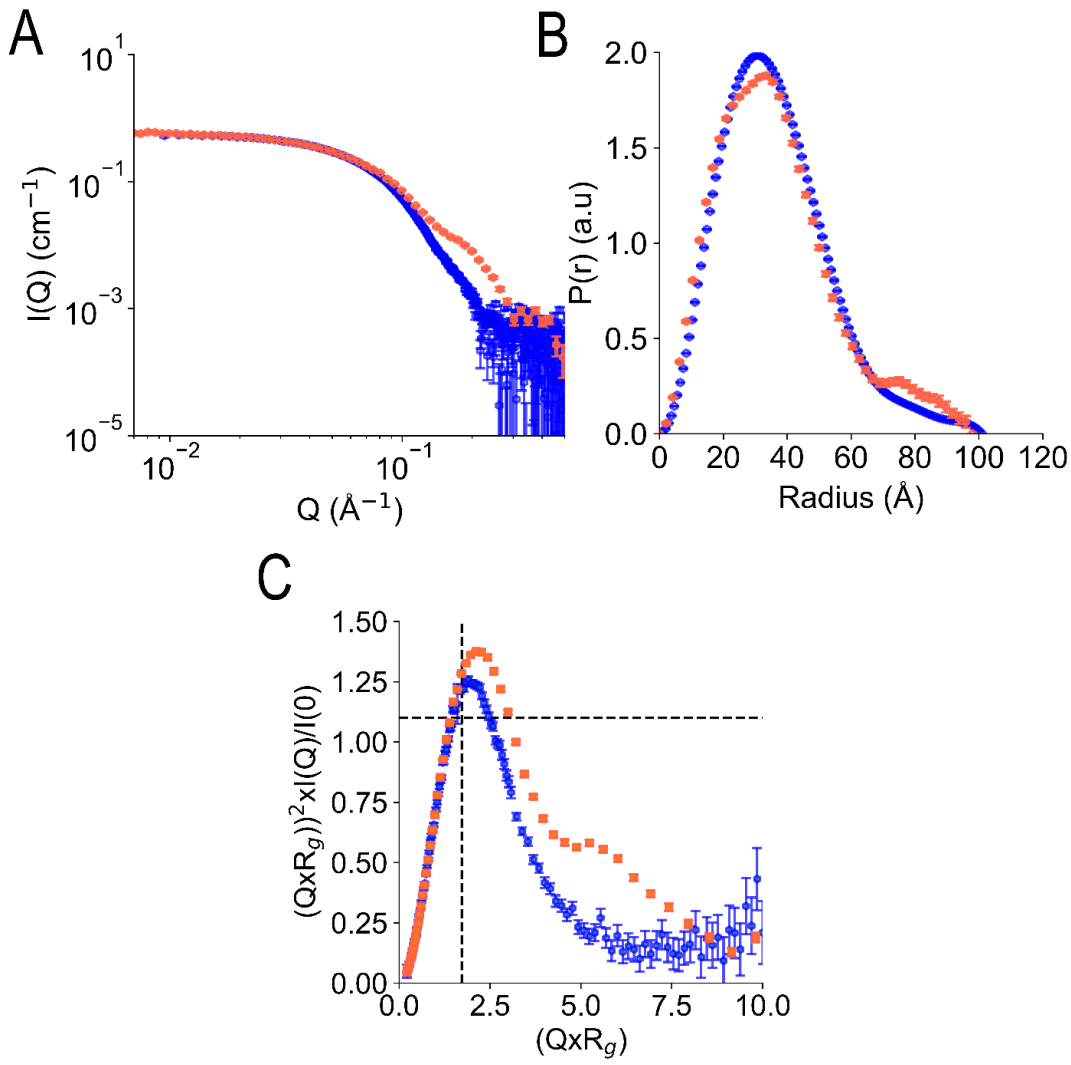

Figure 3:9 Comparison of SAXS and SANS obtained for OX ETF.
SAXS (blue) and SANS (orange) are compared via their scattering profiles (A); P(r) profiles (B); and normalized Kratky plots (C).

3.3. Discussion:

3.3.1 Methodological Considerations and Value of a GA to optimize an ensemble description.

To learn whether conformational changes might be triggered by changes in flavin oxidation state, it was important to have a measurement technique that would not itself produce flavin reduction. This rule out methods that use electromagnetic radiation with wavelengths shorter than 800 nm, and therefore excludes fluorescence, FRET, X-ray-crystallography and SAXS. Cryo-EM is also excluded by the fact that an intense beam of electrons (reductants by definition) is employed. NMR remains possible but is insensitive, difficult to apply to proteins larger than 40 kDa and requiring of concentrations of protein that are prone to aggregation. SANS is another nucleus-based method that does not perturb electrons. However scattering *gains* effectiveness in high molecular-weight complexes, making relatively dilute solutions sufficient for data acquisition. Crucially, SANS is compatible with physiological conditions and applies to proteins in solution where their propensity for conformational change is not constrained by the experiment.

Scattering data yield low detail, in that a three-dimensional conformation change is detected only as a one-dimensional profile of scattering intensity vs. scattering vector. Thus, approaches used to interpret the data are a very important aspect of answering research questions. Our study was critically enabled by the prior publication of crystal structures for the conformations believed to alternate in

turnover and mediate different steps in electron transfer: the closed conformation in which an electron is believed to pass from the Bf-flavin to the ET-flavin, and the open conformation in which further electron transfer from the ET flavin to a recipient flavin of the partner BCD in the case of *Afe*ETF[26, 41, 71, 96]. These and a recent Cryo-EM structure [43] served as invaluable reference points for interpreting the SANS. Indeed, it was immediately obvious that *Afe*ETF in solution either adopts a novel conformation or exists in a conformational ensemble including at least one new conformation not known from prior structural studies. Crucially, all structures solved to-date of ETFs are compact, with R_gs smaller than 26 Å and D_{max} below 80 Å [26, 41-43]. However our SANS data require that a substantial fraction of the protein in solution have D_{max} approaching 100 Å; they are not compatible with existing structures even when unstructured termini or tags are accounted for.

To obtain plausible additional conformations with the goal of accounting for our long scattering distances, we employed two contrasting modeling approaches. Thus we sought to mitigate effects of bias or blind spots associated with one or other method. We used Bilbo-MD[93] which maximizes ability to explore large-scale conformational variation by focussing on rigid body movement of domains relative to one-another rather than smaller scale movements within individual domains. We complemented this strategy with an agnostic MD approach that nevertheless manages to sample widely different conformations by employing the Gaussian accelerated MD protocol ('metadynamics' [97]**REF).

Although Bilbo MD was designed for modeling SAXS data in flexible protein systems, we applied it to interpret neutron scattering data. Moreover once it became evident that our data could not be explained by an individual conformation, even an extended one, we integrated Bilbo-MD-generated conformations and prediction of their scattering with a GA (**REF) to optimize conformational ensembles on the basis of their ability to match SANS data.

The GA succeeded in identifying the simplest ensembles able to produce high quality agreement with data, bringing χ^2 down from 20 to 2, despite the subtlety of this system. The best fitting ensemble combined conformations generated from both Bilbo-MD and metadynamics that individually exhibited relatively high χ^2 values for both oxidized and reduced data. This suggests that a single conformation cannot fully account for the scattering results, and that divergent conformations much be present in the ensemble to optimize its fit to data. Folded proteins' conformations explore an energy landscape that produces an equilibrium population of functional conformations in exchange with other conformations that may or not also be functional[91, 98-100]. Indeed, not all the conformations invoked to explain the SANS data need be catalytically active, so long as the ensemble as a whole harbours at least one conformation that is. To our knowledge. the current study is the first use of this approach. However, we are confident that it will prove valuable for other challenging cases as well.

3.3.2 *Afe*ETF in solution significantly populates an extended conformation not represented by any existing solid-state model:

Use of GA led to a significant improvement in the χ^2 values *vs.* SANS data from both OX and RED samples, achieving χ^2 values close to 1. We used the GA

approach to test conformational ensembles of varying complexities, including 2, 3, and 4-conformer models. However we obtained the best fits with 2 conformations, which we termed 'extended' and 'compact' because they have very different R_gs (as well as D_{max}'). Although our compact conformation is similarly compact to crystallographic structures including both open and closed conformations, our extended conformation is very different. We do not claim that the conformations chosen by the GA to model different experimental results are unique solutions. Instead, we emphasize that all our experiments require (1) a description in terms of a conformational ensemble to explain the solution structure of *Afe*ETF and (2) substantial population of a conformation that is more extended than any previously solved structure. With R_g >30 Å, our extended conformations cannot be equated with the open or closed conformation. Indeed, the virtually identical sizes and shapes (R_gs and D_{max}') of the open and closed conformations make them very difficult to distinguish by SANS (Figure 2). Although neither of these was chosen by the GA as a representative conformation in any of the ensembles, conformations that are very similar were chosen, so our data are consistent with their population.

3.3.3 Observation of a reduction-triggered conformation change.

In the OX state, the optimized two-conformer ensemble had almost equal populations of the extended and compact conformers (Table 2). However, upon reduction of the protein with NADH or dithionite, the population distribution shifted to more compact conformations (Table 2). This documents a shift of the conformational equilibrium in response to the redox states of the ETF's flavins. The compact conformations obtained from the RED ETF data more closely approximate

the crystallographic conformations of ETF than those obtained for OX ETF. However the extended conformations obtained from analyses of OX vs. RED states are different, and this is interesting too. Murray *et al*. used SAXS to characterize redox-driven conformational changes in *T. maritima* ETF, and also employed Bilbo-MD derived conformations to interpret the results. Although their model search was not as extensive as that described here, the fact that they also arrived at ensembles including a compact and an extended conformation confirm that this result is robust. By extending the palette of modeling tools used, we provide confidence that in this case the larger computational investment does not lead to additional fundamental insight, as we agree with the Murray findings. We also reproduce the Murray finding that ETF is more compact overall upon reduction.

Our interpretations differ though. First, we find that both the open and closed conformations are compact, with respect to not only their physical dimensions, but also their abilities to describe scattering data. We emphasize that the extended conformations we find essential to describe ETF are novel conformations not corresponding to previously solved structures. Second, we find that the more compact behaviour of RED states stems mainly from diminished population of the extended conformation. Although the extended conformations selected varied more depending on methodology and means used to produce the RED state, the compact conformers selected by GA were remarkably uniform and tended to resemble crystal structures. Nevertheless, the extended conformation selected always had a R_g at least 4.5 Å larger than that of the compact conformation that co-characterizes the ensemble.

Although reduction of bound flavins produces relatively small immediate consequences, extensive work on CPR documents that this can suffice [91] (The flavin N5 becomes a hydrogen bond donor rather than an acceptor, and the flavin acquires a negative charge). Indeed, Freeman et al. established that flavin reduction triggers large domain-scale motions in CPR, and changes in SAS can be explained in terms of ensembles of extended and compact conformations. Thus, our work follows in a line of thinking that is well grounded in excellent science on diverse systems.

3.3.4 Functional relevance of Models:

A mechanism for ETF-mediated FBEB has been suggested by multiple reports on different ETFs[26]. Demmer *et al.'s* proposal for conformational change as part of ETF turnover [41] has been supported and refined in subsequent papers[42, 43, 101]. In brief, The Bf-flavin is understood to accept a pair of electrons as hydride from NADH. The resulting RED Bf-Flavin then reduces both the ET-flavin and a low-potential acceptor (flavodoxin or ferredoxin) in tightly coupled single-electron transfers. This bifurcating event is held to occur in the closed conformation of Bf-ETFs where the proximity of the two flavins would maximize the efficiency of ET. The ET-flavin likely rests in its anionic semiquinone state (ASQ,[42, 102]) and therefore would assume the HQ state upon reduction, meanwhile the Bf-flavin would have returned to its OX state but might retain bound NAD^+. This state would need to undergo conformational change. The open conformation then presents ET-flavin (now as HQ) to the accepting cofactors of partner proteins that will reoxidize ET-flavin back to its ASQ state. Our current data do not directly address the

electron transfer aspects of the mechanism, but strengthen it by providing experimental evidence for *how* the large conformation changes can occur. We propose that the extended conformation we observe in solution, which has not been captured in crystal or Cryo-EM structures, mediates the conversion of closed ETF to the open conformation, and then back. We suggest that the extended conformations identified by our GA like represent snapshots of a dynamic conformational reservoir that provides conformational paths between the more compact open and closed conformations that have been documented so well.

In both oxidation states, *Afe*ETF populates an ensemble of extended and compact conformations. The hydrophilic interface between the base and head domains allows relatively facile detachment of the head that nonetheless does not result in escape, due to the flexible linkers. Together these provide the energetic plausibility of the extended conformations sampled by Bilbo-MD and metadynamics. In turn, the extended conformations enable the head domain to rotate freely with greatly diminished constraints. Upon re-association with the base, the result can be the closed conformation even if the starting point was open. We propose that the thermal population of extended conformations in solution allows the ETF to carry out electron bifurcation efficiently by facilitating the interconversion of the open and closed conformations.

Since the OX state has a larger population of extended conformations, the rate of interconversion of open and closed conformations may be higher than in the RED state. However we caution that the fully reduced state, wherein both flavins are simultaneously in the HQ state, as well as the fully OX state where both flavins are

OX, may have minor physiological relevance. Nevertheless, our comparison of these two oxidation states demonstrates that the system's conformational ensemble is responsive, and therefore coupled, to the oxidation state of at least one of the two flavins. This may simply reflect greater stability of the head/based domain interface when both flavins are RED. Regardless, our data demonstrate that even then, an extended population persists and is very significant, at approximately 20% of the population.

In the catalytic cycle, association with BCD or other partners is a critical event, which could stabilize the open conformation. However upon oxidizing the ET-flavin, the BCD should detach, and the ETF should return promptly to a conformation receptive of the next NADH. Only the RED state of ETF would be expected to form a strong stable complex with BCD, whereas our sample was OX. This may explain why our complexed ETF retains a large population of extended conformations.

**c7 significance: true B-state ?

3.3.5 SAXS vs SANS

While both SANS and SAXS are grounded in the principles of small-angle scattering, it is essential to recognize that the photoreactivity of flavins with X-rays can produce consequences not present when neutrons are scattered instead. Existing literature demonstrates that X-rays can induce photoreduction of flavins[83-85], which can subsequently influence protein conformation[51]. Nevertheless, carefully conducted experiments can control for this and thereby exploit the shorter acquisitions possible in SAXS[101], for example to perform time-resolved studies.

As part of an effort to understand distinctions between our findings and others in the filed, we compared the results of these two commonly used SAS methods. We do observe differences between scattering profiles of *Afe*ETF in SANS and SAXS. The Kratky plot and P(r) profiles both indicate differences in flexibility of ETF under X-ray irradiation vs scattering by neutrons. Moreover, our visible spectra and the derived intensities at different wavelengths show that our samples populated different oxidation states than those studied by SAXS. It appears that our samples were converted quantitatively to fully RED state, whereas only partial conversion was achieved in other studies. Since additional components of the sample may also be photoreactive under X-rays, including even the reductant or residual dissolved O_2, this is an area deserving further investigation. Thus SANS provides a unique complement to SAXS for small angle scattering studies where it is important to control the oxidation state of the protein. In our work we were able to fully reduce the ETF as confirmed by UV-Vis spectra. This is much more difficult to achieve in SAXS experiments.

3.4. Methods

3.4.1 Protein production:

The expression plasmid, pASG IBA33, containing the genes *Acfer*_0555 (encoding EtfB) and *Acfer*_0556 (encoding EtfA) was provided by R. Hille, with permission of Chowdhury et al. The plasmid was engineered to include a C-terminus 6X-Histag on ETFA. To produce the target protein, this plasmid was transformed into *E. coli* Nico21-DE3 cells.

A preculture was initiated by cultivating 10 mL of LB medium, supplemented with carbenicillin at 100 μg/uL, overnight at 37°C and 220 rpm. For protein expression, a 1 L TB medium, with the addition of 100 μg/uL carbenicillin, was prepared in a 3L Erlenmeyer flask. The medium was then inoculated with the 10 mL overnight preculture. The cells were grown at 37°C and 200 RPM until reaching an OD A600 of 0.6-0.8. At this point, the culture temperature was reduced to 18°C, and protein expression was induced by adding 20 μg/mL anhydrotetracycline. The culture was further grown for an additional 18 hours. Subsequently, the cells were harvested and washed with PBS buffer before being frozen for storage.

To obtain purified protein, frozen cells were thawed and suspended in a binding buffer composed of 50 mM HEPES, 10 mM imidazole at pH 7.5, with the addition of 1 mM PMSF (, 2 mM benzamidine, NaF, DNase I, lysozyme and 1mM FAD. Cell lysis was accomplished through sonication, using a pulse mode with 5 seconds on and 30 seconds off for a total of 30 cycles, while maintaining a cold environment with ice. The resulting cell lysate was loaded onto a gravity column containing Ni-NTA resin, pre-equilibrated with the binding buffer. The column was washed with 20 column volumes (CV) of the binding buffer supplemented with 20 mM imidazole. Finally, the target protein was eluted from the column using 150 mM imidazole in the binding buffer.

The eluted protein was concentrated and subjected to buffer exchange via gel filtration over a DG-10 column, which was pre-equilibrated with 50 mM KPO4(potassium phosphate dibasic plus potassium phosphate monobasic) at pH 7.0. An excess of 1 mM FAD was added in the same buffer for overnight

reconstitution, followed by buffer exchange the following day to remove any excess FAD. The pure protein was concentrated and rapidly frozen for storage.

Before the Small-Angle Neutron Scattering (SANS) experiments, an additional purification step was performed to ensure the highest protein purity. This involved size exclusion chromatography on a AKTA GO purification system using a Cytiva Superdex 200 Increase 10/300 GL column, which was pre-equilibrated with 50 mM KPO4 at pH 7.0. Fractions corresponding to the protein peak were collected and combined for subsequent experiments. Bifurcating *A. fermentans* ETF (*Afe*ETF) produced by *E. coli* eluted as a single peak in size exclusion chromatography (SEC), consistent with the 67.97 kDa mass expected for the heterodimeric protein . Only this fraction was employed for SANS, excluding higher-order aggregates. Such material, prepared and maintained in air-equilibrated buffers, was deemed oxidized (OX) based on its optical spectrum which shows that both flavins were oxidized before as well as after collection of SANS data (<10% of sites are 8fF, Figure 2,[103]). Our *Afe*ETF retained redox reactivity, assuming its fully reduced state (RED) upon treatment with a ten-fold excess of its natural reducing substrate, NADH.

3.4.2 BCD Deuteration:

Gene encoding the BCD cloned into IBA33 was provided by Chowdhury. The plasmid was transformed into *E. coli* BL21(DE3) cell line for expression. For production of deuterated BCD cells were initially grown in LB media at 37 °C until OD600 reached 0.6 – 0.8. The cells were then centrifuged in sterile centrifuge tubes

at 4000 x g for 30 min. The pellets were washed and centrifuged at 4000 x g for 10 min using 85% deuterated Enfors media [104] to remove excess LB media. Cell pellets were again resuspended in 85% deuterated Enfors media with carbenicillin (100 μg/mL) and induced with 20 μg/mL anhydrotetracycline. Expression of BCD in 70% D2O media was carried out for 18 hours at 18 °C.

Purification of BCD was done at 4 °C. Cell pellets were resuspended in the lysis buffer (50mM Tris, 150 mM NaCl, complete EDTA- free protease inhibitor cocktail (Roche), pH 7.5 and catalytic amounts of RNASE and benzoase nuclease and 1mM FAD) at the ratio of 4 g cells/mL of buffer and lysed by sonication using a Branson 450 Digital Sonicator at 40% amplitude for 30 cycles (pulse on for 5 s and off 30 s) on ice. The lysed cells were centrifuged at 17,000 x g for 20 min and the clarified lysate was filtered using 0.45 μm filter. Rest of the purification steps were same as the ETF except the buffer conditions. The final protein was in 50 mM Tris-HCl, 150mM NaCl at pH 7.5.

Complex sample preparation: Staring ETF concentration 4 mg/mL, Starting BCD concentration 3 mg/ mL, 200 microliters of ETF plus 300 microliters of the BCD were combined for the complex reaction. Final concentration of ETF in the complex reaction: 23.55 micro molar, BCD: 25.6 micro molar. The SEC profile of the reaction mixture resulted in a single peak.

3.4.3 SANS experiments:

SANS experiments were conducted at the CG3-BIOSANS instrument located within the High Flux Reactor (HFIR) at the Oak Ridge National Laboratory. A Banjo cuvette, designed with a 1 mm path length, was utilized for all SANS

experiments. These cuvettes were customized in-house by our glassblower shop to include an airtight cap, ensuring that the reduced samples remained in their reduced state throughout the experiments.

For each SANS experiment, a volume of 320 microliters of ETF samples, with a concentration of 3 mg/mL, was loaded into the cuvette. To maintain the reduced state of the samples, all reduced samples were prepared with de-gassed buffers. Excess NADH and dithionite were carefully added to the samples within a glove box, and the cuvettes were tightly sealed to prevent any exposure to oxygen or air, which could potentially cause the samples to become oxidized.

3.4.4 SAXS experiments:

Size exclusion chromatography coupled with Small Angle X-ray Scattering (SEC-SAXS) was employed at the Life Sciences X-ray Scattering Beamline (LIX) at NSLSII. Briefly, 60ul of sample was injected onto a Superdex 200 Increase 5/150GL column (Cytiva) using an Agilent 1260 Infinity II Bio Inert HPLC system. The flow rate was set to 0.35mL/min and eluate from the column was split 2:1 via a passive splitter between the X-ray flow cell and UV branch respectively[105]. SAXS/WAXS images were collected simultaneously on a Pilatus 1M (SAXS) and Pilatus 900K (WAXS) detector. Exposure time was 2 seconds, and 350 frames were collected. Data were merged, scaled and subtracted using LIX beamline software, *py4xs* and *lixtools* as described elsewhere[17].

3.4.5 Crystal Structure Comparison:

AlphaFold2-multimer models of the 4KPU, 6FAH, and 7KOE PDB structures of the ETF domains were made in ColabFold [106, 107]The PDBs were used as the only templates with no MSA with the unpaired-paired setting for alignment.

3.4.6 SANS data analysis and prediction of theoretical scattering

Using BioXTS RAW, CRYSON and PEPSI SANS

3.4.7 Genetic Algorithm: Ensemble optimization for SANS

Genetics algorithms have a long history in meta-heuristic optimization algorithms [Holland, J., UMichPress(1975)]. They are based on the idea of hereditary fitness of genes in that the genes with the fittest phenotypes are selected more often for procreation than poorer phenotypes. A gene is a representation of a particular phenotype or property of the optimization problem at hand. In the case of optimizing a model when comparing the fitness to SAS experiments, the phenotype is the reduced chi-squared, χ_v^2, of your model to the experimental scattering profile. At each generation, the genes are evaluated to their phenotype, evaluated for their fitness, $f_k = 1/\chi_v^2$, and ranked according to their weights, normalized to the total fitness of the population. The parents are paired for procreation to make the next generation, the "children", via their ranking such that those with the highest fitness are more likely to procreate. There are two processes that govern this step with two user-determined probabilities, cross-over (p_c), and mutation (p_m). Cross-over happens by choosing one part of each parent's genes and crossing them to make the new gene. Mutation occurs by randomly mutating one element of the gene to

another. These new children become parents for the next generation. This process is repeated for a number of generations until convergence is reached and theoretically the algorithm will find the best.

In the case of a minimum ensemble search (MES) such as was performed by the original Bilbo-MD algorithm for SAXS, the goal of the optimization is to find the minimum set of conformations, $\{I(q)\}_k$, that best fits the experimental scattering profiles. First, the pool of conformers is randomly grouped into the ensembles, $\{I(q)\}_k$. The gene is the ensemble of conformations and the χ_v^2 is the phenotype, given by the following equations:

$$I_{Mod}(q) = \left(c\sum_{j=1}^{k} w_j I_j(q\ \ ) + b\right); \chi_v^2 = \frac{1}{N-(k+2)}\sum_{i=n}^{N}\frac{\left(I_{Obs}(q_i)-I_{Mod}(q_i)\right)^2}{\sigma_{Obs}^2(q_i)},$$

[Eq 3.1]

where I_{Mod} is the model data with the respective weight of each scattering profile, w_j, in the ensemble, c is a global scaling parameter, b is a background offset, I_{Obs} and

σ_{Obs} are the observed experimental scattering profiles and errors. After each evaluation step in the generation, the fittest ensemble is updated. Then, parents are selected via their fitness as two distinct sets from the total population. Cross-over occurs with a probability, p_c=0.5, and is performed by switching elements of the sets with opposing indices. For example, in a 3-ensemble fit index 1 of parent1 would switch with index 3 of parent2 if the split occurs between profile 1 and 2 in the ensemble. After cross-over, the children are selected for mutation with probability, p_m=0.1, by selecting a random scattering profile in the set and

exchanging it out with another profile from the pool. If the child generated by the cross-over or mutation is not unique, defined as not sharing identical scattering profiles, and valid, having a reasonable fit i.e., all of the weights are greater than a cutoff (1e-6), then the parents become children for the next generation. As stated earlier, the process will continue for a number of generations, but can also be iterated over by randomly selecting different initial ensembles in case the algorithm gets trapped in a local optimum. When comparing ensembles of different size, the Aikake Information Criteria (AIC)[Aikake, IEEE(1974)] was implemented, given by:

$AIC = 2(k+2) - 2\ln(\chi_v^2)$.

[Eq.3.2]

The AIC penalizes the increase in fitting parameters, but as the comparison of X^2 of different models is not statistically sound, this metric provides a rough estimate of the information loss or gain upon adding the extra conformers to model the SANS data.

The initial ensemble of conformers can be generated any number of ways, but in this case two separate MD simulations were performed. The first was generated using the Bilbo-MD MD workflow on the Sibyls computer resources[93]. The MD simulations started from the 4KPU PDB structure. The base domain of ETF, EtfA1-205 and EtfB1-11,36-219, were selected as the fixed domain for Bilbo-MD sampling. The two flexible linkers between the base and head domains in EtfA and EtfB were left flexible, along with the loop,12-35, in EtfB. All other domains were held rigid. These domains were selected because they had

high b-factors in 4KPU. An R_g range from 20 to 35 Å was selected for the R_g restraints and the MD simulation generated 1200 conformations, split into 6 3Å R_g bins, with 200 conformations per bin. Another set of conformations was generated with the full FAD binding in the oxidized state.

The SAXS and SANS profiles were calculated from the two MD ensemble using Pepsi-SAXS/SANS [108] with 500 q-vectors spanning from 0.0 to 0.5 $Å^{-1}$ out to the 50^{th} multipole order. The SANS profile was calculated in 100% D_2O. Individual scattering profiles were fit to the scattering profile using the *scipy curve_fit* module in *python* with the Levenburg-Marquadt algorithm. For the genetic algorithm, the SAS profiles were fit to the model using the *lmfit* python module with restraints such that the sum of the weights were equal to one, i.e. $\sum_{j=1}^{k} w_j = 1$. The algorithm was tested running using 4 numbers of generations (25, 50, 75, and 100) and with up to 5 iterations. We found that it is possible for the algorithm to find an optimal solution equally likely regardless of the number of generations so 75 generations were used in this study for efficiency and sampling. The genetic algorithm was run 5 times for convergence in the ensembles. We ran the minimum ensemble search for ensemble sizes of 2, 3 and 4, but did not see any improvement in the χ_v^2 or fitness after testing 5 conformations.

We ran the GA for 2-, 3-, and 4- conformer ensembles over 75 generations and 5 iterations. These settings were chosen as it was equally likely for the genetic algorithm to find an optimal solution at 25 generations as it was 100 generations, so by running 75 steps and 5 repeat simulations with different initial conditions we

efficiently sample the combinatorial landscape for 2-, 3-, and 4-conformer ensembles.

The distinction remains which ensemble best fits the data. We calculated the Aikake information criteria (AIC) [Eq. 2] for each model and the relative likelihood [Eq 1] between the models, giving how probable a model is at minimizing the information loss relative to the best fitting model (i.e. the one with the lowest AIC). So relative to the 2-conformer model, the three and four conformer oxidized models give relative likelihoods of 0.158 and 0.096 implying they are only 16% and 10% as likely to minimize the information loss by incorporating extra conformers as the 2-conformer model. The same is roughly true for the reduced state, but its likely that a two-conformer model is the best model as the other models give relative likelihoods of 0.421 and 0.078. Thus, at minimum two conformers are necessary to resolve the SANS profiles, but at this level of sampling and resolution more may be necessary to understand the biochemically relevant conformations.

As the AIC was not clear on the particular model, perhaps neither simulation alone sampled the complete conformational landscape necessary to capture the SANS data.

3.4.8. Analysis of Selectively Deuterated Complexes

Bilbo-MD sampling for the BCD/ETF complex was performed by fixing all the BCD atoms and using the same residue selection as the unbound-ETF in the previous simulations. We sampled 1200 conformations with R_g bins between 45 to 55 Å. These 1200 conformations were combined with the combined ensemble from the free-EtfAB for the genetic algorithm analysis. The genetic algorithm was run

as a 3-state model over 5 iterations and 75 generations. The RMSD to the 5OL2 crystal structure was calculated using the pymol *align* function between the backbone and CB atoms.

Structural analysis: The R_FAD distance was measured by calculating the distance between the C7M atoms of the ET-FAD and Bf-FAD in VMD. The radius of gyration and the RMSD of both ensembles were calculated using *mdtraj* . The RMSD was calculated between backbone and CB atoms with the 4KPU crystal structure as reference.

3.4.9 System preparation for MD simulations

We borrowed the high-resolution X-ray crystal structure of the closed conformer (B-state) of Bf-ETF from PDB (code: 4KPU). We discarded water molecules and ions present in the structure. To incorporate His-tag (six consecutive His residues appended along the C-termini of the A-chain of Bf-ETF), we employed MODELLER 10.4. His-tagged Bf-ETF with both the flavins was placed at the center of a dodecahedron box extending a minimum distance of 15 Å from any atom of the protein. The box was solvated by the TIP3P explicit water molecules and requisite ions were added to neutralize the system. The system was minimized using the steepest-descent algorithm to remove any unusual geometry and steric clashes. Further, we performed equilibration of the system at 300 K for 500 ps using the Berendsen thermostat with a coupling time of 0.1 ps followed by equilibration at 1 atm pressure for 5 ns using the Berendsen barostat. All simulations were performed with a time-step of 2 fs with Amber ff14sb force field using GROMACS

2019.6 suite. GAFF compatible parameters for oxidized FAD were adapted from a previous work from our group .

3.4.10 Metadynamics simulations

We have performed metadynamics simulations for 50 ns to sample the energetically favorable intermediates during the transition of Bf-ETF from ‘closed’ to ‘open’ conformer. GROMACS patched with PLUMED 2.9 was employed to study the conformational transition of the protein. In this enhanced sampling method, a history dependent repulsive Gaussian shaped bias potential have been deposited along a collective variable (CV): distance between the center of mass between the non-hydrogen atoms of Bf-FAD and ET-FAD. Gaussians with a height of 0.1 kJ/mol and a width of 0.05 nm were deposited every 0.4 ps to fill the energy wells (local minima) and encourage the system to explore the whole free-energy surface in a reasonable computational time.

CHAPTER 4. STRUCTURE, DYNAMICS AND REDOX REACTIVITY OF AN ALL-PURPOSE FLAVODOXIN

Reproduced from the published paper in the Journal of Biological Chemistry, reference: [109]

.

Contribution from Authors:

Sharique Khan (S.K): Biochemistry work, Fluorine NMR, Assistance with crystallization

Ahmadullah Ansari (A.A): Flavodoxin crystallization and structure solving.

Monica Brachi (M.B): Electrochemistry experiments

Wassim El Housseini (W.E.H): Methodology for electrochemistry experiments

Debarati Das (D.D): Provided WT flavodoxin for electrochemistry experiments.

Shelley Minteer (S.M): Principal investigator & supervision of electrochemistry experiment

Anne-Frances Miller(A-F.M): Bioinformatics work, NMR experiments, Principal investigator, Supervisor

Credit statement from the published paper:

S. K. and A. A. methodology; S. K., A. A., M. B., and A.-F. M. investigation; S. K., D. D., S. M., and A.-F. M. resources; S. K., A. A., M. B., and A.-F. M. validation; S. K., A. A., M. B., and A.-F. M. visualization; S. K., A. A., M. B., D.

D., W. E. H., S. M., and A.-F. M. writing–review and editing; A. A. and A.-F. M. writing–original draft; S. M. and A.-F. M. conceptualization; S. M. and A.-F. M. supervision; S. M. and A.-F. M. funding acquisition

Abstract

The flavodoxin of *Rhodopseudomonas palustris* CGA009 (*Rp9*Fld) supplies highly reducing equivalents to crucial enzymes such as hydrogenase, especially when the organism is iron-restricted. By acquiring those electrons from photodriven electron flow via the bifurcating electron transfer flavoprotein (ETF), *Rp9*Fld provides solar power to vital metabolic processes. To understand *Rp9*Fld's ability to work with diverse partners, we solved its crystal structure. We observe the canonical flavodoxin (Fld) fold and features common to other long-chain Flds, but not all the surface loops thought to recognize partner proteins. Moreover some of the loops display alternative structures and dynamics. To advance studies of protein-protein associations and conformational consequences, we assigned the ^{19}F NMR signals of all 5 tyrosines (Tyrs). Our electrochemical measurements show that incorporation of 3-^{19}F-Tyr in place of Tyr has only a modest effect on *Rp9*Fld's redox properties even though Tyrs flank the flavin on both sides. Meanwhile, the ^{19}F probes demonstrate the expected paramagnetic effect, with signals from nearby Tyrs becoming broadened beyond detection when the flavin semiquinone is formed. However the temperature dependencies of chemical shifts and linewidths reveal dynamics affecting loops close to the flavin, and regions that bind to partners in a variety of systems. These coincide with patterns of amino acid type conservation but not retention of specific residues, arguing against detailed specificity with respect to partners. We propose that the loops surrounding the flavin adopt altered conformations upon binding to partners and may even participate actively in electron transfer.

4.1 Introduction

From shortly after their initial description [110, 111] flavins were recognized as remarkable on account of their ability to execute one-electron (1e) transfers, in contrast to the 2e redox activity of most other dyes known at the time [112]. Indeed, the FMN-containing protein flavodoxin (Fld) was so named based on its ability to serve many of the roles played by ferredoxin (Fd), a [Fe_4-S_4] cluster-containing carrier of 1e equivalents at low reduction midpoint potential (E°) [113]. Use of an organic cofactor instead of Fe would have enabled early life to exploit environments previously uninhabitable due to Fe scarcity. Moreover, diminished reliance on Fe would have become critical when more oxidizing conditions prevailed and made Fe less bioavailable and toxic, by virtue of the Fenton chemistry Fe catalyzes. Modern organisms continue to use Fld instead of Fd under conditions of oxidative stress or Fe scarcity [36, 114-118]. Indeed, Fld is prevalent in our pathogens, enabling them to survive the hoarding used by hosts to deprive pathogens of Fe [117, 119]. Conversely, augmentation of plant genomes with Fld genes has shown promise for increasing plants' tolerance to stress [120].

Flavodoxins comprise a diverse superfamily of 145-180 amino acid long proteins which use a non-covalently bound FMN to mediate electron transfer. (The specialized Fld-like ribonucleotide reductase-activating Nrdl [114, 121], and the more distantly-related WrbA classes display diverging reactivities and are not discussed here [122].) Collectively, bacterial Flds service a wide variety of partner

proteins including Fd (Fld) $NADP^+$ oxido-reductase (FNR, [123, 124]), nitrogenase [125-127], nitrite reductase [128], biotin synthase [129, 130], methionine synthase [131, 132], hydrogenase [133, 134], pyruvate formate lyase [135], pyruvate oxidoreductase and 2-oxoglutarate oxidoreductase [136], cytochrome P450 [137], radical SAM enzymes [138] and likely more. Thus, individual organisms have been found with as many as 14 different Flds [117].

Based on the distributions of Fld and Fd in the tree of life, Fld is inferred to have evolved later than Fd, and gained prominence following the rise in O_2 [117]. However the smaller diversity of Flds vs. Fds suggests that individual Flds may serve more partners . We infer that Fld binding should be less selective than binding of Fd, although Fld's binding motif must be similar to that of Fd, since Fld must bind the same partners. Flds are indeed found to conserve a negatively charged patch surrounding the site where the xylene ring of the FMN protrudes into solution [131, 139]. This resembles the negatively charged binding surface of Fd [140]. However Flds, especially long-chain Flds, are distinguished by the presence of several loops that surround the protruding flavin, and have been proposed to guide Fld's association with partner proteins [119, 141].

While the first protein named 'flavodoxin' was isolated from *Clostridium* [113] and is now known to be a short-chain (SC) Fld, a prior report of a protein named 'phytoflavin' can now be attributed to a long-chain (LC) Fld [142]. The SC and LC Flds diverged early and are distinguished by absence or presence of a ~20 residue 'long loop' (LL) excursion from β strand 5, that buttresses the tyrosine 90 (Y90) which in turn stacks against the *si* face of the flavin ([143] numbering of

Rhodopseudomonas palustris will be the default throughout). Thus the LL provides a flexible cowl [116, 143] that expands the protein's interactive surface surrounding the flavin.

The LL does not appear to confer stability, although it strengthens affinity for FMN [143] and may protect the flavin from oxidative stress [117]. Due to its association with photosynthetic electron transfer to NADH [100, 117], its significant sequence conservation, and its interactions with partner proteins [131, 144], it is widely believed to mediate interactions with partner proteins [141]. Thus, it could play a role in selectivity, location of the correct binding surface and securing Fld in a productive pose for electron transfer. Nevertheless, in organisms which possess only a single Fld, the Fld must interact with several different partners, suggesting that adaptive loop(s) would be a better binding strategy than a fixed surface. This is the situation in the phototroph *R. palustris* CGA009 [145], so we have overexpressed its sole Fld, *Rp9*Fld, to test the hypotheses that the LL will be flexible, other loops surrounding the protruding flavin may also be flexible, but non-loop regions distant from the flavin should not be.

To characterize local dynamics, we are developing ^{19}F NMR probes of individual side chains, especially residues in the surface loops of Fld that may respond differently to different partners [131]. ^{15}N NMR relaxation has shown that surface loops have lower backbone order parameters than the protein as a whole [146] but ^{15}N relaxation focusses on rapid motions in the ps - ns time scale, whereas we want to investigate the LL's role in slower processes of binding to and detaching from large partner protein complexes. For this, it is ideal to place ^{19}F in amino acid side

chains believed to mediate the interactions with partner, or residues that trigger binding/release. Indeed, NMR revealed two environments affecting the conserved Y90 flanking the flavin [147]. However those data also raised concerns regarding the integrity of the protein. To provide better ^{19}F NMR coverage of the Fld while retaining the critical conserved Tyrs, and to assess the generality of our findings, we have now applied the same methods to Fld from a different strain of *R. palustris*. Using a combination of X-ray crystallography and NMR spectroscopy, we confirm that the loops surrounding the flavin access diverse conformations, setting the stage for adaptive binding. Moreover a role in partner binding is supported by analysis of > 500 different Fld sequences.

4.2. Results

4.2.1 Crystal structure of *Rp9*Fld preserves well-known Fld fold

To exploit availability of a fully sequenced genome, the Fld from strain CGA009 of *R. palustris* [145] was employed (Rp9Fld hereafter, = Uniprot ID Q6N7Y7). The Fld we studied previously was from strain 2.1.18 (*Rp8*Fld) and was also the sole Fld in its genome, suggesting that binding surfaces of these two Flds should be conserved. However these two orthologous Flds differ in 25% of their amino acids. 16 distinguishing residues are similar, but 24 are not.

To assess the potential significance of the differences and make sense of previous findings regarding multiple environments for conserved Y90 [147], we solved the crystal structure of *Rp9*Fld. This revealed most of the differences between the two

Flds to be surface-exposed, and predominantly on faces not proposed to interact with partner proteins [148].

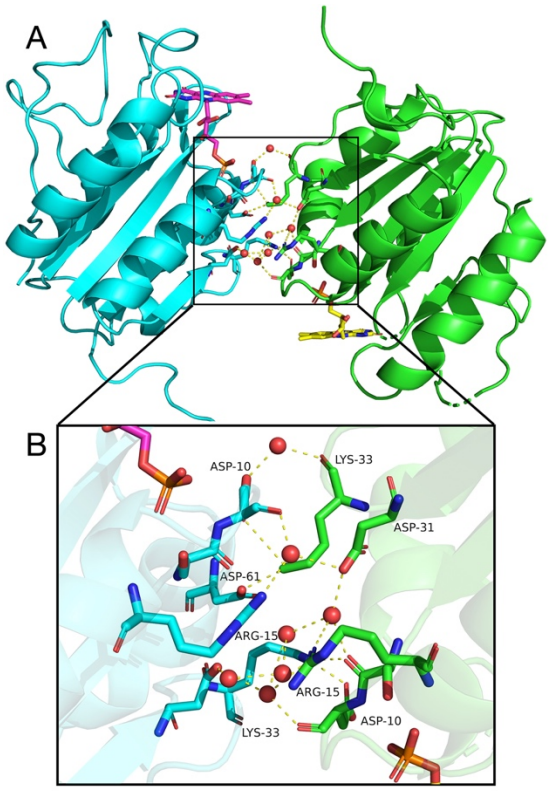

Figure 4:1 Ribbon structure of Rp9Fld, showing the opposing orientations of the two protomers and the interface between them.
Panel A shows complete ribbon diagrams of both protomers, with unresolved backbone of protomer B indicated by dashed lines. The long N termini include a resolved AARGSH peptide representing the LVPRGSH of the thrombin cleavage motif that immediately precedes the natural N-terminus of *Rp*9Fld in our construct. Panel B shows the interactions that bridge the protomers, in this relatively polar interface. The B chain is shown on the right. Figure was generated using Pymol [149].

We determined the structure of *Rp9*Fld at 2.1 Å resolution based on diffraction data collected at 100 K (data collection and refinement statistics appear in Table 1). *Rp*9Fld displays the signature flavodoxin fold, including the 5-stranded β sheet sandwiched in between five α helices (Figure 4.1), however it has a shorter loop

from 59 to 60 than the corresponding loop from 64 to 74 in *Rhodobacter capsulatus* (2WC1.pdb) or *Azotobacter vinelandii* (5K9B.pdb) [150, 151].

Table 4.1: X-ray diffraction data collection and refinement statistics

Wavelength (nm)	1.5406	Number of non-H atoms	2605
Resolution range (Å) a	24 - 2.17 (2.248 - 2.17)	macromolecules	2439
Space group	P 1 21 1	ligands	62
Unit cell (Å) a b c Unit cell (°) a b g	46.6639 58.8915 60.9043 90 105.494 90	solvent	104
Total reflections a	131196 (12875)	Protein residues	329
Unique reflections a	16966 (1672)	RMS(bonds)	0.008
Multiplicity	7.7 (7.7)	RMS(angles)	0.91
Completeness (%) a	99.80 (99.76)	Ramachandran favored (%)	95.22
Mean I/sigma(I) a	22.72 (4.20)	Ramachandran allowed (%)	4.78
Wilson B-factor	30.85	Ramachandran outliers (%)	0.00
R-merge a	0.09818 (0.5687)	Rotamer outliers (%)	0.40
R-meas a	0.1055 (0.6112)	Clashscore	8.57
R-pim a	0.03776 (0.2208)	Average B-factor	35.90
CC1/2 a	0.983 (0.932)	macromolecules	35.90
CC* a	0.996 (0.982)	ligands	37.63
Reflections used in refinement a	16947 (1668)	solvent	34.95
Reflections used for R-free a	825 (78)		2605
R-work a	0.2133 (0.2612)		2439

R-free a	0.2689 (0.3554)		62
CC(work) a	0.944 (0.876)		104
CC(free) a	0.907 (0.748)		329

We modeled our Fld in terms of two protomers per asymmetric unit; chains A and B, because some loops of chain B have Cαs displaced by as much as 4.0 Å from their positions in chain A (RMSD = 0.739 for all Cα) (Figure 1). The two protomers are oriented in opposite directions (the flavin appearing near the top for one protomer when the flavin of the other appears near the bottom, Figure 4.1A). The 726 $Å^2$ interface between protomers constitutes some 10% of the surface area of one protomer based on PDBePISA [152], and is stabilized by 12 hydrogen bonds involving primarily residues from helix 3. Helix 3-mediated association between oppositely oriented protomers is also observed between the two Fld domains in dimeric flavodiiron proteins (FDPs), where dimerization brings together the FMN of one protomer with the diiron site of the other, enabling electron transfer between them (*Moorella*, 1YCH and 4D02, [153, 154]). However consistent with the relatively small interface area shared by its two protomers, *Rp*9Fld is a monomer in solution, based on size exclusion chromatographic elution as a single species with an effective mass of 17.5 kDa ±5%.

4.2.2 Signatures of disorder in crystals at two temperatures.

A second distinction between the two protomers in *Rp*9Fld's crystallographic unit cell is that stretches of chain B could not be resolved whereas the same residues in chain A were well ordered (Figure 4.2A) This indicates that position in the crystal lattice plays a role. Disorder in the loop containing S146 of chain B may reflect a

clash between S146.B and E63.B in an adjacent unit cell. Other unresolved regions in chain B adjoin cavities in the reconstructed lattice, so we speculate that these stretches populate multiple local conformations. Both phenomena could reflect loops that are dynamic and/or adopt a variety of conformations as part of crystallization. To distinguish between mobility vs. static disorder as a basis for the low resolution of these loops, we used another crystal to collect diffraction data at room temperature (RT), with the understanding that at this higher temperature, mobility would be greater due to higher thermal energy.

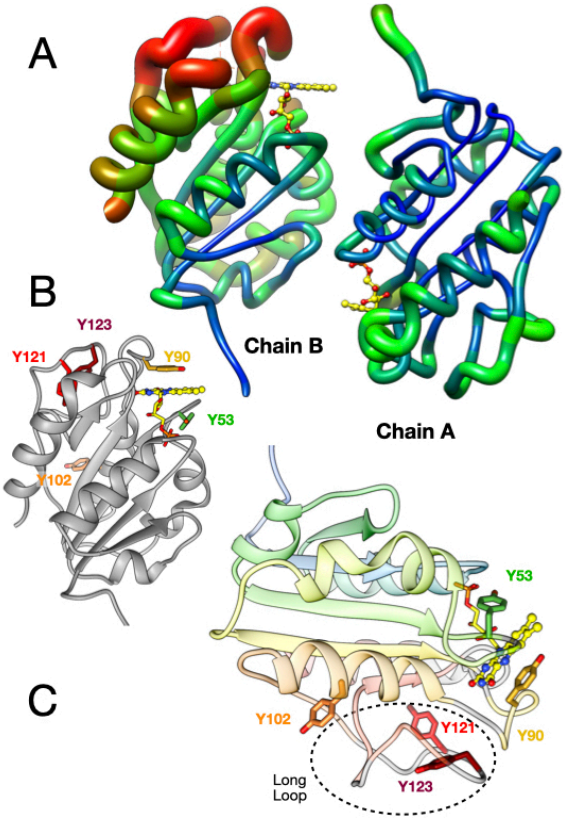

Figure 4:2 Structure and disordered regions of Rp9Fld.
A: B-factors are mapped onto a cartoon of the Fld structure with wide worms (2 Å) presenting regions with the highest B-factors (71 Å2) and narrow worms (0.25) representing regions with the lowest B-factors (19 Å2). A red-green-blue scale reinforces the same information with highest B-factors in red. Panel B reproduces the same orientation used for the B chain in panel A but using the A chain of the low-temperature structure instead to provide the complete backbone and displaying the Tyr side chains colour-coded as in spectra that follow. C. The ribbon topology of the A chain is shown via a blue-to-red N-to-C rainbow in which grey is used to indicate residues that are not resolved in the B chain at RT, residues 54, 86, 93-94, 117-125, 130-132, 141-148. The LL is indicated with a dashed oval and contains Y121 and Y123. The loop containing Y90 is sometimes called the Y-loop while the loop containing Y53 in our case, but W in many others, is known as the W-loop. These molecular graphics and those in later figures were made using Chimera [155, 156].

The structure based on RT data was lower in resolution overall (2.9 Å) and included more unrefinable stretches than the structure based on data collected at 100 K (Figure 4.2A, C). All of these were still in chain B though, and the fully resolved A chain of the RT structure overlayed well with the chain A based on100 K data (RMSD = 0.447 Å across 165 pairs for chains A, vs. 1.718 Å across 134 pairs for chains B). Thus, we interpret the RT model as an additional perspective on the same structure, but one in which local structure is more mobile and/or more amenable to distortion by crystal packing forces. In Figure 2C, regions of chain B that are unresolved at RT are in grey. Most of the chain B residues that were disordered at 100 K were also disordered in the RT data along with numerous additional residues (exceptions are 56, 87-92 and 129). This temperature dependence suggests that causes of the high B-factors and unrefined stretches include dynamics, not just static disorder. The sections not resolved at RT are interesting because three of them occurred in loops observed to dock with the photosystem I protein (PSI) in the cryo-electron microscopy structure of a Fld•PSI complex [144] and/or to have altered

NMR chemical shifts upon binding to methionine synthetase and FNR partner domains [131]. Thus, the dynamics we observe in our crystal structures could have functional significance.

4.2.3 Validation of ^{19}F NMR probes of environment and dynamics

We used NMR spectroscopy to test for conformational dynamics in solution. To compare regions of the protein that were different in the two monomers, or at different temperatures, we targeted Tyr side chains. Incorporation of 3-^{19}F-Tyr enables observation of Tyr side chains via highly sensitive ^{19}F NMR, with essentially no spectral background. We accomplished this by producing *Rp9*Fld in bacterial growth medium containing 3-^{19}F Tyr instead of Tyr, while inhibiting *de-novo* Tyr biosynthesis. *Rp9*Fld has a useful number of Tyr side chains distributed in regions of interest. Y121 and Y123 reside in a region that was unresolved by crystallographic data obtained at RT; Y53 is immediately adjacent to the flavin on its *re* face and Y90 is stacked against the flavin on its *si* face (Figure 4.2C). Y102 provides a valuable control, reporting on integrity of the overall structure, as it is in the centre of a helix that is well resolved in the crystal structures obtained at both temperatures, and not close enough to be directly influenced by the flavin.

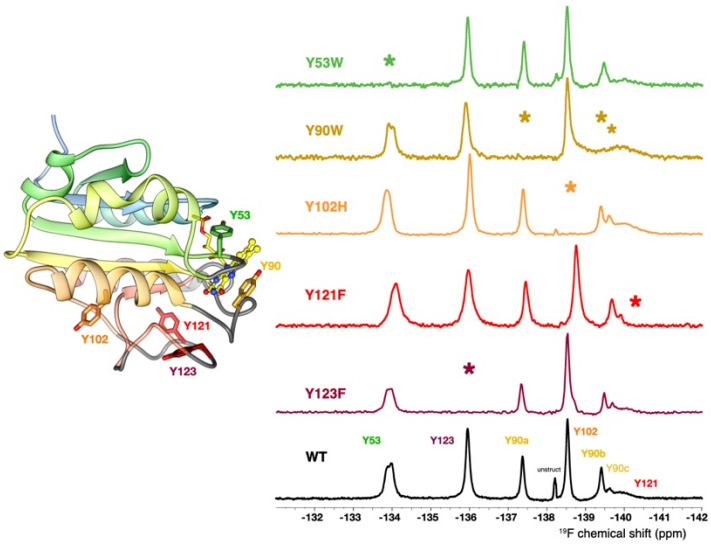

Figure 4:3 NMR spectrum assigned via replacements of individual Tyrs.
In each spectrum, the location of a missing resonance is indicated with an asterisk. The WT resonance at this position is thereby assigned to the residue that had been replaced in the variant. In the case of Y121, replacement with F also caused a slight shift in the resonance of Y123. Spectra were processed with 50 ms Gaussian weighting to maximize signal-to-noise because the samples of amino acid substituted variants were more dilute (weaker signals) than those of WT *Rp9*Fld. A ribbon structure is provided for reference, colored with a blue-yellow-red rainbow (N- to C-terminus) to depict primary structure.

The ^{19}F NMR spectrum is fully resolved but displays *six* strong resonances, although there are only *five* Tyrs (Figure 4.3, bottom spectrum). However two of the resonances have integrals only half as large as those of the other signals (Table 4.2). In the *Rp8*Fld, a pair of half-strength resonances was also observed and was assigned to Y90 [147], so we determined whether the same is true in this case.

We assigned all the resonances by generating a series of singly-mutated *fldA* genes, to remove each Tyr one at a time via single site substitutions (Figure 4.3). For each position, we converted the existing Tyr to a residue that commonly replaces it among other LC Flds. All the variant Flds were well-behaved and one resonance disappeared altogether while others barely shifted, permitting assignment of the

absent resonance to the Tyr that had been changed (Figure 3). Both the half-strength resonances were absent from the Y90W variant, along with a smaller sharp resonance near the upfield half-strength resonance (resonances Y90 a, b, c in Table 4.2). All three were therefore assigned to Y90 and the fact that there are several indicates that FldA populates two dominant conformations that affect Y90, plus a third minority conformation, at least.

Because it is unusual to lose multiple signals at once upon substitution of a single amino acid, we applied a second test of these signal's assignment to Y90. Y90 and Y53 are immediately adjacent to the flavin ring based on our crystal structure, so both should be paramagnetically relaxed when Fld adopts its semiquinone state. In contrast, all other Tyr side chains are ≥ 8.5 Å away. Indeed, Figure 4.4 shows that formation of the neutral semiquinone (NSQ) state made signals attributed to both Y53 and Y90 invisible, presumably due to extreme line broadening. The neutral semiquinone (NSQ) was formed by titration with dithionite and confirmed based on the sample turning blue. The fact that none of the other resonances shifted significantly indicates that flavin reduction did not change the structure in a way affecting environments of Tyrs 102, 121 or 123. The linewidths of Tyrs 102, 121 or 123 were also unchanged. The fact that Y121's resonance retained the same breadth, and indeed remained visible at all, indicates that its breadth reflects dynamics rather than proximity to paramagnetic flavin. Therefore, based on their insensitivities to flavin oxidation state, we can attribute the different linewidths of Y102 and Y123 *vs*. Y121 to different dynamics instead.

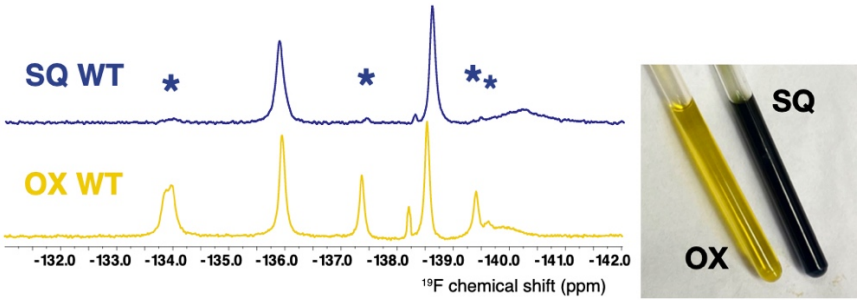

Figure 4:4 Extreme paramagnetic broadening due to FMN semiquinone identifies resonances of residues nearby.

The NMR spectrum of oxidized (OX) *Rp9*Fld in yellow is compared with that of the NSQ state (yellow spectrum) in which the starred resonances are so broad as to be unobservable, in almost the entire population. A small fraction, on the order of 0.1 remains OX and its small signals can be seen in the red spectrum. The inset at right shows the two samples, with the SQ's blue color being so intense as to render the sample almost black. Spectra were processed with 50 ms Gaussian weighting to maximize signal-to-noise and permit use of more dilute samples that facilitate quantitative reduction.

Table 4.2 Chemical shifts, integrals and line widths vs. temperature

Residue	Chemical Shift @ 25 °C (ppma)	Integral Area vs. Y123 @ 6 °C	Deconvoluted Area vs. Y123 @ 6 °C	Temperature dependence of chemical shift rel. to unstructured Y. (ppb/ °C)
Y53	-133.982	1.039	1.09	0
Y90a	-137.381	0.525	0.53	-4
Y90b	-139.424	0.379	0.35	7
Y90c	-139.64		0.03	9
Y102	-138.545	1.038	1.05	2
Y121	-139.9	0.691	0.80	29
Y123	-135.967	1.00	1.0	2
unstructured Y	-138.223	0.183	0.16	-

The assigned resonance positions reveal interesting features. Y53 displays a significant downfield shift relative to the other resonances (Figure 3), consistent with hydrogen bond donation from the phenolic alcohol to the FMN phosphate. Additionally, the resonances vary considerably in width at half-height (Figure 3), suggesting relaxation due to internal dynamics in addition to overall tumbling of the Fld. Thus, our ^{19}F spectrum reports on both residue-specific environments and dynamics that vary over the protein.

4.2.4 Distinguishing slow- vs. fast-exchange dynamics via dependence on temperature

Multiple co-existing conformations can produce sharp NMR lines if the conformations interconvert so slowly that they behave like non-interconverting structures (slow-exchange limit) or if interconversion is so rapid that NMR perceives a single average conformation and reports a single chemical shift that is the population-weighted average of those of the participating states (fast exchange limit). To test whether the multiple sharp lines attributed to Y90 represent alternative conformations that interconvert slowly on an NMR timescale, and whether any other lines might reflect multiple conformations interconverting, we modulated the rate of interconversion by lowering and raising the temperature. Elevated temperature should accelerate exchange and cause broadening and rapprochement of the lines if they correspond to conformations exchanging slowly with one-another. However lines broadened by intermediate exchange should sharpen as averaging accelerates at higher temperatures. Meanwhile a general sharpening is expected at higher temperatures as the decreasing viscosity of water allows more rapid molecular tumbling.

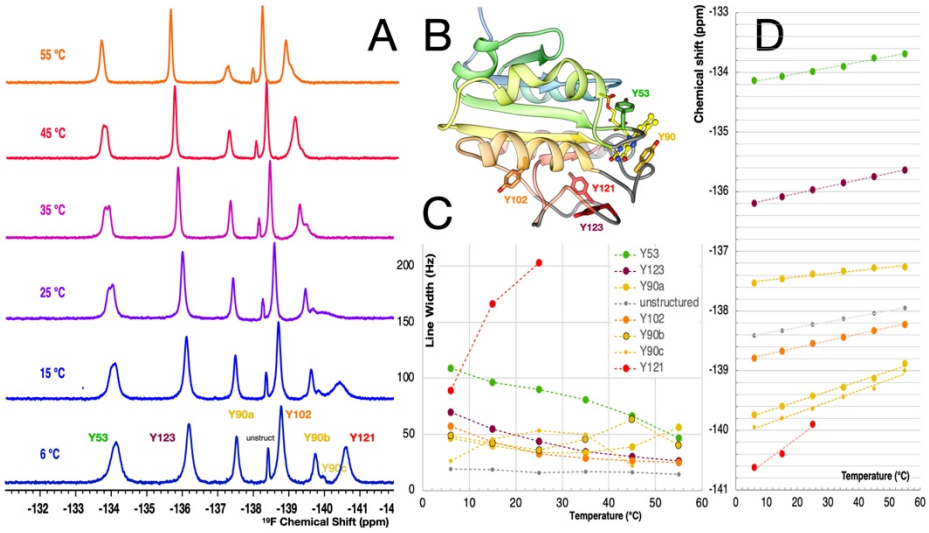

Figure 4:5 Temperature dependence of NMR resonances identifying dynamic regions. A: ^{19}F NMR spectra of ^{19}F-Tyr *Rp9*Fld as a function of temperature. Resonance assignments are provided on bottom spectrum, and the corresponding residues are shown in the ribbon structure of the A chain at 100 K, in panel B. C: Comparison of temperature dependencies o f line widths at half-height as a function of temperature, showing Y121's resonance sharpening at lower temperature (smaller line width), while resonances of Y102 and Y123 sharpen at elevated temperature, as for simple tumbling of a rigid protein. Increased line widths at higher temperature identify Y121 and 90 as affected by slow dynamics at RT. Y90 clearly populates ≥two conformations and Y53's resonance is also seen to be split at intermediate temperatures. All three affected Ys occur in or adjacent to regions not structured in B chain at RT (Panel B). D: Comparison of temperature dependence of chemical shifts. The small sharp resonance near -136.8 ppm corresponds to unstructured ^{19}F-Tyr in unfolded protein and informs on the temperature dependence expected in the absence of conformational dynamics. The chemical shift of Y90a approaches that of Y90b as the temperature increases. Y121's chemical shift also climbs steeply indicating existence of a cryptic resonance at less negative chemical shift with which the visible resonance will eventually merge. The temperature dependence was not extended to higher temperatures due to evidence of slow protein aggregation at 55 °C, where data were collected in blocks so that only blocks completed before onset of significant denaturation / aggregation were pooled. Spectra were processed using a 100 ms Gaussian to diminish its contribution to line widths.

NMR spectra collected at temperatures ranging from 6 °C to 55 °C are shown in Figure 5A. Even the relatively sharp signals of Y102 and Y123 are broader at 6 °C than at 55 °C, due to slowed molecular tumbling at lower temperature. However

the resolved line of Y90a is visibly *sharper* at low temperature, indicating slow exchange that accelerates and contributes to the line width at higher temperatures. Indeed, at 55 °C the widths at half-height of the Y90 signals exceeds those of Y102 or Y123. Thus, Y90 exchanges among multiple conformations, more rapidly at 55 °C (Figure 4.5C). The 'a' and 'b' lines attributed to Y90 approach one-another as the temperature is raised, from 2.2 ppm apart at 6 °C, to 1.6 ppm apart at 55 °C (Figure 5D, note the extra shallow slope of Y90a's temperature dependence and the extra steep slopes of the Y90b and Y90c resonances, as they converge). These observations confirm that Y90's multiple conformations are in slow exchange throughout the temperature we were able to examine. Coalescence temperature is above what we can explore, as protein aggregation occurs above 55 °C.

The resonance of Y121 also became broader and shifted as the temperature was raised, but in a lower temperature range (Figure 5A, C). Indeed, Y121 broadened beyond resolution by 35 °C. The chemical shift of Y121 also became less negative at increasing temperatures, with a slope much too steep to be explained by the temperature dependence of the reference shift alone. This indicates that the unidentified signal of the other conformation is at a less negative chemical shift.

Interestingly, the resonance of Y53 is slightly but reproducibly split at 25 and 35 °C (Figure 5), although the two conformations responsible are evidently in fast exchange at 55 °C and the general broadening due to slow tumbling impedes their resolution at 6 °C. Coalescence occurs near 45 °C. These signatures of dynamics affecting Y53, Y90 and Y121 show that three of the loops that were poorly defined in the B chain of the crystal structures (bearing residues 54, 86, 93-94, and 117-

125) are indeed affected by dynamics. The fact that it is observed in solution exonerates crystal packing as the cause of heterogeneity. These separate stretches of the peptide chain are brought together by Fld's fold, where they surround the flavin head group, and include the long loop (Figure 2C).

4.2.5 Assessing possible functional perturbation of fluorinating Tyr side chains

Although ^{19}F NMR is highly perceptive of dynamics, its utility is limited if it alters Fld's functional properties. Principal among these are the reduction midpoint potentials relating the oxidized and NSQ states ($E°_{OX/NSQ}$) and the NSQ and hydroquinone (HQ) states ($E°_{NSQ/HQ}$). To test for perturbation of these, we employed direct electrochemistry. *Rp9*Fld was immobilized on Toray carbon paper electrodes by entrapment in a cross-linked conductive hydrogel of linear polyethyleneimine functionalized with pyrene groups (pyrene-LPEI). Entrapment in a hydrogel increases the amount of protein that can be immobilized on an electrode [157] and incorporation of pyrene in the polymer improves electrical contact with the electrode, while preserving the protein's natural structure [158]. The pyrene groups also bind to the carbon surface via non-covalent pi–pi interactions [159] and the amino groups of the polymer may interact with the patch of negative surface charge with which Fld binds to its natural partners. Thus, pyrene-LPEI has been shown by Minteer's group to promote direct electron transfer by redox enzymes [157].

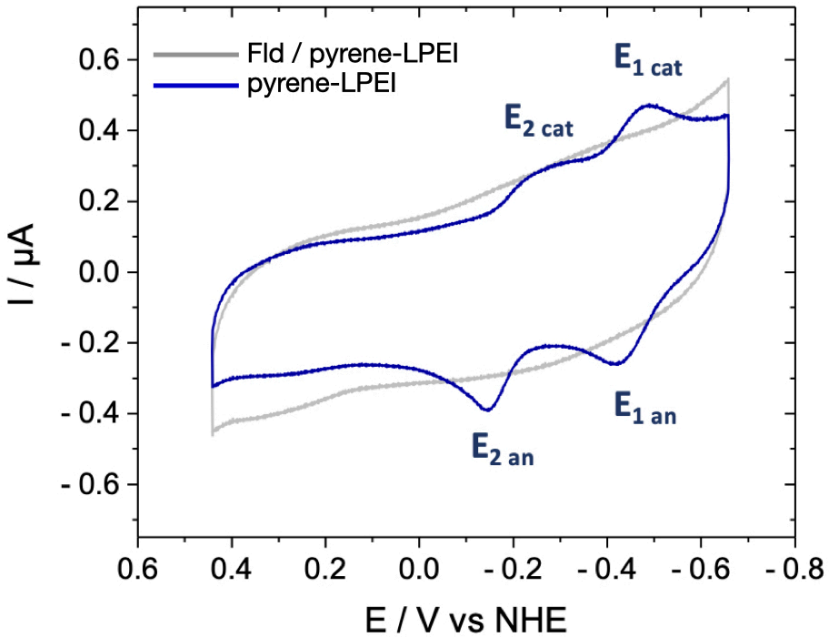

Figure 4:6 Direct electrochemical characterization of Rp9Fld's two E°s.
Cyclic voltammogram of WT *Rp9*Fld immobilized in pyrene-LPEI on a carbon paper electrode (blue) vs. an electrode modified with only pyrene-LPEI (grey) in pH 7.0 phosphate buffer 100 mM) vs. NHE. E_1 and E_2 indicate the potentials for the NSQ/HQ and OX/NSQ redox couples respectively, cat and an indicate the values detected in cathodic and anodic sweeps. The scan rate was 5 mV s^{-1}

WT *Rp9*Fld, ^{19}F-WT *Rp9*Fld and ^{19}F-Y90W-Fld were characterized by cyclic voltammetry (CV) and square wave voltammetry (SWV) over a range of pHs (4.0 - 9.0). In each case, results obtained with replicate electrodes were compared. The CV of WT-*Rp9*Fld shows well-defined and reproducible cathodic and anodic waves, as expected (Figure 4.6). The peak-to-peak separation for the NSQ/HQ couple of about 58 mV confirms the reversibility of the electron acquisition/release process while the larger ΔE (98 mV) observed for the OX/NSQ couple suggests that a quasi-reversible event is coupled to the electron transfer as also reported in literature [150]. We attribute this to the protonation of N5 and rotation of peptide backbone known to accompany the OX → NSQ reduction of Fld [160]. Each redox

couple's anodic and cathodic potential values were averaged to yield E°s, which are compared with literature values in Table 4.3. All values were further confirmed by SWV data.

Table 4.3Comparison of midpoint potentials of native and 19F-Tyr

E° vs. NHE	E°OX/NSQ (mV) at pH 7.0	E°OX/NSQ vs. pH (mV/pH)	E°NSQ/HQ (mV) at pH 7.0	E°NSQ/HQ vs. pH for pH ≤ 7 (mV/pH)	Reference
WT-Rp9Fld	-186 ± 4	-33.5	-450 ± 2	-23.5	
WT-Rp9Fld (19F)	-178	-32.6	-424	-21.9	
Y90W-Rp9Fld (19F)	-181 ± 2	-35.9	-441 ± 5	-31.2	
Rp8Fld	- 241 ± 3		- 434 ± 1		[147]
Rp8Fld (19F)	− 244		− 428±1.7		[147]
A. vinelandii Fld II	-187 ±9		-483 ±1		[150]
A. vinelandii Fld IIa	-164 mV ±12	-34	-454 ±7		[150]
A. vinelandii Fld b			-459		[161]
A. vinelandii Fld			-460		[162]
E. coli Fld c	-254 ± 5		-433 ±6		[163]

[a] adsorbed on DDAB-modified basal plane graphite electrodes, DDAB = didodecyldimethylammonium bromide

[b] with neomycin and glassy carbon electrode.

[c] spectrophotometrically-monitored titrations

Our value of $E°'_{NSQ/HQ}$ = -450 ± 2 mV is very close to the -434 mV obtained for *Rp8*Fld [147], but our $E°'_{OX/NSQ}$ = -186 ± 4 mV is significantly higher than the -241 mV obtained for *Rp8*Fld (E°' is the E° at pH 7.0). *Azotobacter vinelandii* Fld II has $E°_{OX/NSQ}$ = -187 mV and $E°_{NSQ/HQ}$ = -483 mV when free in solution, while $E°_{OX/NSQ}$ = -164 mV and $E°_{NSQ/HQ}$ = -454 were obtained for the same Fld adsorbed to electrodes [150]. The 20-30 mV elevation of the E°s could reflect slight perturbation of the protein's chemical microenvironment including desolvation of the flavin, and the surface chemistry of the electrode. The cationic nature of the LPEI polymer is consistent with elevation of $E°_{OX/NSQ}$ of immobilized *Rp9*Fld vs. free *Rp8*Fld [157] but is larger than expected, so we suspect that the protein itself, in addition to its state in the measurement, raises $E°_{OX/NSQ}$ more in *Rp9*Fld than in *Rp8*Fld. Indeed, the value for *Rp9*Fld agrees much better with literature than did the *Rp8*Fld result (Table 4.3).

Having confirmed the method, we assessed the effect of replacing Tyr with 3-^{19}F-Tyr. $E°_{NSQ/HQ}$ was 26 mV less negative in fluorinated Fld than the natural version, and for $E°_{OX/NSQ}$ the change was 8 mV. These are similar in magnitude to the 17 mV change produced by replacing Y90 with W (Figure 4.7 and Table 4.3, also see [164]), wherein W is present naturally at that position for the LC Fld from *Methylobacte*r locus PPC92239 [165]), and many SC Fld and Fld domains. Thus such variations in E°s appear to be within the functional spectrum.

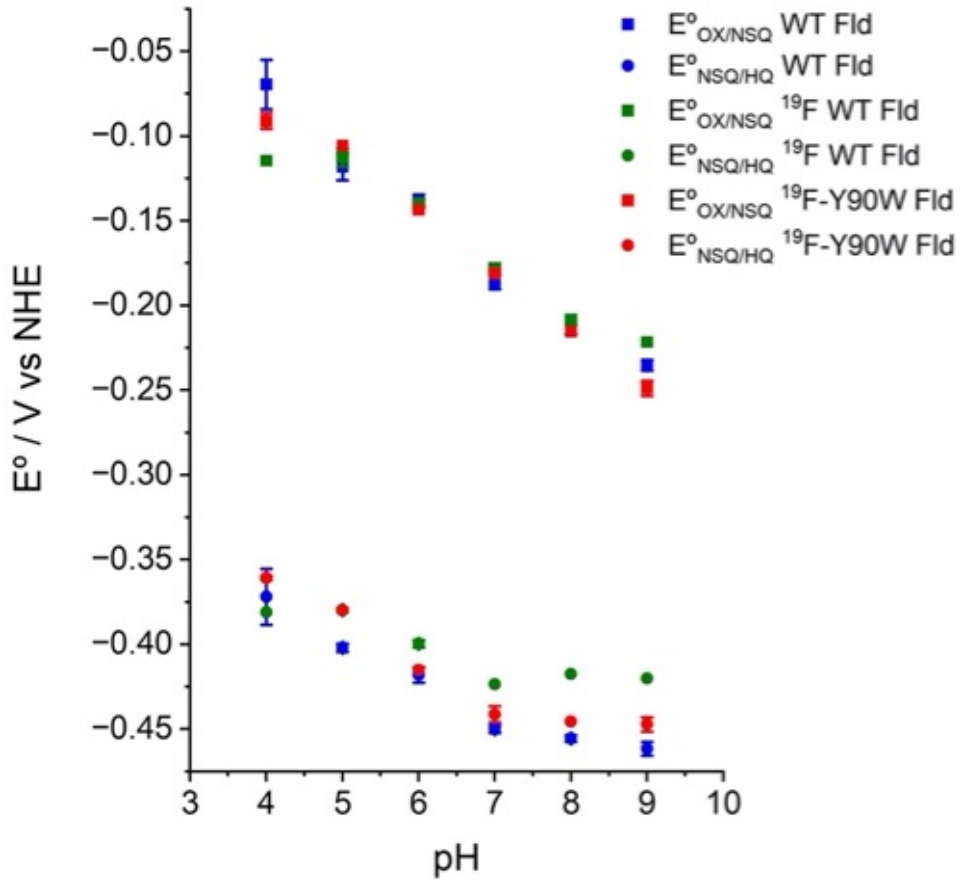

Figure 4:7 pH dependence of the midpoint potentials for the OX/NSQ and NSQ/HQ redox couples for all the three proteins tested.

WT *Rp9*Fld, ^{19}F WT *Rp9*Fld and ^{19}F-Y90W *Rp9*Fld were immobilized in a pyrene-LPEI matrix on carbon paper electrodes (0.25 cm^2 electroactive surface area) and tested in 100 mM sodium phosphate buffer. Reduction midpoint potentials (E°s) were determined from cyclic voltammetry (shown) and square wave voltammetry data, and are presented relative to a normal hydrogen electrode (NHE).

Besides directly affecting redox tuning, the incorporation of 3-^{19}F-Tyr could perturb proton equilibria, since fluorination lowers the pK_a of Tyr from 9.9 to 8.4 [166]. This is not expected to alter the energetics of converting OX to NSQ, which is an electrostatically neutral event wherein the coupled proton is acquired by the flavin itself. Ionization of Y90 at high pH could affect $E°_{NSQ/HQ}$, but Figure 4.7 shows essentially superimposable pH dependence for non-fluorinated *Rp9*Fld, the ^{19}F-Y version and ^{19}F-Y Y90W-*Rp9*Fld. Thus, our electrochemical measurements indicate no significant perturbation of Fld reactivity due to incorporation of our 3-^{19}F-Tyr probes.

4.2.6 A basis for dynamics in the long loop's sequence?

Given that the heterogeneity evident in the crystal structure and the conformational dynamics observed by NMR are in the same region of the structure, surrounding the flavin and flanked by the long loop (LL), we asked whether the amino acid sequence of the LL emphasizes small residues with fewer structural constraints or residues less favored in persistent secondary structure. Moreover given that LC Flds are proposed to employ the LL in binding to partners [143], this would subject the LL to multiple evolutionary constraints in organisms such as *R. palustris* that must use a single Fld for all partners. Therefore we compared amino acid conservation in the LL with conservation of amino acids elsewhere in the sequence of LC-Fld, and looked for conserved flexible residues .

Whereas loops have a tendency to be more variable than core secondary structure [167], we found that the consensus sequence for LL conserved several small flexible amino acids: G, A, S and P which together occurred 6-7 times (3-4, 1, 1

and 1 times, respectively) in this 21-amino acid stretch, yielding 29-33% prevalence compared to an average prevalence of 23% in the entire sequence, and an expectation of 28% for proteins in general.[168, 169] The elevated abundance of G, S, A and P in the LL is consistent with enrichment expected in loops [170]. In contrast, aromatics occurred 4 times in the LL, (2 F, 1 Y and 1 W, 18% prevalence) vs. a prevalence of 8.2 % in the complete sequence, adjusted to an expectation of 6.1% for loops (Figure 4.8A).

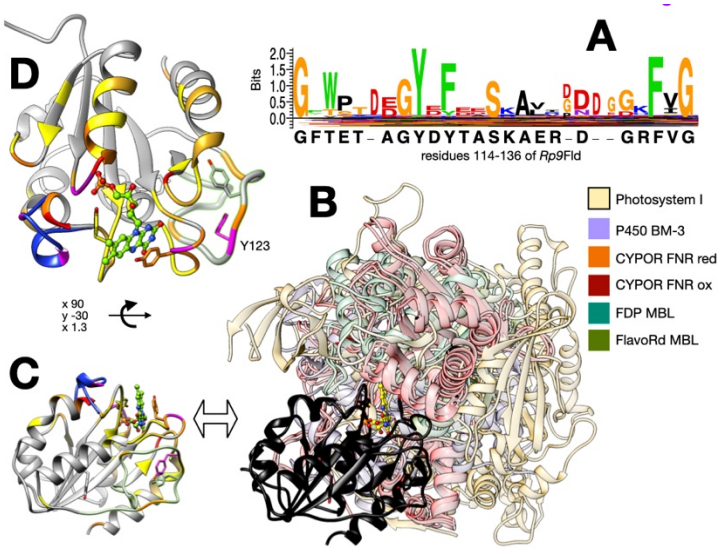

Figure 4:8 Sequence conservation in the LL and Fld residues that interact with partner domains.

A: sequence Logo for the long loop of *Rp9*Fld showing the relative abundances of the different amino acids found at each position, and comparing the consensus identity with the identify in *Rp9*Fld's sequence. Panel B. Overlay of complexes of Fld with a partner protein or partner domain. All complexes were overlayed on *Rp9*Fld using their Fld domain only, however only the other domains, with which it interacts, are shown. These are: wheat = complex with photosystem I based on cryoEM (6KIF.pdb [144]); mauve = cytochrome P450 domain of cytochrome P450-BM3 (1BVY.pdb [171]); red = the Fd $NADP^+$ reductase (FNR) domain of closed conformation NADPH cytochrome p450 OxidoReductase (CYPOR) with $NADP^+$ bound (5URE.pdb [172]); burgundy = oxidized CYPOR FNR domain (1AMO.pdb [76]); blue-green = FprA, nitric oxide reductase's metallo-β-lactamase (MBL) domain (1YCH [154]); olive green = MBL domain of flavorubredoxin, another flavodiiron protein (4D02 [153]). The *Rp9*Fld template is shown in black and its FMN is in yellow. Panel C applies color-coding to the same orientation of Fld to identify residues found to interact with a partner protein in any of the complexes or based on NMR chemical shift changes upon exposure to partner in solution [119, 131]. Grey = non-interacting, yellow, amber, orange, red, pink and purple indicated increasing frequency of interaction, see methods for details. A loop specific to Fld associated with nitrogenase is shown in blue. Panel D shows a second orientation of *Rp9*Fld, from the partners' perspective, facing towards the protruding flavin. In C and D, *Rp9*Fld's flavin is in green, the LL is highlighted in lime green and thickened in D to facilitate reference to the Logo in panel A.

4.2.7 Long loop interactions with Fld's partners.

The LL contains several conserved aromatic residues and an anionic motif (Figure 7A). The Y121 we identified with an exceptionally broad resonance is almost fully conserved (96.0% as Y) and the next most abundant residue at that position is F (3.1%) followed by H (0.9%) demonstrating conservation of an aromatic group smaller than Trp. The position of Y123 is most frequently held by F (77.2%) with Y next (12.9%) and then His (7.6%). Y123 is followed by a conserved pattern of 3-5 D/E/S residues that forms an anionic/polar stretch. These add to a prominent halo of negative charge surrounding the site at which the edge of the flavin protrudes into solvent between Y53 and Y90 .Thus the LL contributes to an extensive anionic interaction surface mimicking Fd [75, 144, 173, 174], and presumably aiding in binding to Fd's partners.

For a counter-test of LL contributions to Fld's affinity and selectivity for partner proteins, we extended our sequence comparison to Flds that do *not* need to identify and capture their partners: Fld sequences that occur within a larger protein (consisting of a Fld domain fused to a second domain). Selectivity is essentially assured by the linker uniting the two domains, and the local concentration of each domain is so high from the perspective of the other that a relatively weak interaction can suffice to bring the two together for long enough for electron transfer. Therefore one expects that evolutionary pressure to optimize selectivity and affinity by other means would be diminished, and any residues in LL selected for these functions in independent Fld would be less strongly conserved in Fld domains.

We analyzed 147 members of the larger protein family identified with *Rp9*Fld's amino acid sequence: IPR008254. Co-alignment of the sequences of their Fld domains with the above set of 222 LC-Flds and 103 SC-Flds reveals that the Fld domains possessed by members of IPR008254 all lack a LL. This is negative information, but nonetheless suggests that the selectable benefits of a LL are not needed by attached Fld domains. This is consistent with a role for LL in partner binding and selection, two challenges unique to independent Flds.

Since LL appears to occur only in independent Fld, whose complexes with partners tend to be short-lived [174, 175], there is relatively little structural information on the placement of LL within complexes. However the Fld residues in contact with partner were identified by a cryo EM structure of a cryo-trapped complex of LC flavodoxin with photosystem I [144], and were implied by NMR resonance shifts upon weak association with nitrogenase in solution for *A. chroococcum* Fld ('NifF')

[119], and more extensive resonance shifts observed for *E. coli* Fld's associations with FNR or methionine synthase [131]. Figure 7B combines these with additional insights from interactions formed by Fld domains in modular proteins, for which crystal structures have been solved. Based on eight studies, each Fld residue was color-coded by the number of complexes in which it interacts with partner. Panels 7C and D show that the partner proteins bind to a shared surface of the Fld protein, with the participant residues contributed mostly by loops, but also subtended by helices 1 and 5. These residues surround the site of flavin exposure (Figure 7D).

4.3. Discussion

4.3.1 Fld has dynamic loops, near the flavin.

NMR and crystallography show that the LC Fld of *R. palustris* CGA009 explores considerable conformational variation in the region that surrounds the isoalloxazine ring. Residues in these surface loops adopt distinct conformations in the two protomers in our crystals' asymmetric unit, they are unresolved in the B protomer, and ordering was dependent on temperature, consistent with dynamic disorder. Although residues in the loop containing Y53 and conserved G54 and G56 are known to accommodate redox-coupled rotation of the backbone at G54 that reverses a H-bond with the flavin N5 [176], our X-ray data collected at 100 K reveal heterogeneity in the LL and the loop containing Y90 as well. The divergence between the two protomers in our unit cell document that these loops can adapt to different partner protein interfaces. This has previously only been observed for apo-FMN Flds [177]. However in our case the FMN density is well resolved.

NMR concurs, providing direct evidence of conformational exchange affecting Tyr side chains in the same loops. We also confirm the utility of ^{19}F NMR in conjunction with incorporation of 3-^{19}F-Tyr for characterizing these dynamics without disrupting the functional redox reactions of *Rp9*Fld. Having assigned the resonances of all 5 Tyrs, and demonstrated retention of LL signals even in the NSQ state, we are now equipped to use ^{19}F NMR to monitor side chains interactions with partner proteins. The LL's conserved anionic motif and its retention in many independent Flds but not Fld domains that are fused to their partners, suggest evolutionary pressure on the LL specifically when affinity and selectivity for a partner are needed.

4.3.2 Loop dynamics are clarified by comparison of Fld from different strains of R. palustris.

Although *Rp9*Fld is 75% identical to the *Rp8*Fld we characterized before [147], comparison of the two resolves some puzzling behaviors seen in *Rp8*Fld and provides a more nuanced view of the LL. Three of the four Tyrs shared by *Rp9*Fd and *Rp8*Fld retain qualitatively the same temperature dependencies in the two Flds, with Y90 in slow exchange but approaching intermediate at 55 °C, Y121 in intermediate exchange, and Y102 in fast exchange, consistent with Y102's surface-exposed side chain that experiences a relatively constant environment. (We rule out occupation of a single conformation because 3-F-Tyr bears a single ^{19}F that would occupy one of two positions if the Tyr ring were immobile.) The lower coalescence temperature for Y121 than for Y90 need not mean a different rate constant for conformational change, as the separation between peaks characterizing the participating conformations also governs the rate sufficient to average the

resonance frequencies. The lower temperature coalescence of Y121 more likely indicates that the slow-exchange Y121 resonances are closer together than those of Y90, and thus that the environments sampled are less different for Y121.

In *Rp8*Fld the two half-strength signals of Y90 could have reflected the two orientations of the Tyr ring. However the three signals we observe in *Rp9*Fld can no longer be explained by slow flipping of the Y90 ring alone, and indicate that Y90 dwells in at least two different well-defined environments, besides the two sampled via ring flips. This could result from repositioning of the LL since it is very near Y90 (Figure 2C), although the 90s loop also displayed backbone dynamics [178]. Given that Y90's signals approach intermediate exchange in a physiologically relevant temperature range, we speculate that transient unstacking of Y90 from the flavin may allow accelerated electron transfer, and that such unmasking of the flavin could be rendered responsive to partner binding by interactions of the LL with both the partner and the 90s loop.

In *Rp8*Fld, only one of Y90's resonances appeared sensitive to flavin paramagnetism raising big questions as to the nature of the conformation in which Y90's side chain appeared distant from the flavin [147]. *Rp9*Fld's behavior is in much simpler agreement with expectations, as formation of the semiquinone state uniformly and almost quantitatively obliterates resonances of all signals of both the two Tyrs that flank the flavin, without affecting the other Tyr resonances. Thus, *Rp9*Fld emerges as a case in which paramagnetism can be a highly selective means of identifying residues that come close to the flavin. This system will provide a promising tool for mapping out the Fld binding footprint on partner proteins [179].

The improved spectra of *Rp9*Fld also reveal a subtle but unambiguous splitting of Y53's resonance consistent with slow exchange dynamics in the 50s loop that has been documented NMR relaxation studies of backbone [116, 178] but was not seen in the ^{19}F NMR signals of side chains. This signature of multiple conformations is particularly interesting given that Y53's side chain H-bonds with the phosphate of FMN, while the backbone of the 50s loop provides redox-coupled H-bonding to the flavin N5 from the amide of residue 55 when the flavin is oxidized but to the carbonyl of G54 when the flavin is reduced [176].

Finally, we get richer insight into dynamics in the LL, thanks to presence of a second probe in the form of Y123. Whereas Y121's side chain is in intermediate exchange at 25 °C, that of Y123 appears to be in fast exchange (see regarding Y102, above). This is consistent with our crystal structures based on greater constraint of the Y121 side chain and more different environments on the two sides of Y121's ring. Based on the A chain in our structure, the side chain of Y121 is more buried than that of Y123 (side chain SASA of 8.95 Å for Y121 vs. 13.46 Å for the side chain of Y123) suggesting that side chain rotation will be slower and chemical shift averaging less efficient.

4.3.3 Interactions with partners.

FNR (Fd-NADP$^+$ reductase) uses electrons from NADPH to reduce Fd or LC Fld; reduced Fld in turn reduces methionine synthase [131]. Changes in NMR chemical shifts identified Fld residues affected by binding to each of these partners, showing that both interact via the face from whence Fld's flavin protrudes, including the

loops that surround the flavin (Figure 7B) [131]. The two partners' footprints are not identical, but they overlap considerably [131].

E. coli's LC Fld also reduces pyruvate formate lyase activating enzyme (PFL-AE), but only binds to it when FMN-replete [180]. Since acquisition of FMN rigidifies Fld's core and loops surrounding the FMN [116, 177], this indicates that partner binding requires some preformed loop structure and implicates the protruding flavin itself as a contributor.

Fld also supplies low-E° electrons to nitrogenase, and titration of *Azotobacter croococcum*'s LC Fld with V-utilizing nitrogenase's Fe protein identified three responsive residues via chemical shift changes [119]. All surround the flavin, but two fall within the loop that is absent from our crystal structure (blue in Figure 7B,C). Indeed, our MSA reveals that the longer loop noted in *A. vinelandii* and *R. capsulatus* is associated specifically with organisms that fix N_2 [151]. This supports a role for surface loops as agents of partner selectivity.

Independent Fld may be evolved to *not* form long-lived associations with partners precisely because slow detachment from a source partner would also diminish the rate at which acquired electrons could be delivered to their destination partner. Besides the cryotrapping employed as part of cryoEM [144], connections between Fld domains in modular proteins serve as natural cross-links enabling crystal structures of complexes, These all involve SC Flds [116], but nevertheless reveal that much of the surface on which the LL lies is not used to interface with partners. By masking this surface, the LL could 'steer' partners to productive surfaces instead, and/or it could enlarge the binding surface. To imagine how the LL might augment

a productive interaction or impede an unproductive one, we looked at interactions of Fld domains in their larger multidomain proteins.

A prevalent combination is fusion of a domain related to FNR (containing FAD) behind a Fld domain (containing FMN) in the so-called diflavin reductases [181, 182] exemplified by the NADPH:cytochrome p450 oxido reductases (CYPORs) [183]. CYPOR's "closed" conformation brings the FNR and Fld domains together to favor electron transfer from FAD to FMN (red and burgundy in Figures 4.8B), however an "open" form is required to enable FMN to reduce the P450 heme [184]. In the self-sufficient P450-BM3, the cytochrome p450 domain is appended to the N-terminus of the CYPOR to produce a modular enzyme with the domain sequence n-P450-Fld-FNR-c [185]. Removal of the FNR domain enabled determination of a crystal structure modeling Fld-P450 interactions (purple in Figure 7B, [185]). (Note that in dimers of P450-BM3, the Fld domain of one monomer may reduce the cytochrome of the other [186].) Similarly, in the paralogous human methionine synthase reductase, the Fld domain reduces a corrin-containing methionine synthase [131]. These three representative structures position the flavin of the FNR domain or the heme of the P450 domain close to the flavin of Fld.

Flavodiiron proteins (FDPs) are a different modular combination that fuse a Fld domain to a preceding metallo-β-lactamase-like (MBL) domain, where a diiron site metabolizes O_2 or NO [187, 188]. The interface between the Fld and MBL domains is seen to occur between protomers of the dimer, associating the Fld domain of one protomer with the MBL domain of the other. Figure 7B also includes the complex of Fld with photosystem I (6KIF.pdb). All are overlayed by superimposing their

bound Fld domains, and Figure 7B, C and D show that the different partners all bind to the face of Fld from whence the flavin protrudes. Additionally, the partners all present their own redox cofactor at the surface or only slightly buried, so that the donor and acceptor cofactors are brought together.

Regarding the LL, it is striking that only the early portion of the LL, bearing Y123 and Y1221, interacts with partners. That portion is near the flavin and increases the size of the binding interface, but does not seem to justify conservation of such a large loop. We speculate that the later portion of the LL may steer Fld into productive orientations by masking a large portion of the Fld surface whose association with partner would place Fld's flavin further away from partner cofactors. Indeed, the LL would impede formation of the dimers observed in the SC Flds of *M. acetovorans* and *D. giga*s (4HEQ and 5WID). Meanwhile, the early portion of the LL contributes to the anionic halo believed to attract Fld to partners [75, 144, 174].

Given the conservation of aromatic residues in the LL, we wondered if they could play a role in electron transfer, in effect providing a conductive interface [174, 180, 189]. If so, they might be enriched in the regions that interface with partners. The 56 residues found to interact with partners includes 5 aromatics, compared to expectation of 4.6 based on the prevalence of Y, F and W overall [169] and 3.4 after correction for their expected prevalence in loops [170]. Thus, the level of conservation we observe on these surface loops supports a role for aromaticity, which we propose to be mediation of electron transfer. By extension, the aromatics at positions 121 and 123 of the LL may do the same. They are folded inward in our

crystal structure, but we propose that the dynamics we have documented enables them to present themselves in the interface upon partner binding.

4.4. Concluding remarks.

Although the LL is structured in the A protomer of our crystal structure, our NMR evidence indicates that it and the two loops containing Y53 and Y90 could adopt different conformations as part of binding with partner protein. All three are exposed on surfaces that interface with partners of Fld, and Y53 and Y90 protect the flavin from solvent in the absence of partner but could unmask it upon partner binding. Indeed, the LL has two conserved motifs that argue for a contribution to function. We speculate that the LL contributes to binding affinity by enlarging the binding interface, while also buttressing Y90. Interactions with partner could reposition the LL, effectively freeing Y90 to interact with the partner and unmasking the flavin. With the ^{19}F-Tyr resonances assigned, we look forward to titrating *Rp9*Fld with its candidate partners, to learn whether Y121 and Y123 are indeed responsive, interacting with partners and possibly even enhancing electron transfer.

4.5. Methods:

4.5.1 Purification of Rp9Fld and incorporation of ^{19}F-Tyr.

The expression system employed a synthetic gene for *Rp*9Fld (GenScript) using codons optimized for *E. coli* and the amino acid sequence of flavodoxin from *R. palustris* strain CGA009 (*Rpa9*Fld). It was provided in the commercial vector

pET28b(+). Primers encoding four of the five desired amino acid substitutions were designed using NEB Base Changer software and ordered from Eurofinn Genomics (Louisville, KY). They were employed along with the WT-Fld bearing parent plasmid as template in polymerase chain reactions (PCR) utilizing Q5 High Fidelity DNA polymerase from New England Biolabs. The fifth mutant gene (encoding the Y90W substitution) was synthesized and cloned into pET28b(+) by GenScript. The WT and mutant plasmids were transformed into competent NiCo21(DE3) *E. coli* cells for protein production (New England Biolabs C2529H) and NEB® 10-beta competent *E. coli* (High Efficiency, Catalog# C3019H) for plasmid propagation.

For protein expression, an initial 10 mL overnight culture of plasmid-bearing *E. coli* strain Nico21DE3 was incubated at 37 °C, and used to inoculate 1 L of Terrific Broth (TB) medium containing 50 μg/mL, and 0.09 μM kanamycin for selection. The culture was agitated at 250 RPM at 37 °C until the optical density (O.D.) reached the range of 0.6-0.8. The temperature was then lowered to 21°C, and protein expression was induced by the addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside) to 0.5 mM. Following 18 hours of post-induction incubation, the cells were harvested by centrifugation, resuspended in phosphate buffered saline pH (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM K_2PO_4 and 1.8 mM KPO_4) and pooled before pelleting again for preservation at -80 °C until use.

To obtain purified protein, frozen cells were thawed and suspended in a binding buffer consisting of 20 mM Tris-HCl, 300 mM KCl, and 10 mM imidazole at pH 7.4, which was augmented with FMN monosodium salt (1mM, Cayman chemicals, catalog #18167), Benzonase nuclease (1 μL from Millipore sigma # 71205-3) and

recombinant lysozyme (1 µL, Millipore sigma # : 71110-4). Cell lysis was achieved through sonication, employing a pulse mode with 5 seconds on and 30 seconds off for a total of 30 cycles, while maintaining a cold environment with ice. The resulting cell lysate was loaded onto a 5 mL His-Trap column (Cytiva, Histrap HP, catalog: 17524802), pre-equilibrated with 20 mM Tris and 300 mM KCl at pH 7.5, using the ÄKTA GO FPLC system. The column was washed with 20 column volumes (CV) of the binding buffer supplemented with 20 mM imidazole. Finally, the target protein was eluted from the column using 250 mM imidazole in 20 mM Tris-HCl, 300 mM KCl, plus 250 mM imidazole, pH 7.4. The eluted protein was concentrated and subjected to buffer exchange by gel filtration over a DG-10 column equilibrated with 20 mM bis-tris propane buffer with 200 mM KCl at pH 7.5. The His_6 tag was not removed. A typical yield from 1 L of culture was 48 mg protein from a 1 L culture.

Fluorination for NMR experiments involved use of initial growth in TB, before transfer of the cells into defined medium [190] withholding Tyr and providing Phe, Trp and ^{19}F-Tyr, as described by Varner et al [147]. Preparation of fluorinated protein was also as per Varner et al. and above, except that sonication was used to lyse cells. Thus we obtained a purified protein yield of 22 mg of ^{19}F-WT-Fld from a 1 L initial culture of TB medium.

Flavin content and protein concentration were determined as per literature [147] and using the Pierce™ 660nm method (# 22660) *vs*. bovine serum albumin standard, respectively [191]. The flavin stoichiometry was determined to be 1.15 ±0.3 FMN per *Rp9*Fld monomer, and optical spectra of the WT, the fluorinated

WT, and variants employed for signal assignment were recorded. To optimize purity of the protein, for crystallography, we employed gel filtration chromatography on a Cytiva Superdex 200 Increase 10/300 GL column (column pressure: 2 Mpa, flow rate: 0.5 mL/min, volume of each elution fraction: 0.5 mL). Fractions corresponding to the protein peak were pooled and concentrated, prior to use in crystallization experiments. Gel filtration chromatography was used as above, in conjunction with a mass-calibration (Biorad gel filtration standard, # 1511901) to determine the effective mass of *Rp9*Fld. A single symmetric elution band was obtained yielding a value of 17.5 kDa ±5% for *Rp9*Fld concentrations up to 1 mM. Thus *Rp9*Fld is >95% monomeric in solution at the 0.55 mM concentration used for SEC.

4.5.2 Crystallization of Rp9Fld.

*Rp9*Fld was crystallized using the hanging drop vapor diffusion method using 8 - 12 mg/mL protein stock solution in 20 mM bis-tris propane buffer, pH 7.5, with 200 mM KCl. Preliminary success was achieved using the Hampton screen (HR2-139) within 96-well sitting vapor diffusion setups. We then refined the concentrations of magnesium formate (0.2M) and PEG 3350 (20%), along with the addition of an additive (Hampton Research HR2-464), and varied the temperature (4-22°C). The mother liquor contained 0.3-0.6M magnesium formate with 15-20% polyethylene glycol (PEG) 3350. Flower-shaped crystal clusters emerged after 2 weeks at 14 °C. These were utilized to generate 'seed-crystals' using Hampton Research's Seed Bead Kit (HR2-320). Diffraction quality and crystal morphology were optimized by iterative cycles of micro seeding and screening of crystallization

conditions. Several morphologies of crystals were obtained, and checked to identify crystals with the best diffraction quality in combination with the lowest mosaicity. This process resulted in the final use of 0.4 M magnesium formate pH 7.5 and 19% PEG 3350, wherein we note that the best crystal morphology was obtained upon supplementation with cadmium chloride at 20 mM. Final crystallization drops contained 1 μL of protein solution and 0.8 μL of 2x crystallization solution with 0.2 μL of micro seed solution (1:1000 dilution).

When data were collected in capillaries at room temperature, radiation damage was observed in the form of decreased I/σ(I) and elevation of R_{sym}. We were nevertheless able to collect a room temperature data set including diffraction to 2.9Å resolution. To obtain a deposition-quality structure, single plate like crystals were exposed briefly to 35% W/V PEG 3350 in mother liquor solution to provide cryoprotection, before exposure to a stream of liquid N_2. Data were then collected at 100 K, yielding diffraction to 2.1 Å.

4.5.3 X-ray Data collection and structure refinement.

X-ray data were collected by using a Rigaku HighFlux HomeLab source equipped with a MicroMax-007 HF X-ray generator and Osmic VariMax optics. Diffraction images were captured using an EIGER R 4M hybrid photon-counting detector. The data were integrated using the CrysAlisPro software suite (Rigaku, The Woodlands, Texas, USA), and then reduced and scaled using *AIMLESS* [192] from the *CCP4 suite* [193]. Initial phases for each structure were obtained by molecular replacement using 1CZN.pdb (flavodoxin from *Anacystis nidulans*, 43% identical, 98% coverage) as a starting model in *Phaser* [194] in the *Phenix* suite [195]. The

model was iteratively refined including manual model building using the *Phenix refinement* module [196] and the *Coot* molecular graphic program [197], respectively. Solvent atoms were added at last step of model building and refinement cycle. These were added only when reasonable 2Fo-Fc density and at least one hydrogen bond to a donor was present. Occupancies and thermal parameters of the solvent were refined in alternating cycles.

The final model was checked and validated with *Molprobity* [198]. Buried surface area and solvent accessible area calculations were done using PISA software [152] and the EPPIC (Evolutionary Protein-Protein Interface Classifier) server [199]. High resolution images were generated using Pymol (http://www.pymol.org, [149]). Statistics of diffraction data processing and the converged refinement statistics for datasets along with R, R_{free}, and other validation measures are given in Table 1, and the final structure coordinates and structure factor amplitudes were deposited in the RCSB Protein Data Bank with the accession code 8SNZ. For completeness, the structure obtained from data collected at room temperature was also deposited, as 8V2Y.

4.5.4. NMR spectroscopy.

NMR samples were prepared in a 20 mM bis-tris propane buffer supplemented with 200 mM KCl at a pH of 7.5. 10-15% v/v D_2O was present to provide a frequency lock signal for instrumental stability. To permit quantitative formation of the neutral semiquinone state of *Rp9*Fld, a 2 mM Fld sample was equilibrated in an inert atmosphere of N_2 inside a glove box (Bell Technologies, U.K.). After stepwise reduction with dithionite to produce the neutral blue semiquinone (inset, Figure 4),

the sample was carefully transferred to an NMR tube, which was then sealed with a septum stopper before removal from the glove box. Immediately upon removal the septum was secured with parafilm and the sample was transferred to the spectrometer. Data collection was complete in 2 hours after which the sample remained blue. Only several days later was a color change to green/yellow visible at the top of the tube. Thus, the sample remained fully in the NSQ state during NMR data collection.

^{19}F NMR spectra were recorded using a Varian Unity/Inova 400 MHz spectrometer, at a constant room temperature of 25 °C or a Bruker NEO 400 MHz spectrometer with SMART probe. 30 ppm wide scans centered at -134.7 ppm were collected following a 10 μs 90 ° excitation pulse with data collected for 0.27 s followed by a 1.5 s delay for recovery between scans. Spectra shown represent averaging of 1024 - 4096 scans, for samples of 0.4 - 2 mM *Rp9*Fld. The above recovery delay was based on our measured T_1 relaxation times, of which the longest was 620 ms. Data were processed using VnmrJ, Topspin and Mnova software, depending on the data source and the analysis to follow. Baseline artifacts from distortion of early time points were eliminated by removing the first 5 data points and replacing them with backward linear prediction based on 256 complex points that followed, modelling the FID in terms of 32 coefficients. Gaussian line broadening dropping to 50% at 100 ms was employed to optimize signal to noise without broadening out reproducible splitting (See figure captions). Baseline correction was repeated multiple times and resulting peak shapes and integrals compared to identify values that were robust to processing parameters. Moreover the spectra collected as a

function of temperature were also simulated by Lorentzian deconvolutions (VnmrJ) to yield a self-consistent set of peak areas, positions and widths, for comparison with those measured directly from the spectra.

4.5.5. Bioelectrode preparation and electrochemical experiments.

Ethylene glycol diglycidyl ether (EGDGE) was purchased from Polysciences Inc. All other chemicals were purchased from Sigma Aldrich and used as received without further purification. All water used in bioelectrochemical experiments was filtered using an Ultrapure MilliQ system. Pyrene-LPEI was previously synthetized using the reported procedure [157]. Stock solutions of protein (WT *Rp9*Fld Fld, ^{19}F-Y WT *Rp9*Fld and ^{19}F-Y Y90W-*Rp9*Fld) 10 mg mL^{-1}, pyrene-LPEI 10 mg mL^{-1} and 13 vol% ethyleneglycol diglycidyl ether (EDGDE) were prepared in water. The crosslinked hydrogel was prepared by gently mixing 70 µL of pyrene-LPEI, 30 µL of protein and finally 4.5 µL of EDGDE. Toray carbon paper electrodes were cut into 2 cm X 0.5 cm stripes and dipped into melted paraffin wax that was successively cooled down and solidified to delimit an electroactive surface area of 0.25 cm^2 (0.5 cm X 0.5 cm) at one end. A total volume of 30 µL of the polymeric mixture containing each protein were drop-coated on the exposed electroactive area of the electrodes, 3 successive depositions of 10 µL aliquots of the mixture were carried out followed by 20-30 min of drying to allow the crosslinking process to occur. Control experiments containing pyrene-LPEI without protein were made by mixing the same amount of polymer and EDGDE with water and depositing the same amount of mixture volume on the surface of the electrodes. After the last deposition, the electrodes were cured under anoxic conditions overnight to allow

stabilization of the polymer matrix. Preparation of the electrodes as described as well as electrochemical measurements have been carried out inside of a glove box under an inert atmosphere of Ar/H_2 (2.8-3.2%) with less than 3 ppm O_2 at room temperature. Bioelectrodes have been tested by cyclic voltammetry (CV) and square wave voltammetry (SWV) with a CH Instruments 1033 potentiostat. CVs were performed at 5 mV s^{-1}. A three-electrode set up was used with a SCE reference electrode, a platinum mesh counter electrode and Toray carbon paper working electrodes. The waxed end was used as an electrochemical connection point for an alligator clip of stainless-steel wire. Electrodes were tested in 5 mL of sodium phosphate buffer (100 mM) at different pHs. Potentials have been recalculated with respect to a normal hydrogen electrode (NHE) reference electrode. Data were elaborated with Origin 2023b software.

4.5.6 Amino acid sequence analysis.

To compare the degree of conservation in the LL with overall conservation in Fld, we constructed a multiple sequence alignment (MSA) of long-chain (LC) Fld sequences from uniref50 and compared it with an analogous MSA of uniref50's short-chain (SC) Fld sequences, because only the former will contain the LL. LC-Fld sequences were gathered using EFI [200, 201] and an alignment score threshold (AST) of 35 based on the interpro family IPR010086, encompassing *Rp9*Fld and *Nif*F from *R. capsulatus* 2WC1.pdb [151]. SC-Fld sequences were similarly obtained based on IPR010087, the interpro family containing the Flds of *D. vulgaris* (3FX2, 1J8Q), *Megasphaera elsdenii* (2FZ5.pdb [202]) and *Lactobacillus* (3EDO.pdb). Sequences were gathered from the Ref50 database, where sequences

are grouped into nodes uniting sequences at least 50% identical over at least 80% of the longest sequence in the node (the so-called 'seed sequence' of that node). Nodes from the ref50 protein database were used to limit the number of sequences considered without altering the structure of the relationships among them [201, 203]. In each of LC- and SC-Flds, multiple sequence alignments (MSAs) were curated to eliminate a few sequences that were incomplete or contained an insert affecting 1% or fewer of the sequences. The separately aligned LC- and SC-Fld sequence sets were combined and coaligned to localize the LL and highlight shared motifs. After culling 3 more LC-Fld sequences based on this expanded perspective, the LC-Fld sequence set consisted of 222 sequences (trimmed from 237) and the set of SC-Flds comprised 103 sequences (trimmed from 106). After co-alignment to obtain a single numbering system applicable to both, the two groups were separated and the quality of residue conservation within each was compared, without further modification of the MSAs .

Sequence Logos were generated by Seq2Logo [204] using the p-weighted Kullback-Leibler method without clustering, based on our alignment of 222 LC Flds.

To learn what aspects of Fld could be most important for orienting Fld vs. selection or affinity for appropriate partners, we compared the sequences of independent Flds with Fld domains that are fused to a partner and thus assured of a locally high concentration. We used the Pfam family PF00258, which subsumes the two interpro families listed above and additionally incorporates multidomain enzymes in which a domain has high homology with Fld. Pfam family PF00258 is

represented by 274 reviewed sequences (75 k total sequences as of Aug 2023 which have been incorporated into IPR008254 representing 444 reviewed sequences, 133 k total as of Aug 2023). We selected enzymes containing a Fld domain plus an additional domain using sequence length cut-offs of 350 to 450 with an AST of 104. A Multiple Sequence Alignment of these sequences along with those of LC and SC Flds caused the larger modular proteins to all align on the basis of their Fld domains. For the modular proteins, we deleted sequence stretches representing the other domains and advanced using the portion of each sequence that represents the Fld domain. Thus we collected 156 Fld domain sequences which we culled to 147 ranging from 144 to 202 residues long. To permit direct comparison with available crystal structures, we augmented this set with the Fld domain sequences corresponding to pdb files 1BVY, 1E5D, 4H2D, 1YKG, and 1AMO.

To map residues involved in partner binding onto the structure of *Rp9*Fld, residues were colored according to the frequency with which they were reported to interact with partner protein or domain, in structures of complexes with partners, where possible. Color coding indicates the frequency with which each residue interacts with partner, based on crystal structures of complexes listed in the caption, changes in NMR chemical shifts upon titration with partner [119, 131], or cryo EM structure of complex [144]. For each instance in which a residue interacted with partner (distance of 3.5 Å or less, or listed as interacting by author, the residue acquired a 'point' since there were 5 structures of SC Fld but only 3 of LC Fld, instances observed for LL residues were possible in only ≈ 1/3 of cases, so they were accorded 3 points each. The sum of points obtained by each residue then determined

its color as follows: 0=grey, 1=yellow, 2=gold**,** 3=orange, 4=red, 5 or 6 = pink, 7=purple. Information was only available for one diazotroph so responsive residues in the loop unique to that case were shown in pink and that loop is in blue.

CHAPTER 5. FUTURE PROSPECTS

FBEB represents a novel phenomenon in the realm of energy conservation, illustrating nature's incredible ability to adapt and evolve to enable chemical reactions that traditional thermodynamic laws deem impossible. Studying FBEB is crucial for comprehending how nature orchestrates these reactions at the molecular level. There are two critical aspects to FBEB: 1) the electron transfer involving the donor, FAD, and the acceptor, and 2) the significant large conformational changes that facilitate this electron transfer. There is substantial ongoing research aimed at understanding the differing reactivities of Bf-FAD and Et-FAD and the role of adjacent residues, but there is also a need to investigate the dynamics of proteins since both processes are interconnected.

Our main objectives behind using SANS were to investigate the structure in solution of the ETF, compare it with existing structures, and check if flavin reduction causes conformational change in solution. We addressed our main goals and demonstrated that extended conformations of the ETF exist in solution with significant populations along with compact conformations. These extended structures are very different from any of the published structures, which are similar in conformation to the compact ones. The presence of extended conformations demonstrates the flexibility of the head domain, which is important for bringing the two ETF flavins into close contact as well as interactions with partner proteins. We hope that our results will bring more attention to understanding the conformation behavior of ETF and its significance.

Another important feature of the research is the technical advancements that can be useful to study proteins with similar folds or functions. New collaborations and discussions can further help in assessing more complementary techniques to support the SANS data. Understanding conformation changes in ETF can help in addressing conformational changes in enzymes such as nitric oxide synthase, methionine synthase reductase, NADP+–pyruvate oxidoreductase, cytoplasmic

NR1 protein, and NADPH-sulfite reductase, which exhibit large-scale conformational change.

Our research has now opened scope for addressing more questions regarding conformation changes in ETFs. Extended conformations have not been reported for ETFs by any other research groups. It is worth investigating the presence of extended structures by other techniques to better understand their functional significance.

It would be interesting to conduct SANS on other Bf-ETFs.

Exploring biophysical techniques to support the SANS can be very helpful. Mass spectrometry in combination with hydrogen/deuterium exchange (HDX) can be used to probe the dynamics and structure of proteins[205, 206]. HDX-MS measures the rate of exchange of amide hydrogens with deuterium in the protein backbone, which can be indicative of conformational changes. If a region of the protein becomes more exposed to solvent due to a conformational change, the deuterium uptake in that region will increase, indicating increased solvent accessibility. Carr-Purcell Meiboom-Gill (CPMG) relaxation dispersion (RD) experiments can be used to characterize conformational equilibria occurring on the μs-ms timescale by solution NMR spectroscopy[207, 208]. The CPMG sequence is designed to refocus dephasing caused by inhomogeneities in the magnetic field, thus extending the coherence lifetime of nuclear spins. It is particularly effective for observing relaxation processes and exchange phenomena between different conformational states of a protein.

To fully utilize the latest advancements in SANS, SEC-SANS experiments can provide the best quality data. Coupling SEC and SANS together will result in real-time data collection and will provide the best quality sample since monodispersity can be a problem with some other ETFs. ORNL has recently acquired this technique, which would be very helpful to the scattering community. UV-Vis spectral changes are important during the ETF catalytic cycle undergoing reduction. If SANS can be collected in different oxidation states of ETF during a sequential reduction, that can tell us if there is a conformational change during oxidation states of the Bf-FAD and the Et-FAD.

The pH of the solution can also have an impact on the conformation of a protein. It would be worth conducting SANS experiments at different pH values. Changes in the charge state of amino acids can alter electrostatic interactions within a protein. For instance, if two nearby residues in a protein are negatively charged at high pH and repel each other, lowering the pH might protonate one or both residues, reducing repulsion and potentially leading to a different conformation.

Our SANS capabilities are maturing into practice, and more experiments are planned in the future by the Miller group. To improve the clarity of conformational dynamics in ETF, EtfB can be partially deuterated. This will effectively isolate the scattering signal from the EtfA subunit, mainly consisting of the head domain. This method is poised to offer deeper insights into the mobility of the head domain compared to a fully hydrogenated complex, although it faces challenges like maintaining the solubility of individual subunits when isolated.

Further investigations using SANS on the ETF-BCD complex under varying conditions, such as deuteration of ETF in lieu of BCD, promise to reveal intricate conformational dynamics distinct from those observed when BCD is deuterated. Such studies would not only detail the conformational shifts within the ETF-BCD complex but also measure the proportion of unbound ETF in solution at a certain percent of deuterated buffer. An exciting extension of this research could involve performing experiments at different ETF to d-BCD ratios to study the equilibrium dynamics and conformational adjustments within the complex. For a thorough analysis of these conformational changes in solution, devoid of any influences from crystallographic packing, comparing uncomplexed ETF to its association with BCD would provide valuable insights. The methodology for this research includes the synthesis of deuterated BCD, its isotopic characterization through mass spectrometry, and optimizing buffer compositions to minimize scattering. Systematically gathering SANS data across a spectrum of BCD to ETF ratios will enable a comprehensive fitting of the data, allowing for the extraction of complex dissociation constants and the identification of distinct scattering profiles for each species in the mix. It is expected that this approach will uncover modified intra-ETF scattering distances that indicate an open conformation, free from

contributions of inter-ETF scattering. Additionally, exploring the complex in different oxidation states might yield further insights, as shifts in these states could induce structural changes. This multifaceted approach not only aims to enhance our understanding of the structural dynamics of ETF but also paves the way for broader applications of small angle scattering techniques in the field of biomolecular research.

Another important future direction is to use SANS for the ETF and Fld complex. Our primary finding was that ETF and Fld do not form a long-lived complex, which suggests that exploring chemical methods to achieve a stable chemical linkage between ETF and Fld could be beneficial. In early attempts to form complexes between ETF and Fld, we explored crosslinking (XL) using bis(sulfosuccinimidyl)suberate (BS3) to form crosslinks between naturally occurring surface amine groups, and bismaleimidohexane (BMH) for crosslinking between surface SH groups modified by the placement of Cys residues.

In our first strategy, we worked with *Afe*Fld that had a TEV-cleavable His6 tag, which was removed after purification. A His6 tag was retained on the ETF to capture complexes formed with BS3. The crosslink reaction was carried out with a 1:2 molar ratio of ETF to Fld. After the reaction, the mixture was purified again using His6 tag affinity purification. It was expected that Fld with the His-tag removed would not bind to the resin and unbound Fld would pass in the flow-through. SEC was performed after elution. The SEC showed three peaks. A Fld peak was seen in the SEC profile, which indicated not a full cleavage of the His6 tag. Nonetheless, AUC was performed on the first two peaks with higher molecular weights expecting to see the complex formation by the presence of a species with a molecular weight ≈ 83 kDa (15 kDa Fld + 68 kDa ETF). Neither of the AUC peaks contained any species in the expected molecular weight range. Despite these efforts, no Fld was co-captured, indicating the absence of a stable complex. This method requires further refinement and optimization of conditions to achieve the desired outcomes, with one main limitation being that BS3 can also induce crosslinking between ETF molecules, resulting in unintended byproducts.

For strategy (2), we began by engineering variants of *Afe*ETF and *Afe*Fld lacking exposed Cys side chains. For *Afe*ETF, C44.A, 93.A, 202.A, 75.A, 73.B were all mutated to Ala. For *Afe*Fld, we constructured C132A. Thus we created *Afe*ETF and *Afe*Fld variants designed to not display sulfhydryl groups on the surface. The potential complexe was identified through ClusPro 2.0 by Dr. Anne-Frances Miller. Based on the model structure (Figure 5.1) , we considered a possible combination of amino acid residue pairs on ETF and Fld that are in close proximity to each other and can form a crosslink. We selected two pairs for complex formation testing: 1) K225C on ETF and A123C on Fld, 2) D122C on ETF and W93C on Fld. All mutants were first tested for expression and purification. SEC was performed for each mutant followed by SDS-PAGE to check for purity. After purifying individual proteins, a crosslink reaction was performed for each pair. The molar concentration of Fld was twice that of ETF for each reaction. The reaction mixture was further subjected to SEC to isolate the complex. Major peaks from SEC were analyzed via analytical ultracentrifugation (AUC). AUC data collection and analysis were conducted at the University of Tennessee by Dr. Ed Wright. There were some interesting observations from the preliminary data.

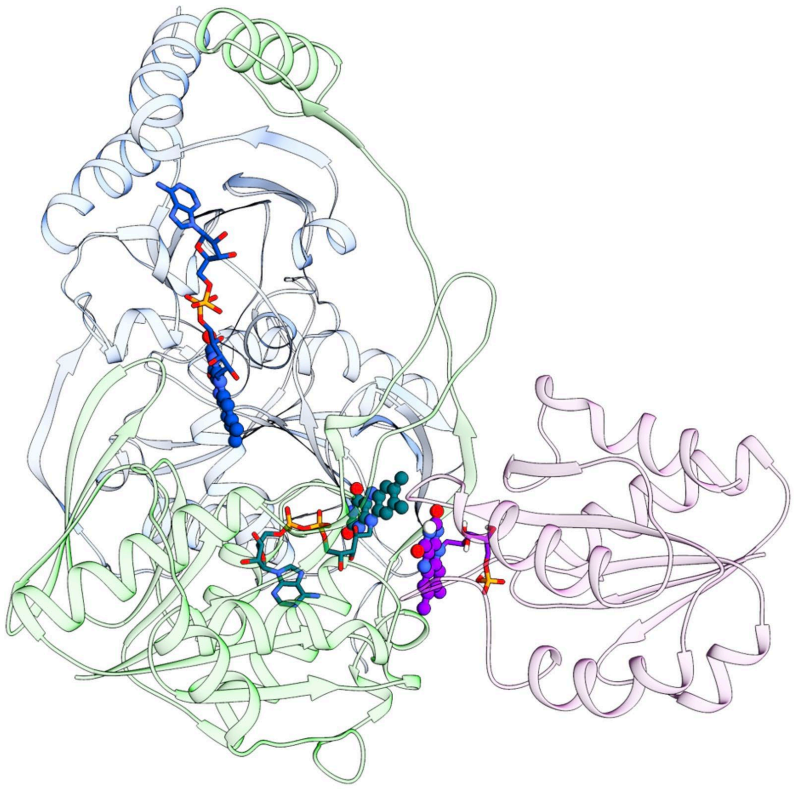

Figure 5:1Model of ETF.Fld complex.

For the reaction between K225C ETF and A123C Fld, the SEC profile indicated two well-separated peaks. Both peaks were analyzed by AUC. The first peak that eluted in SEC was expected to be the complex peak, as the complex has a higher molecular weight than the Fld, which was added in excess. AUC indicated that the majority of the species in the first peak (around 75%) had a molecular mass close to 84 kDa. The expected mass for the complex is close to this molecular weight. The first peak also had some other species (27 kDa: 5%, 58 kDa: 18%, 144 kDa: 2%). The other species observed in AUC could possibly result from protein degradation or non-specific crosslinking. Nevertheless, it is exciting to see the possibility of complex formation between these two mutants. The second peak from SEC, which we expected to be from the free Fld, was also subjected to SEC. This showed a single species with a 100 percent population and a molecular mass of 17 kDa, which is close to the Fld molecular mass. Peak 2 AUC serves as a validation of the technique.

We took a similar approach for D122C ETF and A123C Fld. The results were very different from the other pair. The SEC of this reaction mixture resulted in a single peak. When subjected to AUC, around 81% of the species had a molecular mass of 37 kDa, 16% of 17 kDa, and 3% of 150 kDa species. The AUC did not indicate the presence of an ETF & Fld complex. The numbers obtained from AUC suggest that the different species observed might be from the oligomerization of Fld itself, forming a dimer of 37 kDa and higher-order species. The preliminary results indicate that this approach might work but needs more optimization of reaction conditions and screening of more mutants. Proteomics can assist in characterizing the formation of cysteine linkages.

.

References

1. Mansoorabadi, S.O., C.J. Thibodeaux, and H.W. Liu, *The diverse roles of flavin coenzymes--nature's most versatile thespians.* J Org Chem, 2007. **72**(17): p. 6329-42.
2. Romero, E., et al., *Same Substrate, Many Reactions: Oxygen Activation in Flavoenzymes.* Chem Rev, 2018. **118**(4): p. 1742-1769.
3. Massey, V., *The chemical and biological versatility of riboflavin.* Biochem Soc Trans, 2000. **28**(4): p. 283-96.
4. Fischer, M. and A. Bacher, *Biosynthesis of vitamin B2: Structure and mechanism of riboflavin synthase.* Arch Biochem Biophys, 2008. **474**(2): p. 252-65.
5. Bacher, A., et al., *Biosynthesis of vitamin b2 (riboflavin).* Annu Rev Nutr, 2000. **20**: p. 153-67.
6. Silva, E., A.M. Edwards, and Royal Society of Chemistry (Great Britain), *Flavins : photochemistry and photobiology*. Comprehensive series in photochemistry and photobiology. 2006, Cambridge: RSC Pub. x, 328 p.
7. Macheroux, P., B. Kappes, and S.E. Ealick, *Flavogenomics--a genomic and structural view of flavin-dependent proteins.* FEBS J, 2011. **278**(15): p. 2625-34.
8. Yagi, K., et al., *Effect of hydrogen bonding on electronic spectra and reactivity of flavins.* Biochemistry, 1980. **19**(8): p. 1553-7.
9. Joosten, V. and W.J. van Berkel, *Flavoenzymes.* Curr Opin Chem Biol, 2007. **11**(2): p. 195-202.
10. Woodmansee, A.N. and J.A. Imlay, *Quantitation of intracellular free iron by electron paramagnetic resonance spectroscopy.* Methods Enzymol, 2002. **349**: p. 3-9.
11. Lienhart, W.D., V. Gudipati, and P. Macheroux, *The human flavoproteome.* Arch Biochem Biophys, 2013. **535**(2): p. 150-62.
12. Rossner, R., M. Kaeberlein, and S.F. Leiser, *Flavin-containing monooxygenases in aging and disease: Emerging roles for ancient enzymes.* J Biol Chem, 2017. **292**(27): p. 11138-11146.
13. Barile, M., et al., *Biosynthesis of flavin cofactors in man: implications in health and disease.* Curr Pharm Des, 2013. **19**(14): p. 2649-75.
14. Walsh, C.T. and T.A. Wencewicz, *Flavoenzymes: versatile catalysts in biosynthetic pathways.* Nat Prod Rep, 2013. **30**(1): p. 175-200.
15. Mayhew, S.G., *The effects of pH and semiquinone formation on the oxidation-reduction potentials of flavin mononucleotide. A reappraisal.* Eur J Biochem, 1999. **265**(2): p. 698-702.

16. Macheroux, P., *UV-visible spectroscopy as a tool to study flavoproteins.* Methods Mol Biol, 1999. **131**: p. 1-7.
17. Yang, L., et al., *Solution scattering at the Life Science X-ray Scattering (LiX) beamline.* J Synchrotron Radiat, 2020. **27**(Pt 3): p. 804-812.
18. Toogood, H.S., et al., *Extensive domain motion and electron transfer in the human electron transferring flavoprotein.medium chain Acyl-CoA dehydrogenase complex.* J Biol Chem, 2004. **279**(31): p. 32904-12.
19. Crane, F.L. and H. Beinert, *On the mechanism of dehydrogenation of fatty acyl derivatives of coenzyme A. II. The electron-transferring flavoprotein.* J Biol Chem, 1956. **218**(2): p. 717-31.
20. Beinert, H., *Spectral characteristics of flavins at the semiquinoid oxidation level.* J. Am. Chem. Soc., 1956. **78**: p. 5323-5328.
21. Henriques, B.J., et al., *Electron transfer flavoprotein and its role in mitochondrial energy metabolism in health and disease.* Gene, 2021. **776**: p. 145407.
22. Mayhew, S.G., et al., *Identification and properties of new flavins in electron-transferring flavoprotein from Peptostreptococcus elsdenii and pig-liver glycolate oxidase.* Eur J Biochem, 1974. **44**(2): p. 579-91.
23. Li, F., et al., *Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from Clostridium kluyveri.* J Bacteriol, 2008. **190**(3): p. 843-50.
24. Herrmann, G., et al., *Energy conservation via electron-transferring flavoprotein in anaerobic bacteria.* J Bacteriol, 2008. **190**(3): p. 784-91.
25. Duan, H.D., et al., *Distinct flavin properties underlie flavin-based electron bifurcation within a novel electron-transferring flavoprotein FixAB from Rhodopseudomonas palustris.* J. Biol. Chem., 2018. **293**(13): p. 4688-4701.
26. Chowdhury, N.P., et al., *Studies on the Mechanism of Electron Bifurcation Catalyzed by Electron Transferring Flavoprotein (Etf) and Butyryl-CoA Dehydrogenase (Bcd) of Acidaminococcus fermentans.* J. Biol. Chem., 2014. **289**: p. 5145-5157.
27. Demmer, J.K., et al., *The semiquinone swing in the bifurcating electron transferring flavoprotein/butyryl-coA dehydrogenase complex from Clostridium difficile.* Nat. Commun., 2017. **8**: p. 1577.
28. Leys, D., et al., *Extensive conformational sampling in a ternary electron transfer complex.* Nat Struct Biol, 2003. **10**(3): p. 219-25.
29. Buckel, W. and R.K. Thauer, *Flavin-Based Electron Bifurcation, Ferredoxin, Flavodoxin, and Anaerobic Respiration With Protons (Ech) or NAD(+) (Rnf) as Electron Acceptors: A Historical Review.* Front Microbiol, 2018. **9**: p. 401.
30. Mitchell, P., *The protonmotive Q cycle: a general formulation.* FEBS Lett, 1975. **59**(2): p. 137-9.
31. Mitchell, P., *Protonmotive redox mechanism of the cytochrome b-c1 complex in the respiratory chain: protonmotive ubiquinone cycle.* FEBS Lett, 1975. **56**(1): p. 1-6.
32. Peters, J.W., et al., *Electron bifurcation.* Curr Opin Chem Biol, 2016. **31**: p. 146-52.
33. Garcia Costas, A.M., et al., *Defining Electron Bifurcation in the Electron-Transferring Flavoprotein Family.* J Bacteriol, 2017. **199**(21).

34. Buckel, W. and R.K. Thauer, *Flavin-Based Electron Bifurcation, A New Mechanism of Biological Energy Coupling.* Chem Rev, 2018. **118**(7): p. 3862-3886.
35. Ledbetter, R.N., et al., *The Electron Bifurcating FixABCX Protein Complex from Azotobacter vinelandii: Generation of Low-Potential Reducing Equivalents for Nitrogenase Catalysis.* Biochemistry, 2017. **56**(32): p. 4177-4190.
36. Poudel, S., et al., *Electron Transfer to Nitrogenase in Different Genomic and Metabolic Backgrounds.* J Bacteriol, 2018. **200**(10).
37. Fixen, K.R., et al., *The path of electron transfer to nitrogenase in a phototrophic alpha-proteobacterium.* Environ Microbiol, 2018. **20**(7): p. 2500-2508.
38. Chowdhury, N.P., et al., *Reduction of Flavodoxin by Electron Bifurcation and Sodium Ion-dependent Reoxidation by NAD+ Catalyzed by Ferredoxin-NAD+ Reductase (Rnf).* J Biol Chem, 2016. **291**(23): p. 11993-2002.
39. Roberts, D.L., et al., *Crystal structure of Paracoccus denitrificans electron transfer flavoprotein: structural and electrostatic analysis of a conserved flavin binding domain.* Biochemistry, 1999. **38**(7): p. 1977-89.
40. Roberts, D.L., F.E. Frerman, and J.J. Kim, *Three-dimensional structure of human electron transfer flavoprotein to 2.1-A resolution.* Proc Natl Acad Sci U S A, 1996. **93**: p. 14355-14360.
41. Demmer, J.K., et al., *The semiquinone swing in the bifurcating electron transferring flavoprotein/butyryl-CoA dehydrogenase complex from Clostridium difficile.* Nat Commun, 2017. **8**(1): p. 1577.
42. Demmer, J.K., et al., *Molecular basis of the flavin-based electron-bifurcarting caffeyl-coA reductase reaction.* FEBS Lett., 2018. **592**: p. 332-342.
43. Feng, X., et al., *Cryoelectron microscopy structure and mechanism of the membrane-associated electron-bifurcating flavoprotein Fix/EtfABCX.* Proc Natl Acad Sci U S A, 2021. **118**(2).
44. Vigil, W., Jr., et al., *Spectral deconvolution of redox species in the crotonyl-CoA-dependent NADH:ferredoxin oxidoreductase from Megasphaera elsdenii. A flavin-dependent bifurcating enzyme.* Arch Biochem Biophys, 2021. **701**: p. 108793.
45. Vigil, W., Jr., et al., *Rapid-reaction kinetics of the butyryl-CoA dehydrogenase component of the electron-bifurcating crotonyl-CoA-dependent NADH:ferredoxin oxidoreductase from Megasphaera elsdenii.* J Biol Chem, 2023. **299**(7): p. 104853.
46. Darrouzet, E., et al., *Uncovering the [2Fe2S] domain movement in cytochrome bc1 and its implications for energy conversion.* Proc Natl Acad Sci U S A, 2000. **97**(9): p. 4567-72.
47. Freeman, S.L., et al., *Solution structure of the cytochrome P450 reductase-cytochrome c complex determined by neutron scattering.* J Biol Chem, 2018. **293**(14): p. 5210-5219.
48. Ellis, J., et al., *Domain motion in cytochrome P450 reductase: conformational equilibria revealed by NMR and small-angle x-ray scattering.* J Biol Chem, 2009. **284**(52): p. 36628-36637.
49. Ankner, J.F., et al., *Neutron scattering techniques and applications in structural biology.* Curr Protoc Protein Sci, 2013. **Chapter 17**: p. Unit17 16.

50. Vandavasi, V.G., et al., *A structural study of CESA1 catalytic domain of Arabidopsis cellulose synthesis complex: evidence for CESA trimers.* Plant Physiol, 2016. **170**(1): p. 123-135.
51. Carugo, O. and K. Djinovic Carugo, *When X-rays modify the protein structure: radiation damage at work.* Trends Biochem Sci, 2005. **30**(4): p. 213-9.
52. Kikhney, A.G. and D.I. Svergun, *A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins.* FEBS Lett, 2015. **589**(19 Pt A): p. 2570-7.
53. Sköld, K. and D.L. Price, *Neutron scattering*. Methods of experimental physics v. 23. 1986, Orlando: Academic Press. p. cm.
54. Lombardo, D., P. Calandra, and M.A. Kiselev, *Structural Characterization of Biomaterials by Means of Small Angle X-rays and Neutron Scattering (SAXS and SANS), and Light Scattering Experiments.* Molecules, 2020. **25**(23).
55. Heller, W.T. and K.C. Littrell, *Small-Angle Neutron Scattering for Molecular Biology: Basics and Instrumentation*, in *Micro and Nano Technologies in Bioanalysis,*, J.W. Lee and R.S. Foote, Editors. 2009, Humana Press. p. 293-305.
56. Lindner, P. and T. Zemb, *Neutron, x-ray and light scattering : introduction to an investigative tool for colloidal and polymeric systems : European workshop on neutron, X-ray and light scattering as an investigative tool for colloidal and polymeric systems : Papers*. Delta series. 1991, Amsterdam ; New York: North-Holland : Sole distributors for the U.S.A. and Canada, Elsevier Science Pub. Co.
57. *Small-Angle Neutron Scattering for Molecular Biology: Basics and Instrumentation*. Methods in Molecular Biology 544.
58. Jacques, D.A. and J. Trewhella, *Small-angle scattering for structural biology—Expanding the frontier while avoiding the pitfalls.* Prot. Sci., 2010. **19**: p. 642-657.
59. Fitter, J., et al., *Neutron scattering in biology : techniques and applications*, in *Biological and medical physics, biomedical engineering*. 2006, Springer
Springer Berlin Heidelberg: Berlin ; New York
Berlin, Heidelberg.
60. Zhang, J.X., F.E. Frerman, and J.-J.P. Kim, *Structure of Electron Transfer Flavoprotein-Ubiquinone Oxidoreductase and Electron Transfer to the Mitochondrial Uniquinone Pool.* Proc Natl Acad Sci U S A, 2006. **103**(44): p. 16212-16217.
61. Feĭgin, L.A., D.I. Svergun, and G.W. Taylor, *Structure analysis by small-angle x-ray and neutron scattering*. 1st 1987. ed. 1987, New York: Springer Science+Business Media, LLC.
62. Konig, S., et al., *Synchrotron radiation solution X-ray scattering study of the pH dependence of the quaternary structure of yeast pyruvate decarboxylase.* Biochemistry, 1992. **31**(37): p. 8726-31.
63. Smith, J.C., et al., *Dynamic Neutron Scattering by Biological Systems.* Annu Rev Biophys, 2018. **47**: p. 335-354.
64. Ashkar, R., et al., *Neutron scattering in the biological sciences: progress and prospects.* Acta Crystallogr D Struct Biol, 2018. **74**(Pt 12): p. 1129-1168.
65. Svergun, D.I., *Small-angle X-ray and neutron scattering as a tool for structural systems biology.* Biol Chem, 2010. **391**(7): p. 737-43.

66. Guinier, *La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques.* Annales de Physique, 1939. **11**(12): p. 161-237.
67. Putnam, C.D., et al., *X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution.* Q Rev Biophys, 2007. **40**(3): p. 191-285.
68. Mitchell, P., *Possible molecular mechanisms of the protonmotive function of cytochrome systems.* J Theor Biol, 1976. **62**(2): p. 327-67.
69. Ge, X., et al., *Characterization of the Membrane-Associated Electron-Bifurcating Flavoenzyme EtfABCX from the Hyperthermophilic Bacterium Thermotoga maritima.* Biochemistry, 2023. **62**(24): p. 3554-3567.
70. Watmough, N.J. and F.E. Frerman, *The electron transfer flavoprotein: Ubiquinone oxidoreductases.* Biochim. Biophys. Acta, 2010. **1797**: p. 1910-1916.
71. Toogood, H.S., D. Leys, and N.S. Scrutton, *Dynamics driving function : new insights from electron transferring flavoproteins and partner complexes.* FEBS J., 2007. **274**: p. 5481-5504.
72. Schut, G.J., et al., *The catalytic mechanism of electron-bifurcating electron transfer flavoproteins (ETFs) involves an intermediary complex with NAD<sup/>.* J Biol Chem, 2019. **294**(9): p. 3271-3283.
73. Mohamed-Raseek, N., et al., *Spectroscopic, thermodynamic and computational evidence of the locations of the FADs in the nitrogen fixation-associated electron transfer flavoprotein.* Chemical Sci., 2019. **10**: p. 7762-7772.
74. Demmer, J.K., et al., *Insights into Flavin-based Electron Bifurcation via the NADH-dependent Reduced Ferredoxin:NADP Oxidoreductase Structure.* J Biol Chem, 2015. **290**(36): p. 21985-95.
75. Kurisu, G., et al., *Structure of the electron transfer complex between ferredoxin and ferredoxin-NADP(+) reductase.* Nat Struct Biol, 2001. **8**(2): p. 117-21.
76. Wang, M., et al., *Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes.* Proc Natl Acad Sci U S A, 1997. **94**(16): p. 8411-6.
77. Xia, C., et al., *Structural basis for human NADPH-cytochrome P450 oxidoreductase deficiency.* Proc Natl Acad Sci U S A, 2011. **108**(33): p. 13486-91.
78. Leferink, N.G., et al., *Gating mechanisms for biological electron transfer: integrating structure with biophysics reveals the nature of redox control in cytochrome P450 reductase and copper-dependent nitrite reductase.* FEBS Lett, 2012. **586**(5): p. 578-84.
79. Feng, C., et al., *Intraprotein electron transfer in a two-domain construct of neuronal nitric oxide synthase: the output state in nitric oxide formation.* Biochemistry, 2006. **45**(20): p. 6354-62.
80. Greene, B.L., et al., *Conformationally Dynamic Radical Transfer within Ribonucleotide Reductase.* J Am Chem Soc, 2017. **139**(46): p. 16657-16665.
81. Davidson, V.L., *Protein control of true, gated, and coupled electron transfer reactions.* Acc Chem Res, 2008. **41**(6): p. 730-8.
82. Chen, K., et al., *Atomically defined mechanism for proton transfer to a buried redox centre in a protein.* Nature, 2000. **405**(6788): p. 814-7.

83. Berkholz, D.S., et al., *Catalytic cycle of human glutathione reductase near 1 A resolution.* J Mol Biol, 2008. **382**(2): p. 371-84.
84. Johansson, R., et al., *High-resolution crystal structures of the flavoprotein NrdI in oxidized and reduced states--an unusual flavodoxin. Structural biology.* FEBS J, 2010. **277**(20): p. 4265-77.
85. Kort, R., et al., *DNA apophotolyase from Anacystis nidulans: 1.8 A structure, 8-HDF reconstitution and X-ray-induced FAD reduction.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 7): p. 1205-13.
86. Prevo, B. and E.J. Peterman, *Forster resonance energy transfer and kinesin motor proteins.* Chem Soc Rev, 2014. **43**(4): p. 1144-55.
87. Murray, D.T., et al., *Small-angle neutron scattering solution structures of NADPH-dependent sulfite reductase.* J Struct Biol, 2021. **213**(2): p. 107724.
88. Murray, D.T., et al., *Neutron scattering maps the higher-order assembly of NADPH-dependent assimilatory sulfite reductase.* Biophys J, 2022. **121**(10): p. 1799-1812.
89. Parfait, R., B. De Wolf, and R.R. Crichton, *Protein-RNA interactions in the Escherichia coli ribosome: L24 protein-23 S RNA complex.* Arch Int Physiol Biochim, 1976. **84**(2): p. 404-5.
90. Crichton, R.R., et al., *Contrast variation study of specifically deuterated Escherichia coli ribosomal subunits.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5547-50.
91. Freeman, S.L., et al., *Orchestrated Domain Movement in Catalysis by Cytochrome P450 Reductase.* Sci Rep, 2017. **7**(1): p. 9741.
92. Bodenheimer, A.M., et al., *Structural investigation of cellobiose dehydrogenase IIA: Insights from small angle scattering into intra- and intermolecular electron transfer mechanisms.* Biochim Biophys Acta Gen Subj, 2018. **1862**(4): p. 1031-1039.
93. Pelikan, M., G.L. Hura, and M. Hammel, *Structure and flexibility within proteins as identified through small angle X-ray scattering.* Gen Physiol Biophys, 2009. **28**(2): p. 174-89.
94. Brooks, B.R., et al., *Charmm - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations.* Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
95. Brooks, B.R., et al., *CHARMM: The Biomolecular Simulation Program.* Journal of Computational Chemistry, 2009. **30**(10): p. 1545-1614.
96. Toogood, H.S., et al., *Stabilization of non-productive conformations underpins rapid electron transfer to electron transferring flavoprotein.* J. Biol. Chem., 2005. **280**: p. 30361-30366.
97. Laio, G.B.A., *Using metadynamics to explore complex free-energy landscapes.* Nature Reviews Physics, 2020. **2**: p. 200-2012.
98. Bhabha, G., J.T. Biel, and J.S. Fraser, *Keep on moving: discovering and perturbing the conformational dynamics of enzymes.* Acc Chem Res, 2015. **48**(2): p. 423-30.
99. van den Bedem, H. and J.S. Fraser, *Integrative, dynamic structural biology at atomic resolution--it's about time.* Nat Methods, 2015. **12**(4): p. 307-18.

100. Tzeng, S.R. and C.G. Kalodimos, *Allosteric inhibition through suppression of transient conformational states.* Nat Chem Biol, 2013. **9**(7): p. 462-5.
101. Murray, D.T., et al., *Correlating Conformational Equilibria with Catalysis in the Electron Bifurcating EtfABCX of Thermotoga maritima.* Biochemistry, 2024. **63**(1): p. 128-140.
102. Schut, G.J., et al., *The catalytic mechanism of electron bifurcating electron transfer flavoproteins (ETFs) involves an intermediary complex with NAD+.* J. Biol. Chem., 2019. **293**: p. ASAP.
103. Mohamed-Raseek, N., et al., *Unusual reactivity of a flavin in a bifurcating electron-transferring flavoprotein leads to flavin modification and a charge-transfer complex.* J Biol Chem, 2022. **298**(12): p. 102606.
104. Leite, W.C., et al., *Conformational Dynamics in the Interaction of SARS-CoV-2 Papain-like Protease with Human Interferon-Stimulated Gene 15 Protein.* J Phys Chem Lett, 2021. **12**(23): p. 5608-5615.
105. DiFabio, J., et al., *The Life Science X-ray Scattering Beamline at NSLS-II.* Proceedings of the 12th International Conference on Synchrotron Radiation Instrumentation (Sri2015), 2016. **1741**.
106. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
107. Mirdita, M., et al., *ColabFold: making protein folding accessible to all.* Nat Methods, 2022. **19**(6): p. 679-682.
108. Grudinin, S., M. Garkavenko, and A. Kazennov, *Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles.* Acta Crystallogr D Struct Biol, 2017. **73**(Pt 5): p. 449-464.
109. Khan, S., et al., *Structure, dynamics, and redox reactivity of an all-purpose flavodoxin.* J Biol Chem, 2024. **300**(4): p. 107122.
110. Warburg, O. and W. Christian, *On a new oxidation enzyme (Über das neue oxydationsferment).* Naturwissenschaften, 1932. **20**: p. 980-981.
111. Blyth, A.W., *The composition of cow's milk in health and disease.* J. Chem. Soc. Perkin Trans, 1879. **35**: p. 530-539.
112. Michaelis, L., P. Schubert, and C.V. Smythe, *Potentiometric study of the flavins.* Journal of Biological Chemistry, 1936. **116**: p. 587-607.
113. Knight, E., Jr., A.J. D'Eustachio, and R.W. Hardy, *Flavodoxin: a flavoprotein with ferredoxin activity from Clostridium pasteurianum.* Biochim. Biophys. Acta, 1966. **113**: p. 626-628.
114. Cotruvo, J.A., Jr. and J. Stubbe, *NrdI, a flavodoxin involved in maintenance of the diferric-tyrosyl radical cofactor in Escherichia coli class Ib ribonucleotide reductase.* Proc. Nat. Acad. Sci. U S A, 2008. **105**(38): p. 14383–14388.
115. Mayhew, S.G. and G. Tollin, *General properties of flavodoxins.*, in *Chemistry and Biochemistry of Flavoenzymes*, F. Müller, Editor. 1992, CRC Press: Boca Raton FL. p. 389-426.
116. Karlusich, J.J.P., A.F. Lodeyro, and N. Carrillo, *The long goodbye: the rise and fall of flavodoxin during plant evolution.* J. Exp. Botany, 2014. **65**(18): p. 5161-5178.

117. Campbell, I.J., G.N. Bennett, and J.J. Silberg, *Evolutionary Relationships Between Low Potential Ferredoxin and Flavodoxin Electron Carriers.* Fronti. Energy Res., 2019. **7**: p. 79.
118. Tognetti, V.B., et al., *Enhanced plant tolerance to iron starvation by functional substitution of chloroplast ferredoxin with a bacterial flavodoxin.* Proc Natl Acad Sci U S A, 2007. **104**: p. 11495–11500.
119. Peelen, S., et al., *Possible role of a short extra loop of the long-chain flavodoxin from Azotobacter chroococcum in electron transfer to nitrogenase: complete 1H, 15N and 13C back- bone assignments and secondary solution structure of the flavodoxin.* J. Biomol. NMR, 1996. **7**: p. 315–330.
120. Lodeyro, A.F., et al., *The importance of flavodoxin for environmental stress tolerance in photosynthetic microorganisms and transgenic plants. Mechanism, evolution and biotechnological potential.* Febs Letters, 2012. **586**(18): p. 2917-2924.
121. Bianchi, V., et al., *Flavodoxin is required for the activation of the anaerobic ribonucleotide reductase. Biochem Biophys Res Commun. 1993; 197:792–797.* Biochem. Biophys. Res. Commun., 1993. **197**: p. 792–797.
122. Carey, J., et al., *WrbA bridges bacterial flavodoxins and eukaryotic NAD(P)H:quinone oxidoreductases.* Prot. Sci., 2007. **16**: p. 2301-2305.
123. McIver, L., et al., *Characterization of flavodoxin NADP+ oxidoreductase and flavodoxin; key components of electron transfer in Escherichia coli.* Eur J Biochem, 1998. **257**: p. 577-585.
124. Sibille, N., et al., *Solution Structure of the Sulfite Reductase Flavodoxin-like Domain from Escherischia coli.* Biochemistry, 2005. **44**(25): p. 9086-9095.
125. Perez-Dorado, I., et al., *Structural and Phylogenetic Analysis of Rhodobacter capsulatus NifF: Uncovering General Features of Nitrogen-fixation (nif)-Flavodoxins* . Int. J. Mol. Sci., 2010. **14**(1): p. 1152-1163.
126. Cannon, M., et al., *The nucleotide sequence of the nifJ gene of Klebsiella pneumoniae.* Nucleic Acids Res., 1988. **16**: p. 11379.
127. Shah, V.K., G. Stacey, and W.J. Brill, *Electron transport to nitrogenase: purification and characterisation of pyruvate:flavodoxin oxidoreductase, the nifJ gene product.* J. Biol. Chem., 1983. **258**: p. 2064–12068.
128. Wang, Z.-Q., et al., *Bacterial Flavodoxins Support Nitric Oxide Production by Bacillus subtilis Nitric-oxide Synthase.* J. Biol. Chem., 2007. **282**: p. 2196-2202.
129. Birch, O.M., M. Fuhrmann, and N.M. Shaw, *Biotin Synthase from Escherichia coli, an Investigation of the Low Molecular Weight and Protein Components Required for Activity in Vitro.* J. Biol. Chem., 1995. **270**: p. 19158−19165.
130. Ifuku, O., et al., *Flavodoxin is required for conversion of dethiobiotin to biotin in Escherichia coli.* Eur. J. Biochem., 1994. **224**: p. 173-178.
131. Hall, D.A., et al., *Mapping the interactions between flavodoxin and its physiological partners flavodoxin reductase and cobalamin-dependent methionine synthase.* Proc. Nat. Acad. Sci. U S A, 2001. **98**: p. 9521-9526.
132. Fujii, K. and F.M. Huennekens, *Activation of Methionine Synthetase by a Reduced Triphosphopyridine Nucleo- tide-dependent Flavoprotein System.* J. Biol. Chem., 1974. **249**: p. 6745−6753.

133. Gutekunst, K., et al., *The Bidirectional NiFe-hydrogenase in Synechocystis sp PCC 6803 Is Reduced by Flavodoxin and Ferredoxin and Is Essential under Mixotrophic, Nitrate-limiting Conditions.* J. Biol. Chem., 2014. **289**(4): p. 1930-1937.
134. Blamey, J.M. and M.W.W. Adams, *Characterisation of an ancestral type of pyruvate ferredoxin oxidoreductase from the hyperthermophilic bacterium, Thermotoga maritima.* Biochemistry, 1994. **33**: p. 1000–1007.
135. Blaschkowski, H.P., et al., *Routes of flavodoxin and ferredoxin reduction in Escherichia coli. CoA-acylating pyruvate: flavodoxin and NADPH: flavodoxin oxidoreductases participating in the activation of pyruvate formate-lyase.* Eur J Biochem, 1982. **123**: p. 563-569.
136. Hughes, N.J., et al., *Helicobacter pylori porCDAB and oorDABC Genes Encode Distinct Pyruvate:Flavodoxin and 2-Oxoglutarate:Acceptor Oxidoreductases Which Mediate Electron Transport to NADP.* J. Bacteriol., 1998. **180**(5): p. 119-1128.
137. Hawkes, D.B., et al., *Cytochrome P450 cin (CYP176A), isolation, expression, and characterization.* J. Biol. Chem., 2002. **277**: p. 27725-27732.
138. Bruender, N.A., A.P. Young, and V. Bandarian, *Chemical and Biological Reduction of the Radical SAM Enzyme 7-Carboxy-7-deazaguanine Synthase.* Biochemistry, 2015. **54**(18): p. 2903-2910.
139. Tejero, J., et al., *Surface Charges and Regulation of FMN to Heme Electron Transfer in Nitric-oxide Synthase.* J. Biol. Chem., 2010. **285**(35): p. 27232–2724.
140. Aoki, M., K. Ishimori, and I. Morishima, *Roles of negatively charged surface residues of putidaredoxin in interactions with redox partners in P450cam monooxygenase system.* Biochim. Biophys. Acta, 1998. **1386**: p. 157-167.
141. Wan, J.T. and J.T. Jarrett, *Electron acceptor specificity of ferredoxin (flavodoxin): NADP+ oxidoreductase from Escherichia coli.* Arch. Biochem. Biophys., 2002. **406**: p. 116-126.
142. Smillie, R.M., *Isolation of phytoflavin, a flavoprotein wiht chloroplast ferredoxin activity.* Plant Physiol., 1965. **40**(6): p. 1124-1128.
143. López-Llano, J., et al., *The long and short flavodoxins I. The role of the differentiating loop in apoflavodoxin structure and FMN binding.* J. Biol. Chem., 2004. **279**(45): p. 47177-47183.
144. Cao, P., et al., *Structural basis for energy and electron transfer of the photosystem I-IsiA-flavodoxin supercomplex.* Nat.Plants, 2020. **6**: p. 167-176.
145. Pechter, K.B., et al., *Essential genome of the metabolically versatile alphaproteobacterium Rhodopseudomonas palustris.* J. Bacteriol., 2015. **198**(5): p. 867-876.
146. Ye, Q., Y. Hu, and C. Jin, *Conformational Dynamics of Escherichia coli Flavodoxins in Apo- and Holo-States by Solution NMR Spectroscopy.* Plos One, 2014. **9**(8): p. e103936.
147. Varner, T.A., N. Mohamed-Raseek, and A.-F. Miller, *Assignments of ^{19}F NMR resonances and exploration of dynamics in a long-chain flavodoxin.* Achiv. Biochem. Biophys., 2021.

148. Banci, L., et al., *Molecular view of an electron transfer process essential for iron–sulfur protein biogenesis.* Proc. Nat. Acad. Sci. U S A, 2013. **110**(18): p. 7136-7141.
149. Schrodinger, L.L.C., *The PyMOL Molecular Graphics System.* http://www.pymol.org, 2015. **Version 1.8**.
150. Segal, H.M., et al., *Electrochemical and structural characterization of Azotobacter vinelandii flavodoxin II.* Prot. Sci., 2017. **26**: p. 1984-1993.
151. Pérez-Dorado, I., et al., *Structural and phylogenetic analysis of Rhodobacter capsulatus NifF: uncovering general features of nitrogen-fixation (nif)-flavodoxins.* Int. J. Mol. Sci., 2013. **14**: p. 1152-1163.
152. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J. Mol. Biol., 2007. **372**(3): p. 774-797.
153. Romao, C.V., et al., *The crystallographic structure of Flavorubredoxin from Escherichia coli.* J. Mol. Biol., 2016. **428**: p. 4686-4707.
154. Silaghi-Dumitrescu, R., et al., *X-ray crystal structures of Moorella thermoacetica FprA. Novel diiron site structure and mechanistic insights into a scavenging nitric oxide reductase.* Biochem., 2005. **44**: p. 6492-6501.
155. Meng, E.C., et al., *Tools for integrated sequence-structure analysis with UCSF Chimera.* BMC Bioinformatics, 2006. **7**: p. 339.
156. Pettersen, E.F., et al., *UCSF Chimera - a visualization system for exploratory research and analysis.* J. Comput. Chem., 2004. **25**(13): p. 1605-1612.
157. Hickey, D.P., et al., *Pyrene Hydrogel for Promoting Direct Bioelectrochemistry: ATP-Independent Electroenzymatic Reduction of N2.* Chem. Sci., 2018. **9**(23): p. 5172–5177.
158. Imam, H.T., P.C. Marr, and A.C. Marr, *Enzyme Entrapment, Biocatalyst Immobilization without Covalent Attachment.* Green Chem., 2021. **23**(14): p. 4980-5005.
159. Giroud, F. and S.D. Minteer, *Anthracene-Modified Pyrenes Immobilized on Carbon Nanotubes for Direct Electroreduction of O2 by Laccase.* Electrochem. Commun., 2013. **34**: p. 157-160.
160. Hoover, D.M., et al., *Comparisons of wild-type and mutant flavodoxins from Anacystis nidulans. Structural determinants of the redox potentials.* J. Mol. Biol., 1999. **294**: p. 725-743.
161. Alagaratnam, S., et al., *A crystallographic study of Cys69Ala flavodoxin II from Azotobacter vinelandii: Structural determinants of redox potential.* Prot. Sci., 2005. **14**: p. 2284-2295.
162. Steensma, E., et al., *Redox properties of wild-type, Cys69Ala, and Cys69Ser Azotobacter vinelandii flavodoxin II as measured by cyclic voltammetry and EPR spectroscopy.* Eur J Biochem 1996. **235**: p. 167–172.
163. McIver, L., et al., *Characterisation of flavodoxin NADP+ oxidoreductase and flavodoxin; key components of electron transfer in Escherichia coli.* Eur. J. Biochem., 1998. **257**: p. 577-585.
164. Lostao, A., et al., *Differential stabilization of the three FMN redox forms by tyrosine 94 and tryptophan 57 in flavodoxin from Anabaena and its influence on the redox potentials.* Biochemistry, 1997. **36**: p. 14334-14344.

165. Zhang, Y., *Methylotrophs in a groundwater-derived drinking water system.*, in *Civil and Environmental Engineering*. 2017, University of Illinois at Urbana-Champaign: Urbana-Champaign.
166. Seyedsayamdost, M.R., et al., *Mono- di-, tri and tetra-substituted fluorotyrosines: new probes for enzymes that use tyrosyl radicals in catalysis.* J. Am. Chem. Soc., 2006. **128**: p. 1569-1579.
167. Sitbon, E. and S. Pietrokovski, *Occurrence of protein structure elements in conserved sequence regions.* BMC Structural Bio., 2007. **7**(3).
168. Kozlowski, L.P., *Proteome-pI: proteome isoelectric point database.* Nucleic Acids Research., 2017. **45**(D1): p. D1112–D1116.
169. Proteopedia. *Amino_Acid_Composition*. 2020 24 April 2020 [cited 2023 3 October]; Available from: https://proteopedia.org/wiki/index.php/Amino_acid_composition.
170. Shingate, P. and R. Sowdhamini, *Analysis of Domain-Swapped Oligomers Reveals Local Sequence Preferences and Structural Imprints at the Linker Regions and Swapped Interfaces.* PloS one, 2012. **7**: p. e39305.
171. Sevrioukova, I.F., et al., *1998.* Proc Natl Acad Sci U S A. **96**: p. 1863-1868.
172. Xia, C., et al., *Structural and Kinetic Studies of Asp632 Mutants and Fully Reduced NADPH-Cytochrome P450 Oxidoreductase Define the Role of Asp632 Loop Dynamics in the Control of NADPH Binding and Hydride Transfer.* Biochemistry, 2018. **57**: p. 945-962.
173. Pueyo, J.J., C. Gómez-Moreno, and S.G. Mayhew, *Oxidation-reduction potentials of ferredoxin-NADP+ reductase and flavodoxin from Anabaena PCC 7119 and their electrostatic complexes.* Eur. J. Biochem., 1991. **202**: p. 1065-1071.
174. Saen-oon, S., et al., *A theoretical multiscale treatment of protein–protein electron transfer: The ferredoxin/ferredoxin-NADP+ reductase and flavodoxin/ ferredoxin-NADP+ reductase systems.* Biochim Biophys Acta, 2015. **1847**: p. 1530-1538.
175. Ubbink, M., *Dynamics in transient complexes of redox proteins.* Biochem. Soc. Trans., 2012. **40**(2): p. 415-418.
176. Ludwig, M.L., et al., *Control of oxidation-reduction potentials in flavodoxin from Clostridium beijerinckii: the role of conformational changes.* Biochemistry, 1997. **36**: p. 1259-1280.
177. Liu, S.J., et al., *A dimer-monomer transition captured by the crystal structures of cyanobacterial apo flavodoxin.* Biochem. Biophys. Res. Commun., 2023. **639**: p. 134-141.
178. Hrovat, A., et al., *Backbone dynamics of oxidized and reduced D. vulgaris flavodoxin in solution.* J. Biomol. NMR, 1997. **10**: p. 53-62.
179. Mutoh, R., et al., *X-ray structure and nuclear magnetic resonance analysis of the interaction sites of the Ga-substituted cyanobacterial ferredoxin.* Biochem., 2015. **54**: p. 6052-6061.
180. Crain, A.V. and J.B. Broderick, *Flavodoxin cofactor binding induces structural changes that are required for protein-protein interactions with NADP+ oxidoreductase and pyruvate formate-lyase activating enzyme.* Biochim Biophys Acta, 2014. **1834**(12): p. 2512-2519.

181. Porter, T.D. and C.B. Kasper, *NADPH-cytochrome P-450 oxidoreductase: flavin mononucleotide and flavin adenine dinucleotide domains evolved from different flavoproteins.* Biochemistry, 1986. **25**(7): p. 1682-1687.
182. Ingelman, M., V. Bianchi, and H. Eklund, *The three-dimensional structure of flavodoxin reductase from Escherichia coli at 1.7 Å resolution.* J. Mol. Biol., 1997. **268**: p. 147-157.
183. Xia, C., et al., *Structural and kinetic studies of Asp632 mutants and fully- reduced NADPH-cytochrome P450 oxidoreductase define the role of Asp632 loop dynamics in control of NADPH binding and hydride transfer.* Biochemistry, 2018. **57**(6): p. 945-962.
184. Urban, P. and D. Pompon, *Confrontation of AlphaFold models with experimental structures enlightens conformational dynamics supporting CYP102A1 functions.* Sci. Rep., 2022. **12**: p. 15982.
185. Govindaraj, S. and T.L. Poulos, *The Domain Architecture of Cytochrome P450BM-3.* J. Biol. Chem., 1997. **272**(12): p. 7915-7921.
186. Neeli, R., et al., *The dimeric form of flavocytochrome P450 BM3 is catalytically functional as a fatty acid hydroxylase.* FEBS Lett., 2005. **579**: p. 5582–5588.
187. Ermakova, M., et al., *Heterocyst-specific flavodiiron protein Flv3B enables oxic diazotrophic growth of the filamentous cyanobacterium Anabaena sp. PCC 7120.* Proc. Natl. Acad. Sci. USA, 2014. **111**: p. 11205–111210.
188. Silaghi-Dumitrescu, R., et al., *A flavodi-iron protein and high molecular weight rubredoxin from Moorella thermoacetica with nitric oxide reductase activity.* Biochemistry, 2003. **42**: p. 2806-2815.
189. Warren, J.J., et al., *Electron hopping through proteins.* Coord. Chem. Rev., 2012. **256**(21-22): p. 2478-2487.
190. Muchmore, D.C., et al., *Expression and Nitrogen-15 Labeling of Proteins for Proton and Nitrogen-15 Nuclear Magnetic Resonance.* Meth. Enzymol., 1989. **177**: p. 44-73.
191. Antharavally, B.S., et al., *Quantitations of proteins using a dye-metal-based colorimetric protein assay.* Anal. Biochem., 2009(385): p. 342-345.
192. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallographica Section D: Biological Crystallography, 2013. **69**(7): p. 1204-1214.
193. Winn, M.D., et al., *Overview of the CCP4 suite and current developments.* Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 235-242.
194. McCoy, A.J., et al., *Phaser crystallographic software.* Journal of Applied Crystallography, 2007. **40**(4): p. 658-674.
195. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
196. Afonine, P.V., et al., *Towards automated crystallographic structure refinement with phenix.refine.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 4): p. 352-67.
197. Emsley, P., et al., *Features and development of Coot.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(4): p. 486-501.

198. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography.* Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(1): p. 12-21.
199. Bliven, S., et al., *Automated evaluation of quaternary structures from protein crystals.* PLoS Comput. Biol., 2018. **14**(4): p. e1006104.
200. Zallot, R., N. Oberg, and J.A. Gerlt, *The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways.* Biochemistry, 2019. **58**(41): p. 4169-4182.
201. Oberg, N., R. Zallot, and J.A. Gerlt, *EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools.* J Mol Biol, 2023. **435**(14).
202. van Mierlo, C.P.M., et al., *Tertiary structure of two-electron reduced Megasphaera elsdenii flavodoxin and some implications, as determined by two-dimensional 1H NMR and restrained molecular dynamics.* Eur J Biochem, 1990. **194**: p. 185-198.
203. populations, D.
204. Thomsen, M.C.F. and M. Nielsen, *Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion.* Nucleic Acids Research, 2012. **40**(W1): p. W281-W287.
205. Hodge, E.A., M.A. Benhaim, and K.K. Lee, *Bridging protein structure, dynamics, and function using hydrogen/deuterium-exchange mass spectrometry.* Protein Sci, 2020. **29**(4): p. 843-855.
206. Tsirigotaki, A., et al., *Dynamics and ligand-induced conformational changes in human prolyl oligopeptidase analyzed by hydrogen/deuterium exchange mass spectrometry.* Sci Rep, 2017. **7**(1): p. 2456.
207. Singh, A., J.A. Purslow, and V. Venditti, *15N CPMG Relaxation Dispersion for the Investigation of Protein Conformational Dynamics on the micros-ms Timescale.* J Vis Exp, 2021(170).
208. Mittermaier, A. and L.E. Kay, *New tools provide new insights in NMR studies of protein dynamics.* Science, 2006. **312**(5771): p. 224-8.

VITA

Name: Sharique Ahmad Khan

Education:

PhD, Chemistry (2017-2024)

University of Kentucky, Lexington

M.Sc, Analytical Chemistry (2014-2016)

Savitribai Phule Pune University, India

B.Sc. Chemistry (2011-2014)

University of Mumbai, India

Research Experience:

Graduate Research Assistant (2017- 2024)

Research Associate, Pharmaceuticals (2016-2017)

Syngene International Limited, India

Teaching Experience:

Graduate Teaching Assistant (2017-2020)

University of Kentucky

Honors & Awards:

Neutron scattering research award with stipend, Oak Ridge National Laboratory (U.S. Department of Energy (D.O.E) (2023)

Annual meeting travel award, Biophysical Society (2023)

Graduate Student Researcher Award, American Society for Biochemistry and Molecular Biology (ASBMB) (2023)

Conference award by the graduate student congress at the University of Kentucky (2023)

Outstanding service award to the graduate student congress, University of Kentucky (2021)

First rank in the college at the undergraduate level (2014)

Ranked 7th of approximately 2000 students for the entrance exam, master's program at the Department of Chemistry, Pune University and received scholarship by the national chemical laboratory, Pune (2014)

Publications:

- **Khan, Sharique** *et al*. Structure, dynamics, and redox reactivity of an all-purpose flavodoxin. J Biological Chemistry. 2024 Feb 28;300(4):107122.

- Duan, H. Diessel, **Sharique A. Khan**, and Anne-Frances Miller. "Photogeneration and reactivity of flavin anionic semiquinone in a bifurcating electron transfer flavoprotein." Biochimica et Biophysica Acta (BBA)-Bioenergetics 1862.7 (2021): 148415

Conference abstracts:

- **Khan, Sharique**, et al. "Investigating reduction triggered conformational changes in the bifurcating electron transfer flavoprotein from Acidaminococcus fermentans by small angle neutron scattering." Journal of Biological Chemistry 299.3 (2023): S331-S332.

- **Khan, Sharique**, Anne-Frances Miller, and Ahmadullah Ansari. "19F NMR reveals dynamics not evident in crystal structure of long chain flavodoxin." Biophysical Journal 122.3 (2023): 466a.