2013

Exploration of Hinge Residues among GFP-like Proteins

Binsen Li
Connecticut College, libinsen@gmail.com

Exploration of Hinge Residues among GFP-like Proteins

Binsen Li

Submitted to the Department of Biology

of Connecticut College

in partial fulfillment of the requirements for graduation with

GENERAL HONORS

Honors Project Advisor: Marc Zimmer, Ph.D.

May, 2013

Table of Contents Page

Abstract:

Before the maturation of the chromophore, a fluorescent protein, just like any other three-dimensional protein structure, has to fold into the correct tertiary structure to function. We proposed that the evolutionarily conserved hinge residues, believed to be located on or near the lids of the fluorescent protein, are involved in the folding mechanism and rotation of the β-sheets and lids into the correct geometry. Acting like door hinges, these residues are translationally immobilized but rotationally active. Interference of the hinge sites may lead to allosteric effects and disruption of the protein's functional motions. In the study, significant sequential and spatial conservation was found in conserved lid residues across 28 wild-type fluorescent proteins. Furthermore, the high rotational freedom and dihedral mobility of the conserved lid residues confirmed their behavior as hinge residues in the folding process.

Introduction:

I. General Background

Since Osamu Shimomura purified the first wild-type green fluorescent protein (GFP) from the crystal jellyfish *Aequorea victoria* (*av*GFP, PDB:1GFL) in the 1960s, an expanding family of fluorescent proteins (FPs) has been discovered. Fluorescent proteins, the innovative biomarkers in scientific research, became prevalent within decades. Scientists have sequenced over 150 distinct FPs and crystalized 28 wild type structures, which derived from various marine species including anemones, corals and jelly fish. *(1)* Despite the fact that the proteins can be traced back to 150 distinct species, they share a great deal of amino acid similarity. By simply expressing FPs next to targeted proteins, biological researchers have been maximizing FPs' imaging capability, which has already made revolutionary changes in biological research. For example,

- With the help of the FPs, researchers are able to observe the pattern and distribution of various proteins. For example, various pathological proteins, such as amyloid β-protein that abnormally aggregate during the deterioration of Alzheimer's disease, can be easily monitored by the newly designed fluorescence protein sensor. *(2)*
- Similarly, other proteins, such as matrix metalloproteinases (MMPs) that are involved in stroke, can be visualized through an MMP-activatable probe, which shows significantly higher levels of fluorescence in the stroke affected areas. *(3)*

- The energy transfer between the donor chromophore and an acceptor chromophore nonradiative dipole-dipole coupling, fluorescence (or Förster) resonance energy transfer (FRET) can be used to monitor protein-protein interactions. *(4)*
- Cell cycle activity also can be lit up through fluorescence. Fucci, a fluorescent, ubiquitination-based cell cycle indicator designed by Atsushi Miyawaki from the RIKEN Basic Science Institute, is able to color the progression of the cell cycle with different FPs therefore allowing researchers to visualize the cells in different phases of the cell cycle. *(5)*
- In cancer treatment, the inability to efficiently follow and evaluate cancer metastasis has limited the effectiveness of cancer therapies. By tagging cancer cells with GFPs, researchers are able to trace the metathesis easily. *(6)*

II. Fluorescent Protein Structure

It is widely known that the shape of the *av*GFP resembles a can structure made up of 11 β-sheets from 238 amino acids (Figure 1). Located at the core of the protein, the SYG tripeptide chromophore can be seen as the "heart" of the GFP.

The immature chromophore undergoes a 3-step autocatalytic maturation reaction (cyclization, oxidation and dehydration) of the SYG tripeptide, which means no foreign enzyme is involved in this process. Two mechanisms were proposed by Getzoff and Wachter (Figure 2).

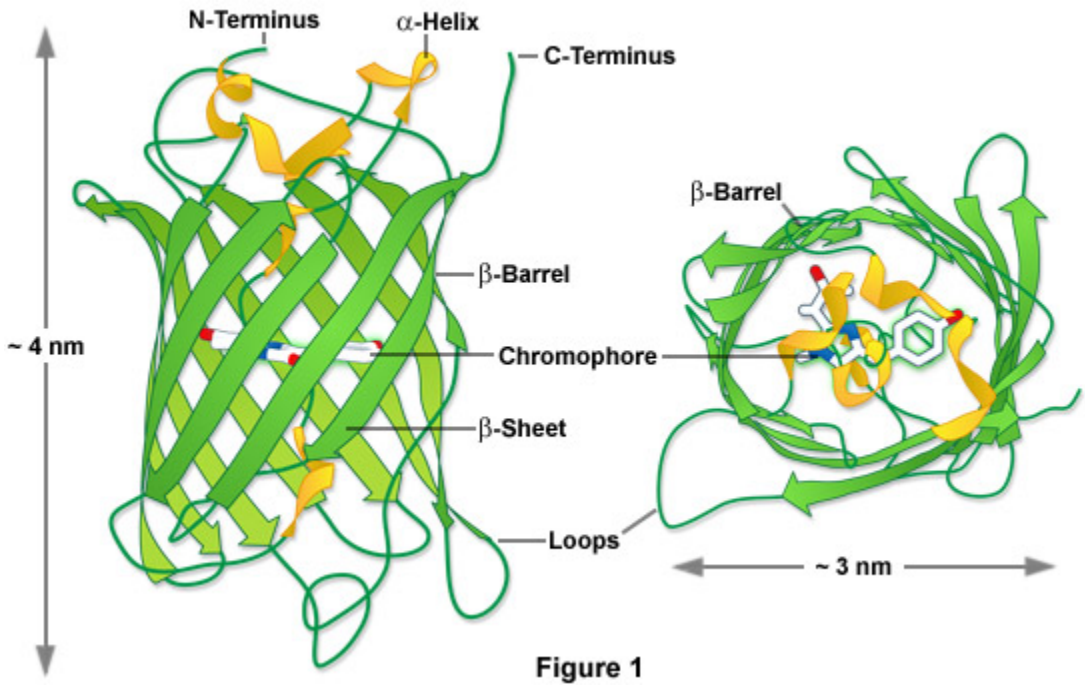

Figure 1. The Architecture of *Aequorea Victoria* Green Fluorescent Protein *(7)*

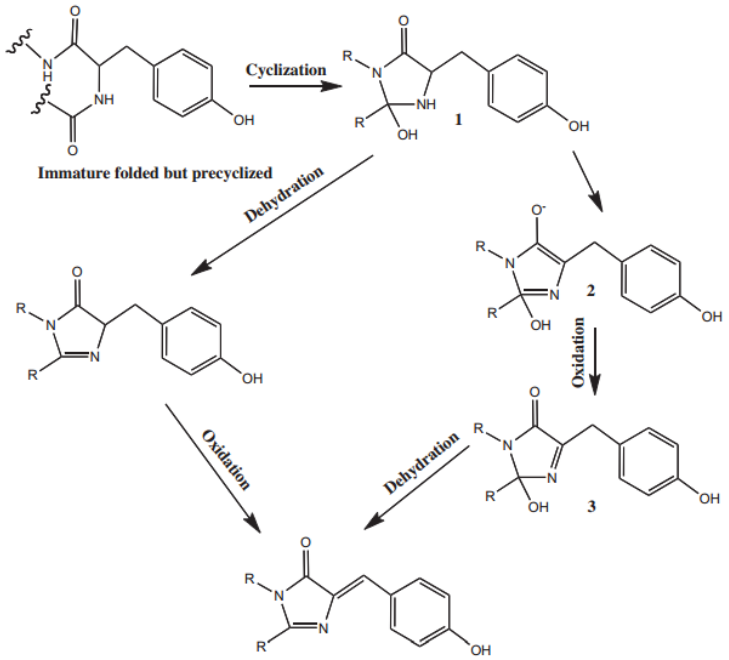

Figure 2. Two chromophore maturation mechanism proposed by Getzoff (left) and Wachter (right). *(8, 9)*

III. Conserved Residues Amongst the Fluorescent Proteins

Over 150 distinct GFP-like proteins are currently known. GFP-like fluorescent proteins (FPs) have been found in marine organisms ranging from chordates (e.g. amphioxus) to cnidarians (e.g. corals and sea pansies). Figure 3 shows the most conserved residues in the structures of the naturally occurring GFP-like proteins. We can divide the conserved residues into 3 groups; residues involved in chromophore formation, residues on and around the "lids" of the β-barrel (underlined on the figure), and centrally located residues with no known function.

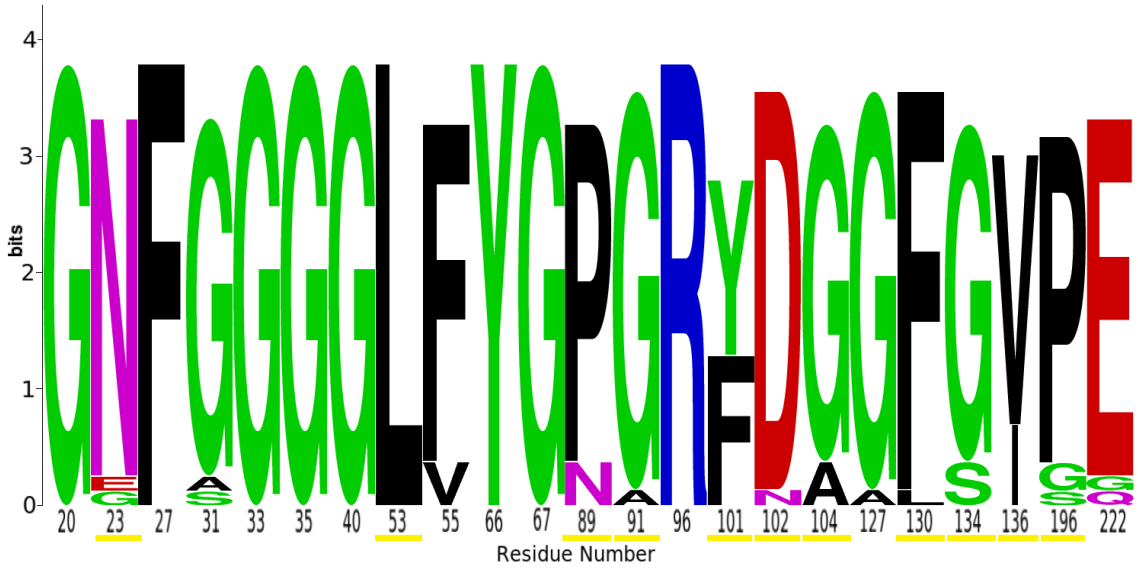

Figure 3: A weblogo representation of amino acid diversity among the most conserved residues of the wild-type GFP-like structures. Underlined are the conserved lid residues. Residues are numbered relative to *av*GFP. A high bit score (y-axis) on the logo plot reflects invariant residues. See Figure 4 for location of some of these residues.

Tyr66, Gly67, Arg96 and Glu222 are involved in chromophore formation and we have recently examined their role in the post-translational green fluorescent protein chromophore formation. (*10*) Gly31, Gly33 and Gly35 are all located on the second β-strand of the 11 β-strand barrel, see Figure 4. There is no obvious reason why these residues are conserved. They are the only conserved residues located on β-strands that are not involved in chromophore formation. The residues are not part of the pore implicated in chromophore maturation (*11*) that is located between sheets 7 and 8. They are also highly conserved in all 266 GFP-like pdb structures and no analogous GXGXG conserved stretch is found in the 16 sheet β-barrels of porins. A 50ns molecular dynamics simulation of *av*GFP revealed that second β-sheet is not significantly more flexible than the other β-sheets. *(1)*

In this thesis we will focus on the function of the conserved lid residues.

IV. The Conserved Lid Residues

The "lid" residues are located at both ends of the GFP's can structure, and they are mainly made up of short turns, helices and ends of sheets. A previous study has shown that the lid residues are notably conserved among all 266 FPs structures in the PDB. *(1)* Furthermore the C and N termini end, which is deemed as the bottom lid in this study, is relatively less conserved in comparison to the top lid. Three conserved residues (89, 91 and 196) are located on the bottom lid and 12 residues (20, 23, 27, 53, 55, 101, 102, 104, 127, 130, 134 and 136) are spread across the top lid. We are curious whether the conserved lids are related to the protein binding function or GFP

folding. In this thesis we will focus on the possibility that the conserved residues are hinge residues crucial to the fluorescent proteins' folding.

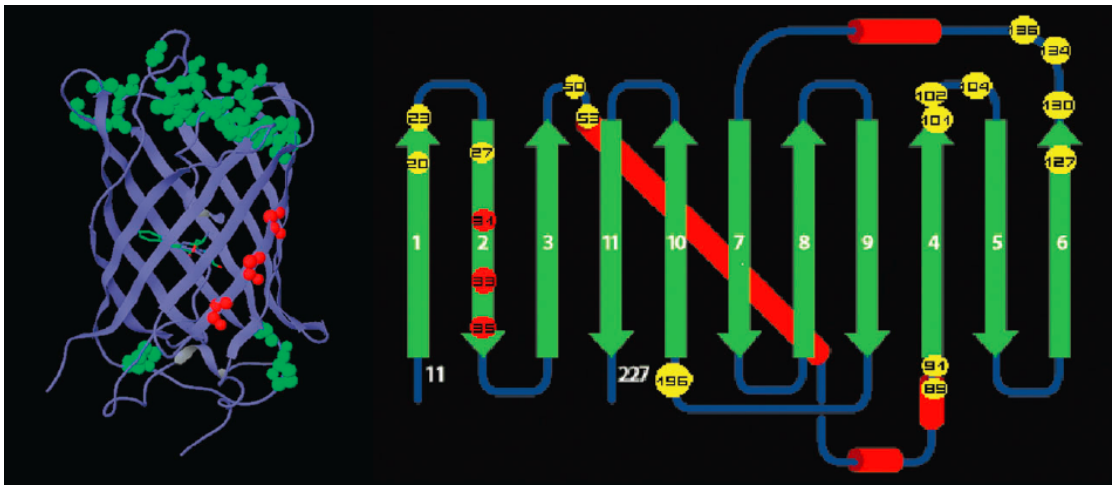

Figure 4. Location of conserved amino acids on β-sheets (31, 33, 35) and lids (23, 50, 53, 89, 91, 101, 102, 104, 130, 134, 136, 196)

V. Folding of Fluorescent Proteins

Before the maturation of the chromophore, the fluorescent protein, just like any other three-dimensional protein structure, has to fold into the correct tertiary structure to be able to function. Different approaches including chemical (urea or guanidinium chloride), thermal and physical methods have been applied to denature the FPs and explore the folding/unfolding.

Bertz et al (*12*) applied a mechanical force to unfold GFP. To collect more evidence about the intermediates, two GFP mutants with disulfide cross-links were pulled apart by forces from opposite directions. Their work revealed the unfolding pathways of GFP shown in Figure 5.

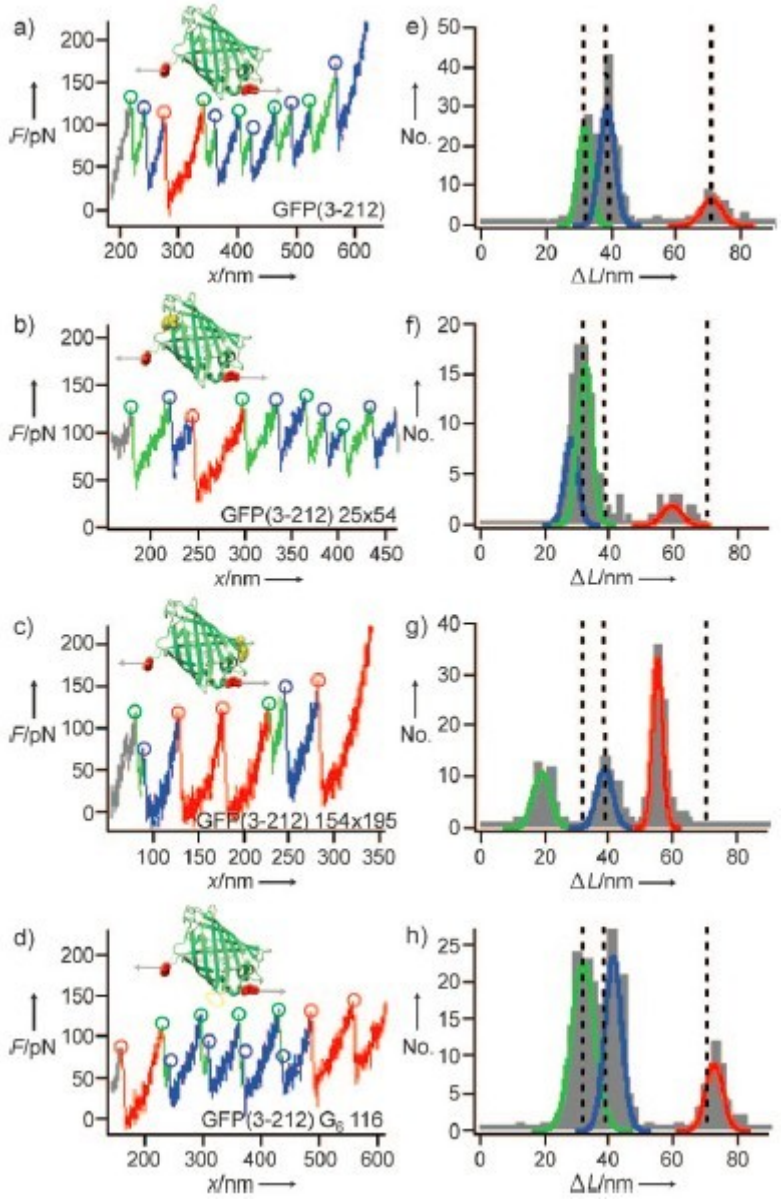

Figure 5. Force-extension graph of two GFP mutants: Green, blue, and red circles indicate partial GFP fracture, unfolding of the intermediate and complete fracture. *(12)*

Reddy et al.(*13*) have combined the results from their coarse grained simulations with reported experimental observations(*14, 15, 16, 17*) to propose a model for GFP folding. It includes multiple pathways, passing through kinetic and equilibrium intermediates as well as misfolded structures. During folding and unfolding the β-sheets move in distinct groups - an N-terminal (β1-3), central (β4-6), and C-terminal (β7-11), group (Figure 6). Rewiring GFP and changing the order of β-strands 1-6 appears to result in misfolding and failure to form a chromophore.

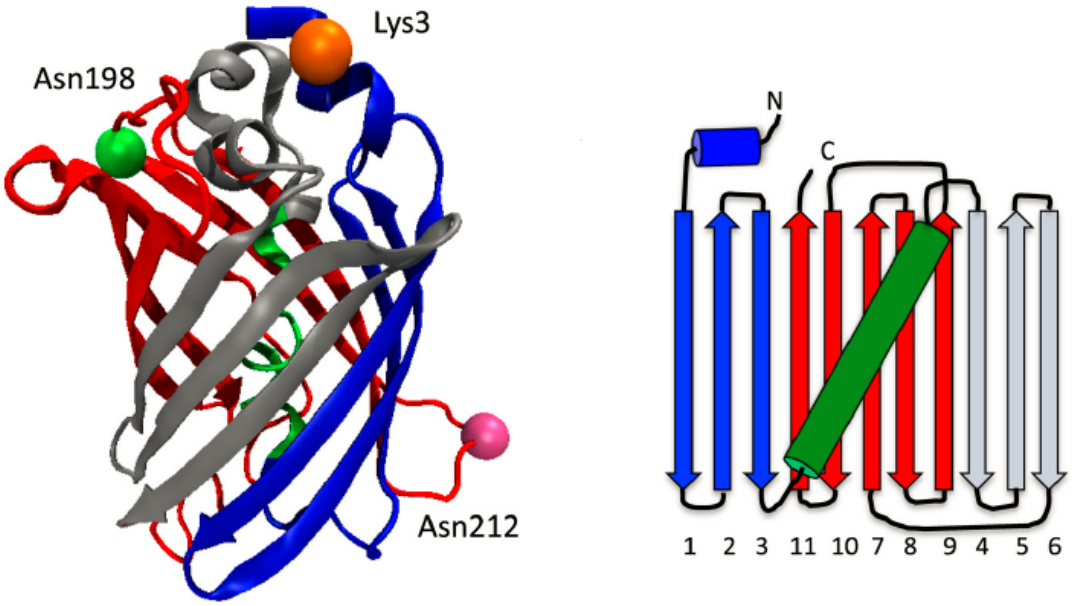

Figure 6. Citrine, a yellow version of fluorescent protein's structure is shown: blue strands are the N terminus β-strands, red are the C terminus β-strands, grey are the other 3 β-strands and the green helix is the central helix that includes the chromophore.

As shown in Figure 7, several equilibrium intermediates in the folding process were found by the simulations, which are similar to the results of chemical denaturant GdmCl experiment. The equilibrium shows fairly slow unfolding kinetics and therefore high energy barriers in protein unfolding.

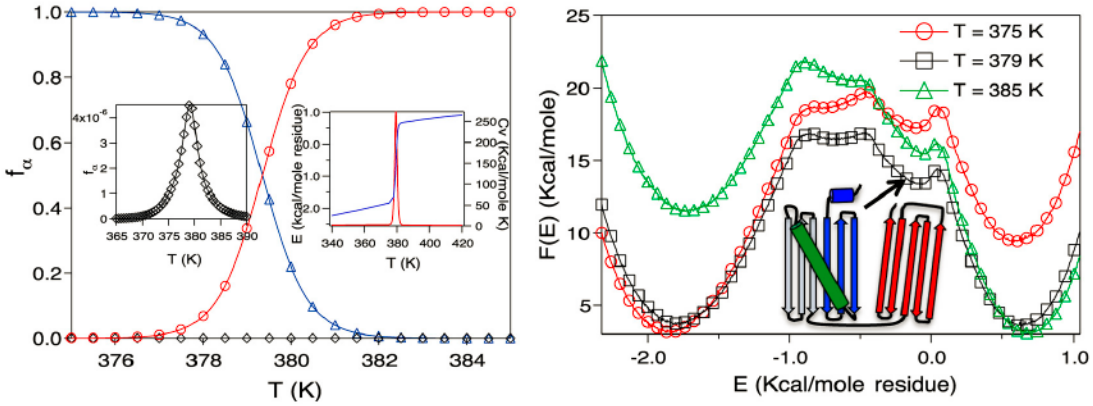

Figure 7. Left: Fraction of GFP in the Native Basin of Attraction (triangles), equilibrium intermediates (diamonds) and UBA (circles). Left right: energy, E, and the heat capacity, Cv as a function of T. Right: Free energy of GFP at temperature around melting point. One of the intermediate is shown.

Summing across all the results, one GFP folding landscape was proposed as in Figure 8, which shows 4 folding pathways between different states.

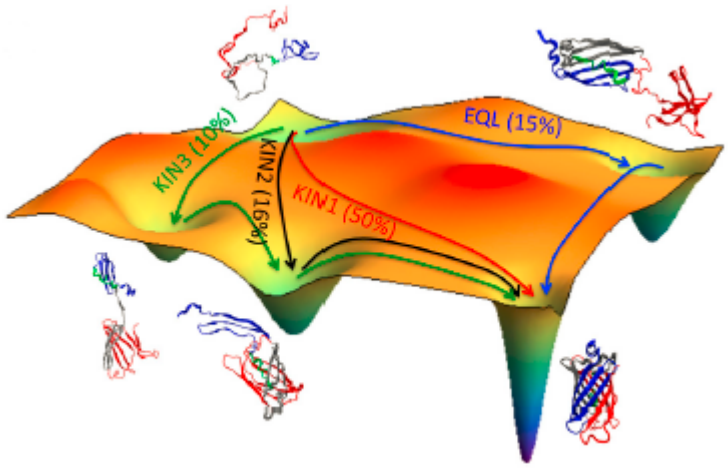

Figure 8. GFP folding energy landscape and network of pathways between intermediates

Due to the fact that GFP tends to misfold with its β-barrel structure, especially when tagged with other proteins, a great variety of mutants were made to enhance its folding ability. Superfolder GFP, which is believed to be the one of the fastest folding GFPs, carries eleven mutations: the cycle-3 mutations (F99S, M153T and V163A), 2 enhanced GFP mutations (F64L and S65 T) and 6 folding enhanced mutations (S30R, Y39N, N105T, Y145F, I171V and A206V). *(18)* Although the cycle 3 mutations do not affect the folding kinetics of the protein, they are able to reduce the hydrophobicity and protein aggregation. S30R and Y39N mutations are believed to increase folding efficiency while Y145F and I174V can reduce the misfolding. N105T and A206V in some way can increase the β-forming propensity. These mutants of Superfolder GFP show great capability in folding and give off bright fluorescence,

even if fused with a poorly soluble protein, which causes problems binding with current GFP variants.

Studies on Superfolder GFP have shown a rough energy landscape in the unfolding and refolding process: the unfolding titration curves constantly change over time but the corresponding refolding curves do not. (*19*)

Additionally it has been found that a GFP mutant without the chromophore shows smoother energy landscape and superimposable refolding and unfolding equilibrium curves than the mature forms. The *de novo* folding indicates that chromophore is somehow involved in the energetic frustration of the unfolding and refolding of the GFP. This is the so-called folding hysteresis.

Circular permutated GFP, in which the C and N termini were joined with a spacer and new C and N termini were formed at residues 144 and 145, has altered pKa values. However, the protein was still able to retain fluorescence, which indicated the occurrence of correct folding and chromophore maturation. GFP's tolerance for circular permutation presents that the folding process of GFP is amazingly robust. (*20*)

In *Aequorea victoria* GFP (*av*GFP) the protein folds quickly and GFP refolding from an acid-, base-, or guanidine HCl-denatured state (chromophore containing but non-fluorescent) occurs with a half-life of between 24 seconds(*21*) and 5 minutes(*22*) and the recovered fluorescence is indistinguishable from that of native GFP.(*23)* The folding of GFP exhibits hysteresis that is due to the decreased flexibility of the chromophore vs. its immature analog(*24, 25*), and the compaction of the β-barrel

upon chromophore formation.(*26*) All signs of hysteresis disappear in mutants that do not form the chromophore.(*25*)

Our goal for this thesis is to explore the existence of hinge residues among the conserved lid residues in FPs.

Materials and Methods:

I. 28 Wild-type Fluorescent Proteins

All the wild-type fluorescent proteins structures imported from the RCSB PDB Bank are listed below. (*27*)

Table 1. The information and sources of 28 wild type GFP-like proteins

Protein	PDB Code	Organism	Reference
Green Fluorescent Protein	1gfl	Aequorea victoria	Yang, F., Moss, L.G. and Phillips Jr., G.N. The molecular structure of green fluorescent protein. (1996) Nat.Biotechnol. 14: 1246-1251
A blue, nonfluorescent pocilloporin	1mou	Montipora efflorescens	M., Ling, M., Beddoe, T., Oakley, A.J., Dove, S., Hoegh-Guldberg, O., Devenish, R.J. and Rossjohn, J. The 2.2 A crystal structure of a pocilloporin pigment reveals a nonplanar chromophore conformation.Prescott. (2003) Structure 11: 275-284
A far-red fluorescent protein, eqFP611	1uis	Entacmaea quadricolor	Petersen, J., Wilmann, P.G., Beddoe, T., Oakley, A.J., Devenish, R.J., Prescott, M. and Rossjohn, J. The 2.0-A crystal structure of eqFP611, a far red fluorescent protein from the sea anemone Entacmaea quadricolor. (2003) J.Biol.Chem. 278: 44626-44631
KikGR	1xss	Favia favus	Tsutsui, H., Karasawa, S., Shimizu, H., Nukina, N. and Miyawaki, A. Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. (2005) Embo Rep. 6: 233-238
A far-red Fluorescent Protein, HcRed	1yzw	Heteractis crispa	Wilmann, P.G., Petersen, J., Pettikiriarachchi, A., Buckle, A.M., Smith, S.C., Olsen, S., Perugini, M.A., Devenish, R.J., Prescott, M. and Rossjohn, J. The 2.1A crystal structure of the far-red fluorescent protein HcRed: inherent conformational flexibility of the chromophore. J. (2005) J.Mol.Biol. 349: 223-237
Discosoma Red Fluorescent Protein (DsRed)	1zgo	Discosoma sp.	Tubbs, J.L., and Tainer, J.A., and Getzoff, E.D.Crystallographic structures of Discosoma red fluorescent protein with immature and mature chromophores: linking peptide bond trans-cis isomerization and acylimine formation in chromophore maturation. (2005) Biochemistry 44: 9833-9840
Green to red photoconvertible GFP-like protein, EosFP	1zux	Lobophyllia hemprichii	Nienhaus, K., Nienhaus, G.U., Wiedenmann, J. and Nar, H. Structural basis for photo-induced protein cleavage and green-to-red conversion of fluorescent protein EosFP. (2005) Proc.Natl.Acad.Sci.Usa 102: 9156-9159
A cyan fluorescent protein, amFP486	2a46	Anemonia majano	Henderson, J.N. and Remington, S.J. Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from Anemonia majano. (2005) Proc.Natl.Acad.Sci.Usa 102: 12712-12717
GFP asFP499	2c9i	Anemonia sulcata	Nienhaus, K., Renzi, F., Vallone, B., Wiedenmann, J. and Nienhaus, G.U. Chromophore-protein interactions in the anthozoan green fluorescent protein asFP499. (2006) Biophys.J. 91: 4210-4220

GFP cmFP512	2c9j	Cerianthus membranaceus	Nienhaus, K., Renzi, F., Vallone, B., Wiedenmann, J. and Nienhaus, G.U. Exploring chromophore--protein interactions in fluorescent protein cmFP512 from Cerianthus membranaceus: X-ray structure analysis and optical spectroscopy. (2006) Biochemistry 45: 12492-12953
a GFP-like protein, CpYGFP	2dd7	Chiridius poppei	Suto, K., Masuda, H., Takenaka, Y., Tsuji, F.I. and Mizuno, H. Structural basis for red-shifted emission of a GFP-like protein from the marine copepod Chiridius poppei. (2009) Genes Cells 14: 727-737
CopGFP	2g3o	Pontellina plumata	Wilmann, P.G., Battad, J., Petersen, J., Wilce, M.C.J., Dove, S., Devenish, R.J., Prescott, M. and Rossjohn, J. The 2.1A crystal structure of copGFP, a representative member of the copepod clade within the green fluorescent protein superfamily. (2006) J.Mol.Biol. 359: 890-900
A GFP turns red with UV radiation, Kaede	2gw3	Trachyphyllia geoffroyi	Hayashi, I., Mizuno, H., Tong, K.I., Furuta, T., Tanaka, F., Yoshimura, M. and Miyawaki, A., Ikura, M.Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein Kaede. (2007) J.Mol.Biol. 372: 918-926
A blue chromoprotein	2ib5	Epiactis japonica	Chan, M.C.Y., Karasawa, S., Mizuno, H., Bosanac, I., Ho, D., Prive, G.G., Miyawaki, A. and Ikura, M. Structural characterization of a blue chromoprotein and its yellow mutant from the sea anemone Cnidopus japonicus. (2006) J.Biol.Chem. 281: 37813-37819
A red fluorescent protein zRFP574	2icr	Zoanthus sp.	Pletneva, N., Pletnev, V., Tikhonova, T., Pakhomov, A.A., Popov, V., Martynov, V.I., Wlodawer, A., Dauter, Z. and Pletnev, S. Refined crystal structures of red and green fluorescent proteins from the button polyp Zoanthus. (2007) Acta Crystallogr.,Sect.D 63: 1082-1093
A photoswitchable GFP, Dronpa	2ie2	Echinophyllia sp. SC22	Wilmann, P.G., Turcic, K., Battad, J.M., Wilce, M.C.J., Devenish, R.J., Prescott, M. and Rossjohn, J. The 1.7 A crystal structure of Dronpa: a photoswitchable green fluorescent protein. (2006) J.Mol.Biol. 364: 213-224
A yellow fluorescent protein, zYFP538	2ogr	Zoanthus sp.	Pletneva, N.V., Pletnev, S.V., Chudakov, D.M., Tikhonova, T.V., Popov, V.O., Martynov, V.I., Wlodawer, A., Dauter, Z. and Pletnev, V.Z. Three-dimensional structure of yellow fluorescent protein zYFP538 from Zoanthus sp. at the resolution 1.8 angstrom. Bioorg.Khim. 33: 421-430
GFP-like fluorescent chromoprotein FP506	2ojk	Zoanthus sp.	Pletneva, N., Pletnev, V., Tikhonova, T., Pakhomov, A.A., Popov, V., Martynov, V.I., Wlodawer, A., Dauter, Z. and Pletnev, S. Refined crystal structures of red and green fluorescent proteins from the button polyp Zoanthus. (2007) Acta Crystallogr.,Sect.D 63: 1082-1093
luciferase's accessory GFP, RrGFP	2rh7	Renilla reniformis	Loening, A.M., Fenn, T.D. and Gambhir, S.S. Crystal structures of the luciferase and green fluorescent protein from Renilla reniformis. (2007) J.Mol.Biol. 374: 1017-1028
Red fluorescent protein mKeima	2wht	Montipora sp. 20	Violot, S., Carpentier, P., Blanchoin, L. and Bourgeois, D. Reverse pH-dependence of chromophore protonation explains the large Stokes shift of the red fluorescent protein mKeima. (2009) J.Am.Chem.Soc. 131: 10356
Photoswitchable GFP-like protein Dronpa	2z6x	Pectiniidae	Mizuno, H., Mal, T.K., Walchli, M., Kikuchi, A., Fukano, T., Ando, R., Jeyakanthan, J., Taka, J., Shiro, Y., Ikura, M. and Miyawaki, A. Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. (2008) Proc.Natl.Acad.Sci.Usa 105: 9227-9232
Orange-Emitting GFP-like Protein Monomeric Kusabira Orange (mKO)	2zmu	Verrillofungia concinna	Kikuchi, A., Fukumura, E., Karasawa, S., Mizuno, H., Miyawaki, A. and Shiro, Y. Structural characterization of a thiazoline-containing chromophore in an orange fluorescent protein, monomeric Kusabira Orange. (2008) Biochemistry 47: 11573-11580
A Cyan Fluorescent Protein, dsFP483	3cgl	Discosoma striata	Malo, G.D., Wang, M., Wu, D., Stelling, A.L., Tonge, P.J. and Wachter, R.M. Crystal structure and Raman studies of dsFP483, a cyan fluorescent protein from Discosoma striata. (2008) J.Mol.Biol. 378: 869-884

A photosensitizer KillerRed	3gb3	Anthomedusae sp. DC-2005	Pletnev, S., Gurskaya, N.G., Pletneva, N.V., Lukyanov, K.A., Chudakov, D.M., Martynov, V.I., Popov, V.O., Kovalchuk, M.V., Wlodawer, A., Dauter, Z. and Pletnev, V. Structural basis for phototoxicity of the genetically encoded photosensitizer KillerRed. (2009) J.Biol.Chem. 284: 32028-32039
Fluorescent protein FP480	3h1o	Entacmaea quadricolor	Pletnev, S., Morozova, K.S., Verkhusha, V.V. and Dauter, Z. Rotational order-disorder structure of fluorescent protein FP480. (2009) Acta Crystallogr.,Sect.D 65: 906-912
Orange-Emitting GFP-like Protein, Monomeric Kusabira-Orange (MKO)	3mgf	Verrillofungia concinna	Ebisawa, T., Yamamura, A., Ohtsuka, J., Kameda, Y., Hayakawa, K., Nagata, K. and Tanokura, M. Crystal Structure of Monomeric Kusabira-Orange (MKO), Orange-Emitting GFP-like Protein, at pH 7.5. Journal: To be Published
Red fluorescent protein eqFP578	3pib	Entacmaea quadricolor	Pletneva, N.V., Pletnev, V.Z., Shemiakina, I.I., Chudakov, D.M., Artemyev, I., Wlodawer, A., Dauter, Z. and Pletnev, S. Crystallographic study of red fluorescent protein eqFP578 and its far-red variant Katushka reveals opposite pH-induced isomerization of chromophore. (2011) Protein Sci. 20: 1265-1274
far-red fluorescent protein Katushka	3pj5	Entacmaea quadricolor	Pletneva, N.V., Pletnev, V.Z., Shemiakina, I.I., Chudakov, D.M., Artemyev, I., Wlodawer, A., Dauter, Z. and Pletnev, S. Crystallographic study of red fluorescent protein eqFP578 and its far-red variant Katushka reveals opposite pH-induced isomerization of chromophore. (2011) Protein Sci. 20: 1265-1274

II. Protein Alignment and Sequence Analysis

Protein alignment and sequence analyses were accomplished with UCSF Chimera software. Twenty-eight mature protein structures were obtained from the PDB. Immature structures were prepared as stated in the next section. The proteins were aligned with the MatchMaker subroutine. They were matched with the structure of avGFP, the best alignment of the other 27 proteins with avGFP was made by using a Needleman-Wunsch algorithm. After the superposition of the 28 proteins, a structural sequence alignment was made through Match -> Align subroutine with a 5 angstroms residue-residue distance cutoff. Three parameters (RMSD of α-carbon, conservation of amino acids and charge variation) were collected by Chimera, exported and summarized using Office Excel. In all calculations residues 21-26, 36-40, 47-60, 75-92, 100-105, 114-118, 128-145, 154-160, 170-176, 208-217, 186-199 and 227

were considered “lid residues”. The position of these residues in bottom and top lids of avGFP are mapped in Figure 9.

III. Immature Form of Fluorescent Proteins

Immature FPs structures were obtained from mutating the mature structures by graphically changing the chromophores to form the original precyclized tripeptide sequence. The conformational searches were conducted by combining the Monte Carlo torsional variation and low mode methods. (*28, 29*) By randomly rotating the dihedral angles of all the sides-chains of residues 64-68 (avGFP numbering) between 0° and 180° and translating all solvent molecules in an 8.00 Å sphere by between 0 Å and 1.00Å in each Monte Carlo (MC) step (*30*), we obtained the lowest energy structures.

IV. Positional and Dihedral Variation of Conserved Residues by Molecular Dynamics (MD)

MD simulations were carried out in the NPT ensemble at 300K and 1 bar with 1.5fs steps using Desmond. (*31*) All molecular dynamics calculations used the OPLS_2005 force field and SHAKE constrained hydrogens. Ten thousand four hundred and eighteen structures were sampled in each 50ns MD simulation. Each structure was in an orthorhombic simulation box of 0.15M NaCl and SPC waters, with a 10Å solvent buffer between the protein surface and the boundary. Minimizations and pre-equilibrium simulations were done using the default

Desmond/multisim relaxation protocol. (*26, 31*) With the scripts from Matthew Zimmer, the dihedral angles φ and ψ of the conserved residues throughout the 10480 frames were measured from VMD 1.9.1.

Results and Discussion:

I. Conserved Residues and Protein Folding

According to the funneled energy landscape theory, there are many folding pathways for an unfolded protein to adopt, although a small number of them dominate the folding process.(*32, 33*) In order to form a correctly folded protein these energy funnels have to be fairly robust under a large variety of conditions.

It has long been known that active site residues are commonly conserved, however recent studies have shown that evolutionary conservation and structural dynamics are also strongly linked. (*34, 35*) Although the protein folding pathways are minimally affected by most mutations,(*36*) folding nuclei that are critically important in helping proteins adopt their three dimensional conformations are highly conserved.(*37*)

Hinge residues are residues that control the movement of the protein like door hinges. The significance of the hinge is the maintenance of protein geometry formation used for protein binding and other functions. Interference of the hinge sites may lead to allosteric effects and disruption of the protein's functional motions. Therefore hinges are spatially and sequentially conserved with less translational mobility, but because of the importance in directing the proteins to adopt the tertiary conformation in protein folding, they have rotational flexibility and dihedral freedom.

II. Hinge Residues and Protein Folding

In the GFP, hinge residues are the immobilized amino acids that control the folding mechanism, and the flexibility and rotation of the β-sheets and lids. Among all the most conserved 18 residues in avGFP, 12 residues are located within the lids' area (Figure 4). Could these residues be hinge residues that are critically involved in β-barrel folding and chromophore formation?

III. Structural Analysis of Wild-Type Fluorescent Proteins

By 2013, a total of 28 wild-type GFP-like proteins' crystal structures have been published on RCSB Protein Data Bank (rcsb.org). Despite the origins from different species, all the proteins had similar sequences and structures. They all have similar eleven stranded β-barrels with a centrally located chromophore. If one assumes that the protein backbone will adopt a low energy conformation in most crystal structures, then the conformational space spanned by the crystal conformations can be equated and compared (qualitatively not quantitatively) with the potential energy surface of the backbone itself. (*38*) With the default settings in Chimera, the superimposition of 28 FPs was done by aligning all sequenced alpha-carbons to *av*GFP's crystal structure. The rigidity of β-sheets, lids structure and conserved residues across the species was obtained from this superimposition.

Figure 10 shows the superimposition of the 28 mature wild-type structures listed in Table 1. In the figure, the values of the charge, conservation and RMSD were mapped with 3 sets of colors. If we take a look at the highly conserved sequence, or

red part on Figure 10, the lid structure seems to be relatively more conserved than the β-sheets. This is confirmed by our previous study that showed most of the conserved residues are located at the lids (Figure 4). The conservation of residues across species is related to the basic cellular function, stability and reproduction. Although evolution may lead to mutations, if the mutations happen to important amino acids that influence the viability of the protein and the host, the mutants will not survive and the genes will be in a highly conserved region.

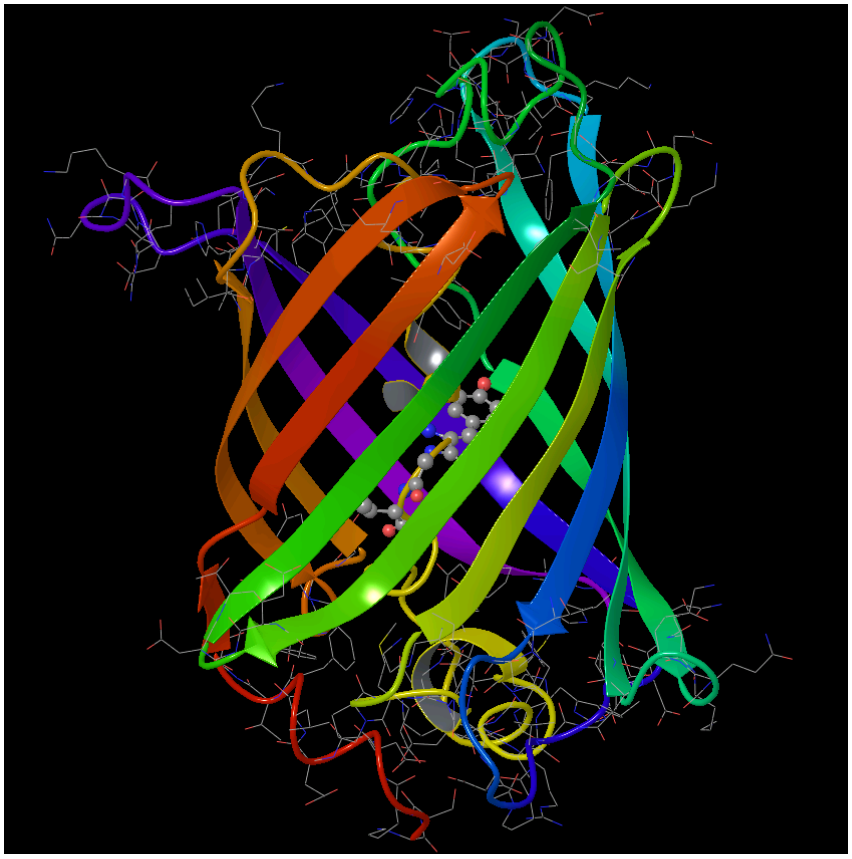

Figure 9. Amino Acids marked in wire representation constitute the bottom and top lids of avGFP (residues 21-26, 36-40, 47-60, 75-92, 100-105, 114-118, 128-145, 154-160, 170-176, 208-217, 186-199 and 227)

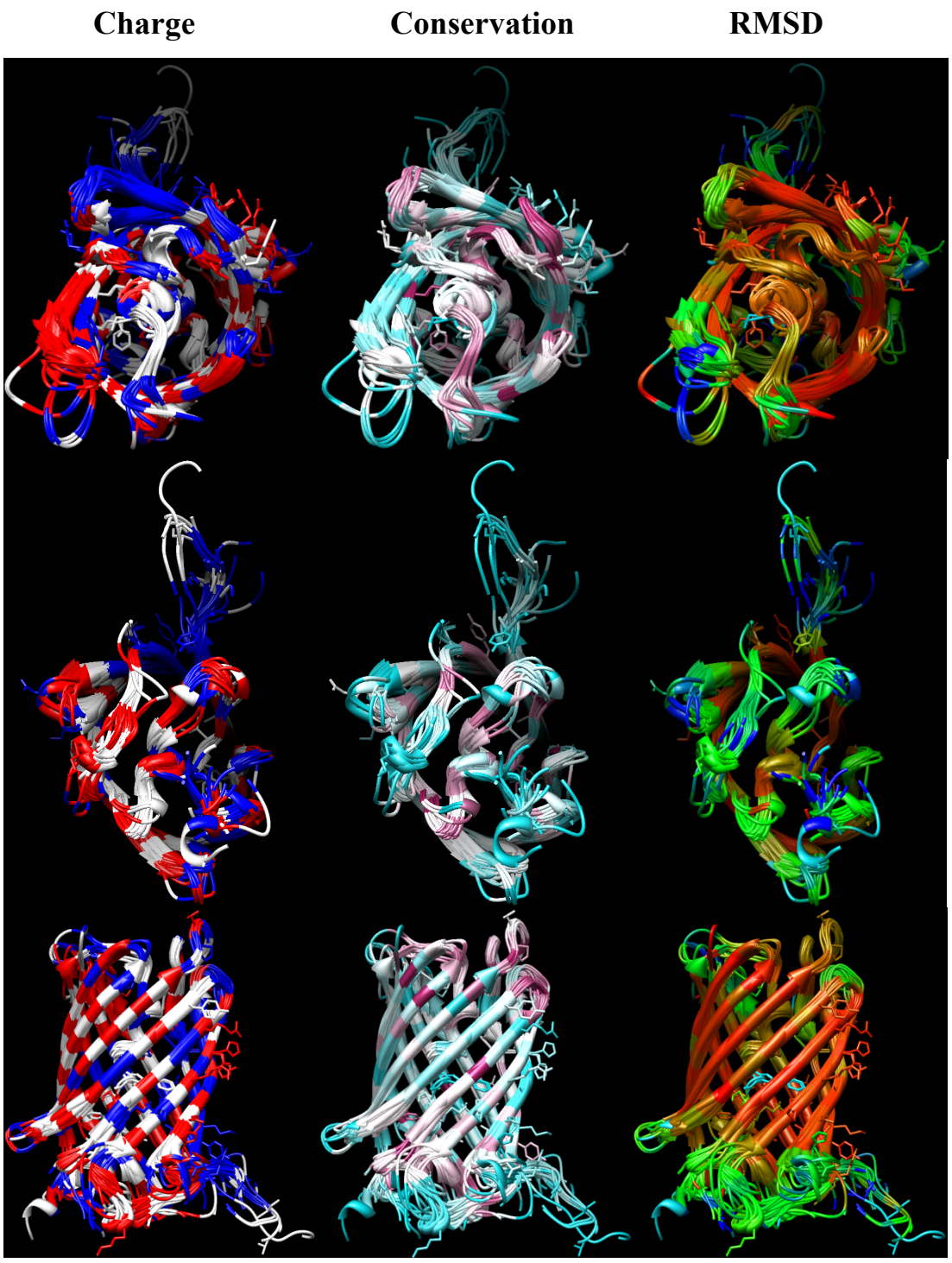

Figure 10. Three views (top, bottom and side) of the superimposed 28 different wild-type FPs structures with 3 properties from the left column to the right: charge variation (white as zero charge variance, blue as one charge variance and red as two charges variance), conservation (cyan as least conserved and maroon as most conserved) and RMSD (rainbow color with red as smallest RMSD and purple as biggest RMSD).

According to Figure 10 and Figure 11 although lids have higher sequence conservation, β-sheets show higher rigidity due to the hydrogen bond between the β-strands. Similarly in Table 2, β-sheets have a RMSD value of 0.66Å, which is almost half of the RMSD value of lid structure (1.213Å). The average RMSD of the entire protein is 0.989 Å, which is, not surprisingly, between the RMSD of the lids and the β-sheets. Across all the 28 FPs, the bottom lid (the end with N and C termini) and top lid show distinctive traits. The top lid shows noticeably higher rigidity than the bottom lid according to the color representation in Figure 9. Since the bottom lid contains the two termini that are flexible and increase the RMSD, the terminal residues were excluded from the RMSD calculation (residues 1-10, 228-230). The RMSD results in Table 2 shows that the top lid has more rigidity (1.044Å) than the bottom lid (1.382Å) even when the two termini are removed from the bottom lid.

The conserved lid residues show more spatial conservation than the lid residues, but more flexibility than β-sheet. Most of the conserved residues are located at the top lid, and it has more spatially and sequentially conservation than bottom lid.

Table 2. The RMSD of residues that make up different structures in mature and immature FPs in the neutral form from the superimposition of 28 wild-type FPs

	Residues	Mature RMSD (Å)	Immature RMSD (Å)
Top Lid	21-26, 47-60, 100-105, 128-145, 170-176, 208-217	1.044	1.755
Bottom Lid*	36-40, 75-92, 114-118, 154-160, 186-199, 227	1.382	1.586
Both Lids	Top + Bottom Lids	1.213	1.671
β-Sheets	11-20, 27-35, 41-46, 61-74, 93-99, 106-113, 119-127, 146-153, 161-169, 177-185, 200-207, 218-226	0.660	0.897
Conserved Residues	20, 23, 27, 31, 33, 35 50, 53, 89, 91, 101, 102, 104, 127, 130, 134, 136, 196	0.889	1.283
Conserved Lids	23, 50, 53, 89, 91, 101, 102, 104, 130, 134, 136, 196	0.989	1.472
All Residues	1-238	0.989	1.404

- For better accuracy, the bottom lid RMSD excludes both N and C termini.

We believe that the conserved lid residues are either hinges crucial to folding and chromophore formation, or the conservation of the lid is indicative of some as of yet unknown protein-protein binding function. Since the top lid has similar structural and electronic properties in all the naturally occurring FPs, naturally a protein may dock on the lid. Figure 10 shows positively, negatively and neutrally charged amino acids scattered over the entire top lid and it does not show any obvious conserved charged area that can act as a binding site. We therefore focus the remainder of this thesis on the possibility that the conserved residues are acting as hinges.

IV. Structural Analysis of Immature Fluorescent Proteins

Before the maturation of the chromophore, GFP, just like any other three-dimensional protein structure, has to fold into the correct tertiary structure to be able to function. Different from the cyclized chromophore in mature GFP, the tripeptide sequence chromophore in the immature structure is in a precyclized form

(Figure 12 C). These are the so-called immature structures. There are many folding pathways for an unfolded protein according to the funneled energy landscape theory, but only a small amount will dominate the folding process. *(32, 33)* Only under some suitable conditions can the energy funnels be fairly robust and form the correctly folded protein. It has long been known that the active sites' residues involving in folding are commonly conserved and recent studies showed strong connection between evolutionary conservation and structural dynamics. *(34, 35)* Although the protein folding pathways are minimally affected by most mutations, folding nuclei that are critically important in helping proteins adopt three-dimensional conformations are highly conserved. *(37)* Local perturbations, or interference of hinge sites can give rise to allosteric effects or even disrupt the entire cooperativity of the functional motions of a protein, and therefore it is not surprising that these sites are conserved. *(39)*

Although GFP is a decent folder, it has a different folding landscape that produces an active chromophore surrounded by 11 β-sheets. *(26)* The fluorescence of fluorescent proteins is only observed after the formation of chromophore by autocatalytic internal cyclization of the tripeptide 65SYG67 and subsequent oxidation of the intrinsically formed structure, which normally takes 90 minutes to 4 hours after protein synthesis. *(40, 41)*

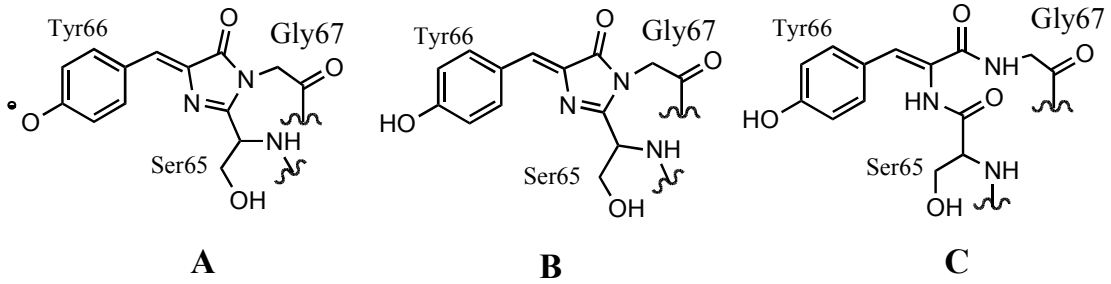

Figure 12. Three forms of Chromophore in *av*GFP: A. Mature ionized form. B. Mature neutral form. C. Immature or uncyclized or precyclized form.

Since the chromophore is involved in an energetic frustration in the folding and unfolding of GFP, we wonder how is the structure that folded with a precyclized chromophore different from the structure folded with a cyclized chromophore. Also to observe the hinge residues' effect on the folding of immature FPs, we computationally created the immature structure from the mature crystal structures of the wild-type FPs. The mature FPs crystal structures were graphically modified by converting the original cyclized chromophore into a precyclized tripeptide sequence in Maestro, and a brief conformational search was conducted to simulate the structure before chromophore formation takes place. Comparing the 28 FPs structures with a mature chromophore to the same 28 FPs structures with an immature chromophore, we see the residues in immature structures are generally more flexible and have higher RMSD value (1.404Å vs. 0.989Å), see Table 2. This is supported by our previous molecular dynamics study which showed that the mature GFP contains a tighter β-barrel than the relaxed precyclized form. *(26)* The β-barrel in FPs may have evolved to be porous during chromophore formation so that it permits the entrance of catalytically important molecules for the chromophore maturation. The tensing of β-barrel upon formation of chromophore may cause the hysteresis observed in FPs'

folding behavior. *(26, 42, 43)* The higher RMSD of the immature structures relative to their mature counterparts is observed at top lid, bottom lid and β-sheets from Figure 13. The trend can also be seen in Figure 14, where the line representing the RMSD of immature structure stays above the lines representing both the mature forms throughout the entire amino acid sequence. However, unlike the mature structures, the immature FPs are more rigid on the bottom lid than the top lid (1.586 Å to 1.755 Å). The conserved residues in immature structures have the RMSD value of 1.283Å, which is less rigid than those in mature structures with RMSD value of 0.889Å. Like other residues in immature GFP, the conserved residues gain flexibility from the folding of structure with a precyclized chromophore while the cyclized chromophore leads to a tighter folding of mature protein structure. It shows some flexibility of the conserved residues relative to the β-sheet, which implies the hinges (if they are hinges) may move slightly and still control the movement of the lids. Despite of the changes in the flexibility between lids, the conserved lid residues in both mature and immature structures are less positionally variant than the remainder of the lid residues, which further reinforces the idea that the conserved residues within the lids are hinge residues.

V. Molecular Dynamics Simulations

In addition to the superimposition of static structures that provides a view into the movement of residues over different species, we have done molecular dynamics (MD) simulations to observe the movement of the conserved residues over time. MD is a

simulation of atoms' physical movements and interactions within a molecule over a period of time. The simulation creates trajectories of atoms and molecules based on Newton's laws of motion.

VI. Molecular Dynamics and Translational Flexibility of Conserved Residues

Fifty nanosecond MD simulations of avGFP in its neutral, anionic mature forms and its immature form were run and analyzed. The residues in avGFP show varied RMSD movement in neutral, ionized and uncyclized forms although the fluctuations of all 3 forms follow similar trends throughout the whole amino acid sequence (Figure 12 and Figure 15). Like the crystallographic comparisons, major RMSD peaks take place in the residues that make up either lid, showing that lids have more movement than the more structured hydrogen-bonded β-sheets. It can be observed from the graph that the uncyclized structure has the highest RMSD value and neutral form has the lowest RMSD value. The immature proteins show less compact structures from the folding with the precyclized chromophore comparing to the more compact mature structures, which indicates the mature chromophore's involvement in folding hysteresis. The ionized form usually falls in the middle. From amino acid 57-69, the uncyclized structure has a major peak compared to very low value on neutral and ionized forms. This is due to the fact that residues 65-67 make up the tripeptide sequence that forms the chromophore. In the immature form the peptide backbone is much more flexible than the rigid chromophore found in the mature forms. In another part of the amino acid sequence, residues 73-87 that make up the bottom lid, the order

from the highest RMSD value to the lowest is also from the uncyclized structure to the ionized structure and finally the neutral structure. As discussed before, the increased flexibility in the immature form indicate that structural changes occur upon the maturation of chromophore that tightens up the protein and decreases the residue's spatial flexibility. *(26)* However in another piece of amino acid sequence, residues 153-161 that make up the bottom lid, the ionized structure has the highest RMSD. Surprisingly the uncyclized structure has the lowest RMSD value in this area. This finding is supported by the previous result (obtained by overlapping mature and immature structures, Table 2 and Figure 15) that the bottom lid in immature structure is more spatially conserved than the top lid, which is opposite to the situation in mature structure. The flexibility difference of top and bottom lids between the mature and immature forms of avGFP indicates a structural change occurs in bottom lid after the chromophore maturation. The hysteresis, resulting from the chromophore formation involves a series of structural changes in both lids. Due to its high spatial conservation in immature form, the bottom lid has also the chance to play a significant role in the protein folding of mature avGFP.

VII. Molecular Dynamics and Dihedral Flexibility of Selected Conserved Residues

In our belief, the conserved lid residues, or the hinge residues, ought to have relatively higher rotational flexibility than the other lid residues, in addition to the inert translational mobility. Therefore the dihedral freedom becomes one of the key elements to determine the existence of hinge residues among the conserved lid

residues. In the 50ns MD simulations of mature and immature avGFP, the dihedral angles (φ vs. ψ) of 3 conserved lid residues (31, 33 and 104) and 3 less conserved lid residues (24,116 and 174) were compared in Figure 16. All six residues are glycines, they were chosen because glycines are the most flexible amino acids and because all six are in similar environments i.e. located in short lid loops that connect β-sheets. Because each dot represents the dihedral angel of a residue in 10 frames out of the entire 10480 frames, the distribution of the 1048 dots of a residue indicates the dihedral angel freedom throughout the 50ns of MD simulation. In Figure 16, the dots representing the dihedral angles of each of the less conserved lid residues are confined within a single region, considering that the 180° equals to -180° mathematically, which indicates the dihedral freedom of less conserved residues is limited to that single region. However, conserved residues have noticeably larger distributions. Conserved lid residue 33, marked in red crosses, spans from one major region to another one in both mature and immature structures. It shows the active dihedral mobility that is typical for a hinge residue. Although conserved lid residue 31 only takes one region in mature structure, it has a very wide distribution in immature structure, so it has a greater rotational mobility in the immature protein. The correct folding is crucial for the later cyclization of chromophore, therefore the hinge residues are more important in the immature structures. Conserved lid residue 104 behaves completely different to residue 31: it covers merely one region in the immature structure but spreads into 3 regions in the mature structure, which shows great dihedral freedom in mature structure but minimal dihedral movement in the immature

structure. All the conserved lid residues' dihedral angles in 50ns MD simulation were listed in Table 3 although the average and standard deviation calculations are insufficient in explanation of dihedral freedom. The high rotational movement, combined with minimal spatial mobility of the conserved lid residues has pointed to their structural function as hinges in the lids, which are able to rotate the β-sheets into correct orientation during protein folding.

Conclusions:

From the observation, the conserved lid residues are the least translationally mobile residues while also being the most active in the dihedral movements among all the lid residues. Due to their spatial and sequential conservation, and rotational freedom, the conserved lid residues fit into our expectation of hinge residues that play a critical role in the correct protein folding of β-sheets and subsequent chromophore formation. Two of my colleagues are looking at the same topic from different aspects: Paola Peshkepija aligned all the crystal structures of wild-type FPs and obtained the dihedral angles of the conserved lid residues from pdb *(27)*; Ramza Shahid used the Anisotropic Network Model *(44)* to study the dihedral freedom with a coarse-grained perspective that allows for longer simulations. She observed the least RMSD fluctuations in the conserved lid residues. The recognition of hinge residues in proteins originated from various species has brought us further in understanding the folding of GFP, which not only provides the ideas for GFP engineering, but will also potentially support the structural study of other similar β-barrel molecules. Also more studies can be made to further explore the behavior of every single hinge residue during GFP folding: for instance, mutations can be made on the hinge residues to provide more substantial evidence for their structural functions.

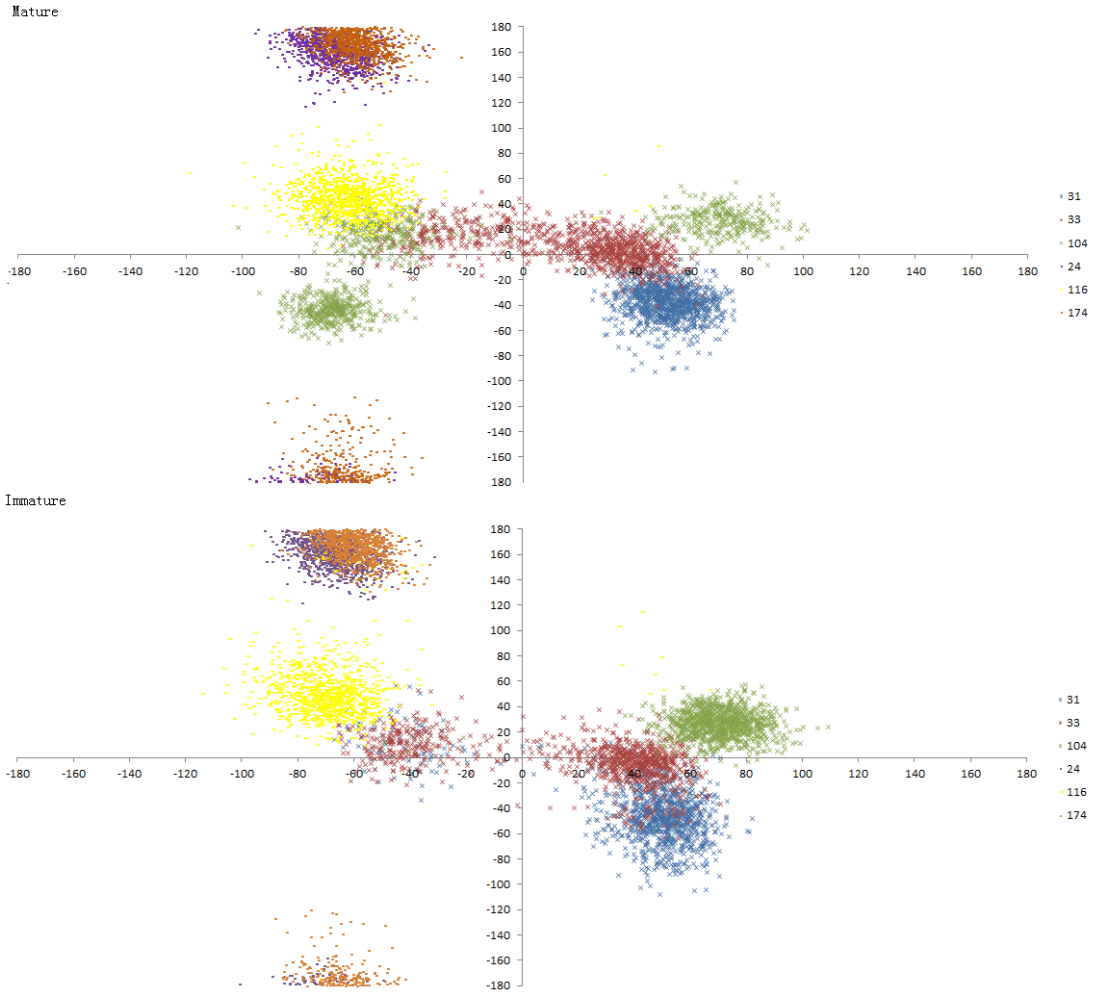

Figure 16. The dihedral angles of 3 conserved residues (31, 33 and 104 marked as X) and 3 less conserved residues (24,116 and 174 marked as -) in mature and immature GFP throughout the 50ns of MD simulations of (x-axis: Phi (φ) angle, y-axis: Psi (ψ))

Table 3. The dihedral angles of conserved lid residues in mature *av*GFP from 50ns MD simulation

	Phi (φ)		Psi (ψ)	
Amino Acid number	Average	Standard Deviation	Average	Standard Deviation
20	161.6018	37.113515	188.99966	17.67146
23	246.9829	39.807777	199.67764	88.679997
27	207.38575	9.0095174	166.21508	9.0106651
31	231.66346	9.3510448	142.37391	13.267286
33	196.81653	29.317618	186.85006	12.997569
35	226.54498	9.6319052	158.91449	11.042868
50	264.80651	17.714279	184.22892	163.3138
53	105.81611	9.1779148	127.15928	7.6447434
89	100.90218	9.4657071	168.17046	102.21596
91	116.57928	8.0527498	189.26594	9.8633345
101	114.82968	13.136535	119.42535	14.072537
102	260.26296	9.9894623	152.15048	133.15336
104	142.62449	61.489049	175.66322	34.116236
127	242.36605	9.5323377	130.88257	14.283819
130	102.00069	15.054824	144.53519	14.207164
134	244.84843	16.5833	206.18548	20.627015
136	90.119651	15.052858	123.8662	10.447013
196	99.873774	7.3759648	167.42944	13.700271
All Conserved	175.334735	18.15868661	162.888521	38.35083878
All	241.901023	51.16638114	172.279948	92.08953841

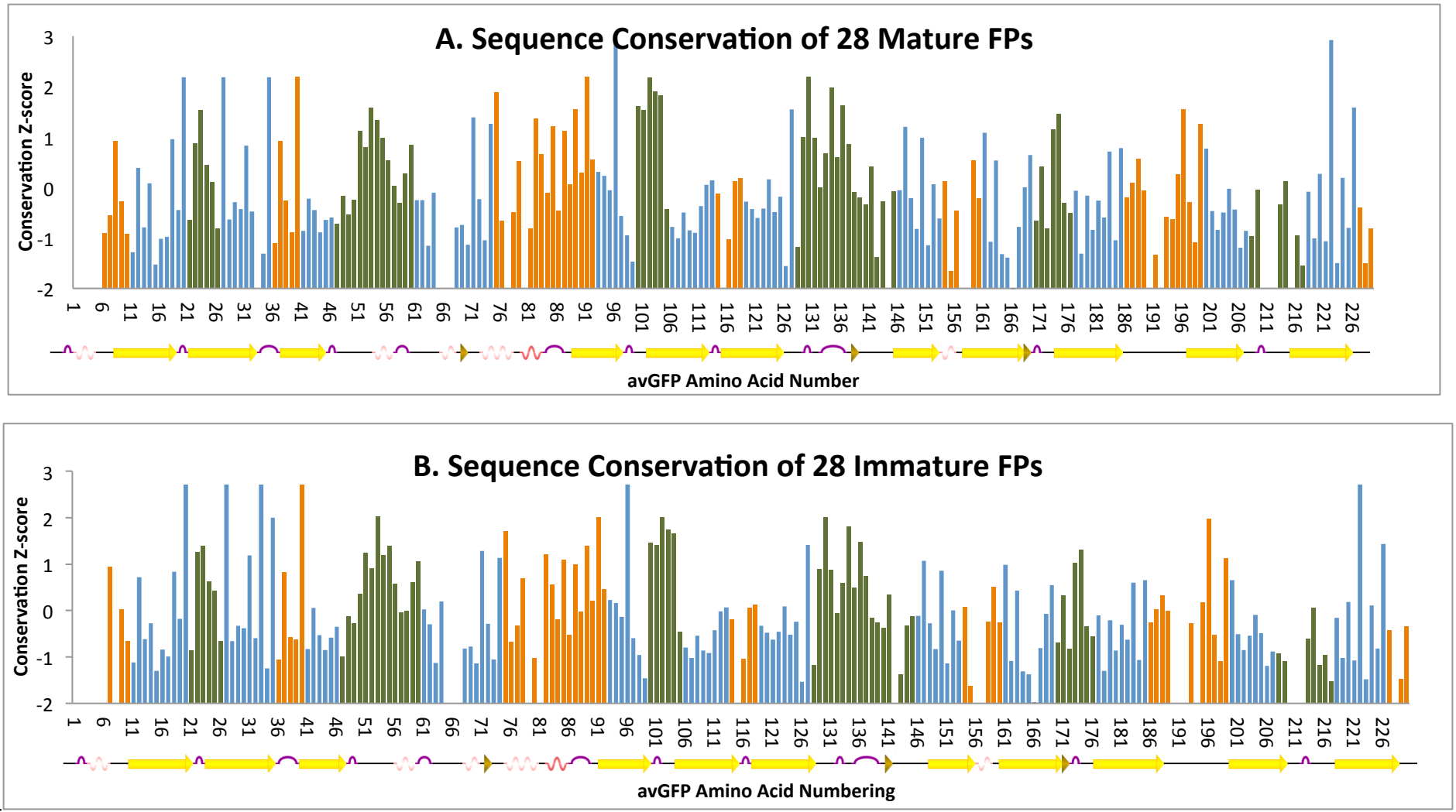

Figure 11. A. the sequence conservation of all residues in 28 mature FPs. B. the sequence conservation of all residues in 28 immature FPs (Red: top lid, green: bottom lid, blue: β-Sheets. Data generated from the superimposition by Chimera)

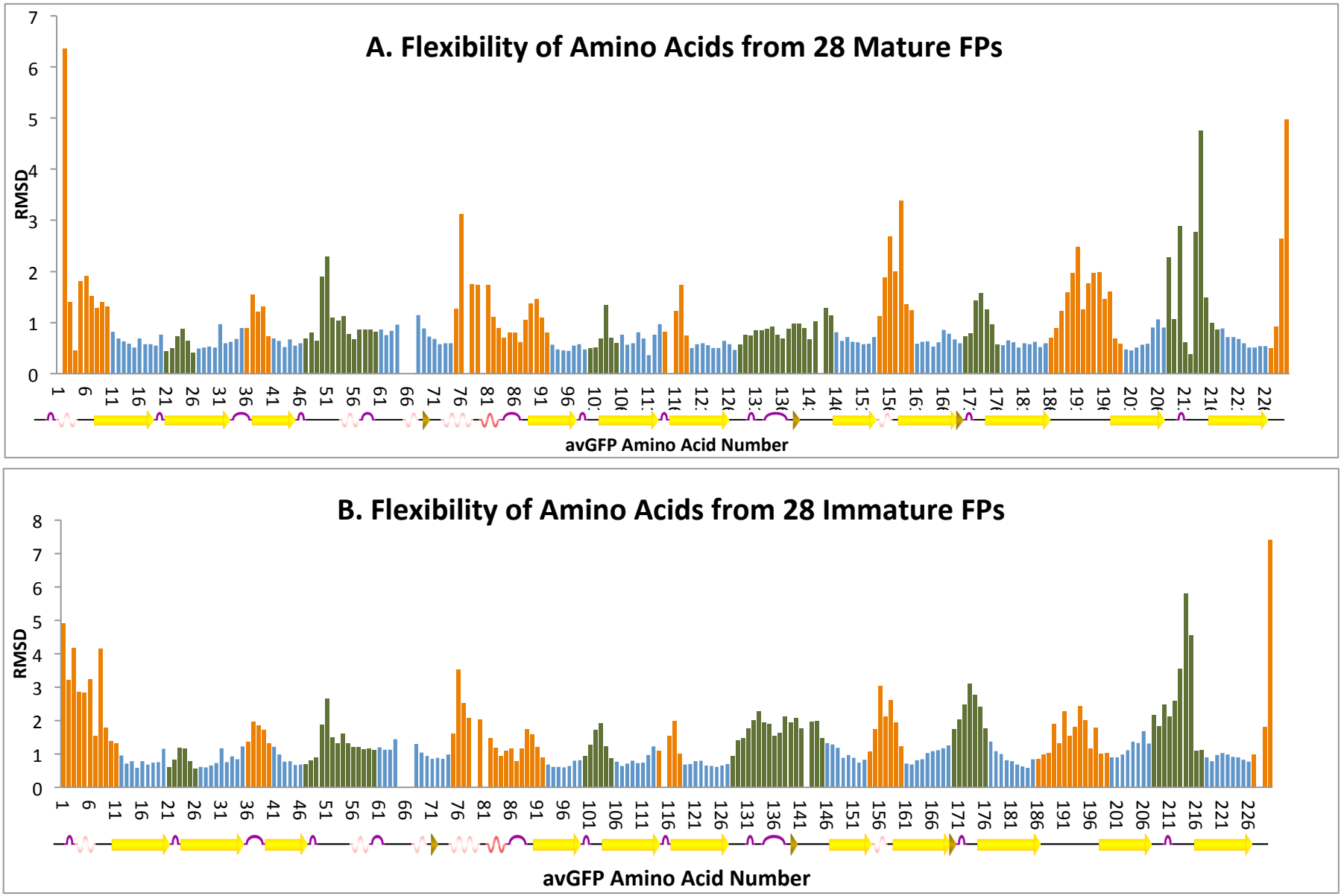

Figure 13 A. RMSD of all residues in 28 Mature FPs according to avGFP numbering. B. the RMSD of all residues in 28 Immature FPs (Red: top lid, green: bottom lid, blue: β-Sheets. Data generated from the superimposition by Chimera)

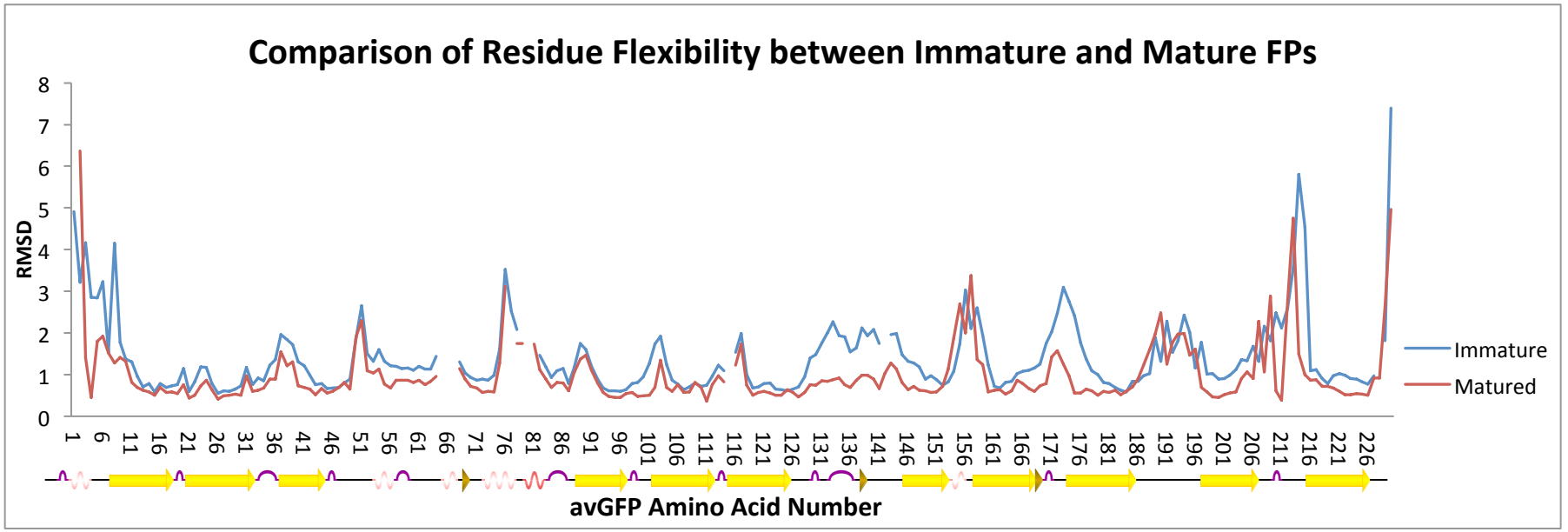

Figure 14. RMSD comparison between mature and immature structures

RMSD of avGFP with Chromophore in 3 forms throughout 50ns of Molecular Dynamics

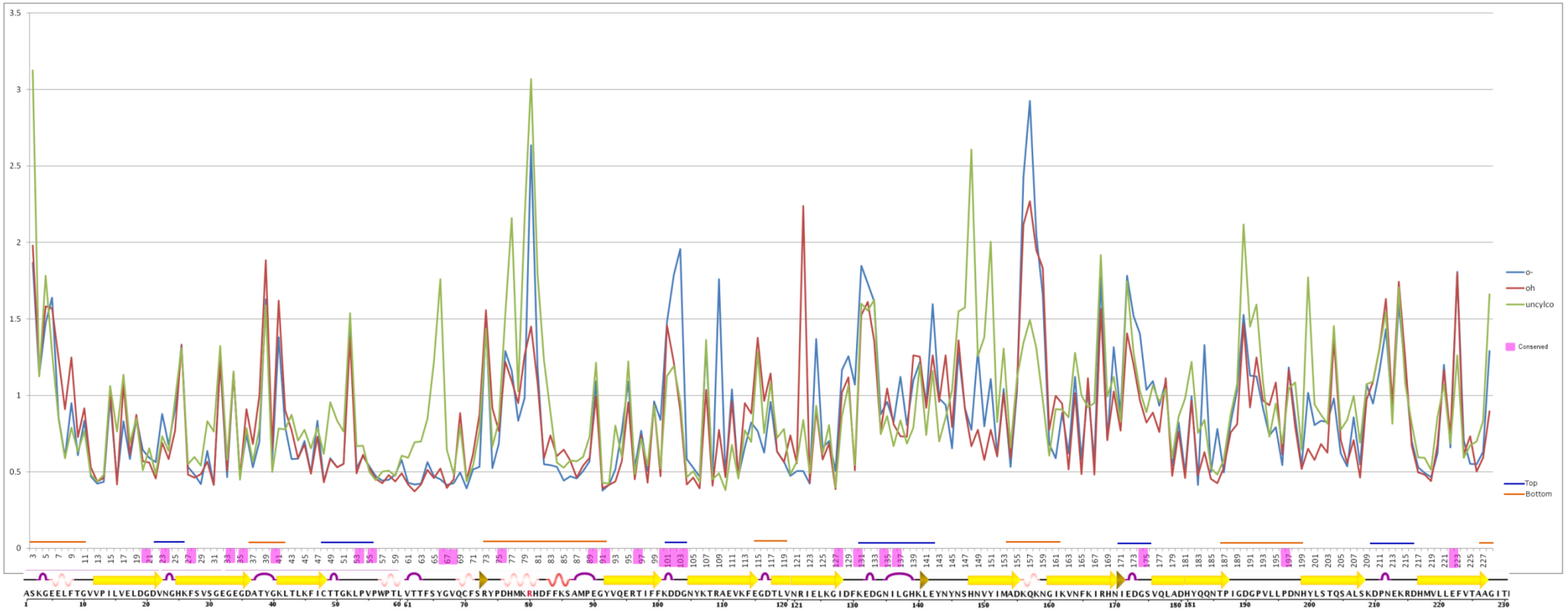

Figure 15. Average RMSD of 228 residues from avGFP structure in neutral (blue), ionized (red) and uncyclized (green) forms in 10418 frames of MD structure.*from Matthew Zimmer's Work

Table 4. Amino acid sequence of 28 FPs

	1	11	21	31	41	51	61	71	81	91	101	111	121
1gfl	ASKGEELFTG	VVPILVELDG	DVNGHKFSVS	GEGEGDATYG	KLTLKFICTT	GKLPVPWPTL	VTTFSYGVQC	FSRYPDHMKR	HDFFKSAMPE	GYVQERTIFF	KDDGNYKTRA	EVKFEGDTLV	NRIELKGIDF
1mou	IAT	QMTYKVYMSG	TVNGHYFEVE	GDGKGRPYEG	EQTVKLTVTK	GPLPFAWDIL	SPQC SIP	FTKYPE DI	PDYVKQSFPE	GFTWERIMNF	EDGAVCTVSN	DSSI QNCFT	YHVKFSGLNF
1uis	IKE	NMRMMVVMEG	SVNGYQFKCT	GEGDGNPYMG	TQTMRIKVVE	GPLPFAFDIL	ATSF SKT	FIKHTK GI	PDFFKQSFPE	GFTWERVTRY	EDGGVFTVMQ	DTSL EGCLV	YHAKVTGVNF
1xss	ITS	EMKMELRMEG	AVNGHKFVIT	GKGSGQPFEG	IQNMDLTVIE	GPLPFAFDIL	TTVF NRV	FVKYPE EI	VDYFKQSFPE	GYSWERSMSY	EDGGICLATN	NITM KNCFV	YEIRFDGVNF
1yzw	LKE	SMRIKMYMEG	TVNGHYFKCE	GEGDGNPFAG	TQSMRIHVTE	APLPFAFDIL	APCC SRT	FVHHTA EI	PDFFKQSFPE	GFTWERTTTY	EDGGILTAHQ	DTSL ENCLI	YKVKVHGTNF
1zgo	IKE	FMRFKVRMEG	TVNGHEFEIE	GEGEGRPYEG	HNTVKLKVTK	GPLPFAWDIL	SPQF SKV	YVKHPA DI	PDYKKLSFPE	GFKWERVMNF	EDGGVVTVTQ	DSSL QGCFI	YKVKFIGVNF
1zux	IKP	DMKINLRMEG	NVNGHHFVID	GDGTGKPFEG	KQSMDLEVKE	GPLPFAFDIL	TTAF NRV	FAEYPD HI	QDYFKQSFPK	GYSWERSLTF	EDGGICIARN	DITM EDTFY	NKVRFHGVNF
2a46	IGD	DMKMTYHMDG	CVNGHYFTVK	GEGNGKPYEG	TQTSTFKV T	GPLAFSFDIL	STVF NRC	FTAYPT SM	PDYFKQAFPD	GMSYERTFTY	EDGGVATASW	EISL KNCFE	HKSTFHGVNF
2c9i	IKE	TMRVQLSMEG	SVNYHAFKCT	GKGEGKPYEG	TQSLNITITE	GPLPFAFDIL	SHAF IKV	FAKYPK EI	PDFFKQSLPG	GFSWERVSTY	EDGGVLSATQ	ETSL QDCII	CKVKVLGTNF
2c9j		NLSVSVYMKG	NVNNHEFEYD	GIGGGDPNSG	QFSLKTKLRG	KPLPFSYDII	TMGF FRA	FTKYPE GI	ADYFKGSFPE	AFQWNRRIEF	EDGGVINMSS	DITY KKVLH	GDVWALGVNF
2dd7	IKT	TFKIESRIHG	NLNGEKFELV	GGGVGE EG	RLEIEMKTKD	KPLAFSPFLL	SHCM FYH	FASFPK GT	KNIYLHAATG	GYTNTRKEIY	EDGGILEVNF	RYTY ENKII	GDVECIGHGF
2g3o		AMKIECRITG	TLNGVEFELV	GGGEGTPEQG	RMTNKMKSTK	GALTFSPYLL	SHVM FYH	FGTYPS GY	ENPFLHAING	GYTNTRIEKY	EDGGVLHVSF	SYRY EGRVI	GDFKVVGTGF
2gw3	IKP	EMKIKLLMEG	NVNGHQFVIE	GDGKGHPFEG	KQSMDLVVKE	APLPFAYDIL	TTAF NRV	FAKYPD HI	PDYFKQSFPK	GFSWERSLMF	EDGGVCIATN	DITL KDTFF	NKVRFDGVNF
2ib5	ISD	NVRIKLYMEG	TVNNHHFMCE	AEGEGKPYEG	TQMENIKVEK	GPLPFSFDIL	TPNC SVA	ITKYTS GI	PDYFKQSFPE	GFTWERTTIY	EDGAYLTTQQ	ETKL DNCLV	YNIKILGCNF
2icr	LTD	DMTMHFRMEG	CVDGHKFVIE	GNGNGNPFKG	KQFINLCVIE	GPLPFSEDIL	SAAF NRL	FTEYPE GI	VDYFKNSCPA	GYTWHRSFRF	EDGAVCICSA	DITV NNCIY	HESTFYGVNF
2ie2	IKP	DMKIKLRMEG	AVNGHPFAIE	GVGLGKPFEG	KQSMDLKVKE	GPLPFAYDIL	TTVF NRV	FAKYPE NI	VDYFKQSFPE	GYSWERSMNY	EDGGICNATN	DITL DDCYI	YEIRFDGVNF
2ogr	LKE	EMTMKYHMEG	CVNGHKFVIT	GEGIGYPFKG	KQTINLCVIE	GPLPFSEDIL	SAGF DRI	FTEYPQ DI	VDYFKNSCPA	GYTWGRSFLF	EDGAVCICNV	DITV SNCIY	HKSIFNGMNF
2ojk	LTK	EMTMKYRMEG	CVDGHKFVIT	GEGIGYPFKG	KQAINLCVVE	GPLPFAEDIL	SAAF NRV	FTEYPQ DI	VDYFKNSCPA	GYTWDRSFLF	EDGAVCICNA	DITV SNCMY	HESKFYGVNF
2rh7	LGL	KEVMPTKINL	EGLVGDHAFS	MEGVGEGNIL	EGTQEVKISV	KGAPLPFAFD	IVSV NRA	YTGYPE EI	SDYFLQSFPE	GFTYERNIRY	QDGGTAIVKS	DISL DGKFI	VNVDFKAKDL
2wht	IKE	QMTYKVYMSG	TVNGHYFEVE	GDGKGKPYEG	EQTVKLTVTK	GPLPFAWDIL	SPQL SIP	FTKYPE DI	PDYFKQSFPE	GYTWERSMNF	EDGAVCTVSN	DSSI QNCFI	YNVKISGENF
2z6x	IKP	DMKIKLRMEG	AVNGHPFAIE	GVGLGKPFEG	KQSMDLKVKE	GPLPFAYDIL	TTVF NRV	FAKYPE NI	VDYFKQSFPE	GYSWERSMNY	EDGGICIATN	DITL DDCFI	YEIRFDGVNF
2zmu	IKP	EMKMRYYMDG	SVNGHEFTIE	GEGTGRPYEG	HQEMTLRV T	GPMPFAFDLV	SHV HRP	FTKYPE EI	PDYFKQAFPE	GLSWERSLEF	EDGGSASVSA	HISL RNTFY	HKSKFTGVNF
3cgl	IKE	EMLIDLHLDG	TFNGHYFEIK	GKGKGQPNEG	TNTVTLEVTK	GPLPFGWHIL	CPQF NKA	FVHHPD NI	HDYLKLSFPE	GYTWERSMHF	EDGGLCCITN	DISL TNCFY	YDIKFTGLNF
3gb3	FQS	DMTFKIFIDG	EVNGQKFTIV	ADGSSKFPHG	DFNVHAVCET	GKLPMSWKPI	CHLI EPF	FARYPD GI	SHFAQECFPE	GLSIDRTVRF	ENDGTMTSHH	TYEL DTCVV	SRITVNCDGF
3h1o	ITE	NMHMKLYMEG	TVNNHHFKCT	SEGEGKPYEG	TQTQRIRVVE	GPLPFAFDIL	ATSF SHT	FINHTQ GI	PDFWKQSFPE	GFTWERVTTY	EDGGVLTATQ	DTSL QGCLI	YNVKIRGVNF
3mgf	IKP	EMKMRYYMDG	SVNGHEFTIE	GEGTGRPYEG	HQEMTLRV T	GPMPFAFDLV	SHV HRP	FTKYPE EI	PDYFKQAFPE	GLSWERSLEF	EDGGSASVSA	HISL RNTFY	HKSKFTGVNF
3pib	IKE	NMHMKLYMEG	TVNNHHFKCT	SEGEGKPYEG	TQTMKIKVVE	GPLPFAFDIL	ATSF SKT	FINHTQ GI	PDFFKQSFPE	GFTWERITTY	EDGGVLTATQ	DTSL QGCII	YNVKINGVNF
3pj5	ISE	NMHMKLYMEG	TVNDHHFKCT	SEGEGKPYEG	TQTMKIKVVE	GPLPFAFDIL	ATSF SKT	FINHTQ GI	PDFFKQSFPE	GFTWERITTY	EDGGVLTATQ	DTSL QGCLI	YNVKINGVNF

(Continued)

	131	141	151	161	171	181	191	201	211	221
1gfl	KEDGNILGHK	LEYNYNSHNV	YIMADKQKNG	IKVNFKIRHN	IEDGSVQLAD	HYQQNTPIGD	GPVLLPDNHY	LSTQSALSKD	PNEKRDHMVL	LEFVTAAGIT
1mou	PPNGPVMQKK	TQ GWEPHSE	RLFARG GM	LIGNNFMALK	LEGGGHYLCE	FKTTYKAKK	P VKMPGYHY	VDRKLDVTN	NK TSVEQ	CEISIARKPA
1uis	PSNGAVMQKK	TK GWEPNTE	MLYPAG G	LRGYSQMALN	VDGGGYLSCS	FETTYRSKK	TNFKMPGFHF	VDHRLERLE	S MFVVQ	HEHAVAKFC
1xss	PANGPVMQRK	TV KWEPSTE	KMYVRG V	LKGDVNMALL	LQGGGHYRCD	FRTTYKAKK	V VQLPDYHF	VDHRIEITS	DK NKVKL	YEHAKAHSG
1yzw	PADGPVMKNK	SG GWEPSTE	VVYPEN GV	LCGRNVMALK	VG DRHLICH	HYTSYRSKK	AALTMPGFHF	TDIRLQMLR	K EYFEL	YEASVARYS
1zgo	PSDGPVMQKK	TM GWEASTE	RLYPRG V	LKGEIHKALK	LKDGGHYLVE	FKSIYMAKK	P VQLPGYYY	VDSKLDITS	NE TIVEQ	YERTEGRHHL
1zux	PANGPVMQKK	TL KWEPSTE	KMYVRG V	LTGDITMALL	LEGNAHYRCD	FRTTYKAKE	K EKLPGYHF	VDHCIEILS	DK NKVKL	YEHAVAHSGD
2a46	PADGPVMAKK	TT GWDPSFE	KMTVCG I	LKGDVTAFLM	LQGGGNYRCQ	FHTSYKTKK	P VTMPPNHV	VEHRIART L	DK NSVQL	TEHAVAHIT
2c9i	PANGPVMQKK	TC GWEPSTE	TVIPRD GG	LLLRDTPALM	LADGGHLSCF	METTYKSKK	E VKLPELHF	HHLRMEKLN	SD KTVEQ	HESVVASYSN
2c9j	PPNGPVMKNE	IV MEEPAEE	TLTAKG V	LVGFCPKAYL	LKDGSYYYGH	MTTFYRSKK	S QPLPGFHF	IKHRLVKTK	EP KMVEQ	AEYATAHVC
2dd7	PSQSPIFKDT	IV KSCPTVD	LMLPMS GNI	IASSYARAFQ	LKDGSFYTAE	VKNNIDFK	N PEFSKPMF	THRRVEET	ENLAM	VEYQQVFNS
2g3o	PEDSVIFTDK	II RSNATVE	HLHPMG DNV	LVGSFARTFS	LRDGGYYSFV	VDSHMHFK	S APSILPMF	AFRRVEEL	TELGI	VEYQHAFKT
2gw3	PPNGPVMQKK	TL KWEASTE	KMYLRG V	LTGDITMALL	LKGDVHYRCD	FRTTYKSRQ	EGVKLPGYHF	VDHCISILR	DK NEVKL	YEHAVAHSG
2ib5	PPNGPVMQKK	TQ GWEPCCE	MRYTRG V	LCGQTLMALK	CADGNHLTCH	LRTTYRSKK	AALQMPPFHF	SDHRPEIVK	SE TLFEQ	HESSVARYCN
2icr	PADGPVMKKM	TT NWEPSCE	KIIPINSQKI	LKGDVSMYLL	LKDGGRYRCQ	FDTIYKAKT	E PEMPDWHF	IQHKLNRE D	SDNQKWQL	IEHAIASRS
2ie2	PANGPVMQKR	TV KWEPSTE	KLYVRG V	LKGDVNMALS	LEGGGHYRCD	FKTTYKAKK	V VQLPDYHF	VDHHIEIKS	DK SNVNL	HEHAEAHS
2ogr	PADGPVMKKM	TT NWEASCE	KIMPVPKQGI	LKGDVSMYLL	LKDGGRYRCQ	FDTVYKAKS	V PKMPEWHF	IQHKLLRE R	SDNQKWQL	TEHAIAFPS
2ojk	PADGPVMKKM	TD NWEPSCE	KIIPVPKQGI	LKGDVSMYLL	LKDGGRLRCQ	FDTVYKAKS	V PKMPDWHF	IQHKLTRE R	SDNQKWHL	TEHAIASGS
2rh7	RRMGPVMQQD	IV GMQPSYE	SMYTNV TS	VIGECIIAFK	LQTGKHFTYH	MRTVYKSKK	P VTMPLYHF	IQHRLVKTN	YVVQ	HETAIAAHS
2wht	PPNGPVMQKK	TQ GWEPSTE	RLFARD GM	LIGNDYMALK	LEGGGHYLCE	FKSTYKAKK	P VRMPGRHE	IDRKLDVTS	NR TSVEQ	CEIAIARHSL
2z6x	PANGPVMQKR	TV KWEPSTE	KLYVRG V	LKGDVNMALL	LEGGGHYRCD	FKTTYKAKK	V VQLPDYHF	VDHRIEIKS	DK NNVNL	HEHAEAHSG
2zmu	PADGPIMQNQ	SV DWEPSTE	KITASD GV	LKGDVTMYLK	LEGGGNHKCQ	FKTTYKAKK	K IKMPGSHY	ISHRLVRKT	E NITEL	VEDAVAHS
3cgl	PPNGPVVQKK	TT GWEPSTE	RLYPRD GV	LIGDIHHALT	VEGGGHYACD	IKTVYRAKK	A AKMPGYHY	VDTKLVIWN	DK MKVEE	HEIAVARHHY
3gb3	QPDGPIMRDQ	LV DILPNET	HMFPHG PNA	VRQLAFIGFT	TADGGLMMGH	FDSKMTFNGS	R IEIPGPHF	VTIITKQM D	TSDKRDHVCQ	REVAYAHSVP
3h1o	PSNGPVMQKK	TL GWEAHTE	MLYPAG G	LEGRADLALK	LVGGGHLICN	FKTTYRSKK	PNLKMPGVYY	VDYRLERIK	A TYVEQ	HEVAVARYC
3mgf	PADGPIMQNQ	SV DWEPSTE	KITASD GV	LKGDVTMYLK	LEGGGNHKCQ	FKTTYKAKK	K IKMPGSHY	ISHRLVRK	E NITEL	VEDAVAHS
3pib	PSNGSVMQKK	TL GWEANTE	MLYPAG G	LRGHSQMALK	LVGGGYLHCS	FKTTYRSKK	PNLKMPGFHF	VDHRLERIK	A TYVEQ	HEMAVAKYC
3pj5	PSNGPVMQKK	TL GWEASTE	MLYPAS G	LRGHSQMALK	LVGGGYLHCS	LKTTYRSKK	PNLKMPGFYF	VDRKLERIK	A TYVEQ	HEMAVARYC

References:

1 Ong, W.J., Alvarez, S., Leroux, I. E., Shahid, R.S, Sama, A. A., Peshkepija, P., Morgan, A.L., Mulcahy, S. and Zimmer, M. Function and structure of GFP-like proteins in the protein data bank. *Mol. Biosyst.***7**, 984-992 (2011)

2 Thirunavukkuarasu S., Jares-Erijman, E.A. & Jovin, T.M. Multiparametric Fluorescence Detection of Early Stages in the Amyloid Protein Aggregation of Pyrene-labeled α-Synuclein, *Journal of Molecular Biology*. **378**, 1064-1073 (2008).

3 Leblond, F., Davis, S.C., Valdés, P.A. & Pogue, B.W. Pre-clinical whole-body fluorescence imaging: Review of instruments, methods and applications. *Journal of Photochemistry and Photobiology B: Biology*. **98**, 77-94 (2010).

4 Helms, Volkhard. Fluorescence Resonance Energy Transfer. Principles of Computational Cell Biology. Weinheim: Wiley-VCH. p. 202. ISBN 978-3-527-31555-0 (2008).

5 Sakaue-Sawano, A., Kurokawa, H., Morimura,T., Hanyu, A., Hama H., Osawa, H., Saori Kashiwagi, et al. Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell* **132** (3), 487-498 (2008).

6 Condeelis, J. S., Wyckoff, J. and Segall, J. E. Imaging of Cancer Invasion and Metastasis using Green Fluorescent Protein. *European Journal of Cancer* **36** (13), 1671-1680 (2000).

7 Aequorea victoria Fluorescent Proteins. Carl Zeiss Microscopy Online Campus. http://zeiss-campus.magnet.fsu.edu/print/probes/jellyfishfps-print.html

8 Rosenow, M.A., Huffman, H.A., Phail, M.E. & Wachter, R.M. The crystal structure of the Y66L variant of green fluorescent protein supports a cyclization-oxidation-dehydration mechanism for chromophore maturation. *Biochemistry* **43**, 4464–4472 (2004).

9 Barondeau, D.P., Kassmann, C.J., Tainer, J.A. and Getzoff, E.D. Understanding GFP chromophore biosynthesis: controlling backbone cyclization and modifying posttranslational chemistry. *Biochemistry* **44**, 1960–1970 (2005).

10 Lemay, N. P., Morgan, A. L., Archer, E. J., Dickson, L. A., Megley, C. M. and Zimmer, M. Chem. Phys. **348**, 152–160 (2008).

11 Evdokimov, A. G., Pokross, M. E., Egorov, N. S., Zaraisky, A. G., Yampolsky, I. V., Merzlyak, E. M., Shkoporov, A. N., Sander, I., Lukyanov K. A. and Chudakov, D. M. Structural basis for the fast maturation of Arthropoda green fluorescent protein. *EMBO Rep*. **7**, 1006–1012 (2006).

12 Bertz, M., Kunfermann, A. & Rief, M. Navigating the Folding Energy Landscape of Green Fluorescent

Protein. *Angewandte Chemie International Edition*. **47**, 8192-8195 (2008).

13 Reddy, G., Liu, Z. and Thirumalai, D. Denaturant-dependent folding of GFP. *PNAS*.**109**, 17832-17838 (2012).

14 Orte, A., Craggs, T. D., White, S. S., Jackson, S. E., and Klenerman, D. Evidence of an intermediate and parallel pathways in protein unfolding from single-molecule fluorescence, *Journal of the American Chemical Society* **130**, 7898-7907 (2008).

15 Mickler, M., Dima, R. I., Dietz, H., Hyeon, C., Thirumalai, D., and Rief, M. Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations, *Proceedings of the National Academy of Sciences of the United States of America* **104**, 20268-20273 (2007).

16 Huang, J. R., Craggs, T. D., Christodoulou, J., and Jackson, S. E. Stable intermediate states and high energy barriers in the unfolding of GFP, *Journal of Molecular Biology* **370**, 356-371 (2007).

17 Enoki, S., Maki, K., Inobe, T., Takahashi, K., Kamagata, K., Oroguchi, T., Nakatani, H., Tomoyori, K., and Kuwajima, K. The equilibrium unfolding intermediate observed at pH 4 and its relationship with the kinetic folding intermediates in green fluorescent protein, *Journal of Molecular Biology* **361**, 969-982 (2006).

18 Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotech*. **24**:79-88 (2005)

19 Craggs, T.D. Green fluorescent protein: structure, folding and chromophore maturation. *Chem. Soc. Rev*. **38**: 2865-2875 (2009).

20 Baird, G. S., Zacharias, D. A., and Tsien. R.Y. Circular Permutation and Receptor Insertion within Green Fluorescent Proteins. *Proceedings of the National Academy of Sciences* **96** (20), 11241-11246 (1999).

21 Makino, Y., Amada, K., Taguchi, H., and Yoshida, M. Chaperon-mediated folding of green fluorescent protein, *J. Biol. Chem*. **272**, 12468-12474 (1997).

22 Ward, W. W., and Bokman, S. H. Reversible denaturation of Aequorea Green-Fluorescent Protein: Physical separation and characterization of the renatured protein., *Biochemistry*. **21**, 4535-4540 (1982).

23 Bokman, S. H., and Ward, W. W. Renaturation of Aequorea green fluorescent protein. *Biochem. Biophys. Res. Commun*. **101**, 1372-1380 (1981).

24 Capraro, D. T., Roy, M., Onuchic, J. N., Gosavi, S., and Jennings, P. A. Beta-Bulge triggers route-switching on the functional landscape of interleukin-1 beta, *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1490-1493 (2012).

25 Andrews, B. T., Roy, M., and Jennings, P. A. Chromophore Packing Leads to Hysteresis in GFP, *Journal of Molecular Biology* **392**, 218-227 (2009).

26 Li, B. S., Shahid, R., Peshkepija, P., and Zimmer, M. Water diffusion in and out of the beta-barrel of GFP and the fast maturing fluorescent protein, TurboGFP, *Chemical Physics* **392**, 143-148 (2012).

27 RCSB PDB Bank. http://www.rcsb.org/pdb/home/home.do

28 Chang, G., Guida, W. C., and Still, W. C. An internal-coordinate monte carlo method for searching conformational space, *J. Am. Chem. Soc.* **111**, 4379-4386 (1989).

29 Saunders, M., Houk, K., Wu, Y.-D., Still, W., Lipton, M., Chang, G., and Guida, W. Conformations of cycloheptadecane. A comparison of methods for conformational searching., *J. Am. Chem. Soc.* **112**, 1419 (1990).

30 Bartol, J., Comba, P., Melter, M., and Zimmer, M. Conformational searching of transition metal compounds, J Comput Chem **20**, 1549-1558 (1999).

31 Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., and Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. , in Proceedings of the 2006 ACM/IEEE Conference on Supercomputing (SC06), ACM Press, Tampa, FL (2006).

32 Onuchic, J. N., LutheySchulten, Z., and Wolynes, P. G. Theory of protein folding: The energy landscape perspective, *Annu Rev Phys Chem **48***, 545-600 (1997).

33 Onuchic, J. N., and Wolynes, P. G. Theory of protein folding, *Current Opinion in Structural Biology,* **14**, 70-75 (2004).

34 Liu, Y., and Bahar, I. Sequence Evolution Correlates with Structural Dynamics, *Molecular Biology and Evolution* **29**, 2253-2263 (2012).

35 Tang, G. W., and Altman, R. B. Remote Thioredoxin Recognition Using Evolutionary Conservation and Structural Dynamics, *Structure* **19**, 461-470 (2011).

36 Dill, K. A., and Chan, H. S. From Levinthal to pathways to funnels, *Nature Structural Biology* **4**, 10-19 (1997).

37 Mirny, L. A., and Shakhnovich, E. I. Universally conserved positions in protein folds: Reading evolutionary signals about stability, folding kinetics and function, *Journal of Molecular Biology*. ***291***, 177-196 (1999).

38 Zimmer, M. Molecular mechanics, data and conformational analysis of first-row transition metal complexes in the Cambridge Structural Database, *Coordination Chemistry Reviews* **212**, 133-163

(2001).

39 Temiz, N. A., and Bahar, I. Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase, *Proteins-Structure Function and Genetics*. **49**, 61-70 (2002).

40 Andrews, B. T., Gosavi, S., Finke, J. M., Onuchic, J. N., & Jennings, P. A. The Dual-basin Landscape in GFP Folding, *PNAS*. **105**, 12283-12288 (2008).

41 Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. C. Improved green fluorescent protein by molecular evolution using DNA shuffling., *Nat. Biotechnol*. **14**, 315-319 (1996).

42 Heim, R., Prasher, D. C., and Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein., Proc. *Natl. Acad. Sci. USA 91*, 12501-12504 (1994).

43 Andrews, B. T., Schoenfish, A. R., Roy, M., Waldo, G., & Jennings, P. A. The rough energy landscape of superfolder GFP is linked to the chromophore. *Journal of Molecular Biology*, **373**(2), 476-490 (2007).

44 Anisotropic Network Model Web Server. Http://ignmtest.ccbb.pitt.edu/cgi-bin/anm/anm1.cgi