# The cofactor-dependent folding energy landscape of a light-sensing protein revealed by single-molecule pulling experiments

**Rodrigo Maillard** ( ram279@georgetown.edu )
Georgetown University https://orcid.org/0000-0001-5078-6775
**Sahar Foroutannejad**
Georgetown University
**Lydia Good**
Georgetown University
**Changfan Lin**
Cornell University https://orcid.org/0000-0002-8119-9845
**Zachariah Ingram**
University of Alabama at Birmingham
**Mahlet Tadesse**
Georgetown University https://orcid.org/0000-0003-2671-1663
**Aaron Lucius**
UAB
**Brian Crane**
Cornell University https://orcid.org/0000-0001-8234-9991

**Article**

**Keywords:**

**Posted Date:** February 24th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-1336063/v1

# The cofactor-dependent folding energy landscape of a light-sensing protein revealed by single-molecule pulling experiments

Sahar Foroutannejad[1], Lydia Good[1], Changfan Lin[2], Zachariah Ingram[3], Mahlet Tadesse[4], Aaron L. Lucius[3], Brian R. Crane[2], Rodrigo A. Maillard[1,#]

[1] Department of Chemistry, Georgetown University, Washington, DC, USA
[2] Department of Chemistry & Chemical Biology, Cornell University, Ithaca, NY, USA
[3] Department of Chemistry, University of Alabama at Birmingham, Birmingham, AL, USA
[4] Department of Mathematics and Statistics, Georgetown University, Washington, DC, USA

# Corresponding author: Rodrigo A. Maillard (rodrigo.maillard@georgetown.edu)

The functional link between cofactor binding and protein activity is well established, but how cofactor interactions with the polypeptide reshape the folding energy landscape of large and multidomain proteins is unknown. Here, we use optical tweezers in combination with a novel analytical framework that integrates clustering, bootstrapping and global fitting of kinetic and thermodynamic data to dissect the folding mechanism of the light-sensing *Drosophila* cryptochrome (dCRY), a 542-residue protein that binds FAD, one of the most common, complex cofactors. We show that FAD binds to multiple dCRY folding intermediates, some of which contain large amounts of unfolded polypeptide. Yet, binding occurs with association kinetics above the diffusion-limit and at sub-nanomolar affinity. Surprisingly, the first parts of dCRY to fold are independent of FAD, but later steps are FAD-driven as the remaining protein folds around the cofactor. Thus, dCRY coordinates cofactor-dependent and independent folding mechanisms to attain its native state. Furthermore, we find that not all the FAD chemical moieties are strictly required for folding: whereas the isoalloxazine ring linked to ribitol and one phosphate group (i.e., FMN) are sufficient to drive complete dCRY folding, the adenosine ring plus the phosphate groups (i.e., AMP and ADP) only allow partially folded structures. Lastly, by combining the results from optical tweezers experiments with structural data, we propose a model for the dCRY folding pathway wherein regions known to undergo conformational transitions during signal transduction are the last to fold. Altogether, our single-molecule experiments and data analysis illustrate the power and broad applicability of optical tweezers to dissect complex mechanisms that couple the folding of large proteins to cofactor binding.

## INTRODUCTION

Since the solution of the first protein structures[1–3] much progress has been made to elucidate experimentally or computationally the tertiary structure and quaternary organization of proteins[4–7]. However, determining protein folding pathways, i.e., identifying the steps and rate constants that connect transient intermediates to end states, has proven more challenging, particularly for large or multidomain proteins[8]. This challenge is further exacerbated when considering that a large fraction of the proteome incorporates cofactors, which are not only important for protein function but can also alter fold and thermodynamic properties[9,10].

Previous bulk studies have investigated the role of cofactor binding on the folding of flavoproteins that contain the cofactor flavin mononucleotide (FMN). These studies showed that flavoproteins can form molten globules or fold into their native state independently of the cofactor[11–13] or fold into an intermediate that is stabilized by FMN binding before reaching the native state[14]. More recently, single molecule manipulation methods have enabled the direct kinetic characterization of intermediates along the folding pathway of small, single-domain proteins that have metal cofactors[15–19] However, these methods have not been applied to proteins that bind to flavin cofactors, let alone multi-domain proteins that bind one of the most common large cofactors known, flavin adenosine diphosphate (FAD). Single molecule techniques like optical tweezers overcomes several challenges in studying the folding mechanism of large proteins in bulk[20,21] as it enables the direct observation of transient intermediates, overcomes the rapid loss of synchronicity among molecules undergoing sequential kinetic steps, provides a direct measure of changes in protein secondary and tertiary structures, and, importantly, avoids protein aggregation[22–26].

In this study, we use optical tweezers to dissect the mechanisms that couple FAD binding to folding of *Drosophila* cryptochrome (dCRY), a multidomain protein of 542 residues. dCRY is a signaling protein that undergoes functionally important conformational changes in response to chemical changes in FAD[27–29]. Thus, dCRY function is fundamentally linked to the coupling between the protein fold and the FAD cofactor. The structure of full-length dCRY displays two large domains, collectively known as the photolyase homology region (PHR; Figure 1a)[30–32]. The N-terminal domain comprises an α/β Rossman fold (residues 1-140)[33], whereas the much larger C-terminal domain (residues 141-542) contains a central 4-helix bundle motif conserved by other DNA-metabolizing enzymes that bind cofactors (residues 361-424)[34], as well as three extended loops that surround the FAD binding pocket: (1) the phosphate-binding loop (residues 249-263) that coordinates a phosphate group close to FAD; (2) the protrusion motif (residue 288-306); and (3) the C-terminal lid (residues 420-446). These three loops, called C-terminal coupled motif (CCM), pack against a 10-residue helical tail called C-terminal tail or CTT (residues 530-539) (Figure 1a)[30,31]. Signaling mediated by dCRY involves displacement of CTT from the CCM upon changes in FAD redox state[27–29].

Herein, we show that dCRY folds into its native state in a stepwise fashion via five intermediate states. We determined all the rate constants connecting these intermediates using clustering, bootstrapping and global fitting analysis. We find that FAD binds to the first two intermediates that are largely unfolded, yet the cofactor binds very fast (above the diffusion-limit) and with sub-nanomolar affinity. It is only after FAD binds and stabilizes these intermediates that dCRY can proceed to the native state. Furthermore, by using a variety of cofactors that contain different FAD moieties, we show that the isoalloxazine ring linked to ribitol and one phosphate group (i.e., FMN) is sufficient for dCRY to natively fold. In contrast, the adenosine ring of FAD with one or two phosphate groups (i.e., AMP or ADP) only forms partially folded structures. By combining the results from optical tweezers experiments with published structural data, we

propose a model in which the dCRY PHR folds first and independently of FAD, followed by several steps of binding and co-folding around the cofactor by the larger C-terminal domain. Thus, the complex topology and domain organization of dCRY seems to require various folding strategies previously seen as mutually exclusive mechanisms for single-domain proteins that bind organic or metal cofactors. Altogether, our single molecule approach allowed us to dissect and quantitate these distinct folding mechanisms for a single protein, underscoring the power and broad applicability of optical tweezers to dissect complex coupling mechanisms between folding of large proteins and cofactor binding.

## RESULTS and DISCUSSION

### Mechanical Unfolding Trajectories of dCRY

To immobilize a single dCRY protein between the two beads in the optical tweezers (Figure 1b), we added the Avi and ybbR tags at the N- and C-termini, respectively[35,36]. The Avi tag was covalently modified with biotin, and the ybbR tag was modified with a 350-bp DNA handle with a digoxigenin in its 5'-end (Supporting Figures S1 and S2). The biotin interacts with a streptavidin-coated bead held in a fixed position on a micropipette tip. The digoxigenin binds a bead coated with anti-digoxigenin antibodies, which is held in a movable optical trap (Figure 1b, Supplementary Information). Importantly, the modified protein displayed the same spectroscopic properties as wild type dCRY, indicating that the addition of the tags did not interfere with the normal fold of dCRY and its ability to bind FAD (Supporting Figure S2).

Molecular trajectories of the mechanical unfolding of single dCRY protein molecules were obtained by moving the bead in the optical trap away from or towards the bead on the micropipette tip. The resulting force-extension curve displayed a gradual increase in force and position, due to stretching of the DNA handle, that is interrupted by one or more rips that correspond to protein unfolding events (Figure 1c, blue line). In this study, dCRY was mechanically unfolded using a constant pulling velocity of 75 nm/s and a force up to 45 pN. Then, the force was reduced to 1 pN to allow dCRY to refold and bind FAD before the next pulling cycle (Figure 1c, red line). Depending on the type of experiment, we varied the refolding time between 0-40 s, the FAD concentration, or both.

### dCRY Requires FAD to Attain its Native State

We first examined whether dCRY was natively folded in the single molecule mechanical assay by determining the total change in contour length, $\Delta Lc_T$, upon unfolding, which reports on the number of folded residues at the start of the experiment. $\Delta Lc_T$ was obtained from the total unfolding rip size ($\Delta x_T$) irrespective of the number of intermediate rips that occurred during the unfolding reaction (Figure 1c, inset). We used the worm-like chain (WLC) model[37] to analyze the observed $\Delta x_T$ as a function of force (Figure 1d). In conditions where [FAD] = 25 μM and the refolding time was 20 s, we obtained a $\Delta Lc_T$ = 205.7 ± 15 nm (mean ± standard deviation, N = 269), which is consistent with the fully folded dCRY structure[30,31] (558 amino acids of native protein and engineered linkers × 0.365 nm per amino acid – the folded end-to-end distance (5.5 nm) = 198.2 nm). This result indicates that dCRY can reversibly fold, bind FAD, and attain its native folded state in the optical tweezers assay.

To study the role of FAD binding on dCRY folding, we tested if the protein could fold into its native state in the absence of the cofactor (Figure 1c). Using the same refolding time of 20 s but with no FAD added into the optical tweezers microfluidic chamber, we found that 25% of

force-extension curves did not display any unfolding rips, indicating that the protein remained in or continued to sample the unfolded state. The other 75% of the force-extension curves displayed small unfolding rip sizes (Figure 1e, N = 132) with corresponding $\Delta Lc_T$ values between 27 nm and 120 nm (Figure 1f). These values are significantly smaller than the observed $\Delta Lc_T$ for the natively folded protein bound to FAD (198.2 nm) indicating that dCRY can only form partially folded intermediates in the absence of the cofactor. Thus, to attain its native folded state, dCRY strictly requires a bound FAD molecule.

## Single Molecule FAD Titration Reveals Multiple Intermediates During dCRY Folding

Given the essential role of FAD in driving the native fold of dCRY, we investigated how dCRY folding depends on FAD concentration. For this experiment, we monitored $\Delta Lc_T$ at FAD concentrations between 1 pM and 25 μM, allowing the protein to refold at 1 pN for 20 s between pulling cycles. The results were plotted as degree of folding from 0 to 1, which reflect the unfolded and fully folded states, respectively. Degree of folding was calculated by dividing the observed $\Delta Lc_T$ at each FAD concentration by 198.2 nm, the theoretical $\Delta Lc_T$ of full-length dCRY (Figure 2a, Supplementary Information).

The plots of degree of folding reveal a complex refolding pattern, with events that correspond to unfolded, intermediate or native states, and whose populations vary with FAD concentration (Figure 2a). The fraction of events with degree of folding near 0 almost completely disappeared as the FAD concentration was increased. In contrast, events with degree of folding around 1 increased gradually with FAD concentration. Events with degree of folding between 0.1-0.9, which represent intermediate states with partially folded structures, were observed in most FAD concentrations but with diverse degrees of folding (Figure 2a). Intermediates with degree of folding < 0.5 were observed more frequently at FAD concentrations below 0.3 nM. Above 0.3 nM, intermediates with higher degrees of folding between 0.5 and 0.9 became more prevalent. We use the fraction of events corresponding to unfolded, intermediates and fully folded states as a function of the log[FAD] to obtain an apparent FAD dissociation constant: $Kd_{app}$ = 0.11 ± 0.04 nM (mean ± 95 % confidence interval) (Figure 2b).

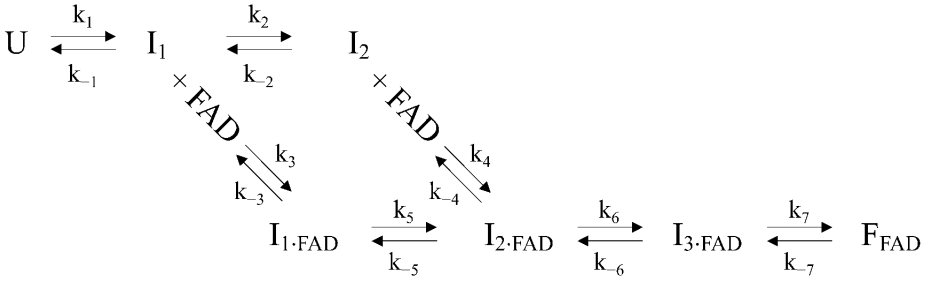
## A Complex dCRY Folding Pathway is Coupled to FAD Binding

The wide-ranging values of degree of folding observed in the FAD titration suggest that folding of dCRY coupled to FAD binding proceeds via multiple intermediate states. We therefore performed experiments to determine the kinetic steps and rate constants that connect these intermediates from the unfolded to the native state, including where in the folding pathway FAD binding occurs. For these experiments, we mechanically unfolded dCRY and allowed the protein to refold at 1 pN for increasing time intervals between 0 and 40 s, using concentrations of FAD of 0, 0.3 nM (~ $Kd_{app}$) and 10 nM (>> $Kd_{app}$). At each time interval and FAD concentration, we calculated the degree of folding from 0 and 1 (Supporting Figure S3).

To determine the number of states along the FAD-dependent folding pathway of dCRY, we applied the Bayesian Information Criterion and Integrated Complete-Data Likelihood tests[38–40] to all of the kinetic refolding data (Supporting Figure S4a-b, Supplementary Information). This global analysis combined with model-based clustering yielded five distinct clusters (labeled C0 through C4) with degrees of folding centered at 0, 0.26 ± 0.01, 0.52 ± 0.01, 0.80 ± 0.01 and 1.0 ± 0.01 (mean ± standard deviation) (Supporting Table S1). The five clusters identified by the statistical analysis are distinctively observed in density plots of the combined refolding data at each FAD concentration (Figure 3a-c). We implemented a bootstrapping procedure to determine

an assignment error of each degree-of-folding data point to one of the five clusters (Supporting Figure S4c). The bootstrapped data was plotted as a function of refolding time and FAD concentration (Figure 3d-f) and globally fitted to various kinetic folding models (Supplementary Information, Supporting Table S2).

Because clusters C0 and C4 with degree of folding 0 and 1 represent the unfolded state and the natively folded, FAD-bound structure, respectively, the remaining clusters C1, C2 and C3 must represent intermediate states. The analysis does not indicate which cluster incorporates FAD-bound and unbound dCRY states. However, the data in Figure 3a provides direct experimental evidence that eliminates possible states along the dCRY folding pathway. For instance, the data at [FAD] = 0 shows that dCRY only samples clusters C0, C1 and C2, indicating that clusters C3 and C4 cannot form without FAD. Moreover, given the extensive inter-molecular interactions established between folded dCRY and FAD[30,31] it is unlikely that the unfolded state binds FAD with high affinity, an observation that is in agreement with the result that the unfolded state is the only cluster that completely disappears with increasing FAD concentration (Figure 2b). Thus, the resulting dCRY folding model includes clusters C1 and C2 as states that can possibly bind FAD (Scheme 1).




(Scheme 1)

In this model, dCRY in the unfolded state (C0 = U) forms a first apo intermediate that can bind FAD (C1 = $I_1$ + $I_{1\cdot FAD}$). A second apo intermediate is formed from $I_1$, which also can form a complex with FAD (C2 = $I_2$ + $I_{2\cdot FAD}$). A third FAD-bound intermediate (C3 = $I_{3\cdot FAD}$) is formed from $I_{2\cdot FAD}$ before reaching the final FAD-bound native state (C4 = $F_{FAD}$). The kinetic parameters in Scheme 1 were globally optimized to describe the time courses for the five clusters at the three concentrations of FAD (Supplementary Information). Binding and conformational equilibrium rate constants were determined between each state, first testing two simpler models in which FAD only binds to $I_1$ or $I_2$, but not both. The resulting chi-squared values and the F-test at a 68% confidence interval were used to determine that the model where FAD binds to $I_1$ and $I_2$ is statistically in better agreement with the data compared to the two simpler models (Supporting Table S2).

The results from the global fitting allowed us to determine the fourteen rate constants that characterize the refolding pathway of dCRY (Scheme 1). We find that $I_1$ is reversibly connected to $U$ with rate constants $k_1 = 0.20 \pm 0.03$ s$^{-1}$ and $k_{-1} = 0.17 \pm 0.05$ s$^{-1}$. The formation of $I_2$ is also reversible with $k_2 = 0.17 \pm 0.03$ s$^{-1}$ and $k_{-2} = 0.06 \pm 0.02$ s$^{-1}$. The reversibility of these two steps along the dCRY folding pathway is consistent with the refolding plot at [FAD] = 0 that shows co-existence between $U$, $I_1$, and $I_2$ at the longest refolding time intervals (Figure 3d). In contrast to the relatively slow forward rate constants of formation of $I_1$ and $I_2$, the FAD binding rate constants for these two intermediates are substantially faster and comparable in magnitude with $k_3 = (2.8 \pm 0.3)\cdot 10^9$ M$^{-1}\cdot$s$^{-1}$ and $k_4 = (10 \pm 1)\cdot 10^9$ M$^{-1}\cdot$s$^{-1}$, respectively. The FAD unbinding rate constants for $I_1$ and $I_2$ are $k_{-3} = 0.82 \pm 0.08$ s$^{-1}$ and $k_{-4} = 1.49 \pm 0.40$ s$^{-1}$, respectively, resulting in microscopic (i.e., state specific) equilibrium dissociation constants ($k_{-3}/k_3$ and $k_{-4}/k_4$) of 0.29 nM and 0.15 nM, which are in quantitative agreement with the apparent dissociation constant obtained in the single

molecule titration experiments (Figure 2b). The conformational rate constant between $I_{1\cdot FAD}$ and $I_{2\cdot FAD}$ are $k_5 = 0.14 \pm 0.03\ s^{-1}$ and $k_{-5} = 0.03 \pm 0.01\ s^{-1}$, resulting in an equilibrium constant of 5.1 that favors the formation of folding intermediates with a larger degree of folding.

After the formation of $I_{2\cdot FAD}$ (either from FAD binding to $I_2$ or via a conformational change from $I_{1\cdot FAD}$) a third FAD-bound intermediate, $I_{3\cdot FAD}$, is formed with $k_6 = 0.9 \pm 0.1\ s^{-1}$ and $k_{-6} = 0.46 \pm 0.09\ s^{-1}$. The fitted parameters for the formation and disappearance of the fully folded state, $F_{FAD}$, are $k_7 = 0.101 \pm 0.008\ s^{-1}$ and $k_{-7} = 0.026 \pm 0.006\ s^{-1}$. We also performed clustering and bootstrapping analysis of the FAD titration data (Supporting Table S1). The statistical analysis and the density plot of the combined FAD titration data (Figure 3g) revealed the same five cluster centers obtained in the kinetic refolding experiments (Figure 3a-c). We fitted the titration data to the model shown in Scheme 1 by itself (Figure 3h) or together with the kinetic refolding data, and obtained rate constants that are in quantitative agreement between the two data sets (Supporting Table S2). Altogether, the statistical and kinetic analysis of the data indicate that dCRY folding is slow and follows a complex pathway, wherein FAD binds fast to early intermediates with partially folded structures. These early intermediates likely represent a minimal structural scaffold for FAD docking to promote the natively folded FAD-bound state.

## Contribution of FAD Moieties to dCRY Folding

To further dissect the mechanism by which FAD promotes the native fold of dCRY, we investigated how the different FAD moieties contribute to dCRY folding. FAD harbors two ring structures that are connected by two phosphate groups in tandem (Figure 4a). The first ring structure is riboflavin, which is composed of the isoalloxazine ring covalently linked to ribitol, a pentose. The other ring is adenosine, composed of ribose and adenine. These moieties have different properties that establish hydrophobic, ionic, and hydrogen bond interactions with dCRY (Figure 4a). In addition, the two phosphate groups in FAD interact with an $Mg^{2+}$ ion. We used optical tweezers to determine the degree of folding of dCRY using various FAD moieties. As in previous experiments, the protein is allowed to refold for 20 s at 1 pN between pulling cycles.

We found that riboflavin (at 0.1 mM) is able to promote a fully folded dCRY structure with a degree of folding of $0.99 \pm 0.01$ nm in 39 % of events. In the other 61 % of events, we obtained a degree of folding of $0.78 \pm 0.01$ (N = 256) (Figure 4b, Supporting Table S3). Interestingly, events corresponding to unfolded or intermediates states with a degree of folding below 0.3 were negligible, indicating the absence of protein in the apo state (Supporting Figure S5). The high percentage of riboflavin-bound intermediates may be due to the fact that riboflavin cannot establish all the intermolecular interactions that FAD allows. We therefore used FMN, composed of riboflavin and one phosphate group, to test the contribution of the phosphate to dCRY folding. At 0.1 mM of FMN, most events displayed a degree of folding of 1 (84 %, N = 158, Figure 4b, Supporting Figure S5). This result indicates that the phosphate group next to riboflavin is critically important for dCRY folding. In agreement with this observation, the dCRY structure shows that this phosphate group establishes two hydrogen bonds with Ser265 (Figure 4a), located in the phosphate binding motif of the protein.

The results obtained with riboflavin and FMN indicate that the isoalloxazine ring is a major contributor to dCRY folding. It may seem, therefore, that the role of the adenosine ring in promoting and reaching the native conformation is minor. We tested the contribution of adenosine (at 0.1 mM) to dCRY folding and observed a wide distribution of intermediates with degree of folding from 0.3 to 0.9 in 75 % of events (N = 286), and the fully folded conformation in 20 % of events (Figure 4c, Supporting Figure S5). Thus, while the adenosine ring is able to promote

complete folding of dCRY, it does so with much less efficiency compared to the isoalloxazine ring in riboflavin or FMN.

**Burial of FAD Phosphates is Energetically Costly but Required for Efficient dCRY Folding**

Because the phosphate group in FMN was found to be critically important in promoting a folded conformation, we tested the presence of one or two phosphate groups in adenosine by using AMP and ADP, respectively. Whereas the distribution of species was similar between adenosine and AMP (Supporting Figure S5), the presence of two phosphates decreased the percentage of fully folded protein from 24 % with AMP to 9 % with ADP (N = 327 for AMP and N =215 for ADP, Figure 4c). Another important effect owed to the second phosphate in ADP is in the observed intermediate states. The events with degree of folding between 0.5-0.9, that reflect bound intermediates based on the kinetic refolding experiments, were much smaller in ADP than in AMP or adenosine (Figure 4c). These results indicate that the second phosphate group has opposing effects depending on whether it is covalently linked to the adenosine ring in ADP (negative folding effects) or to the isoalloxazine ring in FMN (positive folding effects).

Since charges are characteristically incompatible with hydrophobic environments, and only highly stable proteins are found to tolerate engineered charged residues in their hydrophobic core[41], it is possible that the negative effect of ADP on dCRY folding is due to the energetic penalty of partially burying phosphate charges in the protein core. We tested this hypothesis by increasing the number of phosphate groups in adenosine by using ATP. When ATP is 0.1 mM, we observed that dCRY samples the unfolded state (23 %, N = 128) or forms partially folded intermediates (77 %) with degree of folding of up to ~ 0.5 (Supporting Figure S5). No events corresponding to the natively folded state or intermediates with degree of folding > 0.5 were observed (Figure 4b-c). Similar percentages and degrees of folding were obtained in experiments with no FAD, suggesting that ATP at 0.1 mM may not bind to dCRY or can bind and only form early intermediates. Thus, the results with ATP support our interpretation that increasing the number of negative charges has unfavorable effects on the ability of the adenosine ring to promote the native state of dCRY.

It is possible, however, that the additional phosphate group in ATP may not only increase the number of negative charges per adenosine molecule but also generate steric hindrance effects. If steric hindrance were a major force in preventing ATP from interacting and promoting folding in dCRY, then our previous interpretation would be invalid. Therefore, we performed similar experiments with ATP at 0.1 mM but in the presence of saturating $MgCl_2$ = 0.5 mM to reduce the net negative charge of the molecule [42–44]. If ATP-$Mg^{2+}$ and ATP alone have similar effects on dCRY folding, then steric hindrance may be the dominant force in inhibiting ATP to interact and induce dCRY to fold. We found, however, that the percentage of events corresponding to intermediates with degree of folding below 0.5 reduces to 24 % with $Mg^{2+}$ from 77 % with no $Mg^{2+}$ (Figure 4d). And, accordingly, the fraction of events with degree of folding between 0.5-0.8, corresponding to ATP-bound intermediates, increases from 0 to 45 %. Only 11 % of events corresponded to the fully folded state with degree of folding larger than 0.9 (N = 114). These observations indicate that the negative folding effects of the phosphate groups in adenosine can be mitigated by the presence of $Mg^{2+}$. In fact, the crystal structure of dCRY shows a single $Mg^{2+}$ ion interacting with both phosphate groups of ADP (Figure 4a). It is therefore possible that the role of the $Mg^{2+}$ ion is to reduce the energetic penalty of burying a negative charge in FAD. In support of this observation, the addition of $MgCl_2$ (0.5 mM) to ADP (0.1 mM) also increases the fraction of ADP-bound intermediates, albeit to a lesser degree compared to ATP (Figure 4d, Supporting Table S3).

Interestingly, it is well-established that plant CRYs bind ATP and other nucleotides in a cavity close to FAD[45,46]. Binding of ATP to this secondary site increases the quantum yield of the signaling state, facilitates protonation of the flavin when it gets photoreduced, and has been proposed to stabilize conformational changes in the α/β domain important for signaling[47–52]. It is not known whether dCRY has the secondary binding site for ATP. Our studies show that dCRY-ATP interactions mostly induced the formation of partially folded intermediates. Further studies using mixtures of FAD and ATP may help to elucidate if dCRY has a secondary site for ATP that is important for function but less relevant for folding in the native state.

## Mapping Folding and FAD-Bound Intermediates Along the dCRY Folding Pathway

To gain greater insight into the folding pathway between U and $F_{FAD}$ we sought to identify possible structures for the intermediates that were identified in the kinetic refolding experiments (Scheme 1). As in previous mechanical manipulation studies of single molecules[24,53–57] we estimated the intermediate structures by comparing the experimentally determined $\Delta Lc_T$ to values expected from folding different parts of the protein (Supporting Figure S6, Supplementary Information). This approach considers that loops and random coils are unstable against mechanical force, and that secondary structures are stable when they are intact or fully folded, i.e., intermediates ending in the middle of β-strands or α-helices are energetically disfavored[48,50]. For dCRY, the values of $\Delta Lc_T$ for the three intermediates were estimated from the clustering analysis that yielded degrees of folding of 0.26, 0.52, 0.80 for clusters C1, C2, and C3, respectively. These values correspond to 140 amino acids for cluster C1 composed of $I_1$ and $I_{1\cdot FAD}$, 275 amino acids for cluster C2 composed of $I_2$ and $I_{2\cdot FAD}$, and 448 amino acids for cluster C3 composed of $I_{3\cdot FAD}$. (Supporting Table S4, Supporting Figure S6).

The model we propose describes the unfolded polypeptide undergoing a stepwise refolding process to attain its native, FAD-bound structure (Figure 5). We find that cluster C1 ($I_1$ and $I_{1\cdot FAD}$) matches the size of 140 residues in the N-terminal α/β domain of the protein (Figure 5b). Importantly, this is the only structural domain in dCRY that does not make any direct interactions with FAD, and our refolding studies show that this intermediate can fold in the absence of the cofactor (Scheme 1). We propose that cluster C2 ($I_2$ and $I_{2\cdot FAD}$) incorporates the N-terminal domain up to residue Leu275 (Figure 5c). This reasoning is based on the large number of unfolding trajectories in cluster C2 that include a short-lived transition of degree of folding of 0.26 that match cluster C1 or the PHR (data not shown). Cluster C2 includes Arg237 and the phosphate-binding domain (residues 249-263), which make direct interactions with FAD (Figure 4a). Among these residues, Arg237, Lys264, Ser265, and Met266 also interact with the magnesium ion observed in the dCRY structure[30,31].

Cluster C3 or $I_{3\cdot FAD}$ maps to residues starting from the N-terminus to Asp448, which harbor all the motifs that form the FAD-binding pocket in dCRY, including the conserved 4-helix motif common to photolyase and DNA primase[34], and the surrounding loops called the C-terminal coupled motif or CCM (Figure 5d). The 4-helix motif binds the isoalloxazine ring and the CCM completes the interaction network with FAD through Phe280, Gln311, His378, Arg381, Phe404, Asp412, Val415, Cys416, and Asn419 (Figure 4a). As stated for cluster C2, unfolding trajectories of cluster C3 also include short-lived transitions that matched the size of the PHR (data not shown). After the formation of $I_{3\cdot FAD}$, the remaining residues require to attain the native state correspond to the C-terminal linker and the C-terminal tail (CTT). This step is expected to occur at the end of the folding pathway since the C-terminal linker and the CTT dock on top of FAD and in between the scaffold formed by the CCM (Figure 5e). Moreover, the last step of folding of the CTT is consistent with functional studies that showed that signaling in dCRY involves the displacement

of the CTT upon changes in flavin redox state[29]. Hence, its association with the protein core via the CCM is probably the least stable folding step of dCRY, consistent with our analysis of optical tweezers experiments.

Interestingly, both $I_1$ and $I_2$ bind to FAD with similarly high affinities, but their cofactor binding mechanisms are likely distinct. For instance, $I_1$ does not include folded residues that directly interact with FAD, while $I_2$ has pre-formed a first structural scaffold for direct contact with FAD. Cofactor interactions with $I_1$ seem to follow an induced-fit binding mechanism, where $I_{1\cdot FAD}$ may promote the formation of a better-defined FAD binding site, as mapped for $I_{2\cdot FAD}$. Noteworthy, $I_1$ is almost 75 % unfolded, including all residues involved in FAD binding, and cofactor binding still occurs with very fast kinetics ($2.8\cdot10^9$ $M^{-1}\cdot s^{-1}$). Several binding studies involving intrinsically disordered proteins (IDPs) have reported association rates around the estimated diffusion limit for folded proteins ($10^9$-$10^{10}$ $M^{-1}\cdot s^{-1}$)[51–55]. Such fast association kinetics for IDPs has been explained by "fly-casting" effects, where the unfolded polypeptide forms initial interactions with its binding partner at a greater distance than a folded protein[58–60]. It is plausible that the large fraction of unfolded polypeptide seen in $I_1$ binds FAD with fast kinetics following a fly-cast effect, as seen in IDPs. In contrast, cofactor interactions with $I_2$ likely follow a conformational selection mechanism[61], where FAD selects a state with a pre-formed, albeit not fully structured, binding pocket. Given the degree of burial of FAD in the fully folded dCRY structure, and the fact that dCRY dynamics are mostly localized to the CTT upon changes in flavin redox state[27,28], binding of FAD to intermediate states seems a more likely scenario, as opposed to a model in which FAD binds after a fully folded structure forms, which would require large conformational changes or significant unfolding to then incorporate the cofactor.

Altogether, our results indicate that FAD binding to dCRY involves multiple mechanisms, as reported for other proteins[62,63], including proteins that bind FMN, a related cofactor[64]. It is possible that topologically complex proteins with slow folding kinetics as seen for dCRY have evolved to bind a cofactor to different intermediates along its folding pathway to prevent the formation of misfolded states that could accumulate and lead to protein aggregation.

## CONCLUSIONS

A large fraction of proteins harbor inorganic (i.e., metal clusters) or organic (i.e., FMN, FAD, hemes, among others) cofactors in their structure. While the functional role of protein cofactors has been well described and characterized for decades, how cofactor interactions contribute to the protein fold, conformation and stability is not well understood. Previous bulk and single molecule studies have investigated how a cofactor interacts with small, single-domain proteins[15–19]. These studies have shown that a protein can either interact with the cofactor in the unfolded state, forming and stabilizing an intermediate before reaching the native state[65–67], or fold and reach the native state before binding the cofactor[17]. However, mechanistic studies on cofactor interactions coupled to folding for large proteins with multiple domains are lacking. Recently, the application of single molecule optical tweezers has opened new opportunities to study the folding mechanisms of large proteins or protein complexes[22–26]. Here, we used optical tweezers to study folding of dCRY, a multidomain protein of 542 amino acid residues that binds to one of the most common and complex organic cofactors, FAD.

The molecular trajectories of the unfolding behavior of dCRY under force displayed large heterogeneity, suggesting multiple intermediates along the folding pathway of dCRY (Figure 1). Given the complexity of the unfolding trajectories, seen with and without FAD, we established a novel statistical framework based on clustering and bootstrapping procedures (Supporting Figure

S4) that enabled the quantification and mechanistic characterization of such heterogeneous data. While clustering allowed the deconvolution of the optimal number of intermediate states during the folding pathway of dCRY (Supporting Table S1, Figure 3a-c, 3g), the addition of a bootstrapping procedure allowed the determination of an error associated with the population of each observed state under a specific experimental condition (Figure 3d-f, 3h). Both sets of information were used to globally fit[68] different models of folding coupled to FAD binding, which resulted in the quantification of the 14 binding and conformational rate constants that reversibly connect the 7 states shown in the dCRY folding mechanism (Scheme 1). Importantly, the folding mechanism (Scheme 1) provided the best fitting statistics for the kinetic refolding data alone, the FAD titration data alone, or both data sets simultaneously, underscoring the robustness of the statistical procedures and fitting analysis. We envision that the data analysis procedures described in this study will help others in their quantification of single molecule folding data that display high degree of complexity and heterogeneity.

The folding mechanism of dCRY revealed multiple intermediates that can bind FAD and lead to the native FAD-bound structure. Surprisingly, we find that FAD is strictly required for dCRY to attain its native state, i.e., without FAD the protein never folds to its biologically functional structure. In fact, in the absence of FAD, dCRY only forms two intermediates with 26% and 52% of folded polypeptide. WLC analysis indicates that the second intermediate with 52% of folded polypeptide incorporates a minimal folded scaffold for FAD binding. However, the folded structures that map to the first intermediate correspond to the N-terminal $\alpha/\beta$ domain which does not establish direct contacts with the cofactor. Nonetheless, FAD binds to the first and second intermediates similarly fast, displaying association rate constants above the diffusion limit established for globular proteins[69]. We propose that FAD binds to these intermediates following two different mechanisms: an induced-fit binding mechanism for the first intermediate, which has been reported for IDPs that bind ligands or protein partners following fly-casting effects[58], and a conformational selection binding mechanism[61] to the second intermediate, in which FAD selects an intermediate state with a preformed, yet precursory, cofactor binding pocket.

Given the large size of dCRY, it was expected that the unfolded polypeptide in the apo state samples a large number of heterogenous folding pathways, which would reduce in number as the protein progressively establishes native-like contacts[69]. Instead, the data revealed well-defined apo and FAD-bound pathways, with intermediates that share similar degrees of folding with and without the cofactor (Scheme 1). These intermediates may represent cooperative units within the structure of dCRY, which for other proteins have been defined as foldons[70]. Our data, therefore, suggests that dCRY has a defined pathway that progressively folds and binds FAD. It is possible that FAD has evolved to stabilize pre-existing foldons, thereby decreasing the likelihood of misfolding while at the same time promoting intermediate states with larger degrees of folding until the native states is reached.

By testing the effect of various FAD moieties on the degree of folding of dCRY (Figure 4a) we determined that the isoalloxazine ring with ribitol and one phosphate (i.e., FMN) is sufficient to drive the native state of the protein (Figure 4b). In contrast, the adenosine ring covalently linked to one or two phosphates (i.e., AMP and ADP, respectively) does not promote native folding to the same degree, but mostly forms intermediates. Unexpectedly, the adenosine ring without the phosphates is more efficient than AMP or ADP at promoting folding (Figure 4b-c). This finding suggests that the partial burial of negative charges by dCRY is energetically penalized. We show that $Mg^{2+}$ ions that coordinate to the phosphate of ADP overcome such energetic penalty, resulting in a higher percentage of bound intermediates or the folded state. We further tested this model by determining the percent of fully folded events using ATP, with and without $Mg^{2+}$. We found a small, yet significant increase in fully folded events when the ion is

present (Figure 4d). Altogether, this study dissected how the different moieties in FAD contribute to the folding of dCRY in a non-additive fashion. Moreover, the fact that various FAD moieties interact with dCRY and promote folding to varying degrees indicates that initial contacts between unfolded dCRY and FAD can be established with different moieties, increasing the probability of productive folding.

Interestingly, the mammalian clock CRYs have a homologous structure, yet they fold without FAD and carry out their functions void of the cofactor[71,72]. There is experimental evidence *in vitro* that flavin binding is linked to their function, but it is not conclusive[73–75]. However, as a principal photoreceptor, we show that dCRY strictly requires FAD to fold into its functional structure. In this context, our studies allow us to speculate about the evolution of cofactor binding. It is possible that the protein interacted with the cofactor in an unfolded state, and then folded around the cofactor, as seen between FAD and dCRY. As the cofactor locked and optimized this interaction, the protein could fold without the cofactor as functions diverged, which is the case for mammalian CRYs. Thus, the structural template, or CRY fold, became stable without the need of the cofactor. Comparative single molecule studies of folding of mammalian CRYs with and without FAD may help in better understanding how cofactor binding and folding are evolutionarily linked.

## ONLINE MATERIALS AND METHODS

dCRY modified with the Avi and ybbR tags was purified as previously described[30]. Preparation of DNA handles and covalent modification of dCRY is described in Supplementary Information. Optical tweezers experiments were performed in a MiniTweezers instrument[76] in dCRY buffer (50 mM Tris pH 8.0, 150 mM NaCl, 10 mM DTT) supplemented with desired final concentration of FAD or other cofactors (i.e., riboflavin, FMN, ATP, ADP, AMP, and Adenosine with or without $MgCl_2$). For each condition, at least 6 different molecules were sampled for a total of > 150 molecular trajectories. Force ramp experiments were performed at a constant pulling rate of 75 nm/s sampled at 200 Hz to a maximum unfolding force of 45 pN. Formation of a single tether was confirmed by overstretching of the DNA handle up to ~ 65 pN[37]. Sample preparation, optical tweezers protocols and data analysis of titration experiments (with FAD and cofactors) and kinetic refolding experiments are described in detail in Supplementary Information.

## ACKNOWLEDGMENTS

The authors wish to acknowledge Joanne Widom from the Crane lab for cloning and expression of the original tagging enzymes and modified CRY. This work was supported by NSF grants MCB1715572 (to R.A.M.), MCB-1412624 and MCB-1817749 (to A.L.L.), and NIH grants 1R15GM135866 (to R.A.M.) and R35GM122535 (to B.R.C). L.G. acknowledges support from the Clare Boothe Luce foundation.

## AUTHOR CONTRIBUTIONS

S.F. designed, conducted, and analyzed the research, and wrote the manuscript. L.G., Z.I. and M.T. analyzed the research. C.F. conducted research and edited the manuscript. A.L.L. analyzed the research. B.R.C. analyzed the research and wrote the manuscript. R.A.M. designed and analyzed the research, and wrote the manuscript.

## REFERENCES


1. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–666 (1958).
2. Blake, C. *et al.* The three-dimensional structure of hen eggwhite lysozyme at 2 A resolution. *Nature* **206**, 757–761 (1965).
3. PERUTZ, M. F. Structure of hemoglobin. *Brookhaven Symp. Biol.* **13**, 165–183 (1960).
4. Pan, X. & Kortemme, T. Recent advances in de novo protein design: Principles, methods, and applications. *J. Biol. Chem.* **296**, 100558 (2021).
5. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
6. Yip, K. M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).
7. Kuhlman, B. & Bradley, P. Advances in protein structure prediction and design. *Nat. Rev. Mol. Cell Biol.* **20**, 681–697 (2019).
8. Gershenson, A., Gosavi, S., Faccioli, P. & Wintrode, P. L. Successes and challenges in simulating the folding of large proteins. *J. Biol. Chem.* **295**, 15–33 (2020).
9. Wittung-Stafshede, P. Role of cofactors in protein folding. *Acc. Chem. Res.* (2002) doi:10.1021/ar010106e.
10. Culbertson, D. S. & Olson, J. S. Role of heme in the unfolding and assembly of myoglobin. *Biochemistry* (2010) doi:10.1021/bi1006942.
11. Verma, K., Kundu, D., Kundu, L. M., Singh, A. K. & Dubey, V. K. Folding and stability of recombinant azoreductase enzyme from Chromobacterium violaceum. *Enzyme Microb. Technol.* **131**, 109433 (2019).
12. Houwman, J. A., Westphal, A. H., Visser, A. J. W. G., Borst, J. W. & Van Mierlo, C. P. M. Concurrent presence of on- and off-pathway folding intermediates of apoflavodoxin at physiological ionic strength. *Phys. Chem. Chem. Phys.* **20**, 7059–7072 (2018).
13. Bollen, Y. J. M., Sánchez, I. E. & van Mierlo, C. P. M. Formation of on- and off-pathway intermediates in the folding kinetics of Azotobacter vinelandii apoflavodoxin. *Biochemistry* **43**, 10475–10489 (2004).
14. Muralidhara, B. K. & Wittung-Stafshede, P. Thermal unfolding of Apo and Holo Desulfovibrio desulfuricans flavodoxin: cofactor stabilizes folded and intermediate states. *Biochemistry* **43**, 12855–12864 (2004).
15. Cao, Y. & Li, H. Dynamics of protein folding and cofactor binding monitored by single-molecule force spectroscopy. *Biophys. J.* (2011) doi:10.1016/j.bpj.2011.08.051.
16. Lei, H. *et al.* Reversible Unfolding and Folding of the Metalloprotein Ferredoxin Revealed by Single-Molecule Atomic Force Microscopy. *J. Am. Chem. Soc.* **139**, 1538–1544 (2017).
17. Goedken, E. R., Keck, J. L., Berger, J. M. & Marqusee, S. Divalent metal cofactor binding in the kinetic folding trajectory of Escherichia coli ribonuclease HI. *Protein Sci.* **9**, 1914–1921 (2000).
18. Pozdnyakova, I., Guidry, J. & Wittung-Stafshede, P. Copper stabilizes azurin by decreasing the unfolding rate. *Arch. Biochem. Biophys.* (2001) doi:10.1006/abbi.2000.2369.

19. Giannotti, M. I. *et al.* Direct Measurement of the Nanomechanical Stability of a Redox Protein Active Site and Its Dependence upon Metal Binding. *J. Phys. Chem. B* **119**, 12050–12058 (2015).

20. Batey, S., Nickson, A. A. & Clarke, J. Studying the folding of multidomain proteins. *HFSP J.* **2**, 365–377 (2008).

21. Walters, B. T., Mayne, L., Hinshaw, J. R., Sosnick, T. R. & Englander, S. W. Folding of a large protein at high structural resolution. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 18898–18903 (2013).

22. Jahn, M., Buchner, J., Hugel, T. & Rief, M. Folding and assembly of the large molecular machine Hsp90 studied in single-molecule experiments. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1232–1237 (2016).

23. Liu, K., Chen, X. & Kaiser, C. M. Energetic dependencies dictate folding mechanism in a complex protein. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 25641–25648 (2019).

24. Hao, Y. *et al.* Activation of a Protein Kinase Via Asymmetric Allosteric Coupling of Structurally Conserved Signaling Modules. *bioRxiv* (2019) doi:10.1101/611772.

25. England, J. P. *et al.* Switching of the folding-energy landscape governs the allosteric activation of protein kinase A. *Proc. Natl. Acad. Sci. U. S. A.* (2018) doi:10.1073/pnas.1802510115.

26. Bauer, D. *et al.* A folding nucleus and minimal ATP binding domain of Hsp70 identified by single-molecule force spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 4666–4671 (2018).

27. Lin, C., Top, D., Manahan, C. C., Young, M. W. & Crane, B. R. Circadian clock activity of cryptochrome relies on tryptophan-mediated photoreduction. *Proc. Natl. Acad. Sci. U. S. A.* (2018) doi:10.1073/pnas.1719376115.

28. Wang, Y., Veglia, G., Zhong, D. & Gao, J. Activation mechanism of Drosophila cryptochrome through an allosteric switch. *Sci. Adv.* **7**, (2021).

29. Chandrasekaran, S. *et al.* Tuning flavin environment to detect and control light-induced conformational switching in Drosophila cryptochrome. *Commun. Biol.* **4**, 249 (2021).

30. Zoltowski, B. D. *et al.* Structure of full-length Drosophila cryptochrome. *Nature* **480**, 396–9 (2011).

31. Levy, C. *et al.* Updated structure of Drosophila cryptochrome. *Nature* **495**, E3–E4 (2013).

32. Sancar, A. Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem. Rev.* (2003) doi:10.1021/cr0204348.

33. Michael, A. K. *et al.* Formation of a repressive complex in the mammalian circadian clock is mediated by the secondary pocket of CRY1. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1560–1565 (2017).

34. Sauguet, L., Klinge, S., Perera, R. L., Maman, J. D. & Pellegrini, L. Shared active site architecture between the large subunit of eukaryotic primase and DNA photolyase. *PLoS One* **5**, e10083 (2010).

35. Yin, J. *et al.* Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15815–15820 (2005).

36. Chen, I., Howarth, M., Lin, W. & Ting, A. Y. Site-specific labeling of cell surface proteins

with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99–104 (2005).

37. Bustamante, C., Marko, J. F., Siggia, E. D. & Smith, S. Entropic elasticity of λ-phage DNA. *Science* (1994) doi:10.1126/science.8079175.

38. Biernacki, C. & Gilles Celeux, and G. G. Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. *J. Tissue Viability* **8**, 30 (1998).

39. Scrucca, L., Fop, M., Murphy, T. B. & Raftery, A. E. mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J.* **8**, 289–317 (2016).

40. Banfield, J. D. & Raftery, A. E. Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics* (1993) doi:10.2307/2532201.

41. Isom, D. G., Castañeda, C. A., Cannon, B. R., Velu, P. D. & García-Moreno E, B. Charges in the hydrophobic interior of proteins. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16096–16100 (2010).

42. Dudev, T., Grauffel, C. & Lim, C. How Native and Alien Metal Cations Bind ATP: Implications for Lithium as a Therapeutic Agent. *Sci. Rep.* **7**, 42377 (2017).

43. Wilson, J. E. & Chin, A. Chelation of divalent cations by ATP, studied by titration calorimetry. *Anal. Biochem.* **193**, 16–19 (1991).

44. Gupta, R. K., Gupta, P., Yushok, W. D. & Rose, Z. B. Measurement of the dissociation constant of MgATP at physiological nucleotide levels by a combination of 31P NMR and optical absorbance spectroscopy. *Biochem. Biophys. Res. Commun.* **117**, 210–216 (1983).

45. Bouly, J.-P. *et al.* Cryptochrome blue light photoreceptors are activated through interconversion of flavin redox states. *J. Biol. Chem.* **282**, 9383–9391 (2007).

46. Engelhard, C. *et al.* Cellular metabolites enhance the light sensitivity of Arabidopsis cryptochrome through alternate electron transfer pathways. *Plant Cell* **26**, 4519–4531 (2014).

47. Müller, P. & Bouly, J.-P. Searching for the mechanism of signalling by plant photoreceptor cryptochrome. *FEBS Lett.* **589**, 189–192 (2015).

48. Immeln, D., Schlesinger, R., Heberle, J. & Kottke, T. Blue light induces radical formation and autophosphorylation in the light-sensitive domain of Chlamydomonas cryptochrome. *J. Biol. Chem.* **282**, 21720–21728 (2007).

49. Hense, A., Herman, E., Oldemeyer, S. & Kottke, T. Proton transfer to flavin stabilizes the signaling state of the blue light receptor plant cryptochrome. *J. Biol. Chem.* **290**, 1743–1751 (2015).

50. Schroeder, L., Oldemeyer, S. & Kottke, T. Time-Resolved Infrared Spectroscopy on Plant Cryptochrome-Relevance of Proton Transfer and ATP Binding for Signaling. *J. Phys. Chem. A* **122**, 140–147 (2018).

51. Goett-Zink, L., Toschke, A. L., Petersen, J., Mittag, M. & Kottke, T. C-Terminal Extension of a Plant Cryptochrome Dissociates from the β-Sheet of the Flavin-Binding Domain. *J. Phys. Chem. Lett.* **12**, 5558–5563 (2021).

52. Iwata, T. *et al.* ATP binding promotes light-induced structural changes to the protein moiety of Arabidopsis cryptochrome 1. *Photochem. Photobiol. Sci. Off. J. Eur. Photochem. Assoc. Eur. Soc. Photobiol.* **19**, 1326–1331 (2020).

53. Maillard, R. A. *et al.* ClpX(P) generates mechanical force to unfold and translocate its

protein substrates. *Cell* (2011) doi:10.1016/j.cell.2011.04.010.

54. Stigler, J., Ziegler, F., Gieseke, A., Gebhardt, J. C. M. & Rief, M. The complex folding network of single calmodulin molecules. *Science (80-. ).* (2011) doi:10.1126/science.1207598.

55. Dietz, H. & Rief, M. Protein structure by mechanical triangulation. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 1244–1247 (2006).

56. Sen Mojumdar, S. *et al.* Partially native intermediates mediate misfolding of SOD1 in single-molecule folding trajectories. *Nat. Commun.* (2017) doi:10.1038/s41467-017-01996-1.

57. Perez-Jimenez, R., Garcia-Manyes, S., Ainavarapu, S. R. K. & Fernandez, J. M. Mechanical unfolding pathways of the enhanced yellow fluorescent protein revealed by single molecule force spectroscopy. *J. Biol. Chem.* **281**, 40010–40014 (2006).

58. Huang, Y. & Liu, Z. Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the 'fly-casting' mechanism. *J. Mol. Biol.* **393**, 1143–1159 (2009).

59. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8868–8873 (2000).

60. Mollica, L. *et al.* Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Front. Mol. Biosci.* **3**, 52 (2016).

61. Vogt, A. D. & Di Cera, E. Conformational selection or induced fit? A critical appraisal of the kinetic mechanism. *Biochemistry* **51**, 5894–5902 (2012).

62. Daniels, K. G. *et al.* Ligand concentration regulates the pathways of coupled protein folding and binding. *J. Am. Chem. Soc.* **136**, 822–825 (2014).

63. Hammes, G. G., Chang, Y.-C. & Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13737–13741 (2009).

64. Muralidhara, B. K., Rathinakumar, R. & Wittung-Stafshede, P. Folding of Desulfovibrio desulfuricans flavodoxin is accelerated by cofactor fly-casting. *Arch. Biochem. Biophys.* **451**, 51–58 (2006).

65. Pozdnyakova, I. & Wittung-Stafshede, P. Biological relevance of metal binding before protein folding. *J. Am. Chem. Soc.* **123**, 10135–10136 (2001).

66. Bertini, I., Cowan, J. A., Luchinat, C., Natarajan, K. & Piccioli, M. Characterization of a partially unfolded high potential iron protein. *Biochemistry* **36**, 9332–9339 (1997).

67. Pozdnyakova, I., Guidry, J. & Wittung-Stafshede, P. Copper-triggered β-hairpin formation initiation site for azurin folding? [26]. *J. Am. Chem. Soc.* **122**, 6337–6338 (2000).

68. Ingram, Z. M., Scull, N. W., Schneider, D. S. & Lucius, A. L. Multi-start Evolutionary Nonlinear OpTimizeR (MENOTR): A hybrid parameter optimization toolbox. *Biophys. Chem.* **279**, 106682 (2021).

69. Alberty, R. A. & Hammes, G. G. Application theory diffusion-controlled enzyme. **62**, 154–159 (1957).

70. Englander, S. W. & Mayne, L. The case for defined protein folding pathways. *Proc. Natl.*

*Acad. Sci. U. S. A.* **114**, 8253–8258 (2017).

71. Chaves, I. *et al.* The cryptochromes: blue light photoreceptors in plants and animals. *Annu. Rev. Plant Biol.* **62**, 335–364 (2011).

72. Rosensweig, C. *et al.* An evolutionary hotspot defines functional differences between CRYPTOCHROMES. *Nat. Commun.* **9**, 1138 (2018).

73. Xing, W. *et al.* SCF FBXL3 ubiquitin ligase targets cryptochromes at their cofactor pocket. *Nature* (2013) doi:10.1038/nature11964.

74. Hirano, A., Braas, D., Fu, Y.-H. & Ptáček, L. J. FAD Regulates CRYPTOCHROME Protein Stability and Circadian Clock in Mice. *Cell Rep.* **19**, 255–266 (2017).

75. Chaves, I. *et al.* The Potorous CPD photolyase rescues a cryptochrome-deficient mammalian circadian clock. *PLoS One* **6**, e23447 (2011).

76. Smith, S. B., Cui, Y. & Bustamante, C. Optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol.* (2003) doi:10.1016/S0076-6879(03)61009-8.

# Main Figures

# Figure 1

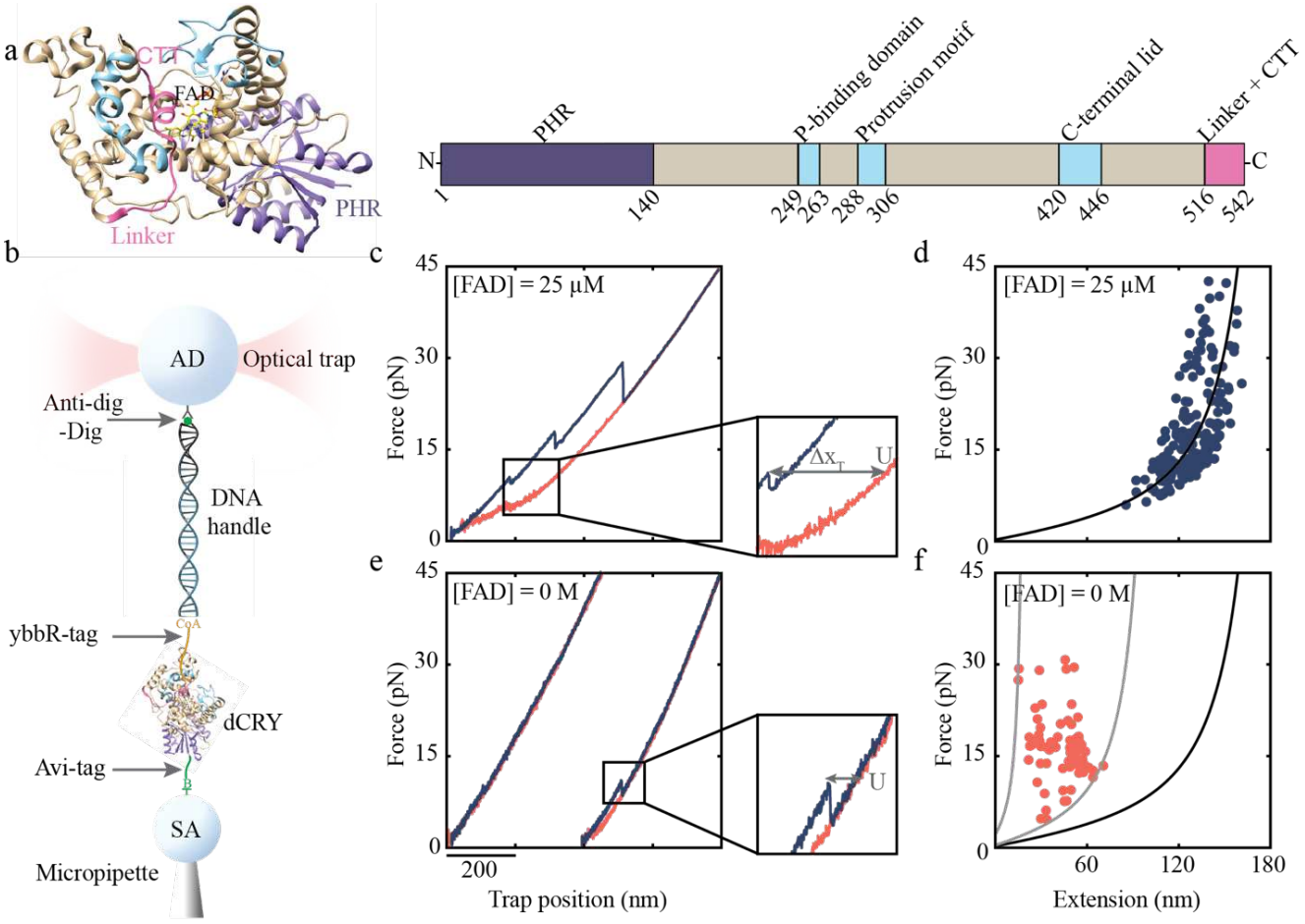



**Figure 1. Experimental design to study dCRY folding with optical tweezers**. a- Structure of full-length dCRY (left) and domain organization (right) (PDB ID: 4GU5). b- Schematic representation of the optical tweezers set up to mechanically manipulate a single dCRY molecule. c- Force-extension trajectory of the mechanical unfolding (blue) and refolding (red) of dCRY in the presence of [FAD] = 25 uM. Zoomed-in is the total change in extension ($\Delta x_T$) measurement from folded to unfolded state. d- Worm-like chain (WLC) analysis of $\Delta x_T$ vs. force of dCRY with [FAD] = 25 µM. The solid line corresponds to the WLC model for full-length dCRY with a contour length of 203.7 nm and folded distance of 5.5 nm (Supplementary Information) . e- Force-extension trajectories of dCRY in the absence of FAD with no unfolding rips (left) and partially folded structures (right). Zoomed-in is the change in unfolding event. f- WLC analysis of $\Delta x_T$ vs. force for unfolding and refolding of the protein in the absence of FAD with three WLC curves using a contour length of 27 nm and 120 nm (gray lines). The black line corresponds to the WLC for fully folded dCRY.

# Figure 2

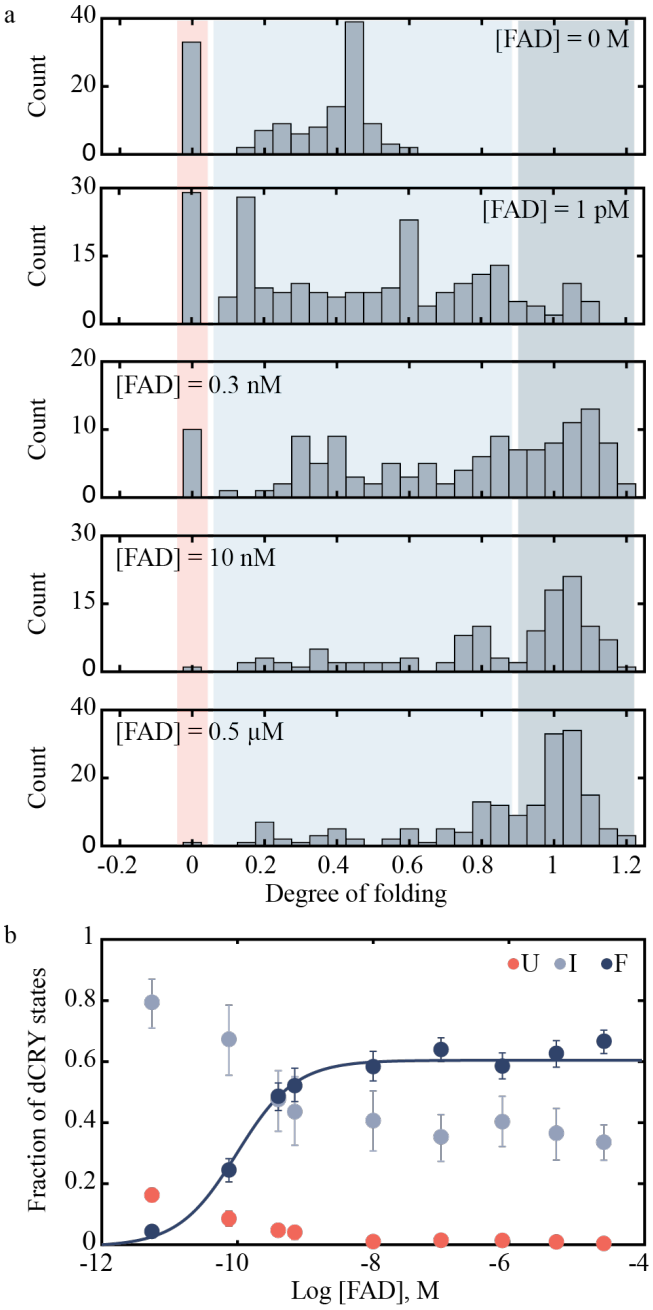



**Figure 2. Single-molecule FAD titration.** a- Histograms of population of dCRY states vs. degree of folding in the presence of varying FAD concentrations. Unfolded, intermediate and folded states are shaded in light red, light blue and grey. b- Fraction of dCRY states (unfolded = U, intermediates = I and folded = F) plotted as a function of FAD concentration. The folded fraction was fitted to a single site binding isotherm with a $Kd_{app}$ = 0.11 ± 0.04 nM.

# Figure 3

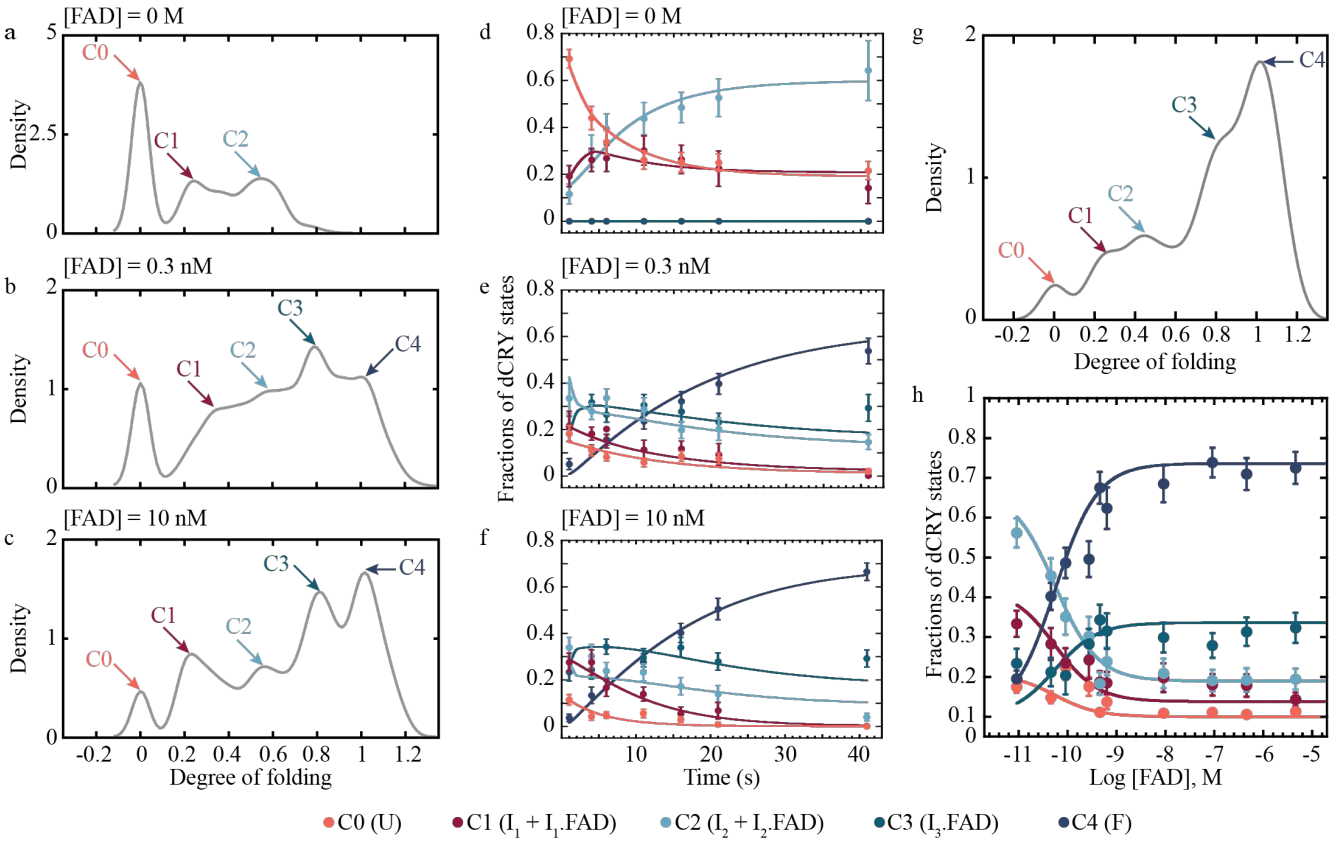



**Figure 3. Kinetic analysis of FAD binding coupled to dCRY folding.** a-c- Density plots of degree of folding of combined data at [FAD] = 0, 0.3 nM and 10 nM. The combined density plots reveal five distinct clusters (C0 thru C4) with degree of folding centered at 0, 0.26 ± 0.01, 0.52 ± 0.01, 0.80 ± 0.01 and 1.0 ± 0.01, respectively (Supporting Table S1). d-f- Plots of bootstrapped refolding data as a function of time at each FAD concentration. The data was globally fitted following the mechanism shown in Scheme 1 (main text). g- Density plot of combined FAD titration data displays the same five cluster centers as in the kinetic refolding data shown in a-c. h- FAD titration plot of clustered and bootstrapped data from FAD titration experiments. The solid lines represent a global fitting of scheme 1. Fitted parameters are listed in Supporting Table S2.

# Figure 4

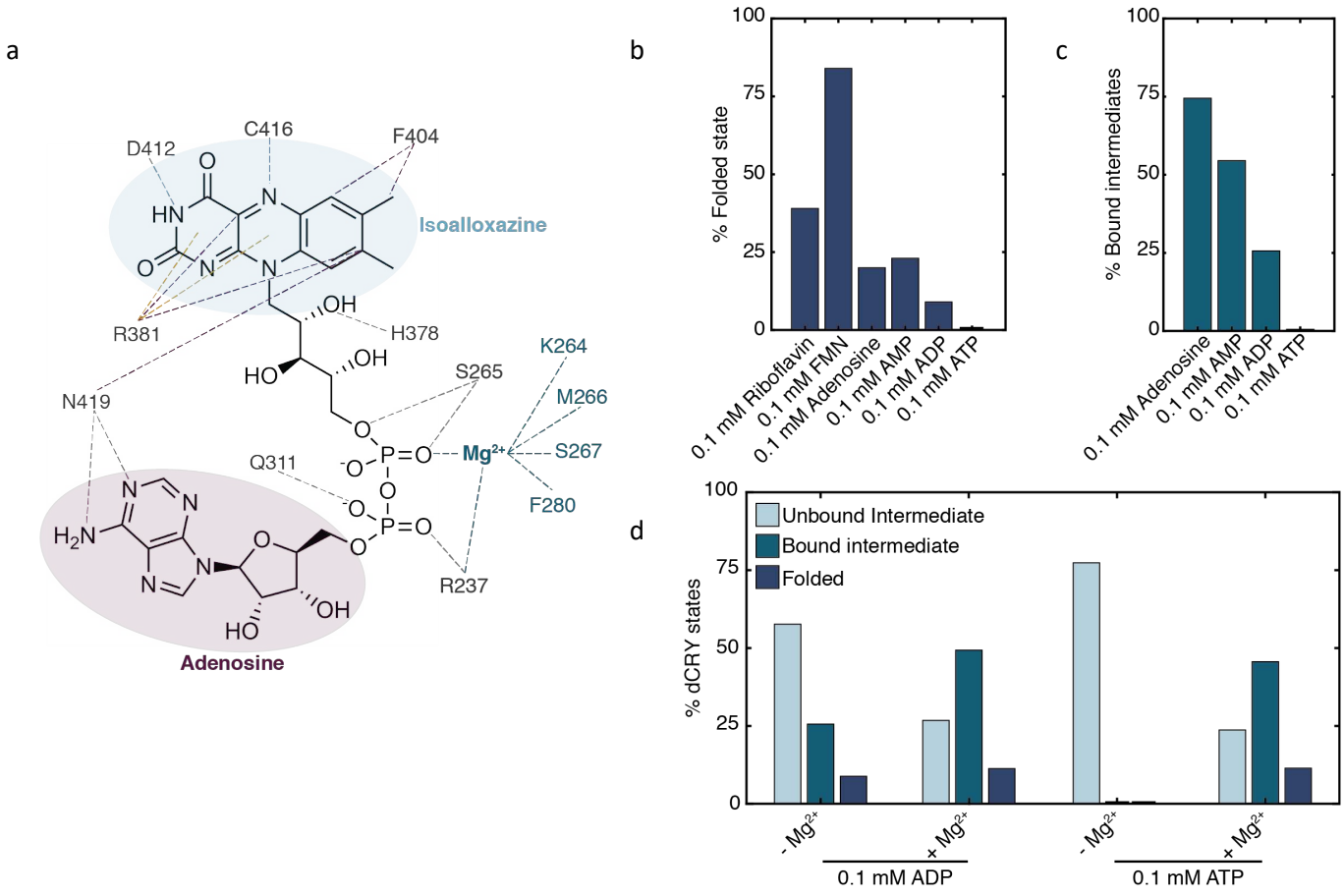



**Figure 4. Contribution of FAD moieties to dCRY folding.** a- Contact map between FAD moieties and amino acid residues in dCRY. Interactions include hydrogen bonding (grey), hydrophobic effect (purple), metal coordination (dark teal) and cation-pi interactions (ochre). b- Bar plot representing the percentage of dCRY in the folded state in the presence of different cofactors. c- Bar plot of percentage of dCRY intermediate states bound to adenosine, AMP, ADP and ATP. d- Bar plots of percentage of dCRY in an unbound intermediate, bound intermediate or folded state in the presence of ADP or ATP, with and without $Mg^{2+}$.

# Figure 5

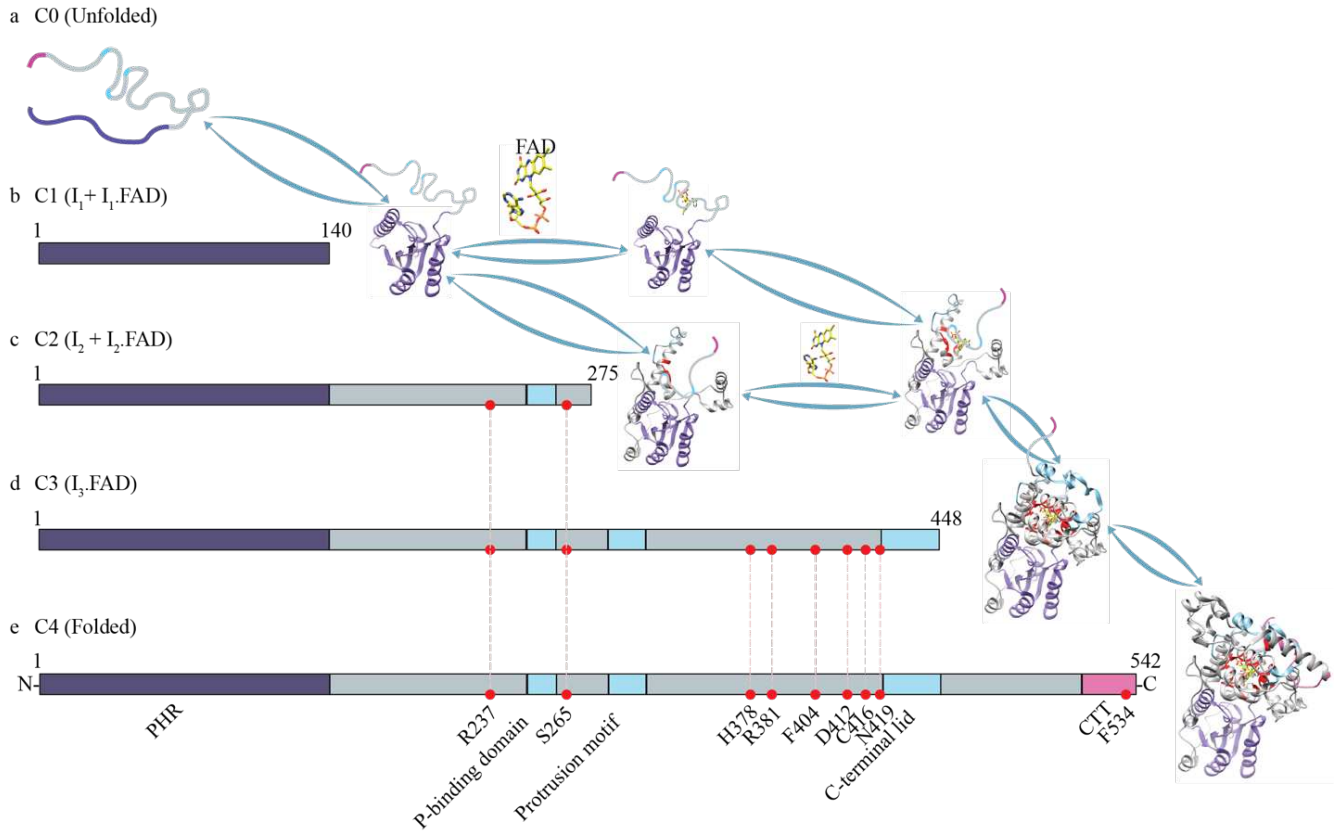



**Figure 5. Mapping intermediate structures and FAD binding events along the dCRY folding pathway.** a-,b- The unfolded polypeptide folds into the first intermediate ($I_1$) that matches with the size of PHR (purple). c- $I_1$ either binds FAD to form $I_1$·FAD or folds into $I_2$, which encompasses the Phosphate binding motif. d- $I_2$ can bind FAD to form $I_2$·FAD, which then folds into $I_3$·FAD that includes folding of the protrusion motif and the C-terminal lid. Together with the phosphate binding motif, $I_3$·FAD has the C-terminal coupled motif (CCM) folded, which include the entire network of residues that interact with FAD. e- In the last step, $I_3$·FAD folds into the dCRY native structure and the C-terminal tail (CTT) docks on top of FAD.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 01312022SupportingInformation.pdf