Development of NMR as a tool for the structural and dynamic high-resolution characterization of phototranformable fluorescent proteins

Nina-Eleni Christou

Université Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **Physique pour les Sciences du Vivant**

Arrêté ministériel : 25 mai 2016

Présentée par

Nina-Eleni CHRISTOU

Thèse dirigée par **Bernhard BRUTSCHER**,
et codirigée par **Dominique BOURGEOIS**

préparée au sein du **Biomolecular NMR spectroscopy group,** et **PIXEL team, Institut de Biologie Structurale**
dans **l'École Doctorale Physique**

Development of NMR as a tool for the structural and dynamic high-resolution characterization of phototransformable fluorescent proteins

Thèse soutenue publiquement le **18/12/2020,**
devant le jury composé de :

Prof., Hideaki, MIZUNO
Professeur, Katholieke Universiteit Leuven, Belgique, Rapporteur
Prof., Bruno, KIEFFER
Professeur, Université de Strasbourg, France, Rapporteur
Prof., Antoine, DELON
Professeur, Université Grenoble Alpes, Président
Dr., Martin, WEIK
DR-CEA , IBS, Grenoble, France, Examinateur
Dr., Carine, VAN-HEIJENOORT
DR-CNRS, ICSN, Gif-sur-Yvette, France, Examinateur
Dr., Michel, SLIWA
DR-CNRS, Université de Lille, France, Examinateur

DEVELOPMENT OF NMR AS A TOOL FOR THE STRUCTURAL AND DYNAMIC HIGH-RESOLUTION CHARACTERIZATION OF PHOTOTRANSFORMABLE FLUORESCENT PROTEINS

By

NINA-ELENI CHRISTOU

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

UNIVERSITÉ GRENOBLE ALPES
École Doctorale de Physique

OCTOBER 2020

ACKNOWLEDGMENT

My PhD thesis was an exciting and ultimately rewarding journey. During it, I was able to learn a lot about science and the tools required to be able to not only tackle questions concerning natural phenomena, but also handle the challenging lifestyle of a scientist. Probably the most important lesson that this experience has taught me is that, no one can achieve success or happiness alone. Therefore, in this part, I would like to thank all the people that were there for me and, together, we were able to overcome the biggest hardships of everyday life and research. Without them, I wouldn't have become the person or the scientist that I am today.

Firstly, I would like to thank Dr. Bernhard Brutscher. Some of the most exciting moments during my time in IBS were the long and lively discussions we had concerning our work. Having someone who was as passionate and curious about unidentified Chromophore NMR peaks as I was, definitely made my PhD experience much more pleasant and wholesome. I learnt a great deal of things working with Bernhard, regarding NMR but also scientific ethos and integrity. So I thank him for guiding me through this difficult journey, as a supervisor, and providing me with the necessary support and solidarity, as a colleague.

I would like to also thank Dr. Dominique Bourgeois, my co-supervisor. His expertise in fluorescent proteins and microscopy always inspired me to ask more questions and further investigate fluorescent proteins by NMR and other techniques. I thank him for his friendly and patient attitude every time I found myself confused, and the thorough explanations he provided me that allowed me to advance my knowledge.

For their help towards understanding NMR theory, I would like to thank Dr. Dominique Marion and Dr. Paul Schanda. By setting up the NMR theory course, they allowed myself

and other PhD students to have a better grasp of our projects, of the NMR experiments we are running, which, ultimately, helped in interpreting our (sometimes) complicated NMR data. Outside of the NMR course, I found that both were keen to know the progress of my work and would always be available for advice and discussions, regarding my project, NMR in general, and beyond. So thanks Paul and Dominique, and I hope to have many occasions to share a beer with you and “pick” your brains...

Furthermore, I would like to thank everyone in the NMR group and the PIXEL team for the incredible support; Isabel and Karine for producing my beautifully fluorescent proteins, Alicia and Adrien for always being there when the spectrometers decided not to work, and Cedric, for when my computer would decide not to work. I would like to thank Cedric, Beate, Catherine for all the advice regarding chemistry related questions, Iva, Tiago and Undina for taking the time (and possibly the risk!) to help me out in the wetlab, Ricarda and Faustine for our interesting scientific conversations during lunch at IBS and Tomas for helping me with almost everything that I can think of. In the PIXEL team, I would like to thank Daniel and Sasha for supporting me through the optical measurements experiments during my project and Sasha for actually taking part in the measurements and data analysis at times. I would like to thank Virgile and Martin for showing me how to work in the optical spectroscopy set-ups and how to extract reliable values for my proteins’ optical properties.

In addition, I would like to thank all the members of my thesis committee who gave me invaluable feedback and advice on how to continue my work; Dr Martin Weik, Prof Ludovic Jullien and Dr Paul Schanda. Their attentiveness and expertise enabled me to guide my project into interesting paths.

I would like to also thank the different PSB platforms that I used, the AUC unit and the microscope (M4D) platform for allowing me to investigate my samples and for providing me with all the necessary practical support I needed.

On a personal note, I would like to thank all my friends and family for being there for me, for the good times and the bad. It is difficult to remain motivated and hopeful in our

current world and the stress of a PhD, with all the discomforts and existential questions that come with it, can sometimes exaggerate this feeling of despair. I realised, after having moved between countries, labs, projects, research fields, etc. that the only important thing that can ensure happiness is having a network of valuable and irreplaceable people in my life with whom I have built a strong connection. Such connection goes beyond common interests, ideology, personality traits or language. It is a true human connection, one that can help me survive through a PhD, a pandemic and anything else that I might be faced with. So thank you for being there for me and with me; Tiago, Iva, Ricarda, Undina, Lefteris, Tomas, Ariel, Daniel, Lisa, Fotini, Sofia, Loli, Dani, Kelly, Simona, Aldo, Alicia, Ana, Laura and towards the end of it, Benoît.

DEVELOPMENT OF NMR AS A TOOL

FOR THE STRUCTURAL AND DYNAMIC HIGH-RESOLUTION CHARACTERIZATION

OF PHOTOTRANSFORMABLE FLUORESCENT PROTEINS

by Nina-Eleni CHRISTOU, M.Sc.
Université Grenoble Alpes
October 2020

Abstract

English version

The discovery of Phototransformable Fluorescent proteins (PTFPs) over the last decades has revolutionized the field of microscopy. Reversibly photo-switchable fluorescent proteins (RSFPs), in particular, are currently routinely used for Super Resolution Microscopy techniques, such as RESOLFT (REversible Saturable OpticaL Fluorescence Transitions). Photo-induced switching between a fluorescent "on"- and a dark "off"-state, in combination with advanced illumination schemes has allowed for imaging nanometer sized compartments in biological cells. Crystallographic studies of such RSFPs have provided useful mechanistic explanations for their photophysical behaviour and has guided fluorescent protein engineering into designing better tags. However, the crystal forms of such proteins studied at cryogenic temperatures fail to capture dynamics present in RSFPs which could potentially play a significant role in their photophysics. So far, only a single NMR study for the RSFP Dronpa has been reported in the literature [147]. During my PhD thesis, I was able to complement crystallographic studies of rsFolder, a green RSFP, with a dynamic perspective using multidimensional solution NMR spectroscopy.

Using a portable in-situ laser illumination device coupled with the NMR spectrometer,

I was able to extract quantitative local dynamic information for both the fluorescent "on"- and "off"-states of rsFolder, characterized by a primarily *cis* and *trans* chromophore, respectively. NMR signatures of residues in the non-fluorescent "off"-state were identified using LASER-driven Exchange NMR experiments. The metastable photo-induced "off"-state of rsFolder appears more dynamic on the millisecond timescale than the fluorescent "on"-state. NMR investigations of the chromophore resulted in the deciphering of four configurations, populated in a pH-dependent fashion. Moreover, pH-induced *cis-trans* isomerization of the chromophore was observed, in the absence of light. NMR-derived values of activation energies for isomerization and free energy differences between the *cis* and *trans* chromophore enabled the mapping of the ground-state free energy landscape of rsFolder at different pH values and buffer compositions. Lastly, comparing NMR observables with optical measurements on rsFolder and mutants highlights the potential role that NMR can play in the field of RSFP engineering. Altogether, my PhD work yielded in not only a reliable in-situ illumination set-up accompanied with relevant NMR experiments to study RSFPs, but also highlighted the importance of dynamics in understanding RSFPs' photophysical properties.

Version française

La découverte de protéines fluorescentes photo-transformables (PTFP) au cours des dernières décennies a révolutionné le domaine de la microscopie optique. Les protéines fluorescentes réversiblement commutables (RSFP), en particulier, sont couramment utilisées pour les techniques de microscopie à super-résolution comme en RESOLFT (REversible Saturable OpticaL Fluorescence Transitions) par exemple. Par photoactivation, les RSFP passent d'un état "on" - fluorescent - à un état "off" - éteint - qui, combiné à des systèmes d'illumination avancés permet d'imager des composants cellulaires préalablement marqués à une résolution nanométrique. De nombreuses études cristallographiques sur les RSFP ont apporté des informations structurelles importantes et ont permis de dresser des hypothèses quant à leur comportement photo-physique. Elles ont également guidé

l'ingénierie des protéines fluorescentes afin d'améliorer leur conception et leur utilisation in vivo. Cependant, les cristaux de ces protéines qui sont étudiées à des températures cryogéniques ne permettent pas de capturer la dynamique moléculaire des RSFP dans le but de comprendre, voir d'améliorer leur propriétés photo-physique. C'est pourquoi au cours de ma thèse, j'ai majoritairement utilisé la résonance magnétique nucléaire (RMN) en solution multidimensionnelle sur une RSFP verte - appelée rsFolder - afin de compléter et améliorer nos connaissances sur ces protéines.

À l'aide d'un dispositif d'éclairage laser in situ portatif couplé au spectromètre RMN, j'ai pu extraire des informations dynamiques locales quantitatives concernant les états "on" et "off" fluorescents de rsFolder qui sont respectivement caractérisés par un chromophore en conformation *cis* et *trans*. Les signatures des résidus dans l'état "off" non fluorescent ont été identifiées à l'aide d'expériences de RMN d'échange induite par le LASER. L'état "off" métastable de rsFolder apparaît plus dynamique à l'échelle de temps de la milliseconde que l'état "on" fluorescent. La RMN a également permis de mettre en lumière quatre configurations du chromophore possible qui sont pH dépendante. De plus, j'ai observé l'isomérisation du chromophore induite par le pH *cis-trans*, en l'absence de lumière. Les valeurs dérivées de la RMN des énergies d'activation concernant l'isomérisation et les différences d'énergie libre entre le chromophore *cis* et *trans* ont permis de cartographier le paysage d'énergie libre de l'état fondamental de rsFolder à différents pH. Enfin, la comparaison de données de RMN et des mesures optiques sur rsFolder ainsi que sur différents mutants a mis en évidence le rôle important que la RMN peut jouer dans le domaine de l'ingénierie des RSFP.

Dans l'ensemble, mes travaux de thèse ont permis non seulement d'établir un système d'illumination in situ fiable, accompagné d'expériences de RMN pertinentes dans le but d'étudier les RSFP mais aussi de souligner l'importance de la dynamique moléculaire dans la compréhension des propriétés photo-physiques des RSFPs.

Contents

Page

ACKNOWLEDGMENT iii

ABSTRACT vi

LIST OF TABLES xiii

LIST OF FIGURES xiv

CHAPTER

1 Introduction 3
1.1 The history of fluorescent proteins 3
1.1.1 The discovery of the Green Fluorescent Protein (GFP) 3
1.1.2 The chromophore of the GFP – formation and electronic properties 5
1.1.3 Phototransformable Fluorescent Proteins (PTFPs) 8
1.1.4 Super-Resolution Microscopy 10
1.1.5 Application of RSFPs beyond SR microscopy 12
1.1.6 Reversibly Photoswitchable Fluorescent Proteins (RSFPs) used in SR microscopy – advantages and drawbacks 13
1.2 Biophysical characterisation of RSFPs 17
1.2.1 Photophysical/Optical characterisation of RSFPs 18
1.2.2 Crystallographic studies of RSFPs 24
1.2.3 Molecular dynamics simulations and Ultrafast vibrational spectroscopy as complementary techniques in the discovery of new RSFPs 29
1.2.4 Our model system of negative RSFP: rsFolder and mutants 32
1.3 Multi-dimensional solution NMR to add a dynamic view on RSFPs 35
1.3.1 Achieving atomistic information – the chemical shift 35

1.3.2 NMR as an ensemble technique . . . 36
1.3.3 The effect of Radiofrequency (RF) pulses and signal detection . . 38
1.3.4 1D versus nD heteronuclear NMR spectroscopy – sensitivity and resolution . . . 39
1.3.5 An example of a two-dimensional NMR pulse sequence: Heteronuclear Single Quantum Coherence (HSQC) spectroscopy . . . 43
1.3.6 Resonance frequency assignment to protein residues . . . 45
1.3.7 Studying protein dynamics using nD solution NMR . . . 45
1.3.8 Limitations of nD solution NMR for studying proteins . . . 65
1.3.9 Using NMR to study fluorescent proteins . . . 65
1.3.10 Thesis objectives: Studying rsFolder and mutants by nD NMR . . . 67

2 Materials & Methods . . . 68
2.1 rsFolder sample preparation . . . 68
2.1.1 Protein purification of Uniformly labelled rsFolder samples – with and without His-tag . . . 68
2.1.2 rsFolder pH series preparation . . . 69
2.2 In-situ LASER illumination in NMR spectrometer . . . 69
2.2.1 LASER box . . . 69
2.2.2 Optical fiber: use and maintenance . . . 70
2.2.3 Modifications in NMR setup . . . 70
2.2.4 Assessing the in-situ illumination setup using rsFolder . . . 71
2.3 rsFolder backbone and side chain assignment – “on”- and “off”- state . . 71
2.3.1 Backbone amide assignment . . . 71
2.3.2 Side chain assignment . . . 75
2.3.3 Structural information from C_α and C_β frequencies . . . 78
2.4 NMR experiments for protein dynamics characterization . . . 78
2.4.1 Fast conformational exchange . . . 78
2.4.2 Intermediate conformational exchange . . . 79
2.4.3 Slow conformational exchange . . . 79
2.5 Optical characterization of rsFolder and mutants . . . 81
2.5.1 Absorption and Emission spectra . . . 81
2.5.2 Extinction coefficient, fluorescence Quantum Yield and Brightness 81

2.5.3 Ensemble microscopy measurements to extract switching rates and switching contrast . . . 82

3 Results . . . 83
3.1 In-situ LASER illumination in NMR spectrometer . . . 83
3.1.1 Choosing the LASER box . . . 83
3.1.2 In-situ illumination of the NMR sample . . . 87
3.1.3 Performance of the in-situ illumination set-up in NMR spectrometer 91
3.1.4 Advantages, limitations and future perspectives of the in-situ illumination setup . . . 94
3.2 NMR studies of rsFolder "on"- and "off"-states . . . 96
3.2.1 Sample and spectral quality control . . . 97
3.2.2 Distinction between "on"- and "off"- states – NMR-observed switching contrast . . . 97
3.2.3 NMR-observed photobleaching . . . 99
3.2.4 Effect of photocycling and LASER power on NMR observables . . . 101
3.2.5 NMR-observed thermal relaxation . . . 102
3.2.6 Establishing the LASER-Driven EXSY experiment used for backbone assignment of rsFolder "off"-state . . . 104
3.2.7 Secondary structural elements prediction from NMR CS using the TALOS-N program . . . 105
3.2.8 Characterization of rsFolder backbone dynamics in the "on"- and "off"-states . . . 107
3.2.9 The "off"-state shows increased water accessibility and barrel instability . . . 110
3.2.10 Characterization of rsFolder side-chain and chromophore dynamics in the "on"- and "off"-states . . . 113
3.2.11 Interpretation of NMR findings of rsFolder conformation and dynamics upon *cis-trans* isomerization . . . 123
3.3 NMR investigation of ground-state chromophore isomerization and protonation in rsFolder . . . 129
3.3.1 NMR deciphers the various chromophore configurations . . . 130

3.3.2 NMR investigation of the *cis*-CRO of rsFolder in the absence of light 131
3.3.3 NMR investigation of the *trans*-CRO of rsFolder 138
3.3.4 Impact of NMR observations of pH-induced effects on the optical properties of rsFolder 147
3.3.5 Interpretation of NMR results on ground-state (de)protonation and isomerization kinetics 150

4 Preliminary work and future perspectives 157
4.1 Correlating NMR observables with photophysical properties of rsFolder mutants 157
4.1.1 Rationale for choosing rsFolder mutants 157
4.1.2 Optical studies on rsFolder2 and mutants 158
4.1.3 NMR studies on rsFolder2 and mutants 160
4.1.4 Comparing NMR observables and optical measurements 160
4.1.5 Interpretation of NMR and optical studies on rsFolder mutants . . 164
4.2 Preliminary NMR and optical studies of rsFolder in different sample conditions relevant to microscopy 166
4.3 Future perspectives for studying rsFolder to complement and/or prove NMR conclusions 167

5 Conclusion 171

REFERENCES 194

APPENDIX

A 196

List of Tables

1.1 Examples of RSFPs and their properties mentioned in the Introduction . . . 17

2.1 Volume of Phosphate buffer and Citric acid required to prepare the McIlvaine pH series 69

3.1 Chromophore NMR chemical shifts 119

3.2 Fitted values from fluorescence photocycles of rsFolder pH=5 and pH=8 . . 151

4.1 Fitted values from fluorescence photocycles of rsFolder2 and mutants 159

4.2 Results from NMR measurements of the properties of rsFolder2 wild-type(WT) and mutants 161

List of Figures

1.1 Crystal structure and excitation and emission profile of GFP(S65T) 4

1.2 Chromophore Formation 6

1.3 Absorption and emission spectra in wtGFP 7

1.4 PTFPs chromophores 9

1.5 Example of an image captured by RESOLFT versus Confocal microscopy . . 11

1.6 Doughnut-shaped beam in RESOLFT 12

1.7 Principle of RESOLFT microscopy using a negative green GFP-type RSFP . . 15

1.8 Absorption spectra of rsFolder 16

1.9 Switching rates of different RSFPs 17

1.10 Photocycles and kinetic model fitting in RSFPs 20

1.11 Example of fixation effect on photophysical properties of RSFPs 21

1.12 Comparison between single molecule and ensemble studies of Dronpa . . . 23

1.13 Zooming in the overlapped crystallographic structures of Dronpa “on”- and “off”-states, in the vicinity of the phenol ring of the CRO 25

1.14 X-ray refinement model comparison 26

1.15 Zooming in the chromophore region of the crystallographic structure of IrisFP 27

1.16 rsEGFP2 XFEL structure 28

1.17 rsFolder vs rsEGFP2 chromophore switching 33

1.18 rsFolder side chains comparison between crystal structures of “on”- and “off”-states 34

1.19 Tautomeric states of Histidines observed by NMR 37

1.20 Net magnetisation of spin system 38

1.21 Effect of 90° RF pulse on net magnetization 39

1.22 Signal detection by NMR 40

1.23 One dimensional vs Two-dimensional spectrum 41

1.24 NMR pulse sequence basics 43

1.25 HSQC pulse sequence basics 44

1.26 NMR Lineshapes 47

1.27 Correlation function 48

1.28 Spectra density function 49

1.29 Differences between τ_c and τ_e 50

1.30 Dipolar couplings and CSA affecting spins while tumbling 52

1.31 Transverse relaxation scheme 53

1.32 relaxation equation graph 54

1.33 N15 T_1 relaxation experiment 56

1.34 N15 T_2 relaxation experiment 57

1.35 N15 CPMG-RD relaxation experiment 60

1.36 NMR EXSY experiment 62

1.37 Real-time NMR scheme 63

2.1 TROSY effect 73

2.2 Example of a 3D triple-resonance NMR experiment 74

2.3 LASER-Driven EXSY 76

3.1 LASER box components 85

3.2 Photo of laptop connected to LASER box in the NMR setup 86

3.3 Photo of TTL lines connected to the IPSO channel of the NMR console . . . 87

3.4 Fiber end scheme 88

3.5 Etched fiber end 89

3.6 Fiber end scheme 90

3.7 NMR observed “on”-to-“off” photoswitching 92

3.8 Differences in “on”-to-“off” photoswitching for a normal versus a pen-shaped plunger 93

3.9 Idea for additional optical set-up 95

3.10 Scheme of two-fiber system for detection of fluorescence during NMR acquisition 96

3.11 Temperature series on rsFolder 98

3.12 Overlaid ^{1}H-^{15}N BEST-TROSY spectra of rsFolder in the “on”- and “off”-state 99

3.13 NMR signal decrease during 488nm illumination 100

3.14 Evaluation of the effect of photocycling on observed NMR signals 102

3.15 NMR-observed thermal relaxation of rsFolder 103

3.16 Principle of LD-EXSY 106

3.17 NMR prediction of secondary structural elements using TALOS-N 107

3.18 Chemical shift and peak intensity differences on rsFolder “on”- and “off”-states amide spectra 109

3.19 ^{15}N T_1, T_2, hetNOE values for rsFolder, line broadening and ^{15}N-CPMG-RD . 110

3.20 Summarised scheme of rsFolder “off”-state dynamics 111

3.21 Summarised scheme of HDX results on rsFolder 112

3.22 rsFolder water channels modelled by PyMol CAVER 3.0 113

3.23 NMR-observed dynamics of methyl side chains upon 488nm illumination . . 114

3.24 Assignment of Histidine-specific NMR spectra 116

3.25 Assignment of Tyrosine-specific NMR spectra 117

3.26 Assignment of chromophore NMR spectra 118

3.27 NMR-observed dynamics of aromatic side chains upon 488nm illumination . 119

3.28 Dynamics in the “pocket” of the *trans* chromophore 120

3.29 NMR observations on the chromophore 121

3.30 Histidine 149 and chromophore H-bonding observed in NMR 123

3.31 NMR signatures of the 4 chromophore configurations 131

3.32 NMR-observed pKa of residues near the *cis* chromophore 132

3.33 ^{15}N-CPMG-RD data for amides near the *cis* chromophore 134

3.34 Summary of ring flipping rates for the two configurations of the *cis* chromophore136

3.35 NMR spectral signatures of “on”-state H149 at different pH values 137

3.36 Summary of exchange kinetics of rsFolder *cis* chromophore as observed by NMR 138

3.37 H_β-C_β NMR cross-peaks of N-*trans* and A-*trans* chromophore 139

3.38 Line narrowing of “off”-state NMR signals upon increase of Hydrogen ion availability 140

3.39 NMR peak line narrowing upon increase of temperature in the “off”-state . . 140

3.40 pH titration in the methyl region of the “off”-state 141

3.41 Thermal relaxation times for rsFolder at different pH 143

3.42 Free energy landscape of rsFolder at different pH 143

3.43 Ground state free energy landscape for rsFolder at different pH and buffer composition 144

3.44 pH titration of “off”-state H149 146

3.45 Light-induced population ratio of major chromophore configurations at different pH values 148

3.46 Optical measurements of rsFolder at different pH values 149

3.47 NMR-derived N-trans population at different pH values after 405nm illumination 155

4.1 NMR pH titration of Threonines of rsFolder2 mutants 162

4.2 Correlation between NMR observables and optical measurements for rsFolder2 mutants 163

4.3 Comparison of fluorescence photocycles of rsFolder with and without MEA . 167

4.4 Comparison of 2D amide and methyl spectra of rsFolder with and without MEA168

A.1 Assigned amide spectra of rsFolder “on”-state at 40°C 196

LIST OF ACRONYMS

A-*cis* Anionic-*cis*

A-*trans* Anionic-*trans*

BEST Band-selective Excitation Short-Transient

CPMG-RD Carl Purcell Meiboom–Gill Relaxation-dispersion

CRO Chromophore

CS Chemical Shift

CSA Chemical Shift Anisotropy

DD Dipole-Dipole

dSTORM direct STochastic Optical Reconstruction Microscopy

EXSY EXchange SpectroscopY

FCS Fluorescence Correlation Spectroscopy

FP Fluorescent Protein

GFP Green Fluorescent Protein

HDX Hydrogen-Deuterium eXchange

HMQC Heteronuclear Multiple Quantum Coherence

HSQC Heteronuclear Single Quantum Coherence

LASER Light Amplification by Stimulated Emission of Radiation

LED Light-Emitting Diodes

MEA β-mercaptoethylamine

N-*cis* Neutral-*cis*

N-*trans* Neutral-*trans*

nD Mutli-Dimensional

NMR Nuclear Magnetization Resonance

OPIOM Out-of-Phase Imaging after Optical Modulation

PAFPs Photo-Activatable Fluorescent Proteins

PALM Photo-Activated Localization Microscopy

PAM Photo-Acoustic Microscopy

PCFPs Photo-Convertible Fluorescent Proteins

PTFPs Photo-Transformable Fluorescept Proteins

RESOLFT REversible Saturable OpticaL Fluorescence Transitions

RF Radio-Frequency

RSFPs Reversibly photo-Switchable Fluorescent Proteins

SOFAST band-Selective Optimized Flip-Angle Short-Transient

SR Super-Resolution

STED Stimulated Emission Depletion

TROSY Transverse-Relaxation Optimized SpectroscopY

WT Wild-Type

Chapter One

Introduction

1.1 The history of fluorescent proteins

1.1.1 The discovery of the Green Fluorescent Protein (GFP)

In the quest to understand the function of biological cells and cellular networks, scientists have historically made use of the optical microscope. Dating all the way back to 1674, with the visualization of red blood cells by Van Leeuwenhoek [91], the light microscope in its many modalities (conventional light, fluorescent, confocal, super-resolution, two-photon etc.) is commonly used to retrieve localized information on intracellular activity or to decipher intercellular communication.

More specifically, fluorescence microscopes revolutionized the visualization of biological cells during the early 1900s [182]. Fluorescence microscopy, as compared to conventional light microscopy, utilizes fluorescent tags to enhance the signal received from cellular compartments of interest. The signal to noise ratio achieved is incomparable to the one of conventional light microscopy due to the use of tags that are excited by high intensity light which are then detected at a specific channel (e.g green-band pass filter for green fluorophores) [186]. Most importantly, the use of such tags ensures significantly improved specificity. Overall, the image sensitivity and resolution achieved depends primarily on the quality of sample staining – labelling specificity, fluorescent tag brightness, lifetime of fluorescence, fluorescent contrast of the tag as compared to auto-fluorescence of the specimen – and secondarily, on the microscope's specifications.

Currently, fluorescent tags used in microscopy can be grossly categorized into two groups; synthetic fluorophores and genetically encoded fluorescent proteins (FPs). The latter consists of proteins that are found in several biological organisms (mainly Hydrozoa and Anthozoa [44]) with the *Green Fluorescent Protein* (GFP) being the first one to demonstrate FPs use as fluorescent biomarkers. In 1962, Osamu Shimomura isolated the nucleotide sequence of the wild-type GFP (wtGFP), a protein found in a species of jellyfish *Aequorea*

victoria [198]. Decades later, in 1992, Douglas Prasher et al achieved the cloning of wtGFP [176] while expression of wtGFP in *E.coli* and *C.elegans* was performed for the first time in 1994 by scientists from the lab of Martin Chalfie [33]. A year later, Roger Tsien's lab reported the first significantly brighter mutant of wtGFP, namely the S65T mutant [98].

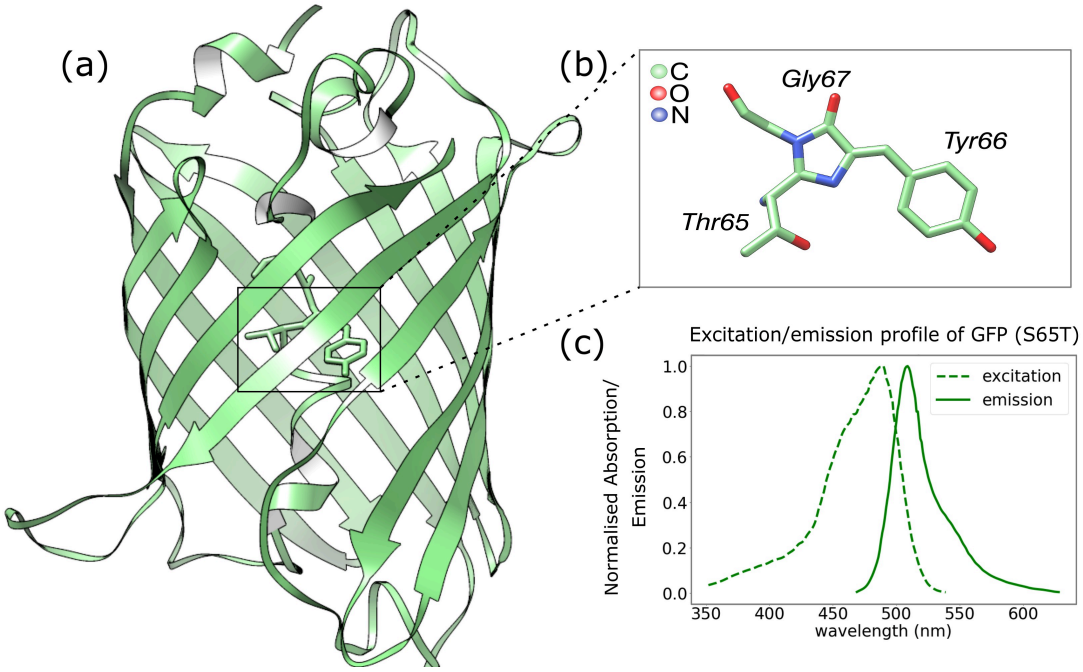


Figure 1.1 (a)Structure of the Green Fluorescent Protein (S65T), PDB:1EMA, in (b) zooming into the chromophore and (c) its corresponding excitation and emission spectrum

.

Notably, the mechanism of GFP's fluorescence was deciphered only after the S65T was introduced, and its crystal structure was published in an issue of Science in 1996 [160]. The structure revealed a 11-strand β-barrel protein, enclosing a chemical moiety which would be identified as the *chromophore* (CRO) of the GFP. Helices capping the barrel and extending inside its centre are disrupted by the CRO. Revealed for the first time, the chromophore responsible for the GFP's fluorescence was described as the product of 3 aminoacid residues (Ser65-Tyr66-Gly67, SYG) encoded in the gene of the GFP (Figure 1.1).

The triad of SYG, when in proximity during the folding of the barrel of the GFP(S65T), undergoes an autocatalytic reaction during the "maturation" step of the protein. This process results in a chromophore consisting of a phenol ring (p-hydroxy-benzylidene-imidazolinone), which, when deprotonated, can emit a green photon (at 509nm) upon the absorption of a blue photon (at around 489nm). The work of Tsien, Chalfie and Shimomura paved the way for many discoveries to come concerning the cellular world. For their work on the GFP, the first gene-encoded fluorescent tag, they were awarded the Nobel Prize in Chemistry in 2008.

1.1.2 The chromophore of the GFP – formation and electronic properties

It is precisely the endogenously synthesized CRO that makes the GFP a crucial addition to the toolbox of fluorescent tags. More specifically, the CRO allows for the molecular labeling of specimens without the need of other co-factors except oxygen (which results in the production of one peroxide molecule per chromophore) [238, 138]. Since its discovery, mutations have been introduced in order to further enhance or change its photophysical properties. Nevertheless, CROs found in GFP-type proteins are encoded as 3 amino acid residues, with a conserved pattern of Xxx – Tyr – Gly [162, 154].

There are two mechanisms of CRO formation suggested in the literature, both supported by several studies [48, 19, 83, 238, 183, 138]. They both propose that maturation follows well-timed steps which allow for the cyclization of the Xxx-Gly residue pair into a five-membered ring and a dehydration/oxidation step which forms the double bond (methine bridge of the CRO) of the Tyrosine ring with the imidazolinone [18, 48]. For wild-type GFP in particular, maturation follows the cyclization-oxidation-dehydration model, with the rate-limiting step being the production of a peroxide molecule [238, 138, 183]. More importantly, the β-barrel is what gives rise to favored conformations and electronical conjugation states in GFP CRO, by applying a type of mechanical compression onto its core, where the CRO lies [18, 139]. This distorted architecture forces atoms of the CRO in proximity which, subsequently, form covalent bonds such as the nitrogen of Gly67 with the carbon of the carbonyl of Thr65 in GFP (TYG CRO). Further evidence of the distortion is the existence of an "atypical" 3_{10} alpha helix, characteristic of β-barrel shaped FP proteins such as the GFP and its variants [18, 17].

Consequently, the CRO consists of two rings - a five-membered imidazolone ring and a six-membered phenol ring originating from Tyr66. Depending on the protonation of these two aforementioned rings, there can be different CRO species: a. deprotonated (or anionic), b. protonated (or neutral), c. zwitterrion, and d. cationic. These correspond to the phenol (at the Tyr Cz position) and the imidazolinone ring (at the Nd position) being: a. deprotonated and deprotonated, b. protonated and deprotonated, c. deprotonated and protonated, and d. protonated and protonated, respectively. The most commonly encountered CRO species are the deprotonated (anionic) and protonated (neutral) form.

Furthermore, the electronic configuration found in a GFP CRO is characterized by a large delocalised pi-electron system which is shared between the two rings of the CRO [209]. The allowed electronic states of this pi-system dictate the wavelength of maximum absorption. When extending the conjugated bond network, for example, the absorption (and therefore emission) wavelength is red-shifted [144]. It is the planarity and the rigidity of the CRO inside the barrel that influence its brightness capabilities [139, 154, 157, 212]; a free deprotonated CRO in solution, not fixed by the barrel, is essentially non-fluorescent [139]. That is because the CRO rings are allowed to freely twist in relation to each other upon absorption of light, and thus the characteristic delocalized pi-system is disrupted. Therefore, the GFP barrel acts not only as a "shield" against the solvent but also as a matrix

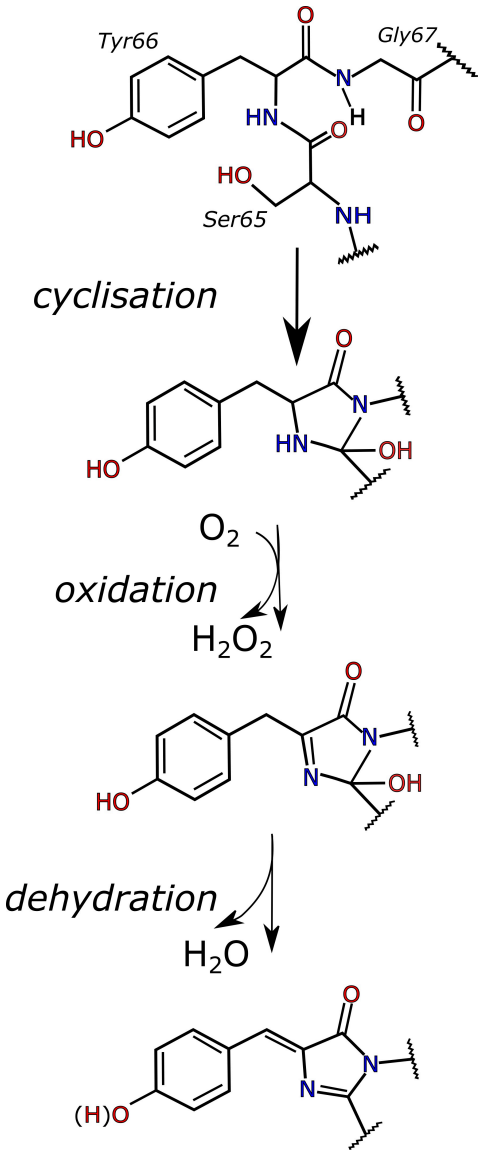


Figure 1.2 Formation of the wtGFP chromophore, from the autocatalytic reaction of Ser65, Tyr66 and Gly67. The first step is cyclisation of the imidazolinone ring, followed by oxidation and finally dehydration that forms the methine bridge between the phenol and the imidazolinone ring of the chromophore. The chromophore in its final form is in equilibrium between two species, characterised by a protonated and a deprotonated tyrosine ring, respectively.

.

that fixes the CRO, decreasing the chances of radiationless decay.

Moreover, the electronegativity of the CRO is of importance. In the original GFP which is relatively dim upon cyan light excitation, the probability of the CRO being in the neutral (protonated) form is quite high – the ratio of neutral and anionic species as retrieved

from optical spectroscopy in Tsien's lab is 6:1 [213]. The neutral phenol CRO absorbs UV light (395nm absorption peak) and is characterizaed by weak broadband emission in the range of 420-470nm [36], which is a wavelength range at which autofluorescence is prevalent [49]. Notably, in GFP UV-radiation can excite the phenol, causing it to lose a proton (Excited State Proton Transfer, or ESPT) and thus recovering the anionic species [51]. Interestingly, UV-excitation can in some cases cause *cis-trans* photoisomerisation of the neutral species [213]. The anionic CRO species in wtGFP are therefore contributing to the green fluorescence emission, with a characteristic maximum absorption wavelength at 470-475nm and emission at 504nm [213]. Progressively, mutations have been introduced to the gene of the wtGFP to further skew the ratio in favor of the anionic species since it is conventionally regarded as the 'fluorescent' species [167].

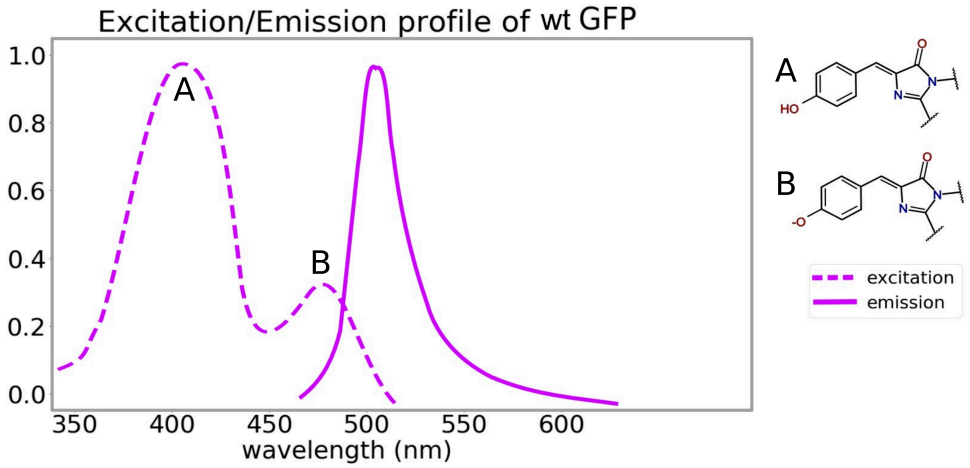


Figure 1.3 Absorption and emission spectra in wtGFP. Neutral CRO species absorb at 395nm while anionic CRO species absorb at 475nm and fluoresce at 504nm.

Since the discovery of the GFP, the development of a pallet of GFP variants as well as other fluorescent proteins originating from other organisms, have given rise to a variety of protein tags characterized by different photophysical properties such as color, brightness, Stoke's shift (i.e difference between the absorption and emission wavelength originating from the same electronic transition), photostability, folding, CRO maturation time, etc. The main advantage of gene-encoding tags is their high labelling specificity as compared to synthetic dyes [44, 154]. Not only are they expressed without the need of other bigger "linkers" (such as antibodies in the case of synthetic dyes) but they are also more bio-friendly (i.e no need to perfuse cells) in the systems where they have so far been utilized (from bacteria to mice) [44]. Nevertheless, their main limitations concern their low brightness as well as low photostability as compared to their synthetic counterparts [4]. All in all, their positive attributes are the main driving force behind the continuing efforts of protein engineers to further enhance FPs photophysical properties.

Interestingly, a synergy between the robust labelling specificity of FPs and the exceptional fluorescent behavior of synthetic dyes has lead to numerous attempts to combine the two

systems in the form of self-labeling enzyme tags (SNAP- [119], CLIP- [75] and HALO-tags [135]), and as of recently, fluorogenic proteins [76].

Self-labeling enzyme tags combine the use of specifically expressed enzymes with synthetic fluorescent dyes. The underlying chemistry consists of the ability of the protein of the system to covalently bind to a fluorescent dye.

Fluoregenic proteins are gene-encoded proteins that can bind to an exogenous synthetic chromophore [76]. This chromophore becomes only fluorescent upon binding in the protein pocket, ensuring high specificity of labelling.

1.1.3 Phototransformable Fluorescent Proteins (PTFPs)

Only a decade after the discovery of the GFP, Phototransformable Fluorescent Proteins (PTFPs) were developed [154, 3]. PTFPs radically change their optical properties upon illumination at certain wavelengths of light. Such photochemical reactions take place in the CRO and its neighboring environment as suggested by crystallography. So far, there are 3 main categories of PTFPs, depending on the effect that light of certain wavelength(s) has on their properties: a. photoactivatable fluorescent proteins (PAFPs), where the CRO is irreversibly entering a fluorescent state after illumination, b. photoconvertible fluorescent proteins (PCFPs), whereby an extension of the pi-electronic system of the CRO causes a redshift of its fluorescence emission wavelength and c. reversibly photoswitchable fluorescent proteins (RSFPs) which allow for the CRO to reversibly alternate between a fluorescent and a non-fluorescent state depending on the wavelength of illumination. [154, 26, 3].

PAFPs have been engineered to have primarily neutral species of the CRO present, therefore not fluorescent upon cyan light excitation (in the case of green-emitting PAFPs) [168]. Upon use of “activating” light (typically UV for green PAFPs) , these species become anionic (in their majority) due to the rearrangement of the hydrogen bond (HB) network around the chromophore [168, 137]. Using the example of PA-GFP, the HB rearrangement is a secondary effect of the photo-induced (upon UV illumination) decarboxylation of the Glu222 residue in the barrel [216, 215, 168]. Therefore, the process can be considered irreversible.

Similarly, PCPFs are FPs that undergo irreversible changes in their FP matrix that result in changes of their optical properties. Under specific wavelengths of illumination, an irreversible process, namely backbone cleavage, results in the extension of the pi-conjugated system of the CRO in PCFPs [10, 145, 228]. Ultimately, these processes translate to PCFPs that irreversibly change their optical properties from eg green-to-red fluorescence in the case of Kaede [10, 145].

Last on the list are RSFPs, proteins that can reversibly switch between a fluorescent and a non-fluorescent state upon illumination of light of two different wavelengths. The mechanism underlying photoswitching is (most commonly) photoisomerization, which

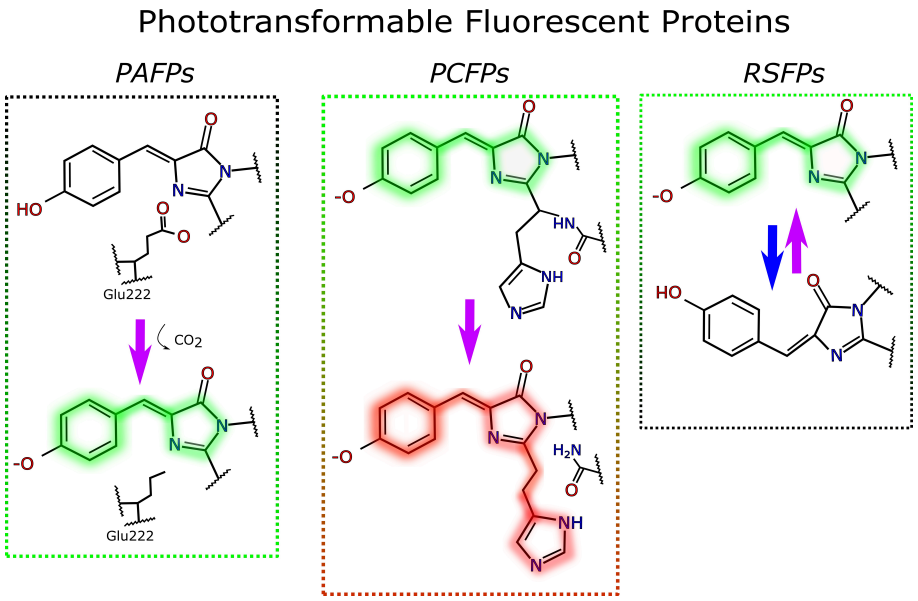


Figure 1.4 Schematic representation of the chromophore modifications upon light absorbance at certain wavelengths of Phototransformable Fluorescent Proteins.

is the result of the rotation of the two bonds consisting the methine bridge of the CRO. Photoisomerisation appears to be reversible (*cis-trans* and *trans-cis*) upon illumination of the protein with another wavelength of light or by allowing for the protein to thermally relax back to its thermally stable state [89, 13]. Exceptions include Dreiklang and SPOON, RSFPs undergoing photo-induced hydration/dehydration of the CRO in a reversible manner [27, 15].

Interestingly, derivatives of the green-to-red PCFP EosFP, have been reported to also undergo reversible photoswitching to a long-lived dark state in their respective green and red state in addition to their photochromic behaviour [5, 54] suggesting that PCFPs show RSFP-behaviour as well.

As mentioned above, the resolution and robustness of fluorescence microscopy techniques are dependent on the properties of the fluorescent tags available. Therefore, with the development of phototransformable fluorescent proteins came a wave of new microscopy techniques (PALM [22, 104], RESOLFT [105], SOFI [55], GSD [70]) that have completely revolutionized the field. While the resolution achieved by conventional fluorescent microscopy is limited to structures of around 250nm, super-resolution microscopy has managed to achieve the resolution of cellular components in the range of a few nanometers [184] (commonly now referred to as “nanoscopy”). Central to these modalities of imaging are fluorescent tags that can change their photophysical properties on demand during acquisition.

1.1.4 Super-Resolution Microscopy

Conventional (fluorescence) light microscopy is limited by the diffraction limit of light. Unless some post-acquisition image processing takes place, the observer cannot distinguish details of a size smaller than about 250nm [22]. This number is dictated by the maximum resolution, i.e the ability of the microscope to distinguish light originating from two sources. When emitting light of wavelength λ, a conventional microscope with an objective lens with numerical aperture of NA ($NA = nsin\alpha$, where n and α are the refractive index and aperture angle of the lens, respectively) can achieve a minimum resolvable distance d, as described by Ernst Karl Abbe:

$$d = \frac{\lambda}{2NA} = \frac{\lambda}{2nsin\alpha} \tag{1.1}$$

Therefore, when working with light of the near UV-visible range (250-700nm), the maximum resolution that can be achieved is 250nm [74].

Super-resolution (SR) microscopy has managed to bypass this physical limit using a combination of novel tags, illumination patterns and post-acquisition processing. Far-field super-resolution microscopy techniques are based on the use of a conventional microscope objective (contrary to more "unconventional" setups of near-field, scanning-tip techniques), thus making them the choice for surpassing the diffraction limit when visualising biological cells [99]. Such techniques aim at temporally distinguishing the signal originating from fluorophores so as to reconstruct the image after acquisition and can be grossly categorised in two groups:

1. Ensemble (or Deterministic) – including Stimulated Emission Depletion (STED), Ground State Depletion (GSD) and Reversible Saturable Optically Linear Transitions (RESOLFT)

2. Single-molecule (or Stochastic) – including Photoactivation Localization Microscopy (PALM), stochastic optical reconstruction microscopy (STORM)

RESOLFT microscopy utilizes reversibly photoswitchable fluorophores (such as RSFPs). More specifically, RESOLFT [105], is established as a more biological sample-friendly alternative to STED, since it requires an order of magnitude less laser power to operate – STED's average laser intensities are in the order of MW/cm^2 while RESOLFT only uses kW/cm^2 [100, 87]. However, the advantage of RESOLFT was only to be proven once RSFPs were used to label the sample [27, 87]. Since then, a combination of advanced illumination patterns and RSFPs properties has allowed the increased resolution and consequently reduced the acquisition time. [40, 41, 224].

Despite the differences with STED, the acquisition principle remains the same. Both STED and RESOLFT aim at retrieving fluorophores' positions by discriminating between two states – a dark and a bright state. In the case of STED, the bright state consists of the fluorophore undergoing spotaneous emission, while the dark state is induced by stimulated

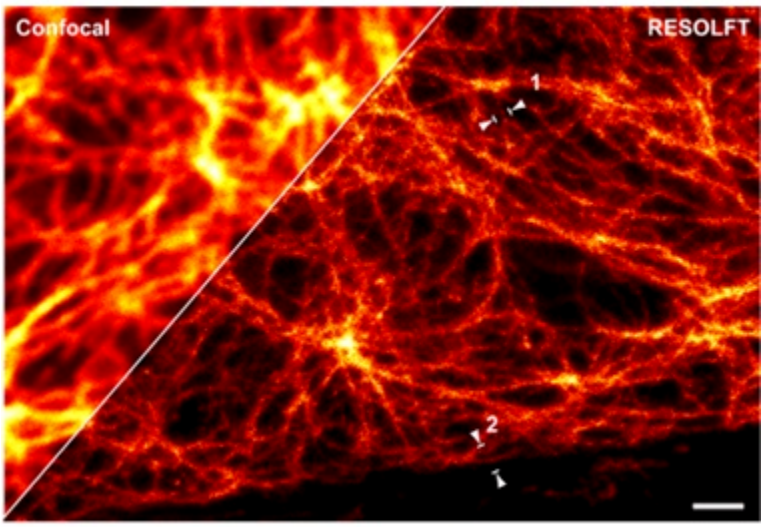


Figure 1.5 rsEGFP2-labelled CV-1 cells were visualised using RESOLFT versus conventional Confocal microscopy, figure reprinted from Ilgen et al. The resolution achieved with RESOLFT ranged between 60-100nm, depending on the cell compartment visualised. Scale bar = 1 μm [110].

emission (via high laser powers), driving the fluorophores to the ground state. RESOLFT, on the other hand, can lower the laser intensity used by making use of the stereochemically distinct dark and bright states of RSFPs.

More specifically for RESOLFT, the addition of a doughnut-shaped laser beam ensures that the majority of the sample area converts to the dark state. The doughnut-shaped beam is the result of a diffraction pattern with high intensity laser power in its outer part (I_{max}) and essentially zero intensity in the middle. The doughnut-shaped beam ultimately, "deplets" the fluorophores around the focal spot for detection – the high intensity laser power increases the probability that the RSFP has switched to the "off"-state. This way, the effective "on"-state RSFPs remaining to be excited (by the excitation laser) can only be in a small spot in the center of the objective's view. Therefore, the minimum distance between features that can be resolved using STED or RESOLFT becomes:

$$d = \frac{\lambda}{2nsin\alpha\sqrt{1 + \frac{I_{max}}{I_s}}} \tag{1.2}$$

Where I_{max} and I_s are the values for the laser intensity at maximum and at the desired intensity on the edges of the doughnut beam, respectively. Ideally, I_{max} is by far higher than I_s, thus making the resolution "infite". In practice, the resolution is still limited to a few tens of nanometers. Overall, by laser-scanning the sample using such unique illumination patterns, one can reconstruct the final image of the sample at nanoscale resolution [99].

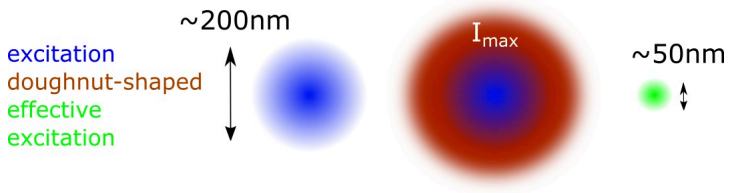


Figure 1.6 Schematic representation of the combination of excitation and doughnut-shaped beam in RESOLFT. While the excitation beam on its own has a diameter of about 200nm due to diffraction, after the doughnut-shaped beam is applied, the effective excitation beam becomes closer to 50nm.

In contrast, single-molecule or stochastic SR techniques reconstruct images after recording the signal received by single molecules. Instead of purposely "activating" or "saturating" fluorescent tags using high intensity lasers, PALM/STORM type techniques use low intensity activation-laser illumination and thus allow for the stochastic photoconversion of molecules. In this way, single-molecule fluorescence, induced by high-intensity excitation lasers, can be temporally disentangled. Consequently, knowing that the fluorescence perceived is the de-excitation of a single fluorophore allows for the projection of the diffracted signal onto a point of the size of the fluorophore used. The latter step, involving mostly image processing, uses a mathematically modelled (Gaussian[173]) prediction (point spread function, PSF[148]) of the ideal emission of a fluorophore in order to sharpen the signal. This way, the position of the emitting fluorophore is accurately determined, and a molecular resolution can be reached.

In single-molecule SR microscopy, the fluorescent tags used are either synthetic tags which are quenched using a buffer and slowly regain their fluorescence, giving temporally distinct signals (dSTORM) or PTFPs that convert between two distinct states over time (PALM)[101].

1.1.5 Application of RSFPs beyond SR microscopy

The use of RSFPs has been demonstrated beyond the scope of SR microscopy, with applications on photodiode [38] and photovoltaic cell [39] technology as well as data-storage devices [6]. Nevertheless, their main applications remain in the realm of microscopy. In this section, there is brief mentioning of the two main types of microscopy (which are not SR) that benefited from the use of RSFPs – Photoacoustic Microscopy (PAM) and Out-of-phase imaging after optical modulation (OPIOM).

Photoacoustic microscopy (PAM), enabling deep-tissue imaging has benefited massively from the reversible switching of RSFPs. As a general concept, PAM, benefiting from the limited scattering of ultrasonic waves inside the tissue, achieves image reconstruction of features at depths beyond those reached by optical microscopy (>1mm)[112]. This

modality can work with native labelling (ie monitoring hemoglobin for example [133]), with highly-absorbing labels (eg chromoproteins [130] with high extinction coefficient at the wavelength of excitation) or using RSFPs [217]. As long as the tag used can be reversibly photoswitched to two distinct absorption spectra, it can be theoretically used to achieve better sensitivity PAM. The unique reversible photoswitching character of RSFPs allows high signal-to-noise ratio since images are acquired of the sample illuminated at two different wavelengths; only the RSFP has two distinct absorption spectra and thus emits two distinct ultrasonic signals. Consequently, overlaying the two images acquired, one expects that the parts of the sample not tagged with RSFPs appear the same and cancel out, while the signal of RSFP-labelled features is still present. Acquisition can take place multiple times until the RSFP is photobleached, since photoswitching is reversible. Overall, PAM using RSFPs can accurately determine quantitative information of the biological sample [217]. Lastly, nanoscopy-inspired post-acquisition image processing can be envisioned when using RSFPs in PAM [236].

Another interesting application of RSFPs outside of the scope of super-resolution microscopy is a technique developed by Jullien's lab (École normale supérieure, Paris) called OPIOM[177] (Out-of-phase imaging after optical modulation) which uses the variety of switching kinetics of RSFPs for multi-color-type imaging. A successor of the Optical Lock-in Detection (OLID) technique, OPIOM can disentangle signals of multiple fluorophores when taking into account discernible properties such as on-off and off-on switching rates. More specifically, the optically modulated signal allows for distinguishing between several green RSFPs which switch at different rates [178]. Wavelength-dependent multi-color microscopy suffers from spectral overlap of fluorescent tags which limits the amount of tags that can be possibly used in one sample. In contrary, OPIOM promises labelling of cellular compartments, using more tags than can be currently used in conventional fluorescence microscopy.

1.1.6 Reversibly Photoswitchable Fluorescent Proteins (RSFPs) used in SR microscopy – advantages and drawbacks

Depending on the application, one can choose from a variety of RSFPs. Primarily, they can be categorised in 3 groups according to how they respond to light illumination: a. positive, b. negative and, c. decoupled photoswitchers [26, 4, 3, 154]. When illuminating a *positive* photoswitcher with the laser required to induce fluorescence, there is also photoisomerisation to the fluorescent species. Positive switchers are characterised by a non-fluorescent thermal ground state, while the thermodynamically metastable state is the fluorescent state. Contrary, in *negative* RSFPs, the chromophore is excited to the electronically excited state before it isomerises to the "off"-state. So accordingly, the thermal ground state is the bright "on"-state while the light-induced metastable state is the dark "off"-state. In the third category, there are so far only two RSFPs available, Dreiklang [27] and SPOON [15], whose isomerisation is induced in an unique manner; reversible photo-induced hydration/rehydration of the CRO is what allows for the existence of the

fluorescent “on”- and the non-fluorescent “off”-states. In Dreiklang, different excitation wavelengths are used for switching the protein “off” (photo-induced hydration at 405nm), “on” (photo-induced rehydration, 365nm) and electronically exciting the protein to acquire fluorescence (511nm).Therefore, Dreiklang is characterised as a *decoupled* photoswitcher. SPOON, used in Single Molecule Localisation Microscopy (SMLM), is initially switched “off” by 488nm illumination and the “on”-state is thermally promoted via dehydration, causing an apparent blinking behaviour.

Among RSFPs, negative RSFPs are primarily used as labels for RESOLFT [60]. This dictates an illumination pattern to be used which includes 4 steps: a. activation, b. switching-off of background proteins (i.e depletion), c. excitation of “on”-state proteins in the focal spot and finally d. detection of fluorescence. In particular, in green RSFPs, these steps are achieved by: a. initially an activation laser (405nm), switching the proteins to the “on”-state, b. a 488nm doughnut-shaped beam which promotes switching “off” of proteins outside of the sample area aimed for detection, followed by c. a 488nm focused laser beam promoting fluorescence emission of the remaining “on”-state proteins and d. detection of the “effective excitation” area. In general, when using negative RSFPs, only two wavelengths are employed. In the exceptional case of Dreiklang, RESOLFT using three different wavelengths (on-, off- and fluorescence-inducing) is possible [41]. The doughnut-shaped beam is running for longer time as compared to the central beam [27, 87, 88]. Therefore, fluorescence emission happening at a faster timescale than off-switching [26], is detected only in a nanosized area in the centre of the beams.

While RESOLFT can achieve high resolution and be used in in vivo studies, decreasing the dwell time (time spent on detecting signal from each pixel) without loss of resolution remains one of its biggest challenges [4]. Consequently, protein engineers have focused their efforts on increasing the “on”- and “off”-switching rates of RSFPs [88, 61] while also maintaining or enhancing their brightness capabilities. Accommodating for both fast switching and high brightness is, however, a great challenge currently [3].

Moreover, all microscopy modes benefit from durable FPs which are less prone to photobleaching, especially RESOLFT – being a laser scanning SR technique, it uses high-intensity focused laser beams as compared to eg PALM. Achieving nanometer resolution involves illuminating each sample area multiple times, as the focal spot moves in smaller increments than typical confocal microscopy. For the sake of high resolution, it is crucial for RSFPs to be able to undergo multiple photocycles before getting bleached – not only due to loss of signal, but also because photobleaching by-products (eg reactive oxygen species, ROS) can be harmful to the biological sample [57, 56].

Additionally, while ideally the absorption spectra corresponding to the on- and off-state of the protein should be decoupled, this is rarely the case. If the RSFP in the “off”-state is allowed to partially absorb the off-laser, this causes a contrast issue during acquisition [61, 4] (Figure 1.8). In practice, this corresponds to a minor population of “off”-state species that can be switched to the “on”-state and thus contribute to the overall background noise inside the region illuminated by the doughnut-shaped beam.

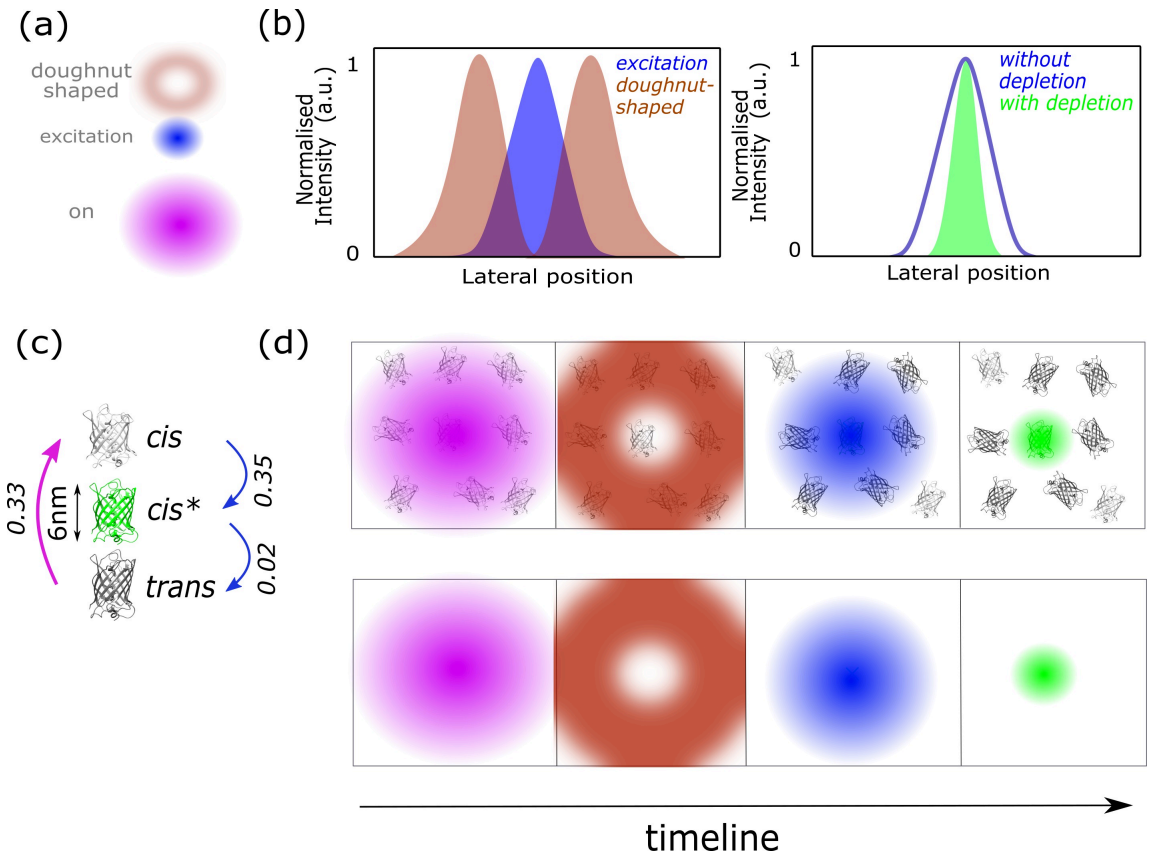


Figure 1.7 Schematic representation of the principle of RESOLFT microscopy using a negative green GFP-type RSFP. (a) Three types of illumination steps are combined, each one characterised by a certain beam shape, intensity and duration. (b) The combination of the doughnut-shaped beam and the excitation beam results in a much narrower effective excitation beam, enabling nanoscale resolution. When considering a negative GFP-type green RSFP, the major possible states are shown in (c), with the corresponding light source inducing them and the quantum yield values of each photophysical process; *cis* and *trans* represent the ground states of the RSFP consisting of a *cis* and *trans*-CRO, respectively, while *cis** is the excited state of the RSFP consisting of the *cis*-CRO after cyal light illumination. A simplified representation of the laser-scanning principle of RESOLFT is shown in (d): on the top panel the protein molecules are depicted, but are not visible in reality, unless when they emit fluorescence, while on the bottom panel the realistic picture of acquisition is drawn. The UV light ensures that all RSFPs are in their fluorescent "on"-state. The doughnut-shaped beam favors "off"-switching of the RSFPs around the focal point of detection, while the focused blue laser beam is of high intensity and short duration to induce fluorescence. Finally, a highly localised fluorescence signal is achieved, leading to nanometer-scale resolution, depending on the doughnut-shaped laser beam intensity profile.

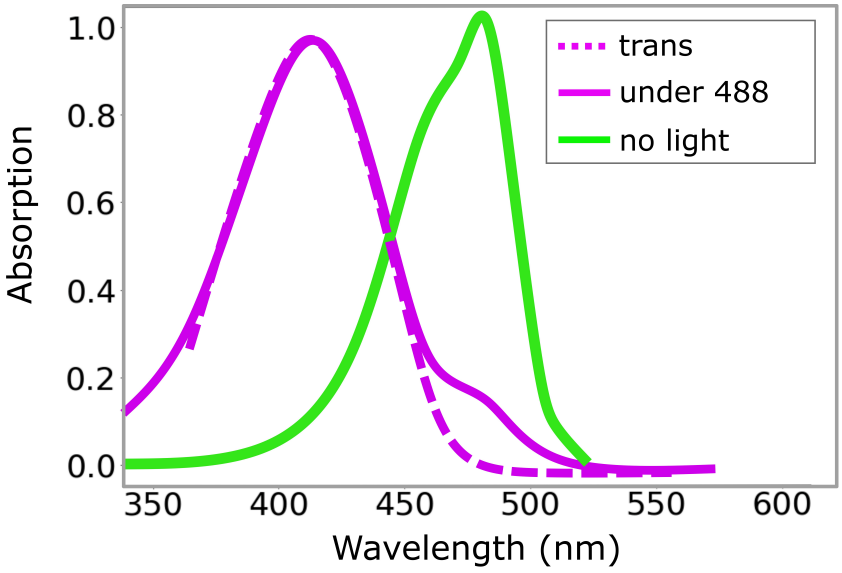


Figure 1.8 Absorption spectra of rsFolder, a negative RSFP, under no illumination (anionic CRO) and at 488nm illumination (protonated CRO), overlayed with an approximation of the absorption of the trans protonated CRO species. At 488nm, the trans-CRO species partly absorbs light and back-switches to the anionic cis-CRO conformation. Overall, this spectral overlap at 488nm leads to a build-up of the anionic cis-CRO conformation as seen from the little "bump" on the absorption spectrum at 488nm, resulting in sub-optimal switching contrast. [61]

Notably, PAM would benefit from the existence of RSFPs with high extinction coefficients (i.e good "absorbers" of light). Engineering RSFPs with high extinction coefficients would translate to an increase in the signal in PAM detection [130, 217]. Moreover, the speed of switching of RSFPs used in PAM should be adjusted to take into account the frequency of the pulsed laser as well as the time required for volumetric imaging [112] specific to each PAM setup.

Lastly, protein engineering efforts should be made to extend the range of switching rates if OPIOM is to be advanced [178]. The discriminatory power of OPIOM and thus the extend of multi-color labeling of samples is dependent on switching kinetics. Currently, the switching rates of RSFPs are grossly categorised in two regimes: slow and fast photoswitchers. A figure of most frequently used green RSFPs reported in literature and their relative switching rates is included below.

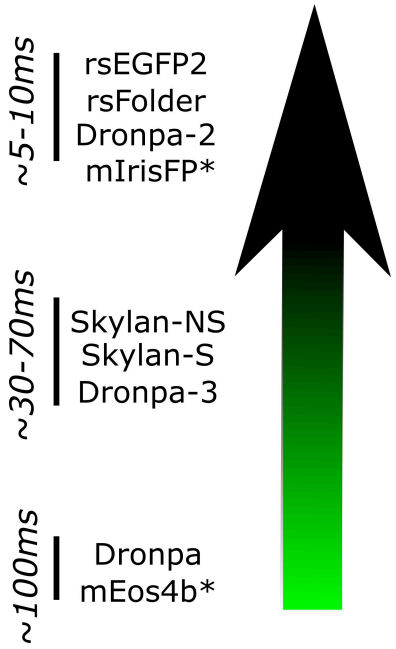


Figure 1.9 Schematic representation of the relative switching rates of a variety of green RSFPs found in literature; separated amongst slow (≈100ms), medium (≈30-70ms) and fast(≈5-10ms) switchers. * indicates proteins that have photochromic in addition to their photoswitching properties. [88, 61, 9, 11, 54, 240, 72]

1.2 Biophysical characterisation of RSFPs

Scientists across different fields of protein biophysics have attempted to characterise RSFPs – in terms of their optical properties, molecular structure and conformational dynamics. This section aims at reviewing the contribution of different biophysical techniques in our understanding of RSFPs.

To aid the reader, I have compiled a table (1.1) of the RSFPs mentioned in the following sections, summarizing their optical properties.

PROTEIN	EM λ (nm)	ORGANISM	SPECIAL FEATURES
rsFolder	503	Aequorea victoria	High stability and expression
rsEGFP2	503	Aequorea victoria	Fast switching
Dronpa	518	Echinophyllia sp. SC22	High brightness
IrisFP (Green state)	516	Lobophyllia hemprichii	Photochromic RSFP

Table 1.1 Examples of RSFPs mentioned in this chapter, alongside some optical properties. EM λ is the emission wavelength.

1.2.1 Photophysical/Optical characterisation of RSFPs

Studies on RSFPs properties and mechanisms of photoswitching

The primary technique used to study FPs is optical characterisation – microscopy as well as spectroscopy. In this way, a direct assessment of the photophysical properties of FPs can be made that is directly correlated to their application. Moreover, the ability to choose between running an ensemble or a single-molecule photophysical characterisation experiment, allows for critically assessing the behaviour of an RSFP for a given application. Bibliography can account for numerous optical studies on RSFPs [88, 87, 89, 9, 45]. However, faced with the complicated photophysics of RSFPs, optical studies come across certain limitations. Here, I will attempt to review some exemplary photophysical/optical studies performed on green negative RSFPs in an effort to highlight the advantages as well as some of the underlying drawbacks of such techniques.

Dronpa was the first green negative RSFP engineered by Miyawaki's team in 2004 [11]. At the ensemble level, when expressed in HeLa cells, they could switch the protein from a fluorescent to a non-fluorescent state in a reversible manner using 400nm and 490nm laser light, respectively. Spectroscopic studies revealed two absorption maxima (400nm, 503nm) that interchange in intensity at pH=7.4 upon illumination using either laser sources, corresponding to a reversible switching between a neutral (400nm) and an anionic species (503nm). The anionic species was identified as the thermally stable species with fluorescence emission in the green region (518nm). Apparently, at low pH, the anionic species could no longer be retrieved using UV light but was recovered only at higher pH values, in a pH-dependent fashion. Dronpa's photophysical properties are summarised as a slow switcher (around 2 mins needed for a complete photocycle in ensemble measurements, 400nm 0.14W/cm^2, 490nm 0.4W/cm^2) with relatively high brightness (80.75$M^{-1}cm^{-1}$ [89] as compared to 33.54$M^{-1}cm^{-1}$ for EGFP [167]). Additionally, the switching rate is correlated to the power of laser illumination used. At the single molecule level, the switching is on the order of μs [235] (as measured by ultrafast spectroscopy, mentioned later on in this chapter). Many derivatives of Dronpa have been engineered since, aiming at speeding up its switching rate as well as overcoming its high photofatigue, which is the characteristic loss of fluorescence upon a number of repeated photo-excitations [9].

Evidently, photophysical studies preceded any other biophysical characterisation of Dronpa. Nonetheless, these early studies could not provide evidence about the structure and oligomerisation state of Dronpa. Using optical spectroscopy, the authors could only retrieve information on the protonation state of the chromophore and thus distinguish between neutral and anionic species. A similar depth of information is reported in the first publications on rsEGFP [87] and rsEGFP2 [88]. Clearly, photophysical characterisation is primarily performed when discovering a new RSFP. Still, it rarely goes beyond the needs of characterisation and into answering questions related to function. Thus, the mechanism of photoswitching still remained unknown until the first crystallographic studies of the "on"- and the "off"-states of Dronpa were published [230, 12].

A thorough study performed by Gayda et al on mIrisGFP, a green negative RSFP[77], aimed at going further into understanding the mechanisms of photoswitching by employing a variety of optical characterisation protocols at different pH values and comparing with published crystallographic data. The study focused on the green state of mIrisGFP – comparing its "on"- and "off"-states for a series of pH values. Using mathematical models accounting for four CRO species (*cis* neutral, *cis* anionic, *trans* neutral, *trans* anionic), they were able to retrieve information on protonation states and pKa values of the *cis* and *trans* chromophore while also suggesting a mechanism of photoisomerisation being coupled with the protonation/deprotonation of the chromophore. Shifts in absorption spectra were interpreted as changes in the pi-conjugation of the chromophore, influenced by protonation of nearby residues.

Generally, Gayda et al [77] demonstrated the multiple ways that results of optical studies can be examined and interpreted. Nevertheless, a strong assumption for the isomerisation state of the CRO is based on crystallographic structures of IrisFP at pH=8.4 [5]. According to these structures, the CRO assumes a *cis* conformation in the dark-adapted "on"-state, while a *trans* conformation is expected in the light-induced "off"-state. As already explained above, absorption spectra cannot accurately distinguish between isomers but only between protonation states of the CRO.

Information coming from circular dichroism results were included to support the existence of certain protonation states. Evidently, these results cannot be deconvoluted between protonation or isomerisation states, once the sample is a mixture of species, and the technique itself reports on multiple changes occurring simultaneously in the protein [82]. In conclusion, the study still is ambiguous on how it can discriminate between the different CRO species, since there is no structural evidence to support their findings at different pH values.

Disparities between optical studies of RSFPs – choice of fixation, illumination and acquisition schemes

With the discovery of RSFPs came the development of unique photophysical experiments that aim at characterising their switching kinetics [58, 11, 45, 27]. These are experiments which measure the fluorescence modulation upon illumination of the two different laser sources needed to promote the protein to its two available states. The CCD camera-recorded fluorescence is plotted over time, its shape being that of characteristic fin-shape photocycles (Figure 1.10). Each photocycle can be fitted to a kinetic model of an exponential decay (on-off) and increase (off-on) rates of fluorescence, while the overarching decay of fluorescence across many photocycles corresponds to photobleaching. Ultimately, once the conditions are optimised, ensemble measurements of this kind can be very instructive in characterising a given RSFP's switching behaviour. However, each group studying optical properties of RSFPs seems to choose different illumination powers, sample fixation techniques as well as general sample conditions (temperature, pH). Overall, there is no standardized protocol for studying the photoswitching properties of RSFPs, causing discrepancies between similar

studies.

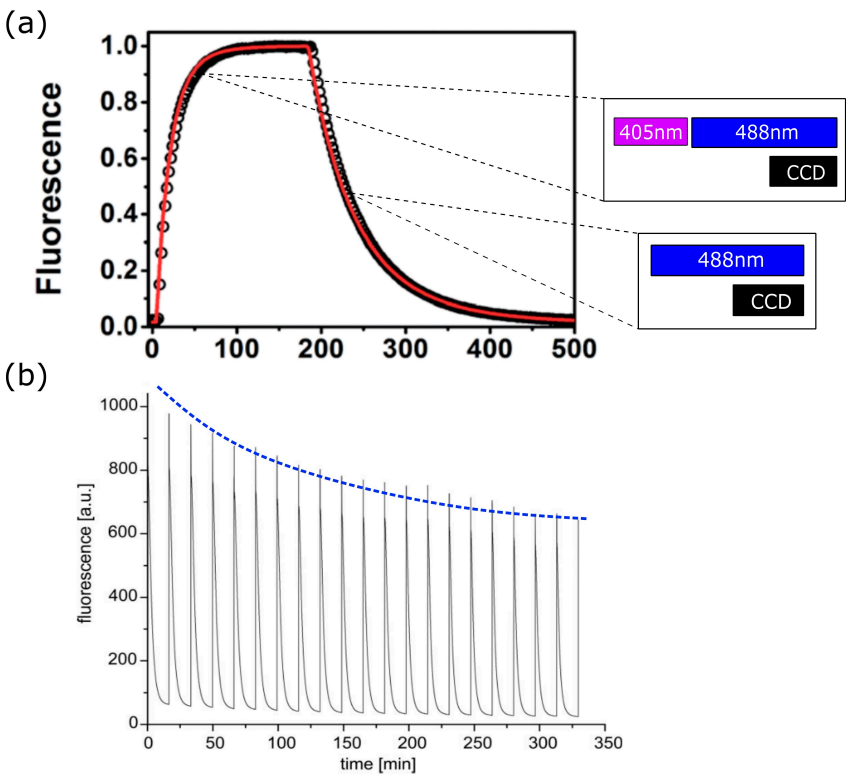


Figure 1.10 (a) Example of fluorescence fin-shape photocycles of IrisFP adapted from Duan et al [58]. Each point in black on the graph is the recorded fluorescence by the CCD camera upon 488nm excitation. At first, a 405nm pulse switches the proteins "on" and a following 488nm pulse induces fluorescence. The rate of fluorescence decay is measured by using only a 488nm pulse which induces excitation and, thus, fluorescence measurement and, gradually, switching "off" of the protein. Finally, the fluorescence built-up (0-200seconds) and decay (200-500seconds) are fitted (red solid line) against a kinetic model to retrieve information on switching rates and switching contrast (fitted against 3-5 cycles). In (b), an example of multiple switching cycles in Dronpa adapted from Andresen et al [12] used to fit (blue dashed line) photobleaching and thermal relaxation rates.

In the case of Dronpa [11], switching kinetics experiments were performed on fixed HeLa cells. In contrast, rsEGFP [87] and rsEGFP2 [88] were fixed in a Polyacrylamide (PAA) gel and promoted by light to undergo multiple photocycles. Notably, depending on the pH and the oxygen available to a given cell compartment or fixating medium, the photophysics can be different [57, 61] (Figure 1.11). Therefore, the conclusions drawn from the studies of Dronpa and rsEGFP/rsEGFP2 cannot be generalised for the behavior of the proteins in different environments. Altogether, optical characterisation needs to be performed in several conditions, including physiologically relevant conditions, like in a cell, in order to

retrieve reliable data concerning their optical properties. Lastly, diffusion of proteins in various environments can significantly alter the observed photophysical properties, making PAA- or PVA-fixed samples more reliable than samples in cuvettes.

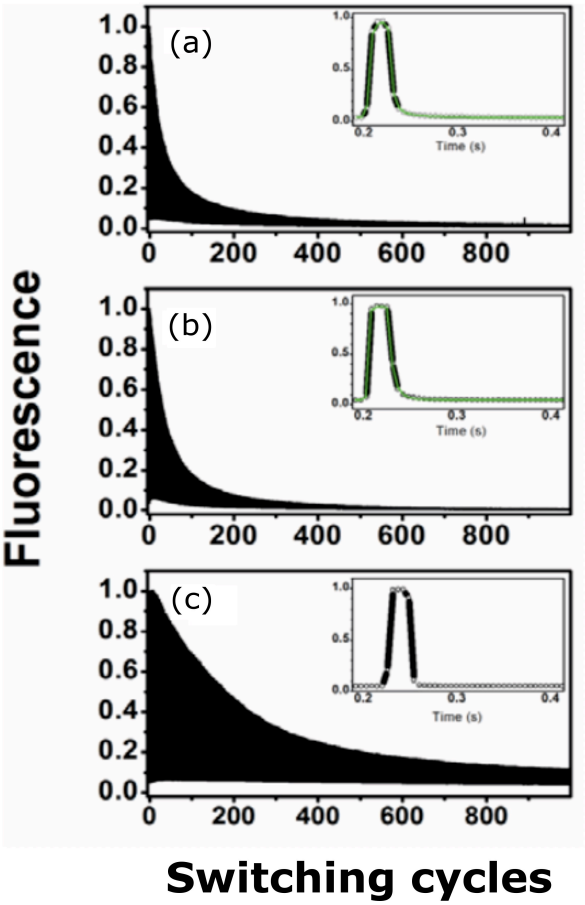


Figure 1.11 (a) Example of fixation protocol affecting the photophysical properties of IrisFP adapted by Duan et al [58]. Comparing PolyVinyl Acetate (PVA) gel (A), fixed(B) and live(C) bacteria, the photobleaching rate is significantly different. Moreover, in insets, individual photocycles are fitted, where the authors calculated about two-fold increase on average for the switching brightness when going from PVA gel to bacteria.

Depending on the property that one wants to measure, optical studies use different illumination schemes. Absorption and emission profiles are essential to characterise an RSFP, but due to differences in the way they are acquired and their sensitivity to certain processes, they can not necessarily be considered complementary to each other.

Measuring light absorption of a protein follows a different illumination scheme (mercury/argon lamp) than what is used for measuring fluorescence (LASER). On the one hand, absorption is a definite measure of the existence or not of certain species (protonation states in the case of the CRO) of a protein, as long as the extinction coefficient of the species is

known. On the other hand, during fluorescence experiments due to the light dependency of negative RSFPs, the same laser light used to detect fluorescence necessarily switches many molecules to the "off"-state before acquisition starts. That can be a severe experimental disadvantage when studying fast switching RSFPs, since the actual fluorescence emitted can be underestimated.

Therefore, while absorption measurements tend to be more reliable given that they are less invasive (do not induce excitation of the CRO), they do not report on the fluorescence properties of an RSFP. To retrieve fluorescence information, LASER induced isomerisation and, thus, change in population of species is required. Therefore, it is often difficult, especially for fast switching RSFPs, to corroborate absorption measurements with fluorescence.

Ensemble vs single-molecule characterisation of RSFPs

When investigating the optical properties of RSFPs at the ensemble level, low laser powers are employed (few mW/cm^2) and the light is distributed among numerous molecules [11] . Notably, the distribution is not homogeneous and thus has to be accounted for. FPs fixed in cells to be studied by microscopy can be assumed to experience no diffusion, while cuvette samples used in spectroscopy are more prone to diffusion-induced phenomena. Moreover, secondary effects caused by protein-protein interactions which could influence their photophysical behaviour (for example oligomerisation) need to be considered. Lastly, at the ensemble level the observed photophysical behaviour is a statistical average of all possible mechanisms that take place at the molecular level.

At the single molecule level, acquisition can be more quantitative. Here, high laser powers need to be used (a few kW/cm^2), since the signal arriving from single molecules is extremely low. Contrary to ensemble level measurements, single molecule studies are instructive in the prediction of "anomalies" i.e when a given RSFP is in another state than the conventional "on"- and "off"-states, such as blinking (Figure 1.12). Such behaviours, despite their infrequent nature, are crucial to predict and possibly control, to avoid discrepancies in single-molecule SR microscopy. Nevertheless, they represent a statistical probability, with each molecule likely to behave differently. Constructing statistical models of RSFP function using single-molecule studies can be trustworthy as long as the number of observations is high enough to minimize the statistical error associated with the measurement.

Taken together, photophysical characterisation of RSFPs is a necessary tool in understanding how to better apply them in microscopy. By adjusting the setup and the experimental parameters, one can go far in predicting the behaviour of such proteins in relevant contexts. Nevertheless, data need to be handled with caution, while interpretation can be misleading unless accompanied by structural studies. Historically, early enough in the field of RSFP engineering [230, 84], X-ray crystallography (alongside optical studies) became the preferred biophysical characterisation technique for many scientists aiming at clarifying the underlying mechanisms for the proteins' photophysical behaviour.

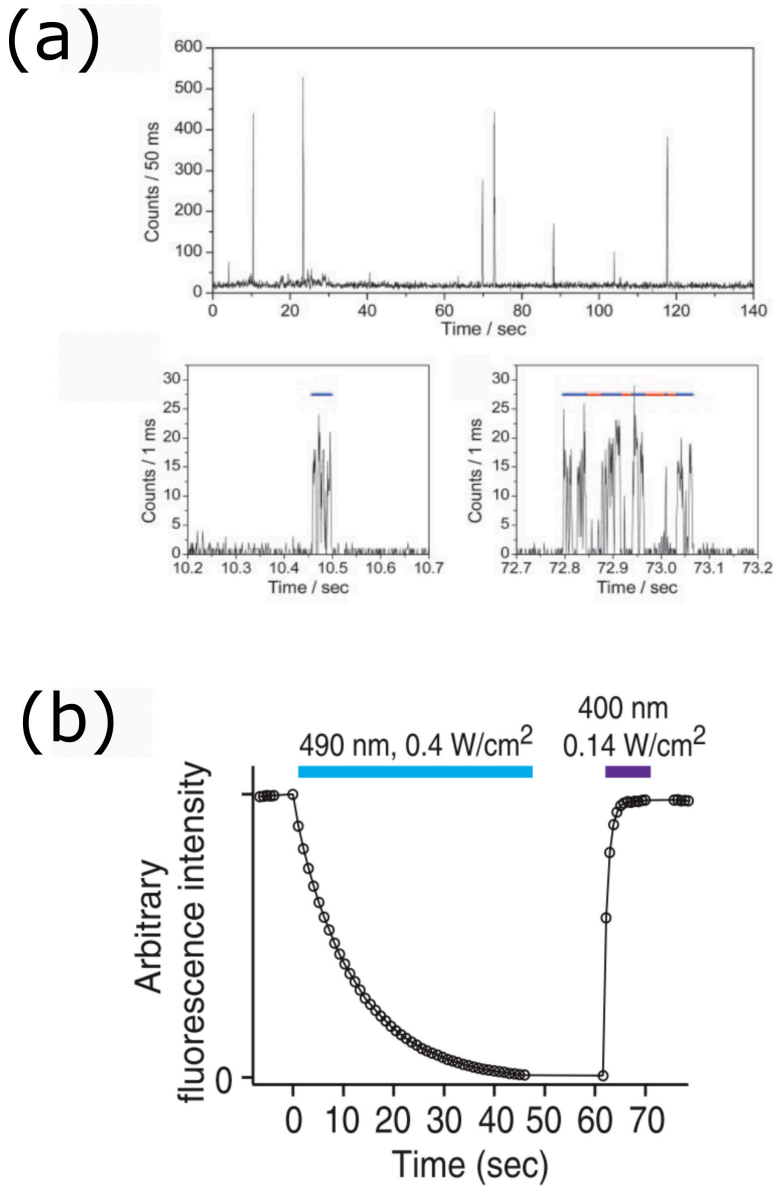


Figure 1.12 Comparison between (a) single molecule and (b) ensemble studies of Dronpa. While the first is more statistically influenced and requires higher laser power (range of kW/cm^2), it reveals instances of fluorescence emission as well as blinking and long off-times – see in blue fluorescence time and red off-time in the 2 bottom graphs (figures adapted by Habuchi et al [89]). Ensemble measurements at a range of hundreds of mW/cm^2 reveal information of switching kinetics, contrast and photobleaching of the protein. Figures adapted from Ando et al [11].

1.2.2 Crystallographic studies of RSFPs

Crystallographic structures of RSFPs have paved the way for deciphering the mechanisms underlying their photophysical properties – switching mechanism [12, 61, 53], photobleaching [57] and partly switching contrast [61, 53]. These include not only standard crystallography of ground state structures, but also kinetic crystallography of transient states of the protein, during or after light illumination. The aim of these studies is to identify the interplay between the chromophore and nearby residues which are responsible for the optical characteristics of RSFPs. Ultimately, information retrieved from X-ray crystallography can inspire mutations for novel RSFPs. Nevertheless, in this section, using the example of key studies, I will attempt to pinpoint not only the historical significance of X-ray crystallography in RSFP engineering, but also unresolved issues which can negatively affect the conclusions drawn about RSFP function.

Structures of the two ground states of RSFPs – "on"- and "off"-states

Starting with Dronpa, as mentioned previously, the mechanism underlying its photoswitching behaviour would only be unlocked after the first crystal structures became available [230, 12]. Initially [230], a simple model of protonation/deprotonation of the chromophore was suggested which is followed by a re-arrangement of residues around the chromophore pocket. Once the light-induced "off"-state structure was resolved by Andresen et al (2007) [12], *cis-trans* isomerisation was revealed for the first time, which explained the thermodynamic stability of the observed light-induced change in protonation of the chromophore as well as the conformational changes of nearby residues. Similarly to the previously reported structures of the positive red RSFP asFP595 [179, 229], Dronpa can occupy two ground states – the dark-adapted "on"-state consisting of a *cis* anionic chromophore, and the light-induced metastable "off"-state including a *trans* neutral chromophore. The authors describe a cascade of chemical events underlying photoswitching. Accordingly, the absorption of a blue photon, forcing the chromophore to isomerise to the *trans* conformation, is coupled to the breakage of H-bonds between the CRO and residue Ser142 as well as the disruption of the previously established pi-stacking with the ring of His193. This is accompanied by conformational changes of some other residues which would sterically hinder the *cis-trans* trajectory. This change in environment favours a certain protonation state – anionic for the *cis*, and neutral for the *trans* CRO. Interestingly, the authors provided hints on residues standing on the way of the CRO when it switches (Val157, Met159 in the case of Dronpa), which could be involved in what would later be identified as intermediate long-lived excited states [47] in the case of rsEGFP2. Their suggestion was that such residues raise significantly the energy barrier needed to be crossed between "on"- and "off"-states.

Notably, Andresen's study investigated the hydrogen-bonding network and CRO planarity to understand the difference between the "on" fluorescent state and the "off" dark state of Dronpa. H-bond networks have been heavily investigated in conventional GFPs and their role in the CRO's ability to fluoresce has been identified [199, 52, 8]. More specifically,

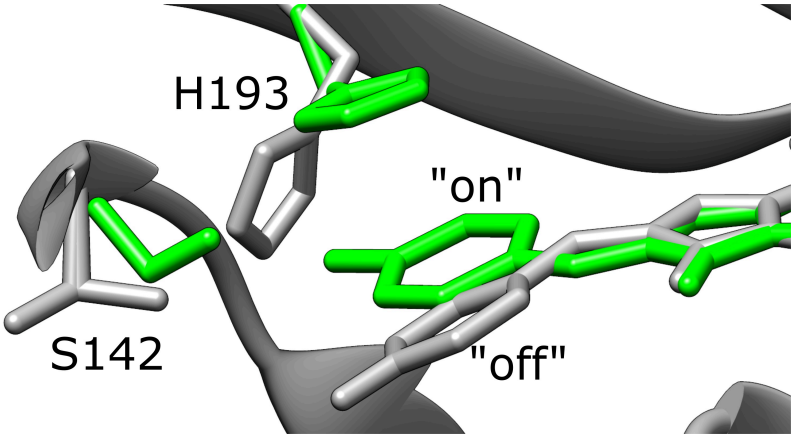


Figure 1.13 Zooming in the overlapped crystallographic structures of Dronpa “on”- (green) and “off”- (gray) states, in the vicinity of the phenol ring of the CRO. His193 no longer forms a pi-stacking with the phenol ring in the “off”-state and Ser142 has been dislocated after the H-bond breakage with the CRO. PDB structures: "on"; 2IE2, "off"; 2POX

the better “anchoring” that the CRO has, the better yield of fluorescence it will achieve. Conformational rigidity of the chromophore ensures that the de-excitation mechanism after the absorption of a blue photon, is fluorescence emission rather than radiationless decay. Moreover, thermodynamically, H-bonds are associated with decrease in enthalpy – a conformation whereby the CRO consists of more H-bonds with the protein matrix can be more thermodynamically stable due to decreased enthalpy. Breakage and/or reconfiguration of H-bonds between the “on”- and the “off”-states are commonly observed in crystal structures of other RSFPs [5, 13, 61, 53], giving rise to thermodynamically stable states of increased entropy. Across the examples found in literature for negative RSFPs, the “off”-state is characterised by a loosening of the H-bond network around the CRO. Therefore, X-ray structures not only provide molecular insight on the photo-isomerisation process [12, 61, 5], but also attempt to link structure with the energetic landscape of a given RSFP [181, 47, 53].

Historically, there is a long standing debate concerning how accurate conclusions from crystallographic structures are in predicting the physiologically expected behavior of a protein [202, 142, 126]. Major pitfalls of crystallography lie in the nature of crystal preparation and the data analysis. Furthermore, by definition, obtaining a picture of a protein in a physiologically more appropriate environment (i.e with the use of buffers mimicking the cell environment) is sometimes contradictory to the generation of an acceptable diffraction pattern. In the next paragraphs, I explain the origins of these problems encountered by crystallographers.

RSFPs prepared in crystals, similarly to other proteins, are forced into a conformation which might not be representative of their native state [117]. More specifically, adopting cryogenic temperatures to avoid X-ray radiation damage and crystal packing necessarily

interfere with the energetic landscape available to the molecule [92, 65, 116]. Thermodynamically the effect is clear – lowering the temperature stabilises the protein which might otherwise act more dynamically. Therefore, when dynamics are not present, heterogeneity of conformations is hindered [117] as well as interatomic bonds and proximity overestimated [202]. Furthermore, there is evidence that flash-cooling [92] as well as crystal packing [65] can distort the structure of the protein in a crystal. Interestingly, a recent study on rsEGFP2 suggests an influence of crystal packing on the preferred photoswitching mechanism of the chromophore [34].

In relation to data interpretation, models of structures are refined based on the acquired electron densities. While they are essential in providing atomic resolution information, refinement methods differ and can lead to different results [113]. Comparing the output of each refinement, does not concern only resolution but overall conclusions on residue orientation, inter-atomic distances and therefore presence of H-bonds (Figure 1.14). More specifically, the challenge of identifying hydrogen atoms combined with the limited discriminating power between atom types translates to a long-known difficulty to define the orientation of Histidines, Asparagines and Glutamines [106]. Overall, RSFP engineers should be aware of such ambiguities especially when it comes to key residues such as Histidines (His148 in rsEGFP2 [61], His193 in Dronpa [230]) and Glutamines(Glu183 in rsGreen0.7 [59]) that are reported to be H-bonded to the CRO.

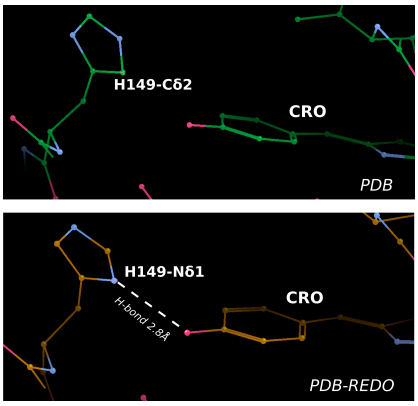


Figure 1.14 Zooming in the chromophore region of the crystallographic structure of the "on"-state rsFolder (5DTZ). Top picture shows the output from the PDB refinement while the bottom is the result of PDB-REDO [113]. H149 rotameric state is equally probable, thus it is not clear whether the H-bond with the phenolate of the CRO is possible or not.

Furthermore, the interaction of X-rays with the conjugated electron system of the CRO requires consideration. As suggested by a study of Adam et al (2009) [7], upon X-ray irradiation IrisFP crystals would induce photobleaching and distortion of the CRO conformation (Figure 1.15). That was justified as a consequence of the electron-transfer susceptibility of the methine bridge of the CRO. The authors concluded that existing crystal structures of other FPs might also be distorted due to such electron interference.

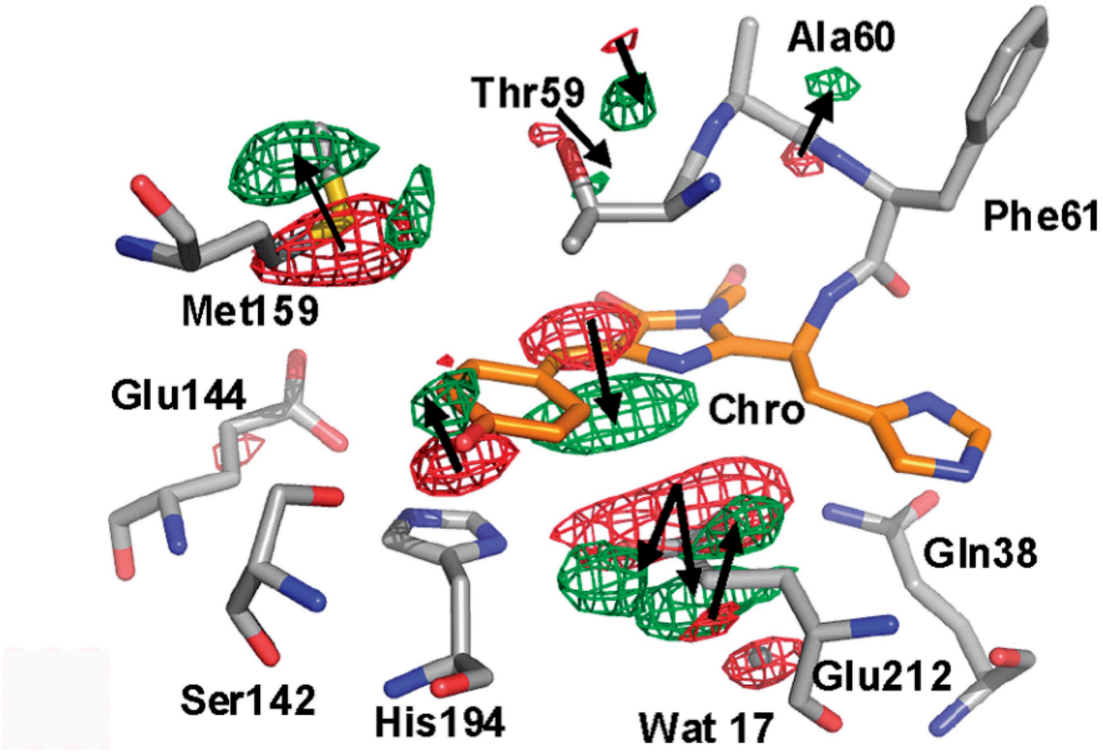


Figure 1.15 Zooming in the chromophore region of the crystallographic structure of IrisFP. Green and Red electron densities represent the difference in the electron maps of low and high power X-ray irradiation. Image adapted from Adam et al [7]

.

Lastly, having to avoid so-called "mosaic" patterns in order to obtain highly resolved data, forces crystallographers to subvert from physiological conditions for the sample preparation [103]. For this purpose, only specific solutions (in terms of salt concentration, pH etc) can be used in crystallography. Nevertheless, the solution surrounding FPs is crucial, since, among other factors, proton transfers from the bulk solvent are reported to be the main mechanism allowing protonation/deprotonation of the CRO in most GFP-variants [153, 158]. Consequently, RSFPs in different media or cell compartments are expected to have different photophysical behaviours. Taken together, environmental factors affecting photoswitching and other photophysical properties of RSFPs cannot be extensively studied using crystallography, due to the limited solutions that can be used in crystal preparation.

Adding a dynamic view using kinetic crystallography

Evidently, the establishment of the switching mechanism as a spatio-temporal process (ie within a given time, CRO and surrounding environment undergo structural changes) unlocks numerous questions regarding dynamics. A more sophisticated attempt to outline

the energetic landscape of RSFPs and to uncover other thermally metastable states that exist at longer or shorter timescales, is kinetic crystallography. Kinetic crystallography aims at capturing intermediate conformations of the CRO and its respective influence to the protein matrix. Within the framework of kinetic crystallography, there are two main categories: (a) Real-time serial crystallography, dominated primarily by X-ray free electron laser (XFEL) studies for the case of FPs [25], and (b) Temperature-trapping based crystallography. Both methodological pathways have informed the field of RSFPs about short-lived excited states as well as long-lived transient states. This area of research aims at explaining photophysical properties such as switching kinetics, lowly-occupied dark states and photobleaching. Despite its growing presence, kinetic crystallography still suffers from unresolved problems which have to be accounted for in RSFP engineering.

Starting with real-time serial kinetic crystallography, XFEL has seen an impressive rise in its use in resolving protein structures over the last few years (185 structures accounted for in PDB since 2014 – April 2020 search). The advantage of such a technique over conventional crystallography is the ultra-bright femtosecond-pulsed free electron lasers that can be used to take snapshots of proteins in motion. More importantly, the short duration of these laser pulses ensures detection before radiation damage becomes prevalent in the sample. Consequently, room temperature crystals can be used. XFEL studies on RSFPs are reported for a limited number of RSFPs so far [46, 47, 107] , but its contribution is expected to rise in the future [25].

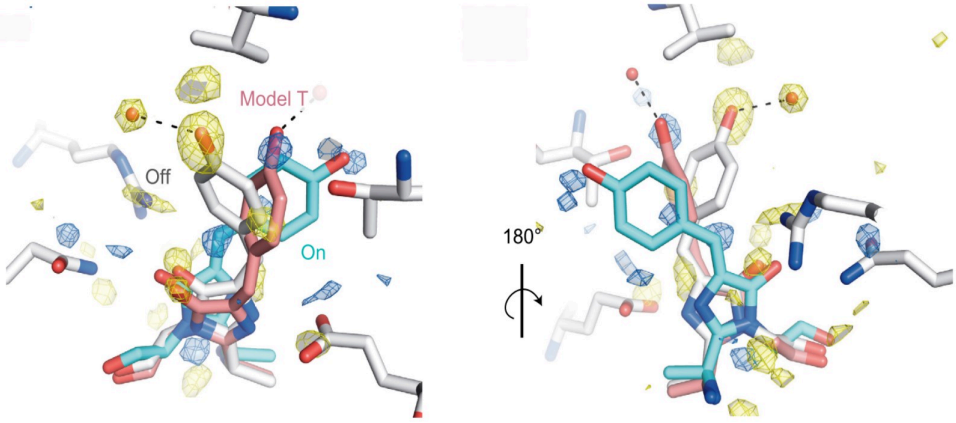


Figure 1.16 Refined models of rsEGFP2 chromophore “on”- and “off”-states (turquoise and white, respectively). In pink, model of intermediate twisted chromophore, 3ps post-excitation, where the phenol ring is perpendicular to the imidazolinone ring of the chromophore. Two perspectives 180° apart, (a) and (b). Image adapted from Coquelle et al [47].

XFEL studies recently captured the CRO occupying a “twisted” conformation within the matrix of rsEGFP2 1ps-post laser illumination [47]. In combination with other biophysical methods, the authors ultimately constructed the step-by-step photoswitching pathway of the CRO in rsEGFP2. The authors addressed concerns related to the necessarily high laser

power and its effect on the final output. Nevertheless, being a crystallographic study, it still faces the issue of limited or distorted dynamics and state heterogeneity due to crystal packing.

Intermediate conformations can also be structurally resolved by trapping based crystallography [25]. This is a technique whereby the protein is temperature-trapped, i.e with flash cooling while or after it transforms from one state to another. Long-lived transient states can be occupied within the crystal at RT and then flash cooled while short-lived states are trapped while at cryogenic temperatures. While this technique has been essential in identifying metastable "off"-states (Dronpa [12], rsFolder and rsEGFP2 [61]), there is only one example whereby intermediate states are captured. Namely, two intermediate states present during photoswitching were discovered in Padron [181], a green positive RSFP, as it converted from its ground "off"- to the metastable "on"-state. Nevertheless, the authors recognised that dynamics might be decreased in the protein matrix at such low temperatures. Thus, while in theory temperature-based trapping could result in interesting information, in practice, crystal packing and the need for flash-cooling is undermining the native dynamics of the protein.

1.2.3 Molecular dynamics simulations and Ultrafast vibrational spectroscopy as complementary techniques in the discovery of new RSFPs

Crystallographic studies have been progressively supported by other biophysical techniques. Apart from optical spectroscopy (sometimes in crystallo, icOS [219]), Molecular dynamics and ultrafast vibrational spectroscopy have enabled a better understanding and interpretation of the static crystal structures. Their contribution to the field of RSFP design is summarised in this section, followed by notes of potential risks associated with interpreting results from these techniques.

Molecular dynamics simulations based on classical or/and quantum mechanical models have attempted to address a variety of questions concerning FPs in general, and RSFPs in particular. Some emblematic studies have paved the way to conceptualising the *cis-trans* isomerisation – what causes it, what are its consequences and how is the protein matrix affecting it. Alongside these answers, MDs have enabled scientists to further map out the energetic landscape of RSFPs and understand their dynamics. However, all such studies are based on crystal structures which as mentioned above, need to be critically used. Additionally, MDs can only represent a model whose role is to explain experimentally acquired data. The risk of MD simulations not supported by experiments is a distorted representation of the native behaviour of proteins. Nonetheless, their dynamic insights made MDs invaluable for understanding photoswitching.

Morozov et al(2016) [150] showed the importance of HB network in photoswitching in a QM/MM study of a Dronpa mutant. Accordingly, photoisomerisation causes the breaking

of H-bonds of the chromophore with the protein matrix. Furthermore, energetic fluctuation of H-bonds is what allows for a relatively low energy barrier for isomerisation, a feature that is distorted in crystal structures due to under-appreciation of dynamics. In contrast, planarity of the chromophore, ensuring fluorescence, is associated to a higher number of available H-bonds "fixing" the chromophore in place. In conclusion, the thermodynamically stable "on"-state is characterised by less dynamics as compared to the metastable "off"-state. Supporting this conclusion, another MD study highlighted increased flexibility of the β barrel upon light illumination [204], while most recently, the work of DeZitter et al [53], showed presence of CRO dynamics in the "off"-state of the green-state in mEos4b, a photochromic protein, associated to the weakening of the H-bonding network of the *trans* CRO with the barrel.

Another complementary technique to crystallographic studies is ultrafast absorption spectroscopy. Ultrafast absorption spectroscopy is a technique whereby kinetics of excited states of photomolecules can be accessed and measured [86]. For negative RSFPs, the intermediate excited states can be distinguished as the protein relaxes from the excited photo-induced "off"-state to the dark adapted "on"-state. Knowledge of such intermediates, supported by the corresponding crystal structures, can be useful for optimising photoswitching quantum yields.

Indicative of the significance of ultrafast absorption spectroscopy is the contribution to serial crystallography studies on rsEGFP2 [47] and IrisFP [46]. In both cases, such spectroscopic techniques allowed deciphering a cascade of events – placing each corresponding crystal structure to well separated time steps with corresponding lifetimes. For rsEGFP2 [47], femtosecond transient absorption spectroscopy revealed two excited state species occurring a few ps after excitation with 405nm light which corresponded to *trans-cis* isomerisation events. Comparing the effect of H_2O and D_2O in the kinetics of these captured states, they suggested that deprotonation of the *cis* chromophore is a ground state process that occurs at longer timescales.

Interestingly, a long-standing debate of the cascade mechanism underlying photoswitching can be addressed by temporally dissociating intermediate states according to their conformational and chemical differences. The advantage of ultrafast optical spectroscopy is its sensitivity to the breaking and forming of chemical bonds, its ability to operate at a multitude of timescales and photo-excitation schemes, its capacity for quantifying ultrafast processes as well as its capability to identify proton transfer steps [86]. Taken together, this technique has highlighted very interesting aspects related to excited state dynamics of RSFPs. Nevertheless, depending on the setup used and the interpretation of data, spectroscopists have reached different conclusions on whether *cis-trans* isomerisation is following or followed by a proton transfer step [71, 136, 235, 47].

Altogether, these two complementary techniques have been instrumental in adding a dynamic dimension to static crystallographic structures. When combined, they have assembled a relatively coherent picture for the mechanism underlying photoswitching. When a given RSFP absorbs light to the metastable "off"-state, the chromophore loses many – if not all – of its H-bonds to the surrounding residues of the barrel [150, 61, 47, 53]. When relaxing back

to the “on”-state, it first isomerises from a neutral *trans* to a neutral *cis* conformation and at a later stage, loses a proton to the solvent. Therefore, the main hypothesis for the difference in protonation for each isomer is that the chromophore is neighbouring different residues in either conformation which make up a different electronegative environment (pKa), thus forcing the chromophore to change its protonation state. Furthermore, intermediate excited states can raise the energy barrier for photoisomerisation, resulting in lower photoswitching rates.

1.2.4 Our model system of negative RSFP: rsFolder and mutants

Calling for a clear dynamic picture characterising RSFPs comes up frequently as a suggestion in recent studies [150, 235, 47]. For this reason, my work consisted of introducing multi-dimensional (nD) solution NMR spectroscopy as a new player in the effort to expand our knowledge on RSFPs' dynamics. For this reason, I chose to study rsFolder [61], and some of its mutants, a green negative RSFP which despite its many promising properties, still suffers from unwanted drawbacks.

First reported in 2016, rsFolder is the marriage of Superfolder GFP and rsEGFP2. Benefiting from its predecessors, it is a fast photoswitcher which can be expressed in unfavourable environments. Therefore, rsFolder can be considered as a variant of rsEGFP2 optimised for a wider range of in vivo applications. Its main advantages over other RSFPs are: relatively fast photoswitching, high photostability preventing photobleaching, a relatively stable "off"-state, and impressive expression and folding durability in 'harsh' environments. Its superiority over rsEGFP2 was demonstrated in the oxidative environment of the bacterial periplasm, where rsEGFP2 was essentially non-fluorescent, due to misfolding.

Many photophysical properties of rsFolder and rsEGFP2 are essentially the same – absorption, emission maxima – and others are similar, such as switching rates. Nevertheless, two main characteristics of their photophysical behaviour distinguish them – rsFolder benefits from an extremely stable "off"-state but it is impaired by a low switching contrast. Furthermore, the value for the pKa of chromophore protonation is different, rsFolder's pKa is 5.5 while rsEGFP2 is higher, at 5.9. Brightness is lower in rsFolder since it has a lower extinction coefficient and quantum yield. Lastly, rsFolder can endure several photocycles demonstrating its low photofatigue, but rsEGFP2 still surpasses it.

Crystallographic structures comparing the two proteins were instructive in understanding the reasons for some of those differences. Slightly faster photoswitching for rsFolder ("on"-"off" switching quantum yield = 0.21%) can be explained by the smaller amplitude rearrangement required for *cis-trans* isomerisation as compared to rsEGFP2 ("on"-"off" switching quantum yield = 0.17%) (Figure 1.17). Moreover, CRO stability in the "on"-state is ensured by 3 H-bonds with the protein barrel; His149, Thr204 and a water molecule. Interestingly, the His149 bond is presumably preserved in the metastable "off"-state which could be the reason for its high thermal stability. Nevertheless, crystallographic investigation could not directly account for the low switching contrast.

Nonetheless, the authors attempted an explanation based on the crystallographic data, in an effort to engineer rsFolder into having high contrast. According to them, the difference between a more hydrophobic phenylalanine in rsFolder and a tyrosine in rsEGFP2 at position 146 can partly be the reason for the limited amplitude rearrangement of the CRO in rsFolder (as compared to rsEGFP2) (Figure 1.17). The tyrosine 146 in rsEGFP2 forms an H-bond with His149 which is 'freed up' upon *cis*-to-*trans* isomerisation ("on"-to-"off" switching) leaving no H-bond with the *trans* CRO. In rsFolder, where a tyrosine is not readily available to bond to the free His149 in the "off"-state, the H-bond with the *trans* CRO is preserved.

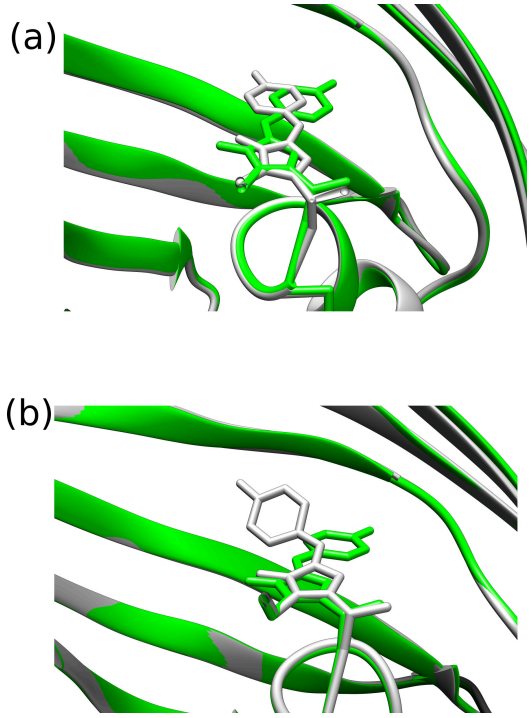


Figure 1.17 Comparing the CRO "on"- and "off"-state, in green and grey, respectively, of rsFolder (a) and rsEGFP2 (b). The rsFolder CRO undergoes smaller amplitude rearrangment compared to rsEGFP2. PDB: 5DTZ (green), 5DU0 (white) rsFolder, 5DTX (green), 5DTY (grey)

This peculiarity is how the authors comprehend not only the increased thermal stability of the "off"-state, but also the lower contrast of rsFolder. The close structural resemblance of the "on"- and the "off"-states could potentially mean that they both occupy equally stable energetic states.

Consequently, a mutation on rsFolder targeting this difference with rsEGFP2 (F146Y) lead to a mutant with higher contrast, rsFolder2. Additionally, its photofatigue ameliorated while its folding capabilities were maintained. However, the brightness dropped to almost half due to a lower extinction coefficient. Thus, while some rational design was possible upon crystallographic observations, the outcome is more complex than expected. Finally, rsFolder2 was regarded as a promising RSFP for RESOLFT, as long as some of its main disadvantages are overcome in the future. Since then, the effort to optimise rsFolder is ongoing with mutants such as rsFolder2 V151A [47] and rsFolder2 V151L.

rsFolder presents itself as an interesting system to study by NMR, not only due to the need to engineer it further but also because crystallographic structures leave us with more questions than answers. Namely, what is the role of His149 in both the "on"- and the "off"-state? How does the difference in pKa affect the photophysical properties of rsFolder as compared to rsEGFP2? What is the role of Phe146 and other residues near the CRO?

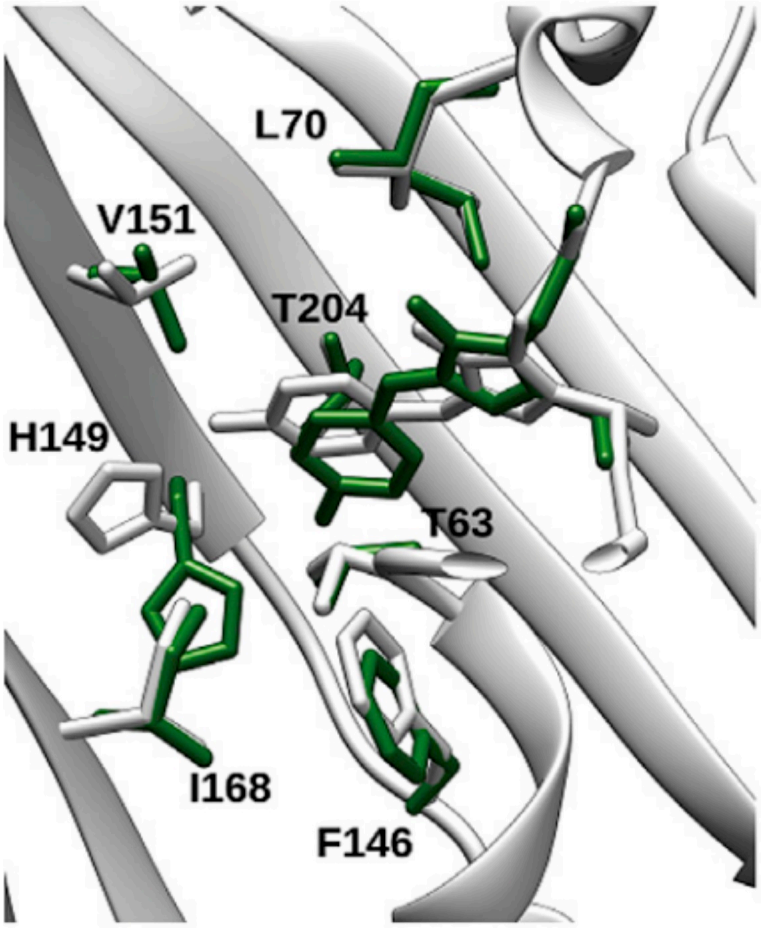


Figure 1.18 Comparing side chain conformation of residues in the vicinity of the chromophore of rsFolder in the "on"- and "off"-state, in green and white, respectively. PDB: 5DTZ (green), 5DU0 (white)

Is it realistic to expect that two such distinct optical states, are so similar structurally or is there dynamics involved that are not accounted for by crystallography? These and more such questions have been the focus of my PhD work and choosing NMR would prove to be essential in tackling them.

Lastly, rsFolder has been a very easy protein to work with, not only due to its efficient expression but due to photophysical properties that matched the NMR setup. Initially, during laser installation and optimisation, experiments could be run in the long-lasting "off"-state, while when combining NMR and laser pulse sequences, the combination of the short RF pulses with short laser pulses benefited from rsFolder's fast switching rates.

1.3 Multi-dimensional solution NMR to add a dynamic view on RSFPs

Multi-dimensional (nD) solution NMR spectroscopy is known for its decades-long contribution to the field of protein biophysics. Solution NMR allows for the study of molecules in *solution*, while multi-dimensional NMR refers to spectroscopy correlating the signals of several (spin-$\frac{1}{2}$) nuclei (i.e either same-type nuclei but with distinct frequencies, eg H_α-H_β or different nuclei types eg ^1H-^{15}N). From protein structure to protein interactions and dynamics, some fundamental questions regarding protein function have been elucidated using nD NMR spectroscopy. For the purpose of this work, NMR was mainly used as a tool to characterise the equilibrium dynamics of each of the respective thermal equilibrium states of rsFolder – the stable "on"- and the metastable "off"-state. Therefore, this part of my thesis introduction will focus on solution protein NMR studies which demonstrate the effectiveness as well as limitations of this technique when studying protein dynamics. But before discussing the impact of such NMR studies, I will give a short overview of the basic concepts underlining nD NMR spectroscopy, insofar as it aids the reader's critical understanding of the subsequent sections and of my thesis work as a whole.

1.3.1 Achieving atomistic information – the chemical shift

When placed in a magnetic field, $\vec{B_0}$, like the one induced by the superconducting coil of an NMR spectrometer, a spin-$\frac{1}{2}$ nucleus with a magnetic moment, μ, will align parallel or anti-parallel to $\vec{B_0}$. This is known as the "Zeeman effect", whereby previously degenerate eigenstates of an isolated nucleus, now split in distinct energy levels under the influence of the $\vec{B_0}$. For a spin-$\frac{1}{2}$, with gyromagnetic ratio γ, the spin may align parallel or anti-parallel to the $\vec{B_0}$, so the energy levels available are $E_\uparrow$ and $E_\downarrow$, respectively. The energy difference, ΔE between the two energy levels is proportional to the strength of the external magnetic field, $\vec{B_0}$ according to:

$$\Delta E = \hbar \gamma B_0 \tag{1.3}$$

where $\hbar$ is the Planck's constant.

Spectroscopically, this corresponds to an energy band gap proportional to a frequency ω_L, called the *Larmor frequency* of the nucleus.

$$\omega_L = \gamma B_0 \tag{1.4}$$

$$\Delta E = \hbar \omega_L \tag{1.5}$$

However, nuclei are not in isolation. The environment surrounding them – electrons orbiting the nucleus as well as nearby nuclei with their respective electrons – slightly influence the local magnetic field experienced by each nucleus, termed as $\vec{B_{local}}$ field. Electrons orbiting the nucleus induce an opposing magnetic field, σ, and are said to *shield* the nucleus. Despite them being small, the differences between Larmor frequencies of nuclei are detectable by NMR. In this way, NMR achieves atomistic resolution by retrieving characteristic frequencies for each nucleus, termed the *chemical shift, CS* of a nucleus. The CS has a characteristic frequency, ω_{CS}, proportional to the magnitude of the $\vec{B_{local}}$, which is the total effect of the external magnetic field, $\vec{B_0}$ and the local magnetic field *shielding* the nucleus, σ:

$$\omega_{CS} = \gamma B_{local} = \gamma(B_0 - \sigma) \tag{1.6}$$

By relying on gyromagnetic ratios, NMR overcomes common problems encountered in crystallography for nuclei of similar electron densities. The characteristic CS of a nucleus in a protein is influenced by its nuclear properties (eg ^1H, ^{13}C, ^{15}N), chemical properties (aromatic, aliphatic, backbone, protonation state etc) and structural properties (part of a flexible or rigid region). For example, depending on the protonation and tautomeric state of Histidine aromatic rings, each aromatic carbon and nitrogen, experiencing a different electronic environment gives rise to a characteristic CS [94]. The same level of precision holds true for other residues.

Moreover, NMR experiments can give insight, which might not be available by other techniques, into protonation states of residues and H-bonds since they are expected to affect the resulting CS (Figure 1.19). Therefore, NMR is a powerful technique for analytical chemistry of residues in proteins, providing atomistic resolution.

Additionally, the influence of solvent composition and other environmental factors are often the focus of solution NMR studies. Not only is this technique performed to study samples *in solution* but the sensitivity of the CS of each nucleus to temperature, pH etc enables precise titration studies [222]. An example is the use of solution NMR in measuring the pKa of aminoacid residues in proteins [50, 159].

1.3.2 NMR as an ensemble technique

NMR is an ensemble technique. Atomistic information about nuclei are averaged for all equivalent nuclei in a sample (eg a hydrogen of a specific position in a protein). Considering a system of magnetic moments in equilibrium under the magnetic field $\vec{B_0}$, the ratio between populations of spins occupying the two available energy levels, follows a Boltzmann distribution.

$$\frac{N_{down}}{N_{up}} = e^{\frac{-\Delta E}{kT}} \tag{1.7}$$

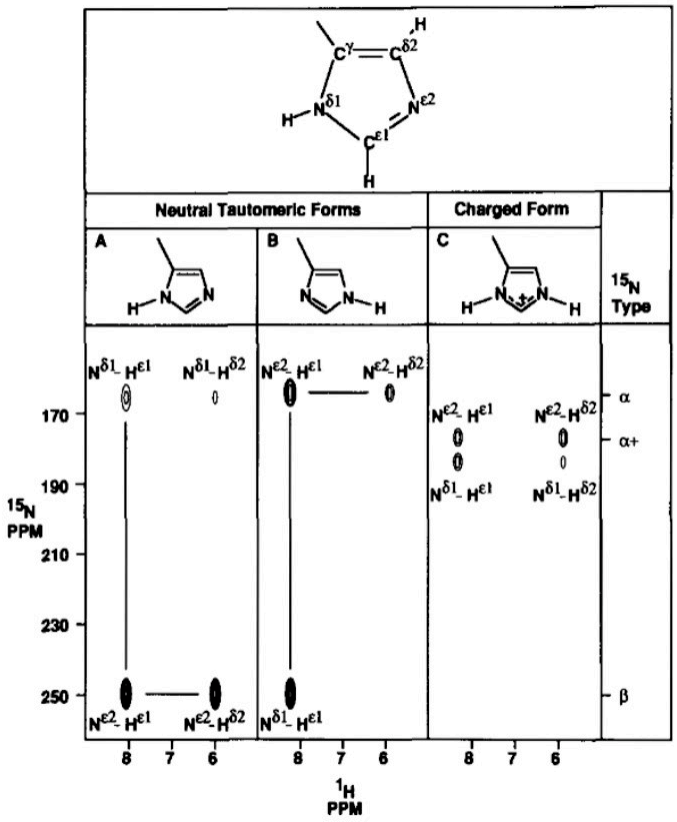


Figure 1.19 ^{15}N -^{1}H spectrum adapted from Pelton et al, focusing on the region of Histidines. From left to right, the three possible protonation states of Histidines and the expected CS patterns of their aromatic ring nitrogens. H_ϵ-N_ϵ cross peaks can be clearly distinguished from H_δ-N_δ, while protonation states of histidines have unique CS patterns. [170]

where N_{down}, N_{up} are the populations of spins anti-parallel and parallel to the external $\vec{B_0}$ field, respectively, k is the Planck's constant and T is the temperature of the system.

Overall, using the ensemble method, the spins considered can be grossly represented by a bulk magnetisation vector, $\vec{M}$ i.e net magnetisation after considering the contribution of the individual magnetic moments (parallel and anti-parallel) of a given nucleus (Figure 1.20). Shifting from a quantum to a classical mechanics representation helps in keeping track of the changes in magnetisation during an NMR experiment, without having to consider complex quantum processes. The net magnetization vector, $\vec{M}$ consists of two vectors containing different information: a z-component, $\vec{M_z}$, termed the *longitudinal magnetization*, which is the classical representation of the difference in populations of spin-up and spin-down nuclei, and a xy-component, $\vec{M_{xy}}$, termed the *transverse magnetization* representing the coherence present between spins. At Boltzmann equilibrium, $\vec{M_z}$ is pointing along the +z axis (if the gyromagnetic ratio of the nucleus is positive), while $\vec{M_{xy}}$ is zero.

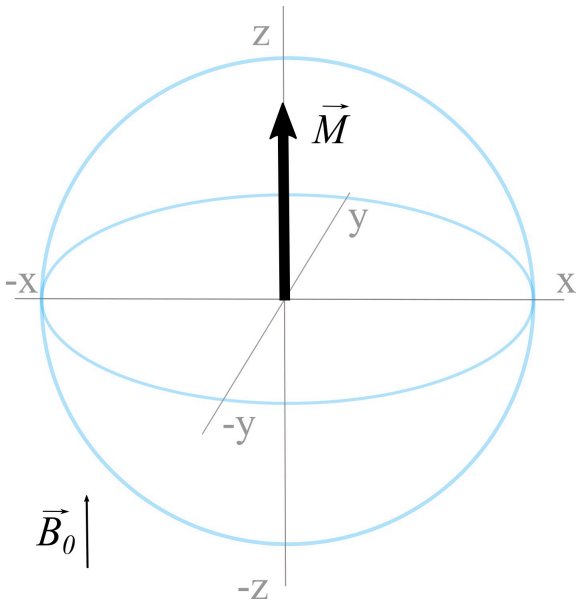


Figure 1.20 Schematic representation of the net magnetisation $\vec{M}$, sum of all the magnetic moments (parallel and anti-parallel) of a given nucleus in Boltzmann equilibrium; along the z-axis, no transverse component of the magnetization present.

1.3.3 The effect of Radiofrequency (RF) pulses and signal detection

By reshuffling the two populations i.e tilting the magnetization vector away from the z-axis, NMR spectroscopists can retrieve the CS corresponding to a given nucleus. To achieve such reshuffling, *radiofrequency, RF* pulses of certain strength and duration are used during an NMR experiment.

Under the influence of the $\vec{B_{local}}$ field, the magnetisation vector $\vec{M}$ is along the z-axis. An additional magnetic field, $\vec{B}_1$, oscillating at a *reference frequency*, ω_0, near the expected ω_L of the nuclei of interest and perpendicular to the $\vec{B}_0$ field, can "tilt" $\vec{M}$ away from its position at equilibrium (Figure 1.21). This oscillating field comes along the x- or y-direction and it is termed as an RF pulse. By being *in resonance*, the RF pulse forces the $\vec{M}$ vector to now precess around the x- or the y-axis (depending on whether it is a x- or a y-RF pulse, respectively). The duration of the RF pulse determines how much the $\vec{M}$ magnetisation has tilted away from the z-axis. Once it is switched off, the $\vec{M}$ precesses around the z-axis and slowly returns to equilibrium, due to *nuclear spin relaxation*.

The xy-component of the magnetization vector is detected. When the $\vec{M}$ vector has titled away from the z-axis, some part of it (if not all) will be in the xy-plane. By precessing in the xy-plane with a frequency ω_L, around the z-axis (once the RF pulse is off), this magnetisation vector induces a small electrical current in the detection coils surrounding the sample (Figure 1.22). The signal recorded has the shape of a *Free Induction Decay (FID)*

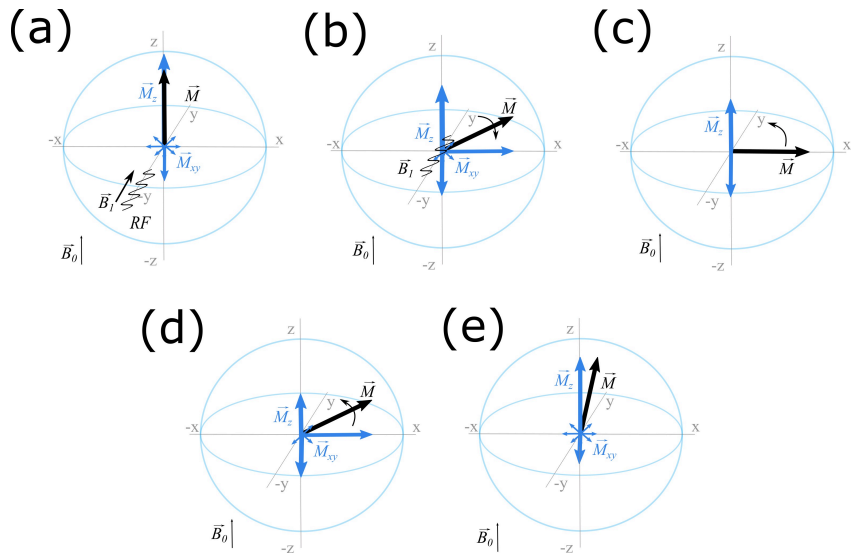


Figure 1.21 Schematic representation of the influence of a $\frac{\pi}{2}$ RF pulse on the net magnetisation $\vec{M}$, as seen in the rotating frame. (a) Starting at Boltzmann equilibrium, (b) the RF pulse starts tilting the net magnetization from the z-axis to the xy-plane. In the case of a *90° pulse*, (c), once the RF pulse is off, the longitudinal component of $\vec{M}$, $\vec{M}_z$, has disappeared while the transverse $\vec{M}_{xy}$ has built up and it precesses on the xy-plane. Consequently at (d), with the switching off of the RF pulse, the $\vec{M}_z$ magnetisation starts returning back to the z-axis, now under the influence of the B_{local}, until (e) Boltzmann equilibrium is achieved.

which is later Fourier Transformed (FT) to obtain the NMR spectrum. The spectrum consists of signals, "peaks", each corresponding to the Larmor frequency of a precessing nucleus present in the sample, i.e the ensemble-averaged CS of the nucleus.

It is convenient to describe an NMR experiment from the point of view of the *rotating frame*, i.e from the point of view of an observer rotating at frequency ω_0. This aids in schemes representing the $\vec{M}$ vector rotation around an xyz-sphere.

1.3.4 1D versus nD heteronuclear NMR spectroscopy – sensitivity and resolution

So far, only one-dimensional spectra have been considered, referring to only one type of nucleus (eg hydrogen). Evidently, a problem arises when recording 1D spectra of complex macromolecules containing many hydrogens – spectral crowding. Discriminating the different hydrogen nuclear spins precessing in a protein sample requires employing *multidimensional* (nD) NMR spectroscopy [64, 118]. I will limit the discussion to nD heteronuclear NMR spectroscopy, correlating different types of nuclei, such as hydrogens with nitrogens, while nD homonuclear spectroscopy (eg correlating hydrogens with other

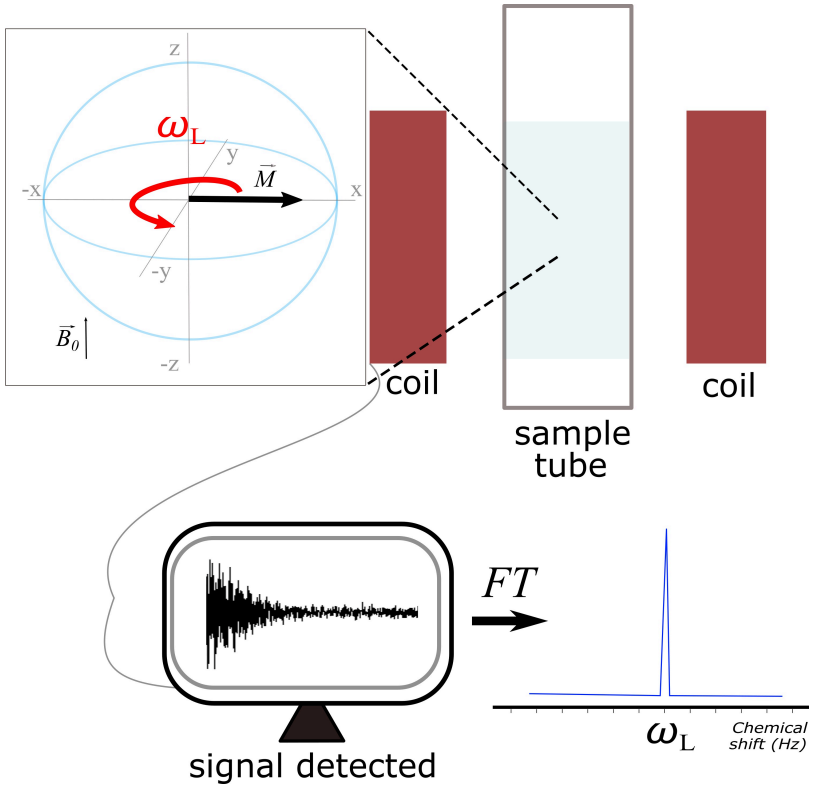


Figure 1.22 Schematic representation of the signal detected by the detection coils of the spectrometer probe, after the excitation by an RF pulse. The net magnetisation vector describing a spin system is precessing on the xy-plane with an frequency, ω_L, perturbing the Alternating Current of the detection coils. Over time the signal decays, therefore the time-dependent signal has the shape of a Free Induction Decay(FID). By applying a Fourier Transform (FT) on the acquired FID, the resulting spectrum consists of a peak in the frequency domain at frequency ω_L.

hydrogens) is not described here.

The development of nD solution NMR was a breakthrough that enabled NMR spectroscopy to contribute substantially to studies of complex macromolecules, such as proteins. Unlike its predecessor, 1D NMR, nD NMR achieves higher signal resolution since it collects information on correlations between nuclear spins. In this way, each signal detected in one frequency dimension (hydrogen dimension) is "connected" to another signal on higher dimensions (for example nitrogen). Therefore, depending on the NMR RF pulse sequence used, different nuclear magnetic interactions are considered which correspond to different information (eg about local or global architecture surrounding a nucleus, within certain timescales). Taken together, the dimensions in an NMR spectrum correspond to the "probing" nucleus i.e the nucleus activated by the RF pulses used, and the resulting peaks represent the correlation between those nuclei "activated" by a given RF pulse sequence.

Hence, if for each resonance frequency of a hydrogen, its correlation to another type of

nucleus is available, hydrogen nuclei with identical resonance frequency in the hydrogen dimension can be discriminated when correlated to, for example, the nitrogen dimension (Figure 1.23). In this case, the signal would be a correlation signal of the interaction between a hydrogen and a nitrogen, making it a 2D spectrum (H-, N-dimension).

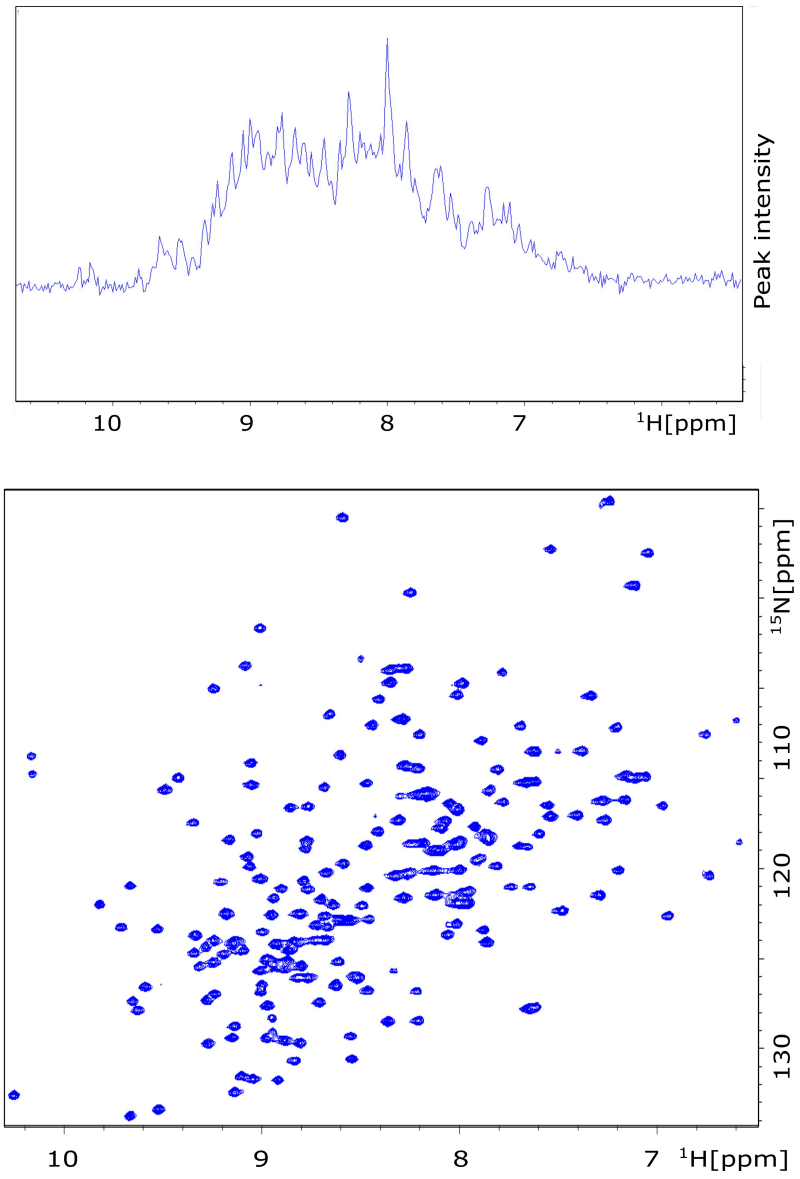


Figure 1.23 Spectra recorded at 850MHz of the "on"-state of rsFolder at 40°C. On top, the 1D spectrum recorded focusing on amide hydrogens, on bottom, the 2D ^{1}H-^{15}N correlation spectrum (i.e focusing on amides) of the same sample. The discriminatory capabilities of 2D NMR are evidently more robust than 1D NMR.

The development of nD NMR not only introduced the need for new RF/detection coils (to include pulses for ^{13}C and ^{15}N for example) and isotopic labelling of samples, but also significantly increased the experimental time required to record such a data set.

A simplistic view of an NMR pulse sequence consists of four consecutive steps; a. preparation, b.evolution, c. mixing and, d. detection (Figure 1.24). By convention, the first dimension is the F1, the second is the F2 etc. During the preparation step, RF pulses perturb the bulk magnetisation in F1 and eventually tilt it to the transverse plane, therefore "activating" it. In 1D NMR, the transverse magnetisation of nuclei in channel F1 is let to "evolve" in the next step, during time t1, as named by convention. Acquiring the resonance frequencies (chemical shifts) just requires repeating these two steps (evolution is the same as detection) as many times as needed to achieve a sufficiently high signal-to-noise ratio. The resulting signal, in the form of an FID is Fourier-transformed to retrieve the resonant frequencies of the active nuclei. In other words, t1 is directly detected, making the t1 or F1 dimension, the direct-detection dimension.

In multi-dimensional NMR experiments, other steps must be added to promote magnetization transfer from the "active" population of nuclei to another population of nuclei, which are said to be "correlated". Transferring of magnetization (or spin polarization) is induced by one or a series of RF pulses, during the "mixing" step. These pulses allow for specific interactions between nuclei to be active. Consequently, depending on the interaction present, RF pulses either achieve magnetisation transfer between nuclei via *scalar couplings (J-couplings)*, or through space via *dipolar couplings*.

A spin is said to be J-coupled to another spin when their electronic orbitals overlap (via one, two or more chemical bonds). The shared electrons can be polarized due to the presence of a nuclear spin nearby and, consequently, "transfer" this spin polarization to the other nuclear spin. On the other hand, spins not necessarily involved in a chemical bond but being in proximity, can induce polarization to each other, via dipolar coupling (or dipole-dipole interaction), since both "feel" the magnetic field produced by their neighbouring nucleus.

In nD NMR, CS evolution is first taking place for the F2, F3 etc dimensions before it does on F1. Simply put, the precession on the xy-plane during detection is the convolution of precessions taking place during the span of the NMR experiment, thus it is affected by the interactions allowed to take place in the "mixing" step. Lastly, *nD* NMR requires the repetition of the second and third step (evolution and mixing) for each nucleus concerned.

Detection occurs during t2 in a 2D NMR experiment, which makes the F2 dimension the direct-detection dimension. Since detection is not taking place during t1, F1 is now the indirect dimension. In order to obtain the information of interest, the sequence is repeated for incremented t1 delays. Thus, in a 2D experiment, both the t2 and the t1 domain need to be Fourier Transformed, to construct the 2D spectrum.

Most commonly used NMR pulse sequences are so-called "out-and-back", whereby magnetisation is transferred from one nucleus type, namely hydrogen, to the nucleus type of interest and back to the hydrogens for detection. Such schemes ensure increased signal, since hydrogens have the highest gyromagnetic ratio, γ, compared to other commonly used nuclei. Their high γ as compared to ^{13}C or ^{15}N means that ^{1}H are ideal for polarization transfer and can achieve an overall higher NMR signal detected [21].

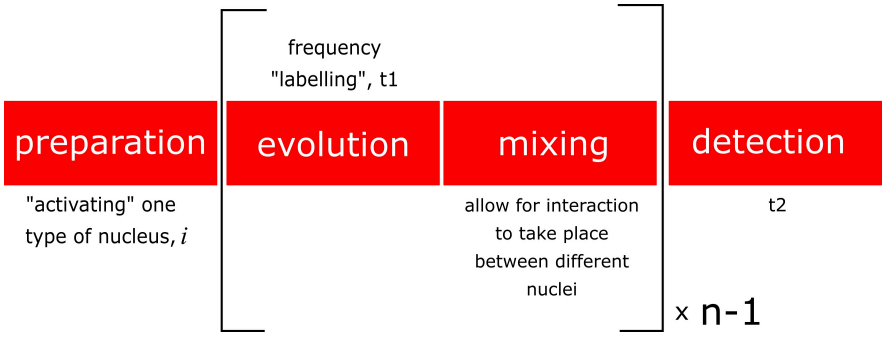


Figure 1.24 Simplified diagram depicting the 4 main steps required to run a multi-dimensional NMR experiment of n dimensions. Initially, a nucleus i is "activated" before evolution and mixing takes place, which need to be repeated $n-1$ times, where n is the number of dimensions of the experiment.

Altogether, while NMR spectroscopy for biomolecules is by far more interesting when more dimensions are employed, it is considerably more time consuming than its 1D counterpart. With the addition of another dimension, the number of necessary scans exponentially increases and so does the required acquisition time. Fortunately, continuous efforts to decrease acquisition time and enhance signal-to-noise ratio [67, 187, 188, 205] have made nD solution NMR less time-consuming and costly.

1.3.5 An example of a two-dimensional NMR pulse sequence: Heteronuclear Single Quantum Coherence (HSQC) spectroscopy

For demonstrative purposes, I shall briefly explain how a conventional *2D* ^{1}H-^{15}N *Heteronuclear Single Quantum Coherence* (HSQC) experiment [23] works which aims at recording the correlation of amide hydrogens with their directly bonded amide nitrogens. For a ^{15}N-labelled sample, an HSQC provides the so-called "backbone fingerprint" of the protein and it is very frequently used in biomolecular NMR – by observing the HSQC of a protein over time, or under different conditions, bound or unbound to a ligand etc, a lot of information can be retrieved about the protein's behaviour on a per-residue basis.

Starting with a 90° pulse on the ^{1}H channel, the signal is first "carried" by the hydrogens during the *preparation* step as shown in Figure 1.25(a). An INEPT [151] step consists of the necessary pulses for magnetisation (or *coherence*) transfer between J-coupled nuclei, in this case between ^{1}H and ^{15}N. The ^{15}N spins are left to precess around the xy-plane, i.e to *"evolve"* during t1 so that the final recorded signal will inherit the precession frequency of the ^{15}N spins. During the *mixing* step, a retro-INEPT step is applied and the magnetisation is transferred back to the ^{1}H. Finally, the retro-INEPT step leaves the ^{1}H spins precessing in the xy-plane for a time t2, whereby *detection* of the ^{1}H frequency, modulated by the correlated ^{15}N chemical shift is recorded.

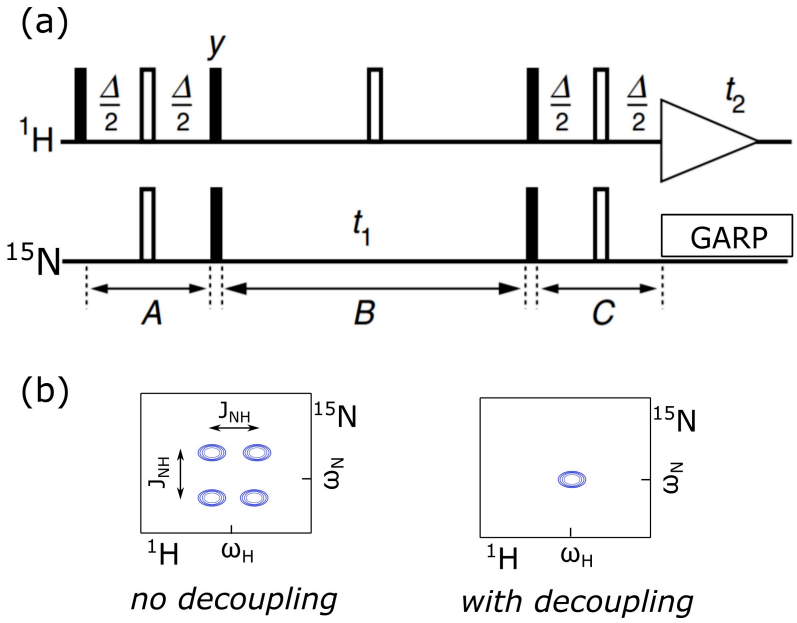


Figure 1.25 Example of a simple 2D ^{1}H-^{15}N correlation HSQC experiment. (a) Picture adapted by Keeler [118] and modified for the pulses taking place in channels ^{1}H and ^{15}N. Filled rectangles represent 90° pulses, and open rectangles represent 180° pulses. $\Delta/2$ is a delay time, with $\Delta = \frac{1}{2J_{NH}}$, ensuring maximum magnetisation transfer for nuclear spins J-coupled with J-coupling strength of J_{NH}. Starting from equilibrium, during (A) an INEPT step is applied to transfer the magnetisation from the amide ^{1}H to the amide ^{15}N. During (B), a 180° pulse refocuses the coupling between the ^{1}H and ^{15}N spins so as to allow for the chemical shift of the ^{15}N to evolve during t1. At (C) a retro-INEPT is applied to recover the magnetisation to the ^{1}H spins which finally evolve and are detected during time t2. The 180° in the ^{1}H channel, during t1, refocuses the J-coupling between the spins, while the GARP element on the ^{15}N channel applies *decoupling* in order to retrieve in (b) a single peak from the quadruplet (at ω_H, ω_N).

Notably, during t1 and t2, the spins are precessing under the influence of not only the $\vec{B}_0$ field and their chemical shift, but also the J-coupling interaction with the covalently bound nucleus. The effect of the J-coupling on the CS adds up or is subtracted in either direction of the ^{1}H and ^{15}N dimensions, resulting in a quadruplet set of peaks for each ^{1}H-^{15}N pair (Figure 1.25 (b)). Fortunately, *decoupling* of the two nuclei from each other in the form of eg the GARP element [195] or a single 180° pulse in the middle of the incremented time delay t1, allows for the averaging out of the J-coupling effect. Consequently, the quadruplet collapses (averages out) onto one single peak.

1.3.6 Resonance frequency assignment to protein residues

The great discriminatory capabilities of multidimensional NMR allow for identifying specific residues in the protein. Once the peaks in the NMR spectrum are assigned to a specific protein residue, the characteristic behaviour and dynamics of this residue can be investigated. Above all, a 2D correlation spectrum (showing amides or side-chains) is considered as the NMR "fingerprint" of each protein. Therefore, interpreting any NMR experiment, an assignment of the relevant spectra is necessary.

More specifically, information on the frequencies of hydrogens, nitrogens and carbons belonging to the same residue or a neighboring residue can be acquired from 3D triple-resonance experiments and the assignment process of the protein can commence. The three dimensions of these experiments consist of either the ^{1}H, ^{15}N and ^{13}C, associating amides to their respective carbons or ^{1}H, $^{13}C_1$ and $^{13}C_2$, when assigning a side-chain HC pair (^{1}H-$^{13}C_1$) to each nearby carbons ($^{13}C_2$). A database of experimentally recorded Cα and Cβ frequencies of aminoacids acquired from a range of proteins allows for the identification of the type of aminoacid in question, while transferring of magnetisation between residues gives information of the aminoacid sequence. Similar to a puzzle, a spectroscopist's task is to identify how residues are linked and what kind of residues they are in order to assign the 2D spectra at hand.

Usually, to assign backbone amide NMR spectra, more than one 3D triple-resonance experiment needs to be recorded and each experiment is focusing on one or two types of carbons. Similarly, when assinging side-chain spectra, triple resonance experiments look for the resonant frequencies of the Hydrogens and Carbons of a side chain. Once all the information is assembled, the "fingerprint" spectra have assigned peaks corresponding to residues in the protein. Only then can these spectra provide site-specific information on structure and dynamics. Evidently, the assignment process is relatively time consuming; due to long experimental times (3-10 days, depending on the protein) as well as data analysis (approximately a week). Nevertheless, once the assignment is available, one can move on to deciphering the dynamical behaviour of the protein.

1.3.7 Studying protein dynamics using nD solution NMR

With the NMR assignment at hand, information can be retrieved concerning the dynamics experienced by specific atoms in the protein. The underlying physical mechanism that allows for such investigation is the relatively "slow" (as compared to other spectroscopic techniques) return to equilibrium of the spin system at hand, i.e the process of nuclear spin relaxation.

As a side note, similarly to photoexcitation in fluorescent molecules, "exciting" the spins to a higher energy level requires electromagnetic radiation (in the form of RF pulses). In contrast to fluorescence emission, however; (a) excitation is in the range of MHz instead of THz and therefore (b) spontaneous emission, which, according to Einstein's coefficient, A_{21},

has a probability proportional to the cube of the frequency of the photon emitted between energy state 2 and 1, ($A_{21} \propto \omega^3$) [20] is comparatively improbable. Overall, the process of re-establishing equilibrium, i.e nuclear spin relaxation, driven by the *local fluctuations* of the $\vec{B_{eff}}$ field is much slower than electronic transitions in fluorescent molecules.

Molecular motions affect NMR signal

The rate of decay of the xy-component of the bulk magnetization is the determining factor affecting NMR signal quality. Interactions between nuclei as well as conformational exchange phenomena over time cause the bulk magnetization vector to return to equilibrium. These time-dependent processes, affected by the motion experienced by each nucleus, can be probed by a series of NMR experiments, each one sensitive to a motion occurring at different timescales. Before choosing to run such NMR experiments, one can appreciate the range of motions characterising a nucleus by observing its signature NMR peak linewidth. More specifically, since variations in motions result in different lineshapes, we conventionally separate motions detected by NMR into three categories: fast ($ps - \mu s$), intermediate($\mu s - ms$) and slow($ms - min$). The quality of the signal does not depend only on the motions, i.e the exchange rate between conformational states, k_{ex}, but also on the NMR spectrometer field strength, the number of available states (two-state model, *A, B*, or more), their respective occupancy (p_A, p_B) and their characteristic CS (ν_A, ν_B). Assuming a two-state model of equally populated states, A, B, with CS difference of $\Delta\nu$, the NMR lineshapes corresponding to fast, intermediate and slow exchange are shown in the figure below (Figure 1.26).

In Figure 1.26, typical NMR experiments probing molecular motion taking place in the NMR-relevant fast, intermediate and slow exchange regime at steady state are shown. The origins of the effect of molecular motion on NMR observables and the ability of each NMR pulse sequence to probe the motions present is explained in the sections below. Firstly, I shall briefly introduce NMR relaxation theory, a process which is intrinsic for all nuclei present in the sample in solution, regardless of the local motions experienced, and leads to the decay of signal over time. This is followed by a description of NMR investigation of fast, intermediate and slow exchange dynamics, as defined above, with each section including few examples found in literature of the use of such experiments to study protein dynamics. More detailed descriptions of NMR relaxation theory can be found in NMR textbooks [118, 233, 32] and reviews [164, 163, 122].

Molecular tumbling in solution and fast local motions

Molecules in solution experience *tumbling*. A concept of paramount importance to NMR relaxation theory, the *rotational correlation time,* τ_c, is the time it takes for a molecule, affected by Brownian motion in liquid, to completely "lose" memory of its initial position, i.e to undergo three dimensional "random walk". The time, τ_c, is dependent on the viscosity

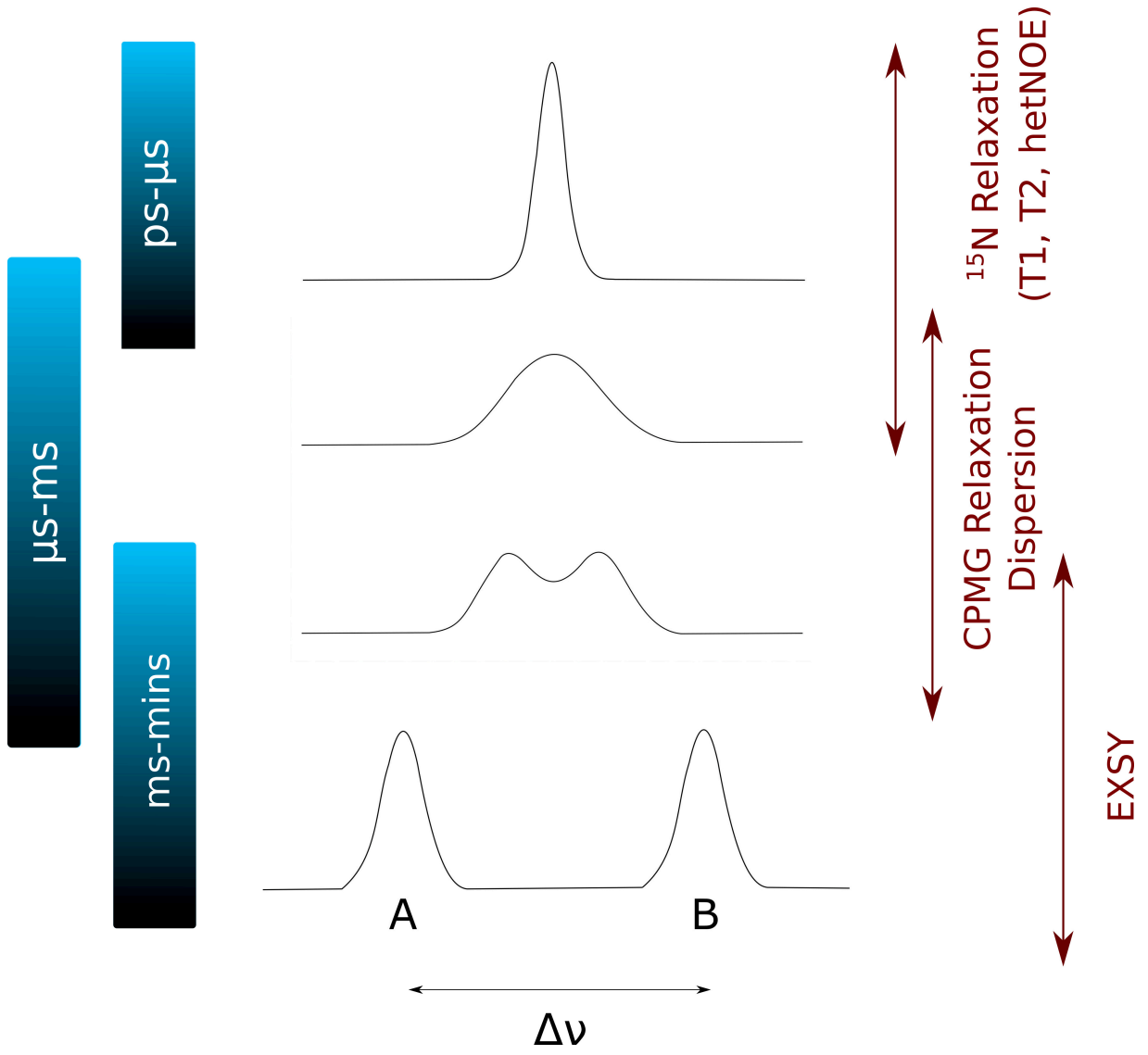


Figure 1.26 The different lineshapes of NMR peaks depending on the timescale of exchange between two conformations of equal populations that the nucleus experiences, from top to bottom; fast, intermediate and slow. On the right, typical NMR experiments sensitive to these motions.

η of the solvent, the radius r of the idealised "sphere" of the molecule and the temperature, T, according to the Stokes-Einstein relation:

$$\tau_c = \frac{4\pi\eta r^3}{3kT} \tag{1.8}$$

This is mathematically conceptualised using a *correlation function*, $G(t)$ which takes the orientation of the molecule at time t=0 as a reference and models how far the molecule is

perturbed from that initial orientation due to stochastic processes. The τ_c is, therefore, the time constant of the exponential decay characterising the "memory" of the molecule of its initial orientation (Figure 1.27).

$$G_t = G_0 exp(-\frac{t}{\tau_c}) \tag{1.9}$$

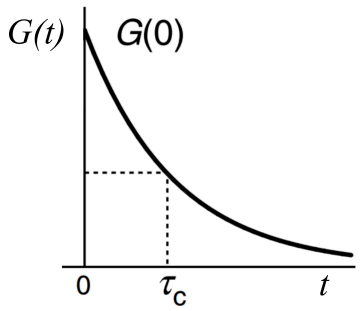


Figure 1.27 The exponential decay over time, t, of the molecule's "memory" of its initial position and orientation, termed "correlation function", $G(t)$ with a time constant τ_c.

For molecules tumbling in similar environments, i.e viscosity, temperature of solution, but having different sizes, the corresponding correlation function varies – a smaller molecule undergoes "random walk" faster than a larger molecule. To have better insight on the occupancy of different frequencies of motion characterising a molecule, one can Fourier Transform the correlation function, $G(t)$, to retrieve the *spectral density function*, $J(\omega)$. As can be seen in the figure below (Figure 1.28), a smaller molecule, with smaller τ_c samples a wider range of frequencies of motion.

Similarly to the comparison of a large versus a small molecule, if a nucleus within a tumbling molecule experiences local motions similar or faster than τ_c, the re-orientation of this nucleus is accelerated. Hence, another time constant, namely the internal correlation time of the nucleus, τ_e should be introduced. As mentioned before, the molecule due to tumbling is allowed to occupy all possible orientations. In contrast, the re-orientation of a nuclear pair within a molecule (eg an amide pair of a protein) is hindered by chemical and steric restrictions. Therefore, the τ_e is the time constant of a correlation function, $g(t)$ which does not necessarily reach zero as is the case for $G(t)$ of the whole molecule, but rather reaches a value defined by an *order parameter,* S^2 (Figure 1.29).

In the case of a nucleus experiencing dynamics whereby τ_e is smaller than τ_c, the correlation function, $g(t)$, decays faster than the $G(t)$, resulting in a different $J(\omega)$.

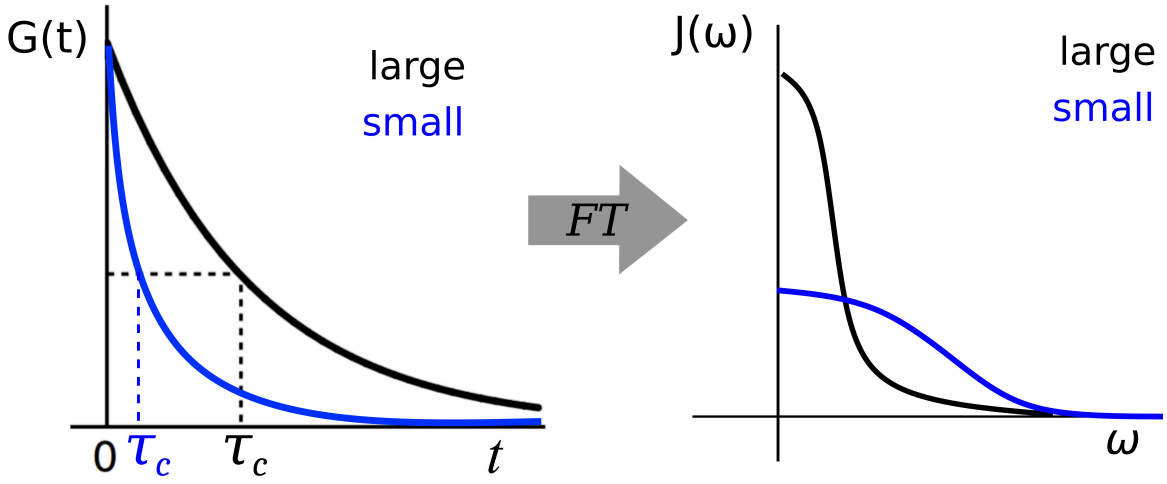


Figure 1.28 On the left, the correlation functions, $G(t)$, corresponding to a small or a large molecule (relative to each other) and on the right, their spectral density functions $J(\omega)$. A smaller molecule, characterised by a smaller τ_c can sample a wider range of frequencies of motion.

Longitudinal and transverse spin relaxation depend on spectral density function of molecular motion

NMR-observed relaxation is influenced by global and local motions. This is because, considering a nucleus within a molecule, motions have an effect on the strength of its magnetization vector. More specifically, orientation-dependent components of the $\vec{B_{eff}}$ field experienced by the nucleus alter over time, if motions are present. The orientation considered is the one of the molecule in relation to the z-axis. More specifically, the $\vec{B_{eff}}$ field is the resulting magnetic field, when all interactions coupling with the nucleus in question are considered. The main interaction mentioned above is caused by the presence of the $\vec{B_0}$ field, called the "Zeeman interaction". In addition, there is CS, dipolar coupling, quadrupolar coupling and J-(or scalar) coupling. The coupling strengths differ for each nucleus, each one giving rise to either *time-dependent* or *time-independent* contributions to the $\vec{B_{eff}}$ field, corresponding to orientation-dependent and orientation-independent interactions. Since only spin-$\frac{1}{2}$ nuclei are accounted for in this section, the quadrupolar coupling (relevant to nuclear spins with higher spin quantum number) is not discussed.

In the case where a nucleus within a molecule undergoes motion, orientation-dependent (or time-dependent) components of the nuclear interactions average out during τ_c (or τ_e, when τ_c»τ_e), while orientation-independent (or time-independent) components remain

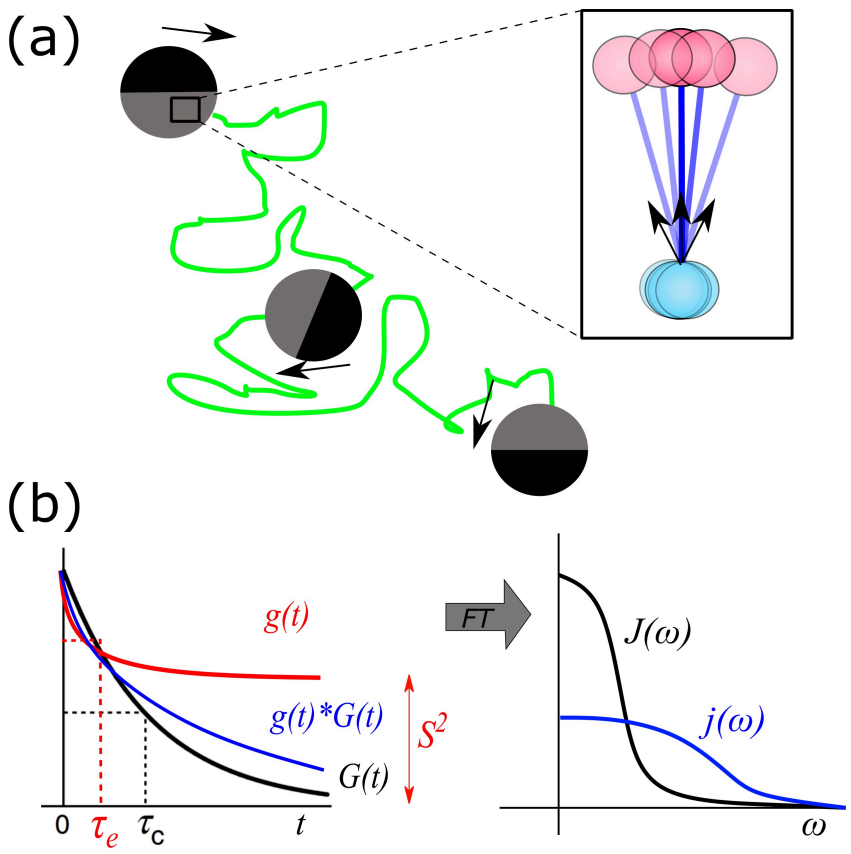


Figure 1.29 Schematic and graphical representation of the global and local motions experienced by a nucleus in a molecule in solution NMR. (a) The idealised molecule in grey-black undergoes "random walk" due to Brownian motion. Zooming into a internuclear bond (blue and pink nucleus), the nuclei have a preferred orientation with respect to each other but are still allowed to occupy a number of orientations. In (b), global, $G(t)$, and local, $g(t)$, correlation functions are depicted, with their respective time constants (τ_c and τ_e). Overall, the internuclear bond vector in question has a correlation function resulting from both re-orientation mechanisms, $G(t) * g(t)$. By applying a Fourier Transform, the spectral density, $j(\omega)$, when both the global and the fast local motion is considered, has a lower starting value, $j(0)$, and extends in higher frequencies as compared to the $J(\omega)$ where only tumbling is considered.

fixed. Accordingly, the two main time-dependent mechanisms considered in protein NMR are the orientation-dependent component of the CS, termed as the *chemical shift anisotropy (CSA)*, and the *dipolar coupling* or *dipole-dipole interaction (DD)*. Ultimately, the contribution of CSA and DD interactions does not alter the final peak position for a given nucleus in the spectrum (Figure 1.30).

Nonetheless small fluctuations on the $\vec{B_{eff}}$ due to DD and CSA experienced by the spins

act as counter-RF pulses, promoting spin transitions from the $E_{\downarrow}$ to the $E_{\uparrow}$. Therefore, after an RF pulse has resulted in a certain population shuffling in the spin system, such spin transitions caused by $\vec{B_{eff}}$ fluctuations over time lead to longitudinal relaxation, T_1 i.e the population ratio of spin-up and spin-down is re-established back to equilibrium. In the classical representation, this is summarised by the bulk magnetization vector re-aligning along the z-axis.

The effect of DD and CSA over time also contributes to the loss of signal i.e diminishing of the net magnetization in the transverse plane, a process called *transverse relaxation*, T_2. To explain how transverse relaxation is caused, without employing the complex quantum mechanical representation, each nucleus contributing to the net magnetization can be thought of as having a given orientation of its magnetic moment at a given time point in the transverse plane (Figure 1.31) and it precesses around the transverse plane with frequency, ω_L. The RF pulse, inducing coherence, results in all same-type nuclei being "in-phase". However, due to differences in local environment and as the $\vec{B_{eff}}$ alters over time, and as a result, so does the corresponding ω_L of each nucleus. In the simplified diagram drawn in Figure 1.31, different ω_L correspond to different orientations of the magnetic moments of nuclei. Therefore, once the RF pulse is not acting on the spins, the nuclei start losing coherence, each one straying away from the others at slightly altered precession frequencies. Consequently, transverse relaxation leads to the diminishing of the bulk magnetization components along the xy-plane.

In summary, it is important to consider the two components of relaxation: longitudinal and transverse. In the ensemble representation, they simply correspond to the return of the M_z and M_{xy} component of the bulk magnetisation vector to equilibrium, respectively. More specifically, longitudinal relaxation accounts for the reshuffling of the population of spins (parallel and anti-parallel) back to Boltzmann equilibrium due to spin transitions and is referred to as $R_1(s^{-1})$ (or time $T_1(s)$), while transverse relaxation $R_2(s^{-1})$ (or time $T_2(s)$), is the rate at which spins lose coherence after the RF pulse is switched off.

The built-up of the bulk magnetisation vector along the z-axis over time, $M_z(t)$, can be described as:

$$M_z(t) = M_{z,0}(1 - e^{-tR_1}) = M_{z,0}(1 - e^{-t/T_1}) \tag{1.10}$$

While the decay of the transverse component of magnetisation, $M_{xy}(t)$, over time is:

$$M_{xy}(t) = M_{xy}(0)e^{-tR_2} = M_{xy}(0)e^{-t/T_2} \tag{1.11}$$

Where $M_{z,0}$, $M_{xy,0}$ are the magnitudes of the M_z and M_{xy} components of the bulk magnetisation at $t = 0$, i.e when the RF pulse is switched off, and R_1 and R_2 are:

$$R_1 = \frac{1}{T_1}, R_2 = \frac{1}{T_2} \tag{1.12}$$

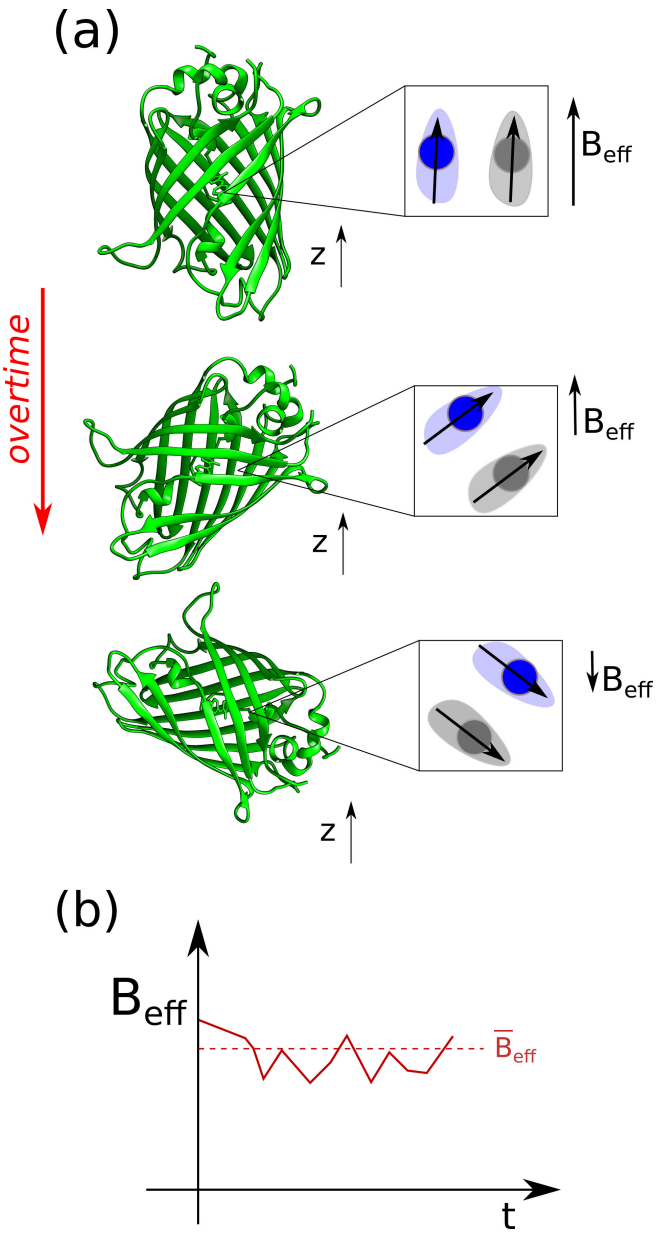


Figure 1.30 Schematic representation of the effect of Dipolar coupling (DD) and Chemical Shift Anisotropy (CSA) on a nucleus within the structure of the GFP. The CSA originates from an anisotropic electron cloud (pale blue and pale grey on blue and grey spin, respectively), while the DD is induced from one spin's magnetic field to the other spin due to proximity. As the GFP tumbles over time, when focusing on the blue spin, the orientation-dependent interactions contributing to the $\vec{B_{eff}}$ alter as seen in (a), causing $\vec{B_{eff}}$ fluctuations, plotted in (b). The average $\vec{B_{eff}}$ is not affected, i.e the peak position of the nucleus in the spectrum will be the same. However, fluctuations on $\vec{B_{eff}}$ lead to relaxation over time.

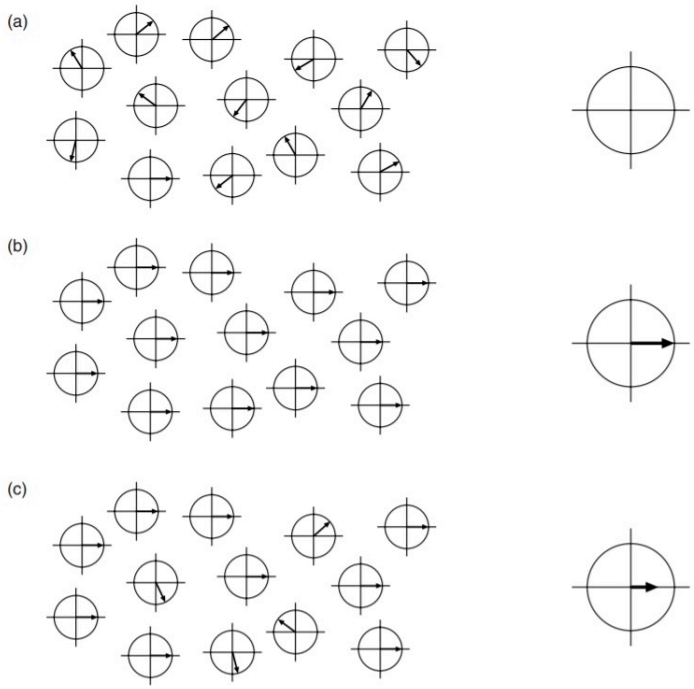


Figure 1.31 Figure adapted from Keeler [118], simplifying the concept of coherence between spins. At (a) spins are out of coherence, or are so-called "out-of-phase"; the net transverse magnetization (on the right) is zero. Here, the spin system is at equilibrium. (b) shows spins "in-phase", after the RF pulse has induced coherence. Finally, at (c), once the RF pulse is switched off, the nuclei start dephasing, each one precessing at a different Larmor frequency, since the $\vec{B_{local}}$ alters over time due to DD and CSA contributions, having a diminishing effect on the net magnetization on the right.

Measuring nuclear spin relaxation to evaluate molecular motions in the fast exchange regime

The majority of NMR experiments that probe molecular dynamics of the backbone at faster timescales (i.e sub-microsecond) measure spin relaxation rates. Particularly in protein NMR, it is convenient to study relaxation mechanisms by probing the backbone amide spin pair, i.e the relaxation induced to ^{15}N due to ^{1}H-^{15}N DD and ^{15}N CSA. Amides are an attractive probe for site-specific dynamics characterisation since they: a. can be assumed to act as an isolated spin pair, b. result in better resolved NMR spectra than other protein parts and c. do not require expensive labelling techniques, since ^{15}N labelling is cheaper than ^{13}C [73, 24].

Since the frequency of fluctuations of the $\vec{B_{local}}$ is directly correlated to the motion due to tumbling of the molecule and local motions of the amide pair, T_1 and T_2 of a ^{15}N nucleus can be mathematically described using the concept of the spectral density function, $J(\omega)$.

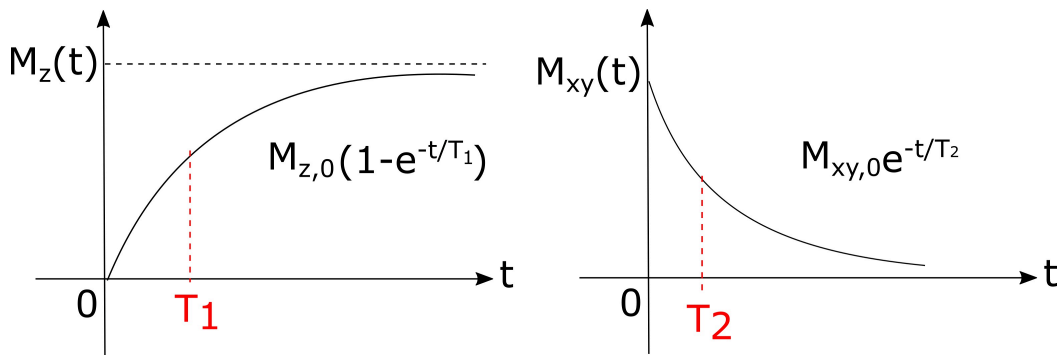


Figure 1.32 Graphical representation of the equations describing longitudinal and transverse relaxation. Exponential built up and decay of the respective components $\vec{M}_z$ and $\vec{M}_{xy}$ of the bulk magnetisation over time, with constants T_1 and T_2, respectively.

Another NMR observable is also affected by global and local motion in the molecule, namely, signal enhancement due to the *heteronuclear Nuclear Overhauser effect (hetNOE)*, related to DD coupling between heternuclei in proximity. Therefore, by retrieving the values of T_1, T_2, and hetNOE the $J(\omega)$ and thus the global and local motions of the molecule can be deciphered. Considering an amide pair, HN, in a molecule the equations describing the T_1, T_2 and hetNOE of the nitrogen nucleus become:

$$1/T_1 = \alpha J(\omega_N) + \beta J(\omega_N - \omega_H) + \gamma J(\omega_N + \omega_H) \tag{1.13}$$

$$1/T_2 = \alpha J(0) + \beta J(\omega_N) + \gamma J(\omega_H) + \delta J(\omega_N + \omega_H) + \epsilon J(\omega_N - \omega_H) + R_{ex} \tag{1.14}$$

$$hetNOE = \alpha J(\omega_N + \omega_H) + \beta J(\omega_N - \omega_H) \tag{1.15}$$

Where $\alpha, \beta, \gamma, \delta, \epsilon$ are constants, whose values depend on the ^{15}N CSA and ^{1}H-^{15}N DD interactions strengths. R_{ex}, related to additional relaxation induced due to exchange in the few hundreds of μs-ms timescale, will be discussed later in this chapter.

This set of three NMR observables, each one differently sensitive to a spectral densities at characteristic frequencies ($J(\omega_N)$, $J(0)$, $J(\omega_N + \omega_H)$ for T_1, T_2 and hetNOE, respectively) allows spectroscopists to probe molecular motions. Typically, once these values are determined, there are 3 ways by which motion can be assessed, either by considering the spectral density function, or by comparing global and local motions:

1. Spectral density mapping – in order to access several frequencies, ω, T_1, T_2 and hetNOE are retrieved from multiple spectrometers (different magnetic field strengths). As a result, the spectral density function, $J(\omega)$ can be "mapped".

2. Model-free approach analysis [132] – using a model described by Lipari-Szabo, T_1, T_2 and hetNOE are used to describe $J(\omega)$ of the HN pair in terms of constants τ_c, τ_e and S^2. This way, the spectral density function is described as:

$$J(\omega) = \frac{2}{5}(\frac{S^2\tau_c}{1+\omega^2\tau_c^2} + \frac{(1-S^2)\tau_e}{1+\omega^2\tau_e^2}) \tag{1.16}$$

3. Qualitative approach – comparing T_1, T_2 and hetNOE values of specific amide pairs to the average values characterising the molecule's global motion. For example, taking a residue experiencing fast motions (usually ps-ns), its T_1 is lower, T_2 is higher, while for a residue experiencing fast-intermediate exchange motions, R_{ex} (few hundred μs), T_2 is lower, while T_1, being more sensitive to higher frequency motions, is the same as compared to the average. Furthermore, the hetNOE values are an indication of the order parameter S^2. Therefore, without the need of further analysis, just obtaining the T_1, T_2 and hetNOE values can be enough to assess local motions present. Notably, fast motions at the ps-ns timescale affect all three observables, while intermediate exchange motions (R_{ex}) affect only the T_2.

Principles underlying ^{15}N-relaxation NMR experiments

The three values, T_1, T_2 and hetNOE can be retrieved by a set of ^{15}N-relaxation NMR experiments, ^{15}N-T_1, ^{15}N-T_2 and ^{15}N hetNOE.

In the case of ^{15}N-T_1 relaxation experiment, the nitrogen increase of magnetisation along the z-axis due to longitudinal relaxation is measured, after it has been purposely inverted to the -z-axis. Simply put, this is done by repeating an NMR pulse sequence which "flips" the M_z of nitrogens using a 180° pulse, with each repetition allowing for different values of the consequent evolution time, T_{relax}. After the varying evolution period, the inverted magnetisation that has been recovered due to relaxation is promoted to the xy-plane for detection, using a 90° pulse. For increased signal resolution, such an experiment is extended to a second dimension, taking into account correlating hydrogens. The corresponding intensities of the peaks present in each 2D ^1H-^{15}N spectrum recorded with varying time delays, T_{relax}, will therefore vary exponentially with time. Figure 1.33 shows the pulses and evolution of magnetisation along the transverse plane for the ^{15}N channel.

Similarly, an ^{15}N-T_2 relaxation experiment follows the decay of nitrogen magnetisation in the xy-plane due to loss of coherence caused by spin transitions (due to the local field fluctuations). Firstly, a 90° pulse is applied to allow M_{xy} ro precess on the xy-plane, under the influence of relaxation. During precession, inhomogeneities in the $\vec{B_0}$ arising across the sample as well as evolution of the CS are eliminated by *spin echoes* repeating in equal temporal increments of τ – 180° – τ pulse elements. Spin echos, however, cannot overcome spin-spin relaxation caused by DD and CSA. Ultimately, the only cause of loss of signal is

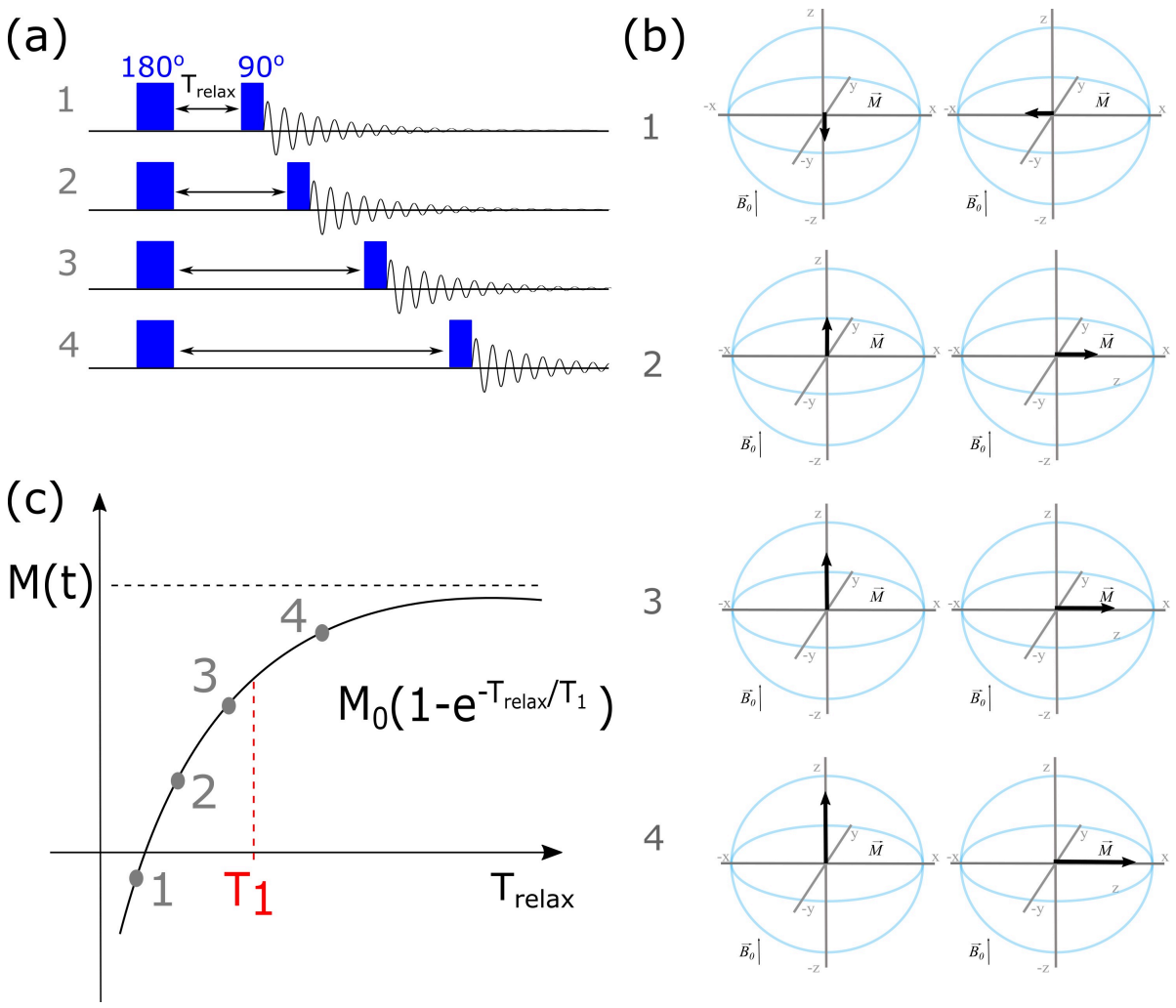


Figure 1.33 Schematic representation of the principle of ^{15}N-T_1 experiment aimed at retrieving the T_1 time of a nucleus in the molecule. Since it is an "out-and-back" experiment, and for simplicity, the hydrogen channel is not shown, only the evolution of the ^{15}N-M_z vector. (a) An 180° pulse tilts the magnetization vector to the -z axis, and an evolution T_{relax} follows of variable length (1,2,3,4), after which a 90° pulse tilts the magnetisation remaining after time T_{relax} to the xy-plane for detection. (b) Depending on the T_{relax} time, the magnetization has recovered more or less along the +z axis, resulting in a series of peak intensities. (c) The peak intensities plotted against time T_{relax} follow an exponential increase with constant T_1.

the spin-spin relaxation, R_2^0, while CS changes are averaged out during T_{relax}. By altering the evolution period and applying spin echoes at equal temporal steps, a profile of the exponential decay of $\vec{M_{xy}}$ can be plotted and T_2 is retrieved. Once again, this experiment is done in a 2D fashion, resulting in a series of 2D ^1H-^{15}N spectra, from which the peak intensities represent the magnitude of $\vec{M_{xy}}$.

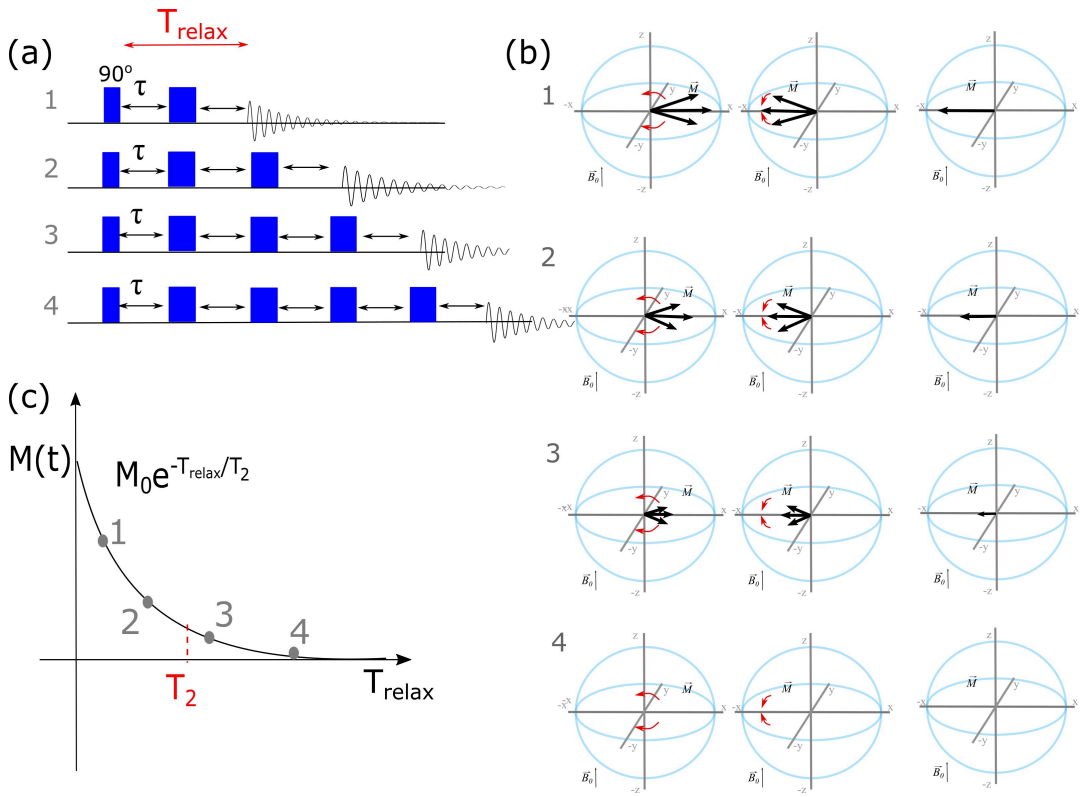


Figure 1.34 Schematic representation of the principle of ^{15}N-T_2 experiment aimed at retrieving the T_2 time of a nucleus in the molecule. Since it is an "out-and-back" experiment, the effect on the Hydrogens is not shown, only the evolution of the ^{15}N-M_z vector. (a) A 90° pulse tilts the magnetization vector to the xy-plane, where over a variable length of time T_{relax} (during which equally temporally spaced 180° pulses refocus the chemical shift), it diminishes due to spin-spin relaxation (b) Depending on the T_{relax} time, the magnetization has precessed for longer or shorter times, during which spin-spin relaxation causes loss of coherence i.e decay of the transverse magnetisation vector. The diagrams represent the magnetisation vector before (left) and after (center) the last 180° pulse, making up the total transverse relaxation time to be recorded on the right. (c) The peak intensities plotted against time T_{relax} follow an exponential decay with constant T_2.

The principle features of these experiments are shown in Figures 1.33 and 1.34. For simplicity, the magnetization transfer from and to hydrogen nuclei is not depicted.

The last experiment type of the ^{15}N-relaxation series is the ^{15}N heteronuclear NOE. In ^{15}N-hetNOE, cross-relaxation, which is the origin of the NOE effect, is evaluated between the hydrogen and the nitrogen of an amide bond pair. Cross-relaxation leads to magnetisation transfer between nuclei, eith the dependence on molecular dynamics given by equation 1.15.

Simply put, the ^{15}N hetNOE experiment evaluates the efficiency of magnetisation transfer from ^{1}H to ^{15}N through space due to the NOE effect (i.e via the cross-relaxation mechanism). Contrary to other experiments, magnetisation transfer between hydrogen and nitrogen is not induced via J-couplings but via dipolar couplings. It turns out that NOE intensity is decreased if the nuclei (or internuclear vector) in question experiences motions that are faster than the overall tumbling [115].

More specifically, the experiment proceeds by recording a spectrum where the hydrogens are saturated (saturation spectrum with peak intensities, I_{sat}) – to achieve the maximum possible polarization transfer through space – and a spectrum of the ^{15}N equilibrium polarization (reference spectrum with peak intensities, I_{ref}). The hetNOE values are then computed as hetNOE = I_{sat}/I_{ref}. In the case of a protein, the hetNOE values are typically 0.8-0.9 for a rigid region as compared to the molecule while a flexible loop or the N- and C-termini can reach hetNOE values close to zero, or even negative values. [115, 24]. The NMR pulse sequence of this experiment is beyond the scope of my thesis, but the reader is referred to the publication of Kay et al [115] for detailed descriptions of all ^{15}N-relaxation experiments.

^{15}N-relaxation studies have been a cornerstone in the effort to access protein dynamics such as IDPs and proteins which include flexible regions [111, 185, 35]. For example, such studies aided at understanding enzymatic catalysis, initiated by a disordered region exchanging conformations in the ps-ns time in adenylate kinase [102]. Similarly, in other proteins, flexible regions are crucial in cellular signalling pathways, such as triggering enzymatic activity [80].

Intermediate exchange regime induces additional contribution to transverse spin relaxation

In the case where the rate of exchange between conformations, A and B (with populations p_A, p_B, respectively), k_{ex} is comparable to the chemical shift difference, $\Delta\nu_{AB}$, the xy-component of magnetization is lost more rapidly than in the case of the fast exchange regime. The interconversion between two states (of different CS) at a time of typically a few hundreds of $\mu s - ms$ causes additional contribution to transverse spin relaxation. In other words, transverse spin relaxation appears to be influenced not only by fluctuations of DD and CSA, (termed here R_2^0), but additionally by a change in the CS of spins that have undergone conformational exchange (termed R_{ex}) during the NMR detection period [163].

Nevertheless, 180° pulses refocusing the CS can be applied to minimize the effect of R_{ex} on the signal. While nuclear spin relaxation, R_2^0, is irreversible, R_{ex} is modulated depending on the rate of 180° pulses applied during a given relaxation time (ν_{CPMG}). For nuclei experiencing conformational exchange in the intermediate exchange regime, the effective rate of decay of their NMR signal, R_{eff}, becomes:

$$R_2^{eff} = R_2^0 + R_{ex} \tag{1.17}$$

where, in the fast exchange limit ($k_{ex}/\Delta\nu_{AB} \rightarrow 0$) :

$$R_{ex} \approx p_B k_{ex} \tag{1.18}$$

in the slow exchange limit ($k_{ex}/\Delta\nu_{AB} \rightarrow \infty$ [165]):

$$R_{ex} \approx \frac{p_A p_B \Delta\nu_{AB}^2}{k_{ex}} \tag{1.19}$$

Note that R_{ex} is not only dependent on the rate of chemical exchange, k_{ex}, but also on populations, p_A and p_B, and chemical shift difference, $\Delta\nu_{AB}$.

Carl Purcell Meiboom–Gill Relaxation-dispersion, CPMG-RD, experiments to probe intermediate exchange dynamics

For line-broadened peaks where signal is still detected, *Carl Purcell Meiboom–Gill relaxation-dispersion, CPMG-RD* experiments can be used to follow the relaxation behaviour of the nuclei of interest. The main objective of CPMG-RD experiments is to deconvolute R_{ex} from R_2^0 by recording 2D spectra with varying number of 180° refocusing pulses within a fixed amount of time, T_{relax}.

Similarly to the ^{15}N-T_2 NMR pulse sequence, in an CPMG-RD experiment, the net magnetization is initially tilted along the xy-plane where it begins to precess and progressively loses coherence between spins. A CPMG pulse train i.e series of spin echos, is applied during a constant time T_{relax}, described as τ-180-τ. During τ, spins are let to dephase (due to CS evolution), while a 180° pulse refocuses the CS. A simplified description of ^{15}N-CPMG-RD experiment is depicted in Figure 1.35.

Contrary to the ^{15}N-T_2 NMR pulse sequence, the time between 180° pulses, τ, varies – a long τ during which the majority of nuclei have converted between A and B, results in complete loss of coherence ($\vec{M_{xy}} = 0$), while a short τ allows only of loss of $\vec{M_{xy}}$ due to faster interconversions i.e happening during τ_c, or spin-spin relaxation. In this way, during a fixed amount of evolution time; low CPMG frequencies corresponding to less spin-echos (and therefore greater τ delays) allow for more dephasing due to CS changes, while at high CPMG frequencies, part of the CS modulation is refocused.

By fitting the R_2^{eff} at different CPMG frequencies and different magnetic field strengths (since magnetic field strength is another parameter affecting R_2^{eff}), we can deconvolute the difference of CS $\Delta\nu_{AB}$, assuming a two-state model, as well as the exchange rate, k_{ex}, and occupancy of these two states, p_A and p_B in steady-state conditions. Notably, CPMG-RD is capable of characterising lowly-occupied states of only about 1% occupancy [66].

In a recent study published in Scientific Reports, the authors achieved in describing protein dynamics in a range of timescales from ps to ms, by employing numerous experiments,

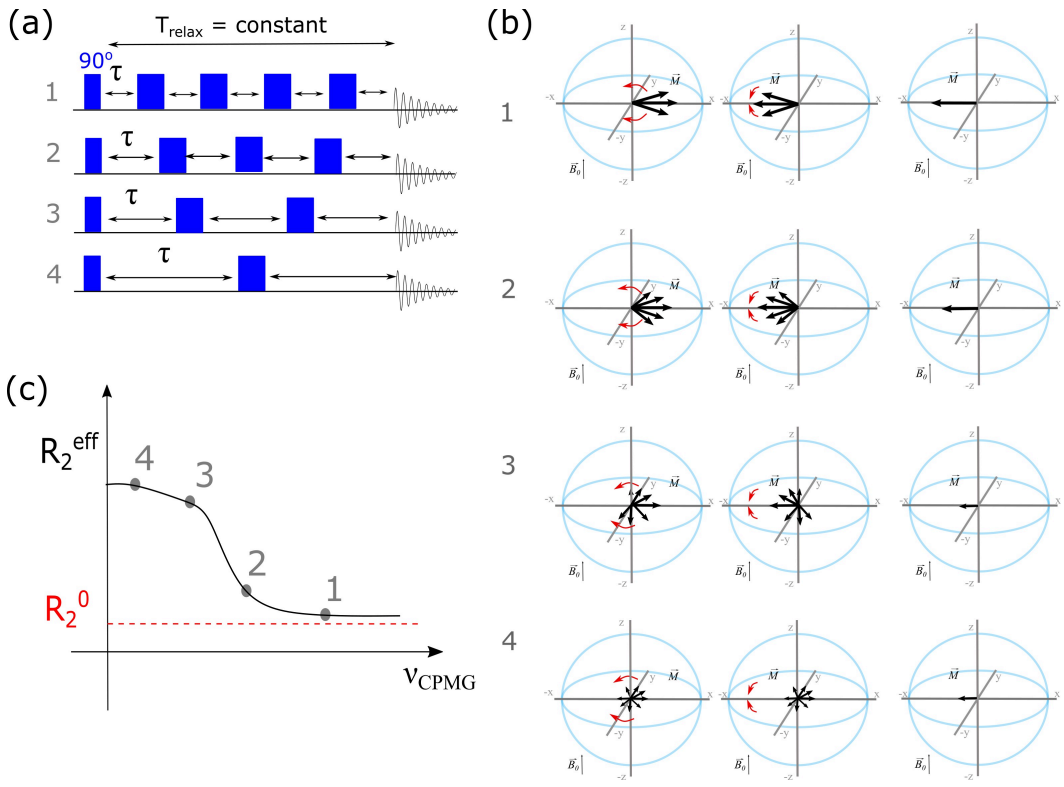


Figure 1.35 Schematic representation of the principle of ^{15}N-$CPMG-RD$ experiment aimed at retrieving the k_{ex} rate of an atom exchanging between two states at a timescale of a few hundred μs to ms. (a) A 90° pulse tilts the magnetization vector to the xy-plane, where over a variable length of time τ between 180° pulses aimed at refocusing the spin magnetization (1,2,3,4), more or less dephasing takes place, causing the signal to decrease (b) Depending on the τ time (or the frequency ν_{CPMG} of the CPMG train pulses), dephasing is allowed or suppressed while spin-spin relaxation equally impacts the magnetization in all 4 cases. In (c), the apparent relaxation rate, R_2^{eff} plotted against CPMG frequency, ν_{CPMG} for an atom exchanging in the intermediate regime. Peak intensities extracted at low ν_{CPMG} tend to be of low signal-to-noise, causing this "bump" on the graph. The plateau of the curve is R_2^0, the rate of spin-spin relaxation taking place within time T_{relax}.

including ^{15}N-CPMG-RD. Working with β-lactamase and its homologues, they identified motions occurring at the μs-ms timescale with exchange rates k_{ex} ranging from around $300s^{-1}$ to $2000s^{-1}$ depending on the homologue. In contrast, faster motions were similar in all cases. Most importantly, crystallographic b-factors were not sensitive to these changes in intermediate exchange dynamics, essentially showing the same structure in all chimeras of β-lactamase [81]. In another study using CPMG-RD experiments, the energetic landscape of several activity-related regions of the *B. subtilis* ArsC was explored, despite its complexity [239]. Taken together, CPMG-RD experiments have been increasingly used to understand

protein function, implying conformational dynamics at the few hundreds of μs-ms timescale.

Exchange spectroscopy, EXSY and real-time NMR to probe slow exchange dynamics

Lastly, for slow exchange processes (slower than a few ms), where the populations of conformers do not reshuffle during NMR data acquisition, the different conformers give rise to distinct NMR signals. There are two peaks for each nucleus in state A and B, respectively. Moreover, the linewidth of each individual peak, is mainly influenced by R_2^0, but a small "leak" of populations from state A to state B during detection translates to a small R_{ex}, as previously explained for the case of the intermediate exchange regime [122]. Hence, in the case of slow chemical exchange the equation describing the effective transverse relaxation is equation 1.17 (note that in this case R_{ex} is small).

Exchange spectroscopy, EXSY is a way to quantify kinetics of molecules on the ms-s timescale. Taking the example of a molecule occupying two states, A and B, a series of EXSY experiments of varying *mixing time* can follow their interconversion. Each EXSY experiment therefore quantifies the effective exchange which takes place during a given mixing time. In this way, the fastest process that can be quantified is the one where enough "snapshots" (i.e mixing time values) of good quality (i.e signal-to-noise ratio in our spectra) can be obtained within the lifetime of the exchange. This typically translates to chemical exchange times of a few ms-s.

In particular, the EXSY experiment makes use of the longer T_1 as compared to the T_2 of nuclear spins in proteins, which is usually a few hundreds of milliseconds to seconds. Briefly, considering a molecule changing from A to B, F1 labelling during t1 occurs while the molecule is either in the A or the B state. Consequently, a mixing step takes place whereby the magnetization vector in question is aligned along the z-axis. This way, magnetisation is only subject to longitudinal relaxation and the magnetization is said to be "stored". If during the mixing time, the molecule has switched to the B and A state, respectively, the evolution step, during t2, will result in detecting the CS of the nuclei in the B and A state in the F2 dimension.

Every residue exchanging on the slow timescale will result in 4 peaks, the two corresponding to unchanged populations of A and B and two *exchange peaks*, AB and BA. The intensity of the exchange peaks is proportional to the populations that have exchanged during the mixing time. In the end, the relative intensities between main and exchange peaks of each EXSY experiment are plotted against mixing time and fitted to a kinetic model of interconversion between the two states [122].

An interesting application of the EXSY experiment in protein NMR demonstrated its ability to quantify Tyrosine ring flip rates in Cytochrome *c* [125]. Under different temperatures and denaturant concentrations, the 180° flipping of the side-chain ring of Y97 ranged from 0.7-33s^{-1}. In particular, NMR is able to distinguish between symmetrical carbon atoms of side chains (eg ring carbons in Tyrosines) as long as they exchange in the slow regime where distinct peaks are present [140]. In conclusion, the authors suggested that while

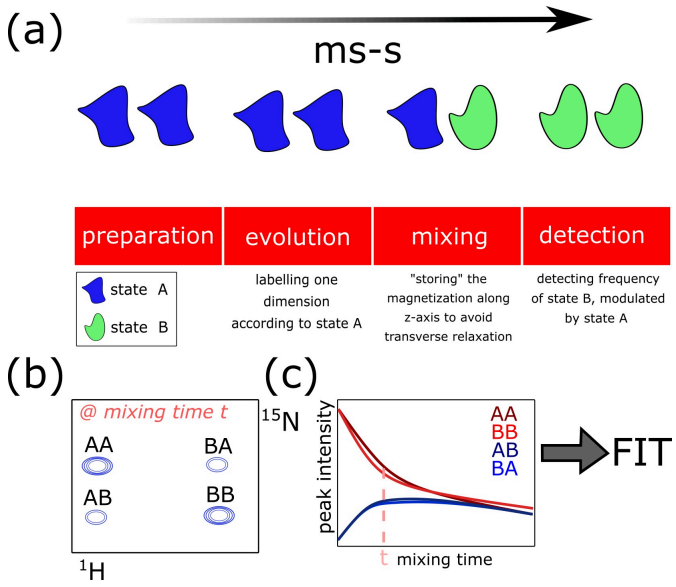


Figure 1.36 A schematic representation of the principle of the EXSY experiment. (a) For two molecules initially at state A exchanging to state B, during t1, frequency "labelling" of the nuclei at state A takes place, followed by a storing of the magnetization along the z-axis and lastly the CS evolution of the nuclei that have exchanged to state B. Considering that at the same time, molecules at state B interconvert to A, the result is a spectrum similar to (b), where 2 major peaks (AA, BB) corresponding to the CS of conformers A and B are detected, alongside minor exchange peaks, with CS originating from both conformations (AB, BA). (c) Plotting the intensities of the major and minor peaks against mixing time results in a characteristic decrease and built-up of the major and minor peaks respectively. Data need to be fitted against a kinetic model [125] to retrieve the conformational exchange rates.

slower flipping is caused by motional rigidity of the Tyrosine, faster flipping is the result of a loosening of the protein matrix due to the denaturing conditions applied.

In the case of even slower timescale motions, a series of nD correlation experiments can be recorded over time, termed *Real-Time nD NMR*. If each nD spectrum gives the characteristic "fingerprint" of a protein conformer and the population of molecules is constant, interconversion between conformers will be reported as variations in the spectra recorded. Depending on how fast the acquisition of each nD spectrum is, the time between identifiable states of the protein can range from seconds [188] to minutes [67]. The series of nD spectra recorded during a kinetic process reports on the changes in conformation during the experimental time.

In an early study of the Photosensory LOV2 domain of *Avena sativa* phototropin (AsLOV2) [93], real-time 2D NMR of backbone amides followed the build-up of its dark-adapted state after light illumination. Consistent with the expected values from optical and CD

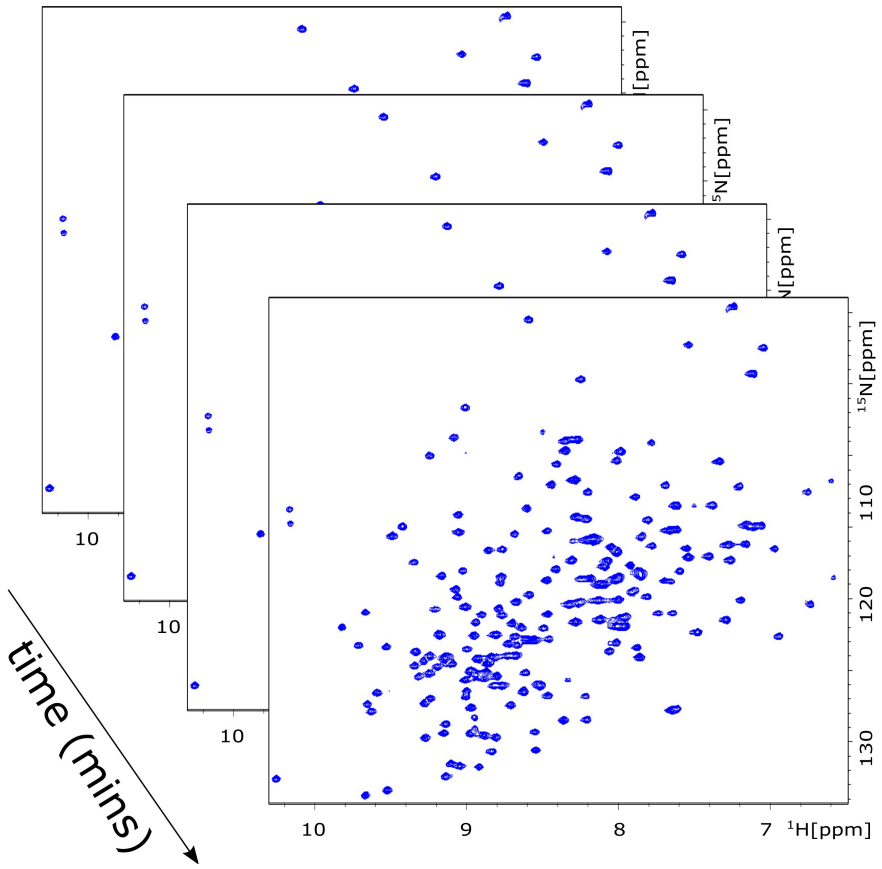


Figure 1.37 A series of ^{1}H-^{15}N 2D correlation spectra recorded over time make up for a real-time NMR experiment aimed at characterising molecular motion in the timescale of minutes.

spectroscopy, the authors estimated the recovery rate to be between 40-50s. Additionally, the energetic landscape could be plotted and the activation energy initiating the signaling pathway for which AsLOV2 is responsible was estimated to be 14.5kJ/mol. Numerous studies have employed real-time NMR for similar purposes [203, 223, 78].

Hydrogen-Deuterium Exchange experiments complementing dynamics characterisation

Hydrogen/Deuterium exchange, HDX experiments can provide information on solvent exchangeable hydrogens in the protein structure. Deuterium is not detectable in the ^{1}H channel. In this way, signal coming from exchangeable hydrogens of proteins that are eluted in D_2O will decrease over time, as more hydrogens are replaced by the "invisible" deuterium. Depending on the pH and the temperature, HDX can be studied in two regimes, EX1 and EX2. To observe HDX, the protein is initially lyophilized and then eluted in

deuterated water. In this way, real-time (RT) recordings of spectra over time (eg series of ^1H-^{15}N spectra), show a decrease in signals of residues at different rates.

A protein is studied in the EX1 regime when the observed deuteration rate, when dissolved in D_2O, k_{obs} is the same as the rate of the "opening" of the NH bond and the replacement of H with D, k_{open} [122].

In the EX2 regime, The k_{obs} of each residue is dependent on pH and it corresponds to the equilibrium between "open", k_{open} and "closed" bonds, k_{closed}, i.e the probability of a bond to "open" up and exchange with its environment, p_{open} [122].

Considering backbone amides, typically, under native-like conditions, they have a relatively stable HN bond where hydrogen atoms do not exchange with the solvent. Therefore, in the EX2 regime, k_{obs} represents the *probability* of an "open" HN bond, k_{open}. Taking into account the intrinsic exchange rate of the hydrogen involve in the bond (when the bond is not protected), k_{int}, the probability of an "opening" event increases either due to bond flexibility caused by the existence of motions or due to increase in solvent accessibility (represented in this equation as a factor, p_{open} accounting for exchange between populations of molecules occupying different conformational states).

$$k_{obs} = k_{int} p_{open} \tag{1.20}$$

HDX in the EX2 regime can be studied with various NMR experiments [237, 134, 234], depending on the timescale of exchange probed. Nevertheless, NMR experiments measuring rapid HDX, such as CLEANEX [108, 109], are beyond the scope of this thesis. In my project, we focused on measuring slower exchange rates for which real-time NMR is the most appropriate technique. A breakthrough on real-time NMR studies of HDX came with the invention of faster NMR pulse sequences such as the BEST-TROSY [67] and the SOFAST-HMQC [188] experiments, able to quantify HDX events happening in few seconds-minutes at the ensemble level.

Running real-time NMR to study HDX rates, Faustino et al [68] mapped out the per residue dynamics and solvent accessibility of two model proteins: dengue virus (DENV) capsid (C) which consists of intrinsically disordered and unfolded regions and B1 immunoglobulin binding domain of streptococcal protein G (GB1) consisting of α-helices and β-sheets. Altering the pH to the extent that folding and overall conformation are not significantly affected, they measured a series of ^1H-^{15}N 2D correlation spectra to follow the per residue backbone amide hydrogen exchange with deuterium. In particular, pH series were employed since following how exchange differs with pH gives insight on the solvent accessibility of flexible regions. In conclusion, the authors expect HDX to answer similar questions to a range of proteins from IDPs to well-folded structures.

In conclusion, real-time NMR studying HDX rates in the EX2 regime for amides probes the free energies of the opening of amide bonds ($\Delta G = -RTln(p_{open})$, where R is the gas constant and T is the temperature), or highlights regions of increased solvent accessibility

[68, 122]. When juxtaposed to the intrinsic behaviour of the amino acids investigated, HDX rates can complement other NMR experiments deciphering protein dynamics.

1.3.8 Limitations of nD solution NMR for studying proteins

Above, I explained how nD solution NMR is able to characterise protein motions occurring at different timescales. Nonetheless, the exchange processes mentioned concern steady-state phenomena, except for the case of processes studied using real-time NMR. Unfortunately, the probing of "off" equilibrium processes that occur on extremely fast timescales (<seconds) remains a challenge for NMR spectroscopy. That is because, when acquiring an NMR spectrum, the NMR pulse sequence is repeated in order to achieve high signal to noise ratio – less in 1D but significantly more in the case of nD NMR. Hence, observations of "off" equilibrium are usually too short to be captured during detection. Additionally, population-weighted averaging in the sample renders most "off" equilibrium processes undetectable by NMR.

Practically, NMR experiments can only capture intermediate states with lifetimes longer than seconds (during protein folding for example [188, 152]). In contrast, "on" equilibrium processes of conformational interconversion, at a range of timescales, are detectable by NMR with an impressive sensitivity to otherwise "invisible" populations [14, 66].

Furthermore, being an insensitive technique, sample concentration must be relatively high to acquire a decent NMR spectrum. While multiple scans can significantly increase the signal-to-noise ratio, the total number of molecules present is the main factor influencing the total signal detected [30]. Samples of a few hundred micromolar concentration can be challenging to achieve for some proteins.

1.3.9 Using NMR to study fluorescent proteins

Literature only accounts for a handful of NMR studies on fluorescent proteins. Despite the limited evidence, the benefits of complementing static crystallographic structures by NMR have been demonstrated. Unfortunately, so far only one RSFP has been studied using NMR which nevertheless inspired my PhD thesis. Due to the lack of literature regarding RSFPs, in particular, in this section, I will briefly mention NMR studies on FPs, in general, to give a historical perspective on the contribution of NMR in this field.

The first dynamical characterisation of a GFP-type protein by NMR came with the studies of Holak's group on mutants of Cyan FP (CFP) and GFP [194]. NMR-Lineshape analysis and HDX experiments suggested slow exchange dynamics taking place in the region of the β7 and β10 sheets, where the CRO ring is facing, supported by further studies in a GFP variant [193]. Particular focus was given to His148 with the authors suggesting slow exchange originating from motions on the ms-s timescale of the CRO and/or His148. Furthermore,

pH titration highlighted the pH dependent behaviour of His148 which they interpreted as being solvent exposed and flexible at lower pH.

Another NMR study employing pH titration aimed at deciphering how protonation occurs in different GFP variants [158]. Once again, β7 was suggested to be responsible for proton gating (i.e transferring protons from solvent to the CRO) in GFP mutants, given its increased flexibility and pH dependence. In particular, the authors performed NMR experiments characterising slow exchange on the ms-s timescale to decouple fast protonation/deprotonation events retrieved via Fluorescence Correlation Spectroscopy, FCS, and slow conformational changes of CRO neighbouring residues. Finally, they concluded on a two-site model explaining the appearance of two exchange regimes, demonstrating the complementarity of NMR and FCS.

Altogether, NMR has enabled the characterisation of GFPs in terms of their pH dependent dynamics as well as suggested mechanisms of CRO protonation. Therefore, not only NMR provides a dynamic dimension to the static structures of crystallography, but has been proven to be used to study environmental influences and distinguish between CRO species [95]. Furthermore, protonation events can be pinpointed to specific residues. While not included in these studies, NMR is commonly used to study pKa of titratable groups as well as distinguish between protonation states of amino acids [159, 50, 85] such as Histidines [94, 170]. Therefore, one can envisage the advantages of NMR over crystallography in studying fluorescent proteins at physiologically relevant conditions.

Furthermore, NMR studies of GFPs and model CRO compounds have reported values concerning the energetic landscape and chemical shifts of the CRO under steady-state conditions, for the different CRO species [1, 96]. Confirming those values to the ones calculated in quantum mechanical and MD simulations, these studies show how NMR can not only highlight the kinetics and energetics of CRO conformations in the steady-state, but can also clearly distinguish between CRO species in an non-invasive manner (i.e without the use of excitation light or X-rays).

So far in the RSFP field, only Dronpa has been investigated using NMR [147, 146]. Hideaki Mizuno is one of the founders of Dronpa, the first engineered green negative RSFP [11]. In two publications dating 2008 and 2010, Mizuno and co-workers, described their attempt to characterise the dynamics of Dronpa in its dark-adapted “on”- and light-induced “off”-state. Namely, by using an optical fiber inserted in the plunger of the NMR tube, they could illuminate the sample of Dronpa while inside the NMR spectrometer. Their work revealed light-induced flexibility of the Dronpa barrel, on the side where the CRO ring faces in its *trans* conformation. Moreover, they reported CS of the carbons of the Tyr63 of Dronpa’s CRO in both the “on”- and the “off”-state. Since its publication, this work has been cited in numerous papers assessing how RSFPs structure and dynamics affect their photophysical behaviour [26, 3, 204].

In spite of its scientific popularity and novelty, Mizuno’s work on Dronpa lacks some crucial information. Commonly encountered issue in NMR studies of proteins, conformational exchange on the ms timescale results in line-broadening (LB), sometimes to the point of

total loss of signal. Consequently, while LB could indicate intermediate exchange processes taking place, this claim would benefit from support from various perspectives. Otherwise, lack of signal could be blamed to unfinished assignment or overlap of signals in the ^{1}H-^{15}N spectrum. Dynamics that are supported by ^{15}N-relaxation experiments are also faulty since there is no peak to retrieve relaxation information from.

Nonetheless, NMR as a key player in understanding RSFPs' dynamics was introduced by Mizuno's work. Due to this breakthrough, my project was not only conceptually conceived but also technically facilitated.

1.3.10 Thesis objectives: Studying rsFolder and mutants by nD NMR

This project was inspired mainly from the work of Mizuno et al [147, 146]. Nevertheless, the investigation of rsFolder and mutants using NMR went beyond the aforementioned Dronpa studies. The objective of my PhD work was to not only characterise the dynamics of rsFolder but also put together a consistent NMR toolbox for studying RSFPs. While various questions arose during the PhD progress which lead the work into different paths, the main objectives were two-fold. On the one hand, it was essential to establish the necessary protocols for using NMR to study RSFPs. On the other hand, the ability of NMR to contribute to the RSFP engineering field was to be shown for the example of rsFolder. Therefore, the objectives were the following:

1. Establishing a reliable in-situ illumination device to be used with NMR.

2. Create a series of NMR experiments suitable for studying RSFPs and report on rsFolder dynamics in native-like conditions.

3. Identify NMR experiments for the study of different CRO species – decipher species variability in rsFolder.

4. Show how NMR observables can be associated to photophysics of RSFPs, using the example of rsFolder and mutants in different pH and light conditions.

Hence, my thesis is constructed based on these four pillars. The results of my work are presented alongside a discussion which includes interpretation, significance and contribution to the field as well as potential limitations of the techniques used.

Chapter Two

Materials & Methods

2.1 rsFolder sample preparation

2.1.1 Protein purification of Uniformly labelled rsFolder samples – with and without His-tag

Protein samples of rsFolder and mutants were mutated, expressed in bacteria, and purified by Isabel Ayala and Karine Giandoreggio-Barranco from the Biomolecular NMR group at IBS.

From our publication in 2019 [42];

The full-length rsFolder complementary DNA was cloned in the ampicillin- resistant expression plasmid pET15b (Addgene, Teddington, UK), and constructs were designed to bear a six-residue N-terminal His-tag linked to the protein coding region via a thrombin cleavage sequence. For initial NMR experiments detecting aromatic side-chain protons, the His-tag was cleaved using a thrombin protease. Otherwise, all experiments were recorded without cleavage of the His-tag because those samples showed a higher time stability. Escherichia coli BL21(DE3) cells were transformed with the plasmid described above for protein expression. An isotopically en- riched minimal medium, containing 15NH4Cl (1 g/L) and C-glucose (2 g/L), was used to grow the cells until the cultures reached an OD at 600 nm between 0.6 and 0.8, at which point protein expression was induced by adding isopropyl b-D-1-thiogalactopyranoside. For uniformly $^{15}N/^{13}C$ labeled samples, 1 mM isopropyl b-D-1-thiogalactopyranoside was added and induction was performed overnight at 20°C. The cells were harvested by centrifugation and subsequently lysed by sonication in lysis buffer (50 mM HEPES, 100 mM NaCl (pH 7.5)). The lysate was centrifuged at 4°C for 30 min at 46000g. The protein present in the supernatant was purified using a Ni-NTA column (Qiagen, Hilden, Germany). The protein was eluted by increasing the concentration of imidazole. The resulting pro- tein was further purified using size-exclusion chromatography with a Superdex 75 column (GE Healthcare, Chicago, IL) equilibrated with the NMR buffer (50 mM HEPES (pH 7.5))

pH	Phosphate buffer (ml)	Citric Acid (ml)
4.2	207	293
4.5	225	275
5.0	258	242
5.5	280	220
6.0	315	185
6.5	350	150
7.0	412	88
7.5	460	40
8	486	14

Table 2.1 Volume of Phosphate buffer and Citric acid required to prepare the McIlvaine pH series

2.1.2 rsFolder pH series preparation

For the purpose of pH titration studies, the McIlvaine buffer was chosen due to its wide range of buffering (pH=2-8). The various pH values were achieved by dialyzing the protein sample (Dialysis tubes purchased from, Gene Bio-Application LTD. 8kDa cut-off Midi tubes Cat. number: D022) against 500ml of McIlvaine buffer of the desirable pH, overnight at 4°C. Dialysis buffers were prepared by mixing 50mM Phosphate buffer (Disodium Phosphate) and 25mM Citric acid at different ratios (Table 2.1).

The published values regarding McIlvaine buffers [141] of different pH correspond to a mix of 200mM of Phosphate buffer and 100mM of Citric acid. However, phosphate buffer affects the overall signal-to-noise ratio in NMR. HCN-cryoprobes are used to overcome thermal losses due to resistances in the RF coil and in the sample (due to temperature, salt concentration and presence of conductive ions) [124]. Phosphate ions significantly increase the resistance of the sample, requiring longer p1 values, corresponding to the 90° RF pulses for Hydrogen nuclei. Thus, overall the sensitivity is reduced and from one point onwards, the cryoprobes cannot overcome the thermal losses originating from the sample [69]. Therefore, the concentration of the Phosphate buffer was minimized to 50mM, for our NMR studies.

2.2 In-situ LASER illumination in NMR spectrometer

2.2.1 LASER box

For alignment purposes, the fiber coupler consists of 6 small screws which according to how far screwed in or out they are, would tilt the coupler at a different angle. Due to the

frequent transporting of the LASER box, alignment made this way was often necessary, while the output power level would be evaluated using an optical powermeter (reference number: PM100D, Thorlabs).

2.2.2 Optical fiber: use and maintenance

For guiding the LASER light into the 5mm NMR Shigemi sample tube, an optical Multi-mode fiber was used with specifications: 0.22 NA, 105 μm core diameter, 2 mm diameter protective cover and 5 m length (reference number: FG105UCA, Thorlabs). The fiber was purchased from Thorlabs alongside Ceramic scribes used for fiber cutting (reference number: CSW12-5, Thorlabs).

When dimmed necessary due to losses of light by the end of the fiber, the ceramic scribes were used to cleave the bare end of the fiber and overcome surface irregularities. More specifically, the scribes were placed at 45degree angle perpendicular to the bare end of the optical fiber (immobilised on a clean table surface using tape) and pressed gently until approximately 30$\%$ thickness is cut. After such a cut, the bare end is pulled apart by hand, in order to create a "clean" cut on the fiber.

2.2.3 Modifications in NMR setup

Overall, the addition of the optical fiber into the plunger resulted in an increase of the weight of the sample. Generally, specific spinners used in NMR ensure the smooth injection of the sample via air flows, thus they account for an average weight of a conventional sample, in order to guide it correctly inside the probe. Hence, the added weight originating from the fiber was counter-balanced with an additional half-spinner i.e a spinner cut in half.

Moreover, the effect of the LASER in terms of temperature and electromagnetic radiation changes had to be considered. In particular, when preparing for the experiment, calibration (i.e shimming of the sample to counteract magnetic field inhomogeneities) was performed once the sample reached equilibrium after continuous LASER illumination (identified as a steady NMR *lock* signal). Nonetheless, for alternating illumination, recorded spectra would show signs of "bad" shimming over time, due to the temperature fluctuations created by the high-power LASERs in an inhomogeneous fashion.

Lastly, to counteract the inhomogeneity of the illumination – higher on top of sample, dispersed across the sample height – I attempted two different modifications. Firstly, a reflective material was sprayed on the outer surface of the Shigemi tube, in order to "keep" most of the illumination within the sample. Secondly, in a collaboration with an NMR group from BMRZ-Goethe University, Frankfurt (group of Dr. Harald Schwalbe), the plunger of the Shigemi was carved into a pencil tip shape. What this achieved was the insertion of the plunger mid-height of the sample, while the light was radially emitted from the sides of the

plunger tip. In both cases, shimming remained possible and LASER light distribution was improved.

2.2.4 Assessing the in-situ illumination setup using rsFolder

In order to evaluate the capabilities of the in-situ illumination setup, a set of experiments was performed investigating the coupling of the LASER to the NMR experiment, the speed of switching of the sample (and how, in turn, this is reflected in the spectra) as well as overall considerations of sample stability under different LASER powers and durations of illumination.

Initially, series of real-time ^{1}H spectra were recorded over time where the sample, starting from an "on"-state equilibrium, would be switched to the "off"-state by turning on the 488nm during the experiment. Switching back to the "on"-state was similarly recorded by turning on the 405nm LASER (after turning off the 488nm LASER) during the experiment. This series of 1D spectra had a two-fold purpose: firstly, to see whether the characteristic spectra are different for either state and secondly, to measure the ensemble switching rates of the sample when found in typical NMR experiment conditions (inhomogeneous illumination, high concentration, 40°C).

Furthermore, after acquiring the uniformly labelled protein (labelled with ^{13}C and ^{15}N), 2D spectra were recorded to assess the stability of the sample over long periods of illumination. As already mentioned in bibliography, depending on the illumination scheme, GFPs tend to photobleach over time. Therefore, recording 2D spectra over time could reveal differences related to potential photobleaching processes.

Lastly, as is the case for all NMR studies, 2D spectra were recorded before each experiment, pre- and post- illumination as well as by the end of long experiments, to confirm the integrity of the sample.

2.3 rsFolder backbone and side chain assignment – "on"- and "off"- state

2.3.1 Backbone amide assignment

Backbone "fingerprints" of rsFolder "on"- and "off"-state – the principle of BEST and TROSY

Typically 3D triple resonance experiments aim at associating backbone carbon frequencies with frequencies of amide pairs belonging to the same residue. Assigning of the backbone amides of rsFolder in the "on"- and in the "off"-state therefore translates to appointing a

residue number and type to each cross-peak found in the so-called backbone "fingerprint" spectra of rsFolder in the "on"- and "off"-state, respectively. More specifically, these are 2D amide spectra recorded using the *Band-selective Excitation Short-Transient Transverse-Relaxation Optimized SpectroscopY* (BEST-TROSY) pulse sequence.

The BEST-TROSY is an impressively fast experiment achieving high sensitivity and resolution in protein NMR at high magnetic field strengths [67]. More specifically, it combines the BEST element [190], resulting in short d_1, as well as the optimised transverse relaxation properties of the TROSY NMR pulse sequence [171]. The BEST pulse element ensures reduced d_1 recycle delays (or inter-scan delays) due to shorter longitudinal relaxation times. The latter is achieved by selective excitation of only the nuclei within the spectral window. Overall, the BEST element allows for more scans to be recorded within a given time-frame, thus increasing the signal-to-noise ratio and overall rendering shorter experimental times possible.

The TROSY element is particularly useful for relatively big molecules (like proteins) at high magnetic field strengths. In the case of big molecular size, the T_2 relaxation time gets shorter with longer tumbling times, leading to line broadening of peaks, while higher magnetic fields result in the CSA becoming comparable to the DD contribution to relaxation. Both DD and CSA contribute to T_2 relaxation. However, they are cross-correlated relaxation mechanisms which means that their contribution in either ^1H or ^{15}N relaxation either adds up (more line broadening) or is subtracted (less line broadening). Overall, in an HSQC without decoupling one would expect a total of 4 peaks with different widths along the ^1H and the ^{15}N dimension for each ^1H-^{15}N pair i.e a multiplet, as a result of the relative contributions from DD and CSA.

As mentioned in the Introduction, in a conventional ^1H -^{15}N HSQC, decoupling results in one cross-peak, originating from the averaging of all four peaks. In contrast, the TROSY experiment picks the one peak from the multiplet of peaks where the contributions of DD and CSA to relaxation are smallest. In this way, the TROSY allows for sharper peaks, shifted approximately by 45Hz (half the J-coupling strength of an amide pair) in each dimension as compared to a ^1H -^{15}N HSQC experiment (Figure 2.1).

Conventional 3D triple resonance experiments

Assignment of the thermally stable "on"- and metastable "off"-state of rsFolder was performed initially for the backbone amides. Employing conventional BEST-TROSY triple resonance 3D experiments (HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, HN(COCA)CB), the chemical shift frequencies of Carbonyls, C_α, C_β and the respective amide cross-correlation peaks were retrieved [205]. Briefly, for a triple resonance experiment, the RF pulse sequence ensures magnetisation transfer via J-couplings to the nuclei of interest, starting from the amide hydrogens, to the amide nitrogens followed by one or more magnetisation transfer steps to the backbone carbons of interest, finishing with the transferring back to the hydrogens for detection. The principle of magnetization transferring via J-coupling in

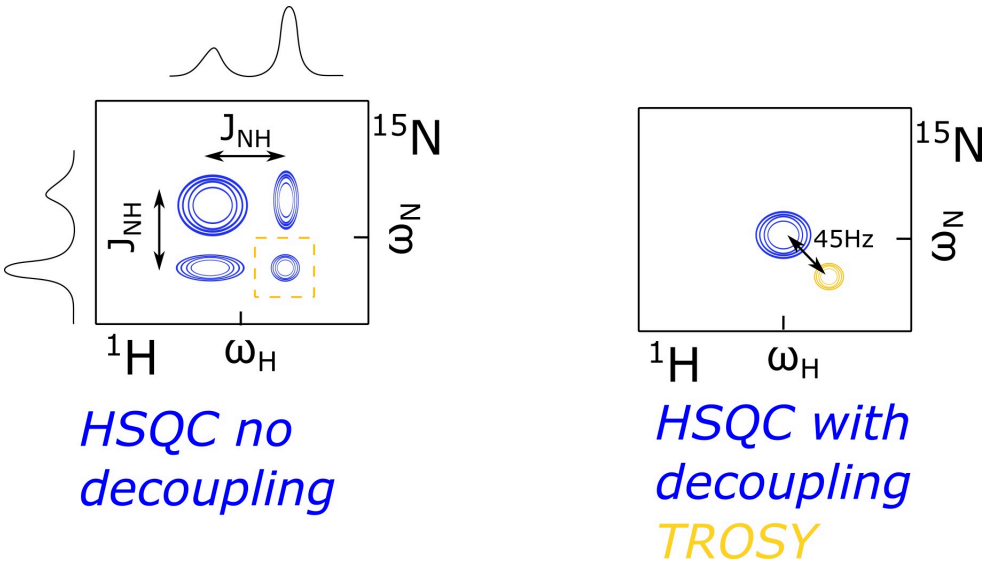


Figure 2.1 Schematic comparison of a ^{1}H -^{15}N HSQC and a ^{1}H -^{15}N TROSY spectrum for one peak. Without decoupling, the 4 peaks of the multiplet resulting from the HSQC experiment are visible (on the left). Each one has a different width along the ^{1}H and ^{15}N dimension, due to line-broadening caused by cross-correlation of DD and CSA (adding up or cancelling out). HSQC with decoupling results in a broad peak while TROSY picks out the sharper peak of the multiplet, shifted 45Hz in both dimensions as compared to a conventional HSQC peak.(on the right)

an HNCA is demonstrated in Figure 2.2 alongside the resulting 3D spectra expected (for one residue).

The spectra were recorded on the 850MHz spectrometer at 40°C in samples of 200μM. Spectra were processed using TopSpin v3.5 (Bruker BioSpin).

With the spectra at hand, sequential assignment was performed using CCPNMR v2.0 software [220]. The assignment was partially confirmed by comparing the spectra of uniformly labelled samples with ones of samples where either the His-tag was cleaved, or specific labelling was used. In addition, Tyrosine-unlabelled and Histidine-unlabelled samples confirmed the assignment.

Using the LASER-Driven EXSY for the assignment of the "off"-state spectrum

As described in the Introduction, the EXSY experiment records how the frequency of a given nucleus is modulated upon conformational exchange in the ms-s timescale. However, in the case of assigning the "off"-state backbone amide spectra, the EXSY pulse sequence was conveniently used in combination with the new "LASER channel" integrated in the

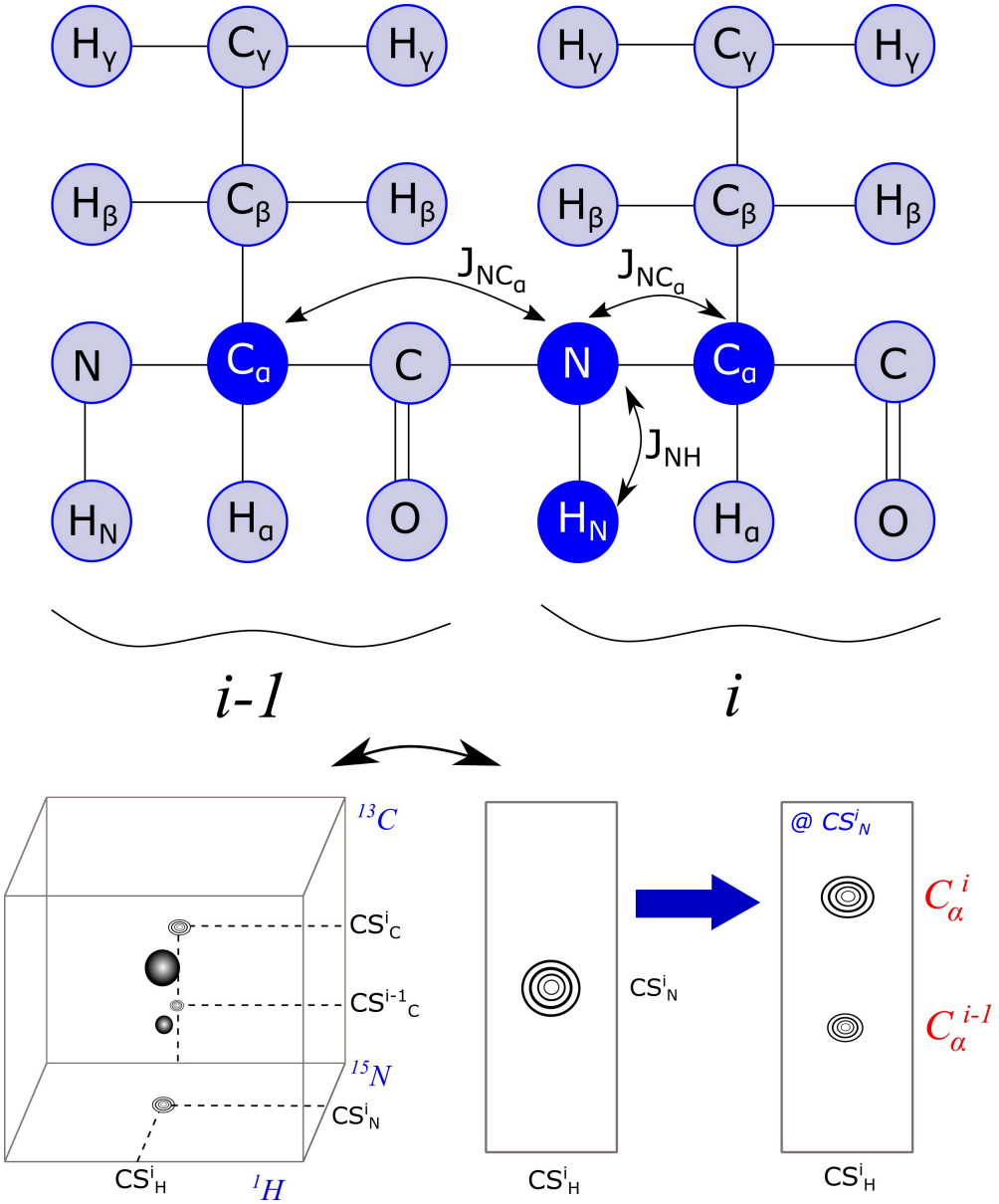


Figure 2.2 Schematic representation of the principle of 3D triple-resonance HNCA experiment, transferring magnetization via J-couplings from ^{1}H to ^{15}N, to the nearby $^{13}C_{\alpha}$ of the same residue (i) and the previous residue in the sequence ($i-1$). On the bottom, the corresponding 3D spectrum of the 2 peaks is shown with coordinates of (H^{i}, N^{i}, C_{α}^{i} and H^{i}, N^{i}, C_{α}^{i-1})

NMR pulse sequence to "transfer" the completed assignment of peaks from the "on"-state spectra to the "off"-state spectra. In the case of no CS change of a residue between the two states, there are no exchange peaks (no "off"-diagonal peaks). In contrast, when the CS of a peak has changed upon 488nm light, the result is an exchange peak inheriting the spectral coordinates of: in one dimension, the "on"-state CS, and in the second dimension,

the “off”-state CS.

Briefly, during $d1$, the 488nm LASER is on at full power, ensuring that the sample is in the “off”-state. The following $t1$ evolution time allows for the CS editing of the frequency of the amides when rsFolder is in the “off”-state. During storing of magnetization along the z-axis, time $d8$, the 405nm LASER is switched on at maximum intensity. Therefore, during the mixing step, a part of the protein sample photoswitches to the “on”-state. Lastly, during $t2$ the CS after frequency modulation is recorded. As long as the characteristic spectrum of the “on”-state differs from the “off”-state, the resulting spectrum includes exchange peaks as molecules undergo conformational exchange between “off” and “on” during $d8$.

Nevertheless, relying on 2D spectra which include the original cross-peaks and the exchange peaks would result in spectral crowding, hence the experiment was extended to a third dimension. As a consequence, two experiments needed to be recorded for the assignment of the “off”-state, a ^{1}H-LD-EXSY and a ^{15}N-LD-EXSY, giving the “off”-state CS frequencies of the amide hydrogens and nitrogens respectively.

Altogether, the LD-EXSY experiments performed shortened the experimental time needed for assignment as compared to using the standard 3D-triple resonance assignment, from approximately a week, to 24 hours. Assignment of the “off”-state was concluded using the LD-EXSY experiments.

2.3.2 Side chain assignment

Methyl side chains

Methyl side chains were assigned using a combination of various experiments. The characteristic 2D ^{1}H-^{13}C spectra used were acquired using the SOFAST-HMQC (band-Selective Optimized Flip-Angle Short-Transient heteronuclear multiple quantum coherence) pulse sequence, allowing short experimental times [187, 189]. SOFAST-HMQC makes use of the band-selective excitation which significantly decreases the longitudinal relaxation time, thus the length of interscan delays, as well as unwanted scalar coupling. Moreover, using multiple quantum coherence, the necessary spin transitions can take place in fewer steps than in an HSQC. Lastly, SOFAST-HMQC makes use of the Ernst-excitation angle, which allows for a greater number of available spins in steady-state prior to RF pulse excitation. Overall, SOFAST-HMQC results in high sensitivity and resolution [187].

Starting with the 3D-triple resonance experiments, correlating amide frequencies with Carbons and Hydrogens of methyl side chains was possible using the (H)C(CCO)NH-TOCSY and H(CCCO)NH-TOCSY experiments [149]. Both these experiments aim at transferring magnetisation from the methyl side chains to the backbone amides of the preceding residue via J-coupling. Unfortunately, the many necessary magnetisation transfer steps render this experiment relatively insensitive, leading to longer experimental times for increasing the signal-to-noise ratio. Nonetheless, most resonant frequencies for the methyl side chain

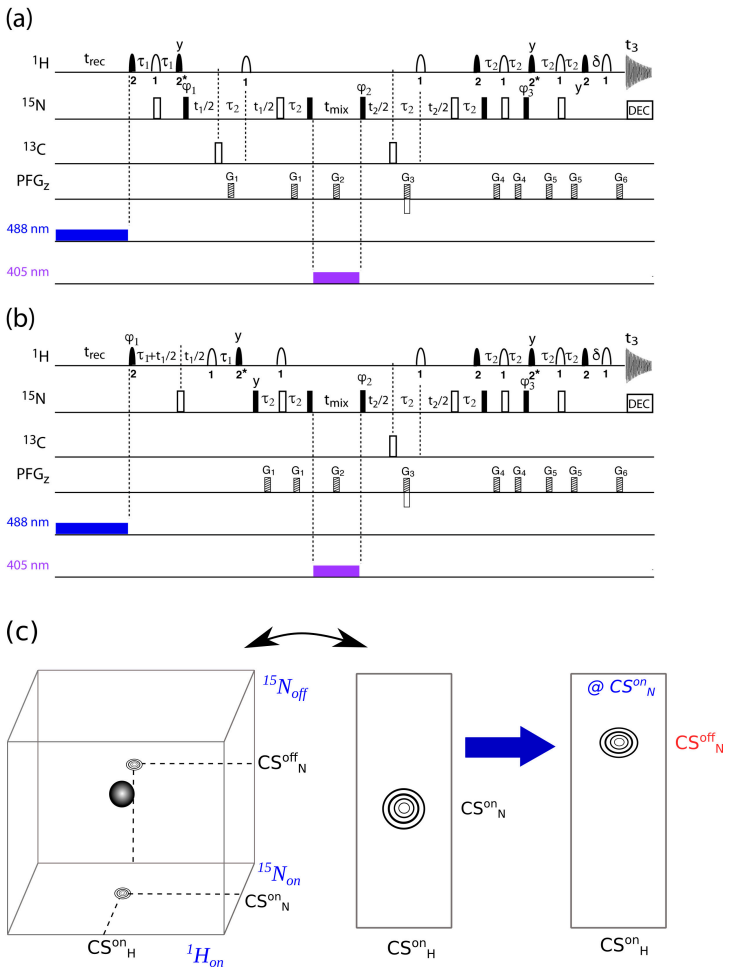


Figure 2.3 3D LD-EXSY experiment: pulse sequences drawing of (a) 15N-edited LD-EXSY and (b) 1H-edited LD-EXSY. Filled and open pulse symbols indicate 90° and 180° rf pulses. Unless indicated, all pulses are applied with phase x. All selective ^{1}H-pulses are centered at 8.5 ppm, covering a bandwidth of 4.0 ppm, with the following shapes: (1) REBURP, and (2) E-BURP2. A star indicates a flip back pulse obtained by time and phase inversion. Delays are set to: τ_1=2.4 ms, τ_2=2.7 ms, t_{rec} = 2 s, and t_{mix}= 600 ms. The delay δ corresponds to the length of G6 gradient. These settings do not take into account the durations of the 180° rf pulses applied during the transfer delays that need to be subtracted. Pulsed field gradients, G1–G8 are applied along the z-axis (PFGz) with durations of 200 μs and field strengths ranging from 5–50 G/cm. The relative strength of G3 and G6 are given by the gyromagnetic ratios G3/G6=$\frac{\gamma_H}{\gamma_N}$. The phase cycling is: ϕ_1 2 (x,-x), ϕ_2 = x, x, -x, -x, ϕ_3=y, and the receiver phase ϕ_{rec}= -x, x, x, -x. Quadrature detection in t1 is obtained by time-proportional phase incrementation of ϕ_1 according to TPPI-States. For quadrature detection in t2, echo-antiecho data are recorded by inverting the sign of gradient G3 and phase ϕ_3.

carbons and hydrogens were collected using these experiments.

3D-triple resonance experiments were complimented by aminoacid-type-edited 2D spectra, specifically aminoacid-edited-SOFAST-HMQC experiments. As reported [143], a frequency-selective filter discriminates between aminoacid types with or without an adjacent carbon (C_{next}) of a given CS range. The signals are then projected as positive or negative, depending on the aminoacid type – methyl side chain with a C_{next} outside or inside the CS range of interest respectively. Overall, aminoacid-edited-SOFAST-HMQC experiments distinguish between 3 categories: (a) Alanines, (b) Threonines and (c) Valines and Isoleucines.

Furthermore, selective isotopic labelling was used to further confirm or extend the assignment of methyl side chains. More specifically, samples were either uniformly labelled, labelled only on one Isoleucine Carbon-δ and Valine Carbon-γ (Pro-S) [120], or unlabeled only on Leucine and Tyrosine [180].

Lastly, an experiment correlating the methyl carbon to the adjacent carbon was performed, namely an HCC-COSY. By exciting the methyl hydrogen, magnetisation is transferred to its J-coupled carbons. In this way, a 3D spectrum is recorded where the resonant frequency of the methyl hydrogen is correlated to its methyl carbon and the consecutive (adjacent) carbon. By knowing the latter's frequency from the 3D triple resonance experiments performed (either for backbone or for side-chain assignment), one can conclude about the methyl residue at hand. Consequently, the HCC-COSY was used to confirm and complete methyl side chain assignments.

Aromatic side chains and the chromophore

Aromatic side chains and the chromophore were assigned using a range of NMR experiments. The 2D spectra of interest to assign were retrieved using ^{1}H-^{13}C Aromatic HSQC experiments – a constant time (CT) and a non-CT version.

From the one hand, the CT version of a ^{1}H-^{13}C HSQC refocuses the J-coupling of homonuclear ^{13}C-^{13}C, in this way resulting in sharper peaks and thus higher resolution and sensitivity. Moreover, depending on the number of nearby carbons (i.e J-coupled carbons) the residue peaks appear as negative (even number of coupled carbons) or positive (odd number of coupled carbons).

From the other hand, the non-CT Arom HSQC is less resolved but manages to record low-intensity signals.

The assignment of the 2D Arom-HSQC spectra was concluded, similarly to the methyl side chain assignment using information from the (H)C(CCO)NH-TOCSY and H(CCCO)NH-TOCSY experiments [149]. Furthermore, BEST and SOFAST type experiments specific to Tyrosines and Histidines side chains allowed for the unambiguous assignment of the 7 Tyrosines and 11 Histidines present in rsFolder [43]. The latter set of experiments

outperforms others targeting the aromatics region [62, 227] not only by making use of longitudinal relaxation enhancement elements but also counteracting loss of signal due to excitation of unwanted nuclei.

Chromophore assingments were possible after assigning the Tyrosine residues. The CRO is a Tyrosine-type residue, which, nevertheless, is characterized by unique CS, reported already in literature for other GFP proteins. Therefore, the rsFolder CRO assignment was further confirmed after comparing the CS retrieved to the already published ones for GFP [158, 95] and Dronpa [146].

2.3.3 Structural information from C_α and C_β frequencies

Once the assignment of rsFolder "on"- and "off"-state was concluded, the frequencies of C_α and C_β were used to predict secondary structural elements using the TALOS-N software [197].

CS originating from C_α and C_β of residues are sensitive to the local environment. Consequently, the distinct dihedral angles between C_α and C_β leading to an alpha helix, a beta sheet or a flexible loop give rise to distinct CS. TALOS-N takes into account a database where values of C_α and C_β CS for different proteins are archived in order to predict with high accuracy secondary structural elements using NMR.

2.4 NMR experiments for protein dynamics characterization

In this section, I include some experimental details concerning the NMR experiments performed to study the dynamics of rsFolder in the "on"- and "off"-state. Unless stated otherwise, the spectra were processed in TopSpin v3.5 and the analytical tools used to retrieve information were mainly CCPNMR v2.0 in combination with customised Python scripts for fitting data into mathematical models.

2.4.1 Fast conformational exchange

In order to study possible ps-ns motions in rsFolder, ^{15}N-relaxation experiments were recorded in the 700MHz spectrometer at 40°C for both the "on"- and the "off"-state. Standard NMR pulse sequences were used for recording the T_1 and T_2 with relaxation delays (during t1) ranging from 0-1.8s. ^{15}N heteronuclear NOE experiments were performed in an interleaved manner (saturation and reference spectrum) with a $d1$ delay of 5s and a saturation delay of 3s.

Out of the 200 non-proline peaks assigned, 181 and 162 were resolved sufficiently enough to extract T_1, T_2 and hetNOE values in the "on"- and "off"-state, respectively. The average correlation time was calculated from the average T_1 and T_2 values using the formula:

$$\tau_c = \frac{\sqrt{\frac{6T_1}{T_2} - 7}}{4\pi\nu_N} \tag{2.1}$$

2.4.2 Intermediate conformational exchange

For studying conformational exchange processes in the range of few hundred μs to ms, ^{15}N backbone as well as ^{13}C methyl and aromatics CPMG-RD experiments were run at two magnetic field strengths (700MHz and 850MHz). The temperature and pH of the sample was chosen according to best signal-to-noise ratio for residues where line broadening was observed. The CPMG frequencies ranged from 50-1000Hz and the total relaxation delay was set to 40ms. Data were fitted for the ^{15}N-CPMG-RD datasets using the *n-trosy* experiment from the Chemex algorithm. ^{13}C methyl and aromatics CPMG-RD had low signal-to-noise ratio and were therefore not analysed further.

2.4.3 Slow conformational exchange

EXSY experiments for quantifying ring flip motions

In order to quantify ring-flips of the phenol ring of the *cis* chromophore, a series of EXSY experiments with varying mixing times focusing on the aromatic region of the ^1H-^{13}C 2D HSQC spectrum (Arom-EXSY) were performed using the 850MHz spectrometer. Arom-EXSY experiments were recorded on rsFolder and mutants at different temperatures and pH values. The resulting peak intensities of the diagonal and exchange peaks of the H_δ-C_δ pair of the phenol ring of the chromophore were plotted against mixing time and fitted using a kinetic model accounting for the interconversion between the two states (180° flips of the phenol ring), to retrieve the rate of exchange $k_{ex} = 2 * k_{rf}$, where k_{rf} is the flipping rate of the phenol ring, like so [125]:

$$\ln(-2\Gamma + 1) = -2k_{ex}\tau_m \tag{2.2}$$

Where $\Gamma = \frac{I_c}{I_c + Id}$

Real-time 2D NMR experiments to retrieve thermal relaxation and hydrogen deuterium exchange rates

To extract thermal relaxation and HDX rates of rsFolder and mutants, real-time NMR was performed by acquiring a series of 2D ^{1}H-^{15}N BEST-TROSY experiments over time at 850MHz magnetic field strength. Measuring thermal relaxation rates required to first illuminate the sample using the 488nm LASER until the 2D ^{1}H-^{15}N BEST-TROSY spectrum retrieved confirmed that the sample has switched to the "off"-state, after which the LASER was turned off and acquisition of a series of ^{1}H-^{15}N BEST-TROSY experiments, each lasting 2.53mins, started. On the other hand, to perform HDX experiments, the sample was first lyophilised and then diluted in D_2O and inserted in the magnet either from the dark or after illumination using the 488nm LASER, to record the HDX rates of the "on"- and "off"-state, respectively. More specifically, in order to minimize the time between the deuteration of the sample and the acquisition, a Shigemi tube of an rsFolder sample of similar volume, concentration and buffer, was used to calibrate the spectrometer, i.e shimming, tuning, matching, pulse calibration. This way, the "dead" time before acquisition was minimized to 5 mins. In an effort to capture the differences between "on"- and "off"-state HDX rates, one more series of 2D ^{1}H-^{15}N BEST-TROSY spectra was recorded, starting from the sample in the "on"-state and after 30mins, the 488nm LASER was turned on to switch the sample to the "off"-state.

On average, such "kinetics" experiments were recorded for approximately 14-20 hours, depending on the sample. Moreover, thermal relaxation experiments were performed for different temperatures and pH values. Consequently, the data was fitted to exponential fits in order to extract the value for the thermal relaxation and HDX rates.

Performing thermal relaxation experiments at different temperatures also allowed for extraction of energetic barriers for ground-state photoisomerization of the CRO between *cis* and *trans* conformation, after fitting the data in an Arrhenius plot.

$$\ln k = -\frac{E_a}{RT} + lnA \tag{2.3}$$

Where k is the thermal relaxation rate, E_a is the activation energy of photoisomerization, R is the gas constant, T is the temperature and A is the Arrhenius constant.

In the case of HDX rates, protection factors were extracted after measuring the titration of residues' peaks at different temperatures, to take into account the effect of the intrinsic hydrogen exchange of each residue on the observed HDX rate. Using the equation below, the exchange-competent populations could be estimated:

$$k_{obs} = k_{int} p_{open} \tag{2.4}$$

2.5 Optical characterization of rsFolder and mutants

2.5.1 Absorption and Emission spectra

Using optical spectrophotometers (absorption and emission) I was able to retrieve the characteristic absorption and emission spectra of rsFolder and mutants at different pH values. The sample used was placed inside a quartz cuvette.

2.5.2 Extinction coefficient, fluorescence Quantum Yield and Brightness

Using standard protocols, the extinction coefficient, ϵ, fluorescence quantum yield, Φ, and consequently the resulting brightness of rsFolder and mutants were measured at pH=7.5, under room temperature conditions.

The extinction coefficient at a specific wavelength (488nm) was retrieved by measuring the absorption spectra (i.e optical density (OD) against wavelength) of the protein folded and unfolded. More specifically, rsFolder and mutants were gradually unfolded using small quantities of concentrated NaOH, until a point were the maximum amount of "free" chromophores in solution is present, before the protein starts precipitating. Knowing the ϵ of a free GFP chromophore (44,000cm^-1) [225], absorbing at 447-448nm, the ϵ of the cis-anionic form of the chromophore protected by the protein barrel at pH=7.5 can be back-calculated, taking into account the total amount of molecules present. The spectrophotometer used recorded the absorption spectra of the sample placed in a cuvette over the range of 300nm-600nm.

The values of Φ for rsFolder and mutants was calculated by recording absorption and emission spectra of the protein at different concentrations. To record the emission profiles, the spectrophotometer excited the protein at 488nm and recorded the emission intensities from 492nm until 650nm. The OD of the protein at 488nm was then plotted against the integral of the whole emission spectrum, to extract the Φ values. For rsFolder mutants that are not 100$\%$ A-*cis* populated at pH=7.5, the OD was corrected to take into account all the molecules present, using NMR-derived values of *cis* and *trans* populations. The Φ of Fluorescein was measured for normalisation purposes.

Consequently, the brightness for each protein was evaluated as the product of ϵ and Φ, as compared to EGFP[37].

2.5.3 Ensemble microscopy measurements to extract switching rates and switching contrast

In order to measure switching rates and switching contrast of rsFolder and mutants at different pH values, ensemble measurements were carried out at the microscope set-up. The protein samples were fixed in polyacrylamide (PAA) gels that were "sandwiched" between two microscopy slides. PAA ensures that the protein molecules are not allowed to diffuse during recording, ensure no diffusion artefacts. In total, 50μl of approximately 40μM molarity protein in the buffer and pH of interest was diluted against 50μl of acrylamide and polymerisation was initiated using 1μl of APS (10$\%$) and 1μl of TEMED (>99$\%$).

To ensure a uniform sample thickness for all samples, a plastic frame was glued on the bottom microscopy slide. Consequently, 40μl of the aforementioned mix was inserted within the frame and covered by the top microscopy slide. Shortly after, it would polymerize.

After having recorded the beam shapes of the 405nm and 488nm LASERs, each sample was placed in the PALM microscope for recording. Depending on the sample (protein type, pH) the sequence of illumination pulses and their respective strength was different. The intensity of the LASER output could be modulated using an AOTF filter and OD filters. Briefly, the sequence consisted of 3 main steps, each one consisting of a number of frames.

Firstly, a "clean" pulse of 405nm LASER before the camera starts recording ensured that the sample has converted to the "on"-state. Secondly, a series of frames where each one included a 488nm pulse of certain duration coupled with a small exposure time of the camera, enabled the detection of fluorescence and recording its decay over time. Lastly, a third step included frames whereby a 405nm LASER pulse was followed by a short 488nm pulse, coupled with the camera detection. This last step enabled the return of the protein to the "on"-state (due to the 405nm LASER) while its fluorescence response was measured upon 488nm illumination.

The data recorded were fitted against a kinetic model taking into account 2 "dark" states. One "dark" state is the conventional "off"-state, containing the neutral *trans*-CRO species state while the second one needs to be accounted for in order to properly fit the decay and build-up curves (represents the effect of photobleaching and blinking events on the overall switching kinetics).

Chapter Three

Results

The majority of the information included in parts 3.1 and 3.2 have been presented in two publications [43, 42], included in the Appendix. Hence, in order to avoid referencing almost every paragraph, I shall disclose here that the majority of data and the resulting figures can also be found in the above cited publications.

3.1 In-situ LASER illumination in NMR spectrometer

In the beginning of my project, I had to establish a reliable and effective illumination set-up that could be easily used alongside the NMR spectrometers. For an extensive review of the different methods of combining NMR with light for investigating photo-induced processes, the reader is referred to the recent work of Nitschke et al [156]. Here, I will briefly describe some of the ways by which one can construct an illumination set-up for NMR studies and the rationale behind choosing the in-situ illumination set-up using a LASER box coupled to an optical fiber which can guide the LASER light into the NMR tube in the context of my project.

3.1.1 Choosing the LASER box

In order to illuminate an NMR sample tube, one can choose between Light amplification by stimulated emission of radiation (LASER) or Light Emitting Diode (LED) sources. The differences lie on bandwidth of emitted wavelength, beam divergence, and cost.

LASERs principally achieve monochromatic light, i.e a single wavelength of light produced, while LEDs are characterized by a broader wavelength range of illumination, typically about 20nm [156]. This advantageous specificity has set them apart from LEDs, especially in experiments where the photoinduced phenomena studied are very sensitive to the wavelength of excitation. By observing the 20-30nm broad absorption spectra of RSFPs, such as rsFolder, one could conclude that a larger bandwidth of emitted wavelength could be advantageous in exciting more molecules.

However, LEDs suffer from high light divergence. This can be partially fixed by collimating the light using special collimating lenses. However, due to Etendue conservation, the collimated light has a larger beam spot and, thus, lower power densities [208]. Moreover, reflection of LED light on lenses necessarily causes power losses. On the other hand, LASERs achieve highly focused beams, enabling for high power densities. For the needs of this project, a focused LASER beam was essential in transmitting a sufficient amount of power through an optical fiber. Therefore, LASERs were more appropriate than LEDs.

Furthermore, since RSFPs are used for (mainly) super-resolution microscopy, where LASERs are used, it seemed more relevant to study them under LASER illumination conditions.

Portability was essential for this project, since experiments were run on several NMR spectrometers. Therefore, a LASER box hosting two LASERs (405nm, 488nm) was the preferred choice, containing all the necessary optical elements required for alignment and power modulation, which remain fixed even during LASER transport. For the purposes of fitting in a box, the LASERs are much smaller, achieving lower powers (in the order of a few hundred mW) as compared to conventional LASERs (in the order of Watts).

The one-channel output L4Cc LASER combiner box distributed by Oxxius S.A. was purchased (Oxxius S.A., Lannion, France) (Figure 3.1). Inside the box lied two LASER diodes, emitting monochromatic light at 405nm and 488nm wavelength, respectively. The advantages of using this device were mainly its ease of transport and variable operative modes.

The LASER box consists of all the necessary parts for operating the two diode LASERs within a small 20x25x10.8cm metallic box. There is no need for manual interference by the user in terms of adding or removing optical components in order to make use of all of its modalities.

More specifically, the principle elements included in the LASER box are:

– 2 LASER diodes aligned parallel to each other with a voltage-dependent power output

– dichroic mirrors (DM) to direct the LASER light towards the output

– 3 electromechanical shutters: one in front of each LASER source and one at the output of the box

– a fiber coupler for the output of the LASER

– a remote controller for the manipulation of the electromechanical shutter situated in the output

All these elements ensure that the two LASER diodes can be operated at different power levels, independently from each other. The power output can be adjusted manually by a user interface (Analog Modulation), by adjusting the current input (0-100% at a frequency of up to 200Hz, equivalent to 300ms pulses) of either laser. At a fixed power output, the

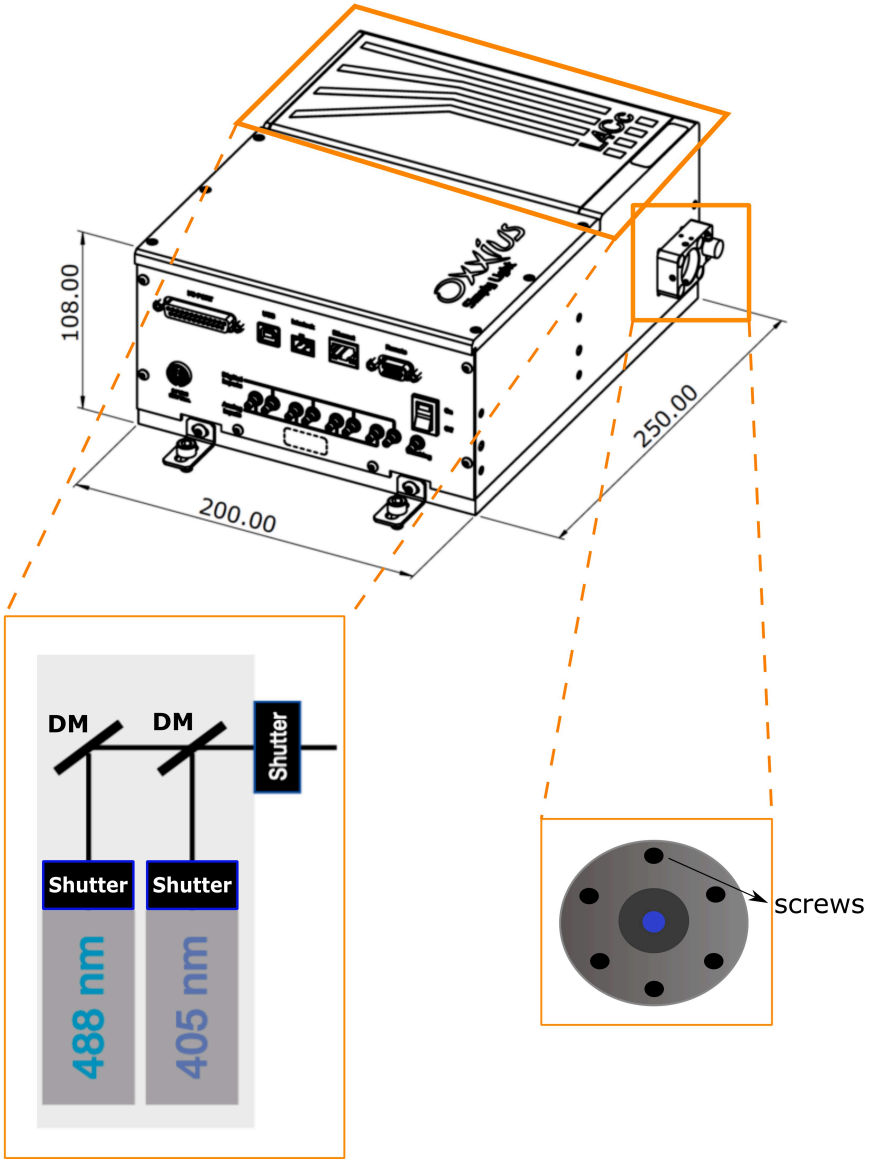


Figure 3.1 On the top, a drawing by Oxxius S.A. of the LASER box L4Cc combiner model. Bottom left, zoomed in the region of the two LASERs within the box and bottom right, the LASER box output consisting of a fiber coupler and the corresponding screws needed for alignment.

LASERs can be switched “on” or “off” digitally via TTL lines (Digital Modulation). The maximum output is 200mW for each LASER as recorded in the output of each LASER.

Manipulating the LASERs manually or via the NMR console

Photoswitching rates of the NMR sample were LASER-power-dependent, as will be discussed later. Hence, in rsFolder, when "quick" photoswitching was needed, short LASER pulses of 100% (200mW) of the total power of each LASER would be used, while when continuous illumination was required, the LASER power was typically reduced to 30% (60-70mW), in order to minimize temperature gradient effects on the sample. Depending on the requirements of the experiment, manual (via the laptop) or remote (via TTL lines) switching of the LASERs was chosen.

No power modulation was used for illumination over long periods of time. Briefly, a user interface installed in a laptop connected to the LASER box (Figure 3.2) allowed for changing the power output manually for each LASER independently, in terms of percentage of input current. The percentage of power allowed was manually adjusted depending on the experiment and LASER used, after ensuring that the total output power is relatively similar as in previous experiments.

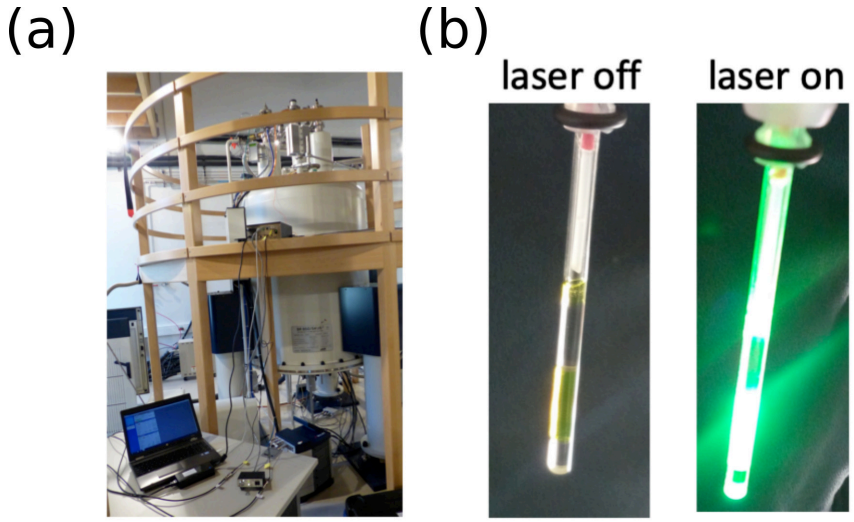


Figure 3.2 (a) A photograph of the in-situ illumination set-up installed next to the 850MHz spectrometer in the IBS NMR platform. A laptop computer is connected to the LASER box which allows for manual adjustment of the LASER output and the LASER box is mounted on the mezzanine (upstairs). (b) A photograph of the Shigemi tube containing rsFolder (200μM) in the absence and presence of 488nm LASER light.

Digital switching "on" and "off" of the two LASERS was employed when the LASER channel was integrated within NMR pulse sequences used. More specifically, the LASER box was connected via TTL lines to the NMR console, in the IPSO panel (Figure 3.3). The IPSO panel synchronizes RF pulse generation from different channels operating in active low logic. By default, the transmission is allowed (Open) with an output of +5V, while blanking of transmission is equivalent to 0V and needs to be actively applied via a command

(Closed). Consequently, the LASER channel integrated in the NMR pulse sequence could take the values of “LASER-on” or “LASER-off” for each LASER, and the LASER “pulse” and “delay” duration between LASER pulses could be specified.

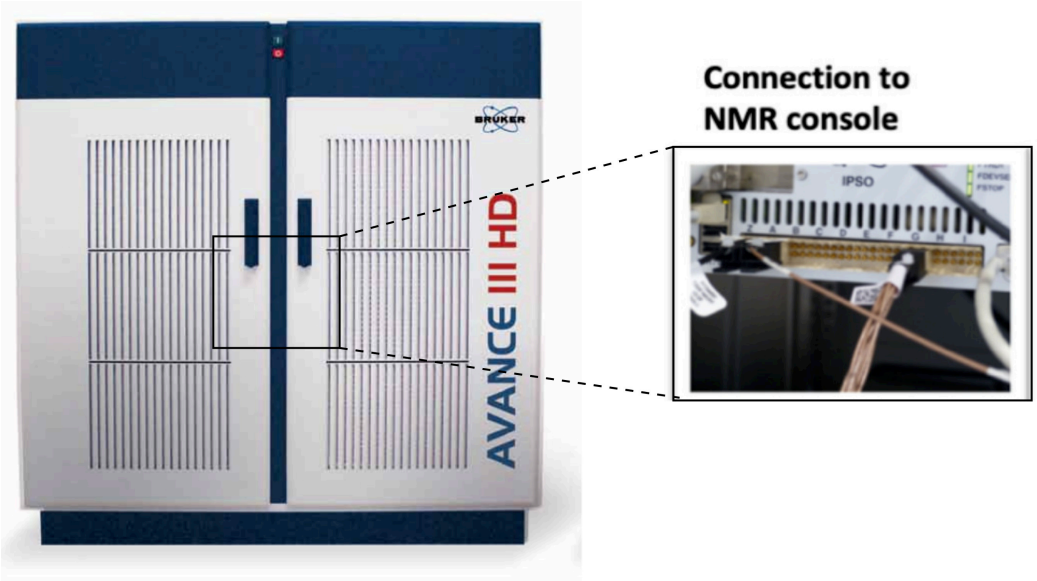


Figure 3.3 Photograph of TTL line connected to the IPSO channel of the NMR console, allowing the integration of LASER pulses in the NMR pulse sequence.

3.1.2 In-situ illumination of the NMR sample

Options for illumination of the NMR sample: ex-situ versus in-situ

A number of previously reported studies combining NMR with light, used ex-situ illumination of the sample [172, 231]. More specifically, the systems studied would remain at a steady state after illumination for relatively long periods of time, from minutes to days. Before the LASER box was purchased, we were able to illuminate the NMR tubes containing rsFolder with 488nm LASER outside the spectrometer and still record spectra of the “off”-state, due to its high thermal stability [61]. However, when recording long NMR experiments lasting over a day (eg 3D triple-resonance experiments), at 40°C, a continuous illumination of the sample in-situ was necessary.

Moreover, in-situ illumination enabled us to study the protein’s response to light on-demand. While for the most part, steady-state phenomena were studied, sometimes alternating between the two available conformations of rsFolder was required. This could not have been possible using ex-situ illumination.

Optical fiber specifications: efficiency and resulting homogeneity

An optical fiber was used to guide the LASER light into the sample in the 5mm NMR tube, which had a connector end and a bare end. The connector end would be screwed in the fiber coupler while the bare end was necessary in order to fit the fiber inside the plunger of the Shigemi sample tube (Figure 3.4). A Shigemi tube and its corresponding plunger were used since they are made of materials with matching magnetic susceptibility as an average solution sample resulting in overall better *shimming* (related to magnetic field homogeneity) of the sample.

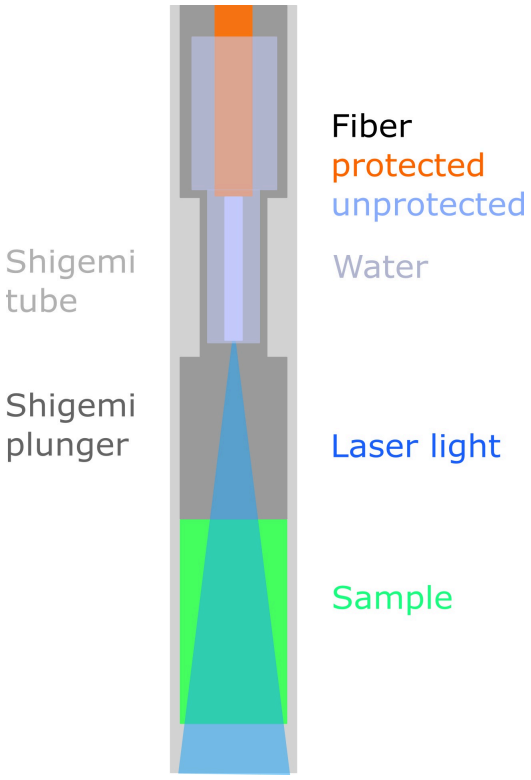


Figure 3.4 Schematic representation of the bare end of the optical fiber inserted inside the Shigemi tube plunger and immersed in water. The sample is illuminated from the top, in a cone-shaped fashion.

Another aspect of using an optical fiber with a bare end inside the plunger, as already highlighted by Mizuno et al [146], is the sudden difference in refractive index between the fiber and the plunger glass material due to an air gap. For reasons of avoiding further light loss due to scattering, the plunger was filled with water which has a refractive index closer to glass (refractive index of water 1.3, glass 1.5) as compared to air (refractive idex 1), thus minimizing transmission losses from the fiber end to the plunger glass bottom. The resulting beam transmitted by the end of the plunger had (ideally) a conical shape, corresponding to a Gaussian distribution of light. This ensured illumination across the whole sample, though, in an inhomogeneous manner.

Improving illumination homogeneity in the sample

One way to partly counteract illumination inhomogeneities was demonstrated in an NMR study using an optical fiber with an etched bare end which was directly inserted in the sample, without the need of a plunger [128]. Etching the 3mm-long end of the fiber ensured that light is produced homogeneously across the sides of the fiber which was placed in the middle of the sample tube. Spectral resolution was not affected by this modification, i.e the presence of the "foreign" fiber end inside the sample. While this would be a very attractive technique to use, the cleaning and maintenance of the fiber would be a disadvantage i.e it would be impossible to protect the etched fiber tip from getting broken or scratched.

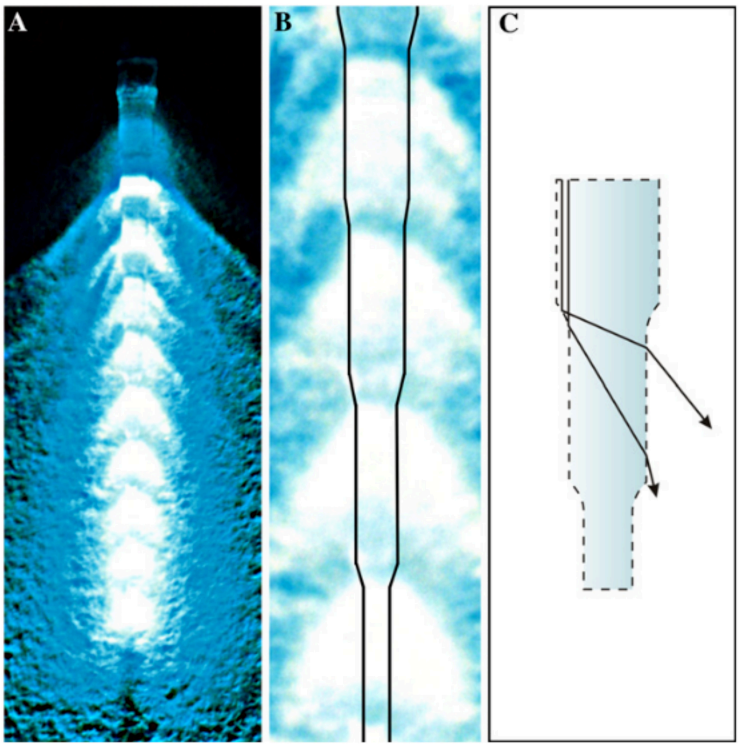


Figure 3.5 Figure adopted from Kuprov et al [128], showing the etched fiber end achieving greater illumination homogeneity. In (A), a picture taken of the emission of argon lamp against white background in the dark, in (B) the shape of the etched fiber is drawn, while at (C) the light reflections resulting in the apparent homogeneity are depicted.

Therefore, increasing homogeneity without modifying the bare end could be achieved by modifying either the plunger tip or the NMR tube. More specifically, in a study investigating the refolding kinetics of α-lactalbumin upon photolysis, the authors used a pen-shaped quartz plunger tip which enabled more homogeneous distribution of light in an optically dense sample [127]. Therefore, we decided to acquire such pen-shaped plunger and test its advantages over a regular flat-bottom plunger. Another modification tested consisted of

spraying the exterior of the NMR tube with a reflective material. The spray, conventionally used for surfaces on road signs in order to make them reflect car lights during the night, seems to be as efficient as the pen-shaped plunger in terms of light distribution, as will be explained in section 3.1.3.

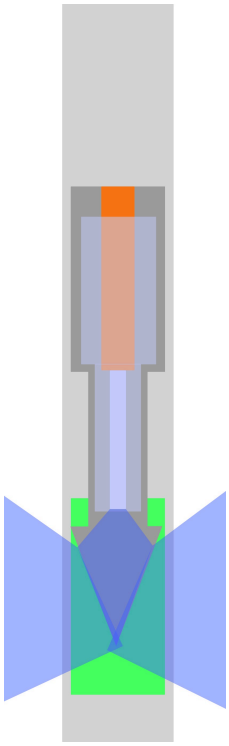


Figure 3.6 Schematic representation of the bare end of the optical fiber inserted inside the Shigemi tube pen-shaped plunger and immersed in water. The plunger is inserted further in the sample to achieve homogeneous illumination of the LASER light, shown in blue.

Maintenance of the optical fiber and the plunger

Unfortunately, a bare end of a fiber is extremely sensitive to scratches and breaks which can significantly lower the total output power of the fiber [79, 206, 121]. The maximum power output achieved at the end of the fiber is, therefore, correlated to the "smootheness" on the fiber end surface. Typically, the "cleanest" surface is achieved by cutting the fiber using diamond knives. However, they tend to be costly, thus, a cheaper alternative is ceramic scribes.

Approximately, the highest output desired after ceramic scribe cleavage, compared to the nominal power achieved at the fiber coupler, was 70% as measured at the glass bottom of the flat-bottomed plunger. In addition, the beam power would be ideally spherically distributed, to ensure homogeneity of illumination. Accordingly, the ceramic scribes were

used to ensure "clean" cuts and reduce imperfections at the fiber bare end that would accumulate over time. Moreover, in rarer cases, the water inside the plunger had to be substituted since salts and dust would build up causing scattering of light. Using an optical powermeter, the total output power of each LASER was monitored, while the beam shape was regularly inspected visually against a white background.

3.1.3 Performance of the in-situ illumination set-up in NMR spectrometer

To test the efficiency of photoswitching of our rsFolder NMR samples, a time-series of 1D experiments targeting methyl hydrogen frequencies was used (as described in the Materials & Methods section regarding illumination homogeneity). The methyl hydrogens were chosen as the most sensitive probes to enable fast data acquisition times without the risk of a low signal-to-noise ratio. That is because the symmetry of methyl hydrogens alongside their fast dynamics results in high-intensity NMR signals. Each individual spectrum of the series lasted approximately 0.5s while the whole time-series was recorded for 5 mins. More specifically, for the "on"-"off" switching, the sample was in the dark at the beginning of the time-series. After 1 minute, the 488nm LASER was manually turned on to switch the sample to the "off"-state. In the case of the "off"-"on" switching, the sample was illuminated using the 488nm LASER at the beginning of the experiment for approximately 1 minute and immediately after, the 405nm LASER was switched-on. LASER manipulation was performed manually at this point.

Initially, we wanted to assess the LASER-power-dependence of photoswitching of our NMR samples. Consequently, we tested the in-situ illumination set-up on a sample of rsFolder at pH=7.5 HEPES, at 40°C, with OD=33, using the conventional flat-bottom Shigemi plunger. The recorded intensities of methyl hydrogens experiencing a chemical shift change upon switching was plotted against time. The switching curves representing the built-up of methyl hydrogen peaks of the "off"- and the "on"-state, respectively, altered according to LASER power (Figure 3.7). At maximum 488nm LASER power, the switching "off" of the protein sample lasted a little over two seconds while the switching "on" of the "off"-state sample upon maximum 405nm illumination was twice as fast. As expected, for negative RSFPs, the "on"-"off" switching is slower than the "off"-"on" switching [3, 26]. Nonetheless, it is important to note that switching times are heavily affected by experimental conditions (optical density of the sample, LASER output alignment, fiber end surface degradation etc) resulting in up to two-fold differences of the time required for the sample to completely photoswitch between different experiments.

Overall, the geometry of the Shigemi tube as well as the high concentration and volume of the samples used (200μM, 3cm height, resulting in OD=33) only allowed for relatively slow arrival at steady-state conditions. While the switching time per molecule is reported to be in the range of μs [61], the time it takes for all the molecules in the NMR tube to absorb a photon and switch to another state is significantly longer (seconds). In order to

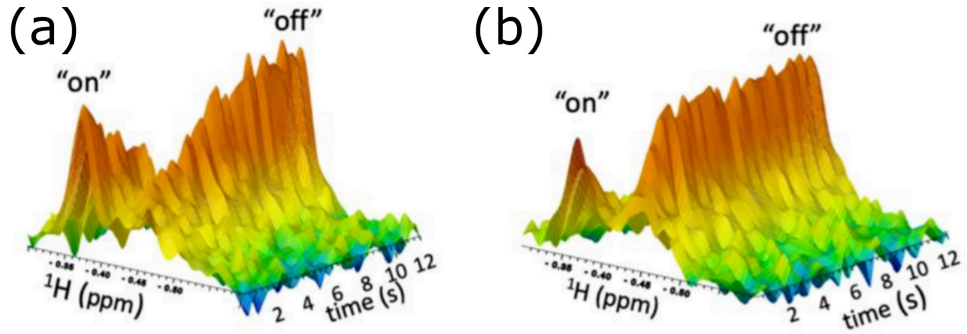


Figure 3.7 NMR observed "on"-to-"off" photoswitching. 3D representation of the timeseries of 1D ^1H-spectra, recorded at 850MHz at 40°C, to assess the photoswitching behaviour of the rsFolder sample in the NMR setup under different power levels. Focusing on methyl hydrogens and zooming at a frequency of a methyl hydrogen undergoing CS change upon photoswitching, intensities can be plotted over time showing the "off"-state peak increasing after (a) low-intensity and (b) high-intensity LASER illumination.

accurately characterize the "on"- and the "off"-states of rsFolder, it was crucial to reach a steady state situation, after all the molecules have undergone photoswitching.

The main factors affecting such slow switching of the sample were the high OD of the sample and the uneven distribution of the LASER illumination onto the sample. Photon absorption is maximum at the top of the tube while at the bottom, photons arrive only once the rest of the protein sample has switched to a "light-transparent" state i.e a state which does not absorb at the given wavelength of illumination. The delay in the arrival of LASER light at the bottom of the tube could be visually observed.

To speed up photoswitching of the optically dense NMR sample, the number of photons arriving per molecule needed to increase, i.e homogeneity had to be ameliorated. This could be achieved either by decreasing the amount of molecules, i.e the OD (lower concentration or sample volume/height) or by increasing the amount of photons readily available for switching. The first option is sub-optimal because it leads to lower signal-to-noise ratio of our NMR spectra, making the second option more favorable.

Illumination homogeneity was improved by the use of either the pen-shaped plunger or the spray-treated Shigemi tube. Moreover, the sample was reduced to 100μM concentration and 2.5cm sample height (OD=13.75). 2D ^1H-^{15}N BEST TROSY spectra were recorded of the same sample using the two shapes of plungers. The sample was initially in the "on"-state and during a recycle delay of 1.5s, a full-power 488nm LASER pulse was applied. The resulting "on"-"off" switching of the NMR sample was almost three times more efficient when using the pen-shaped plunger as compared to a flat-bottom plunger (Figure 3.8).

Similar results were obtained when comparing the untreated and spray-treated Shigemi tube.

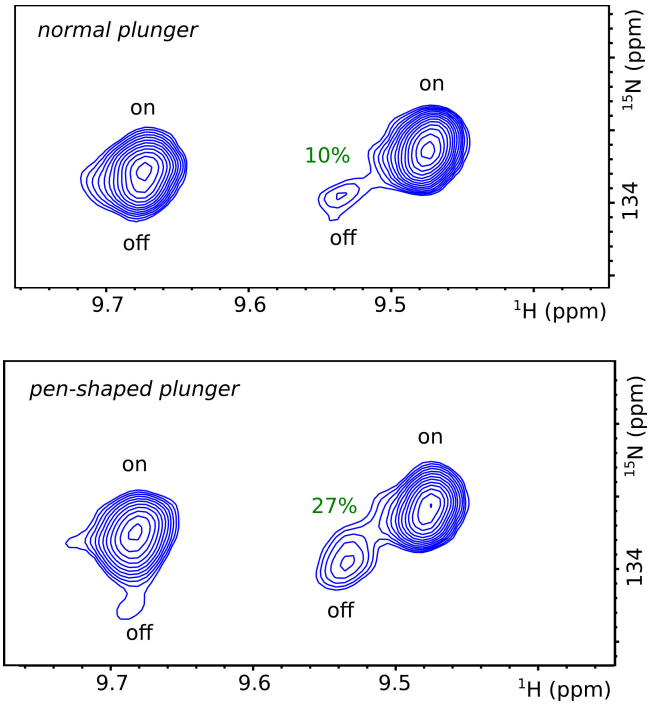


Figure 3.8 2D ^{1}H-^{15}N BEST TROSY spectra, recorded at 850MHz at 40°C, to assess the photoswitching behaviour of the rsFolder sample in the NMR setup using the conventional flat-bottom versus the pen-shaped plunger. The sample, before starting the NMR experiment is in the "on"-state. During a recycle delay time of 1.5s, a 488nm LASER pulse (of 100 % power) partially switches it to the "off"-state. Focusing on two well-resolved peaks in the amide spectrum, the apparent "off"-state peak intensity compared to the "on"-state increased from 10 to 27% occupation, when using the pen-shaped plunger. Indicated are the peaks corresponding to the "on"- and "off"-states.

Moreover, temperature gradients could possibly be induced upon illumination, especially at higher LASER powers. These temperature gradients could cause artifacts on spectra due to sub-optimal shimming. In order to ensure that experiments are run under steady-state conditions, NMR experiments were initiated only once the Lock signal was stabilised. Lastly, to further ensure that illumination does not interfere with NMR acquisition, spectra of the model protein Ubiquitin were recorded with and without illumination (either 405nm or 488nm). We expected no difference to occur due to photo-induced processes in the protein, since it is not photo-sensitive. The spectra recorded for the sample illuminated with either wavelength as compared to the sample in the dark showed no differences.

3.1.4 Advantages, limitations and future perspectives of the in-situ illumination setup

The portable LASER box which can be manually and remotely manipulated and has adjustable power output was central to my thesis work. It allowed for a variety of illumination schemes, of different power and duration, while also enabling measurements on multiple NMR spectrometers.

As discussed in the previous sections, homogeneity of illumination is crucial in acquiring fast and reliable NMR spectra. Future adjustments on the set-up should aim at increasing homogeneity, either by the use of unconventionally-shaped plungers (eg pen-shaped plunger) or by the treatment of the Shigemi tube, either from the outside or on the inside.

Another feature which would be interesting to add is to couple NMR with optical measurements on the set-up. During my project, I had to compare NMR data with optical measurements performed at different set-ups, sample concentration, temperature etc. Therefore, NMR and optical results could not be directly comparable, in some cases.

This limitation could potentially be overcome with the addition of an optical spectrometer which would detect the fluorescence emission profile of the NMR sample in-situ, during NMR acquisition. More specifically, refracted, reflected or emitted light from the sample can be transmitted via the optical fiber, onto a dichroic mirror placed perpendicular to an optical spectrometer. A bandpass filter could separate between wavelengths, in order to measure only fluorescence in the green and block UV or cyan emission originating from the LASER box. The addition of such a set-up would require a disruption of the path of light from the LASER box to the fiber and onto the sample. A sketch of an "optical box" including all the necessary elements making up a portable set-up for optical measurements is shown below (Figure 3.9). Such an "optical box" needs to be placed on the way of the light path going in the sample.

It was not clear whether this set-up would be sensitive enough to measure the fluorescence emission profile of the sample. Additionally, LASER transmission to the sample would be significantly hindered. Many air gaps and optical elements on the light path to and from the sample would lead to various transmission losses, resulting in potentially weak LASER power arriving at the sample and low fluorescence signal detected by the optical spectrometer.

One way to overcome the light path disruption would be the use of a second "detector" fiber (Figure 3.10). This way, the "optical box" would not interfere with the excitation light originating from the LASER box, since it would be receiving the light from an external fiber, independent from the excitation light path. In practice, a second fiber would need to be attached to the one already used, in order to fit inside the plunger. One way of keeping them attached would be to have a unique protective tubing for both. The light transmitted through the second fiber would get through a bandpass filter to separate the LASER light reflected from the glass of the Shigemi tube and the fluorescence emission originating from

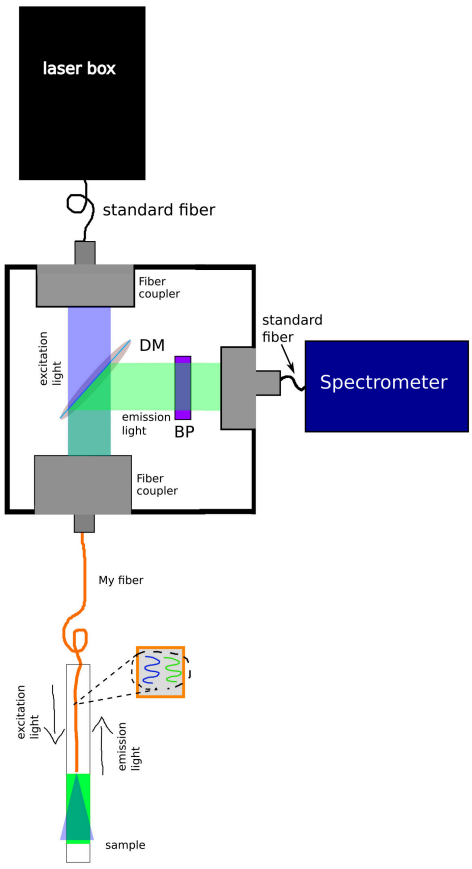


Figure 3.9 A sketch depicting the idea of an "optical box" to be included in the in-situ illumination NMR set-up providing optical measurements during acquisition. The LASER light (in blue) is guided from the Oxxius box to the sample, via two optical fibers coupled to the "optical box". The light coming from the sample (in green) is guided into the "optical box" via the second fiber, where a dichroic mirror reflects it into a bandpass filter allowing only the emission .

the RSFP molecules. However, such a set-up would still suffer signal loss on the detection of fluorescence.

Possibly the most efficient way to detect fluorescence emission without causing disruptions on the path of LASER light to the sample or losses on the light path aimed for detection would be to interfere with the components of the LASER box itself. More specifically, a dichroic mirror could be placed inside the LASER box in such a way to direct the reflected and fluorescence emission light from the sample onto an optical spectrometer. This solution would require for the LASER box to be bigger and, thus, heavier. Nevertheless, it would provide crucial information regarding the fluorescence response of the sample during NMR acquisition.

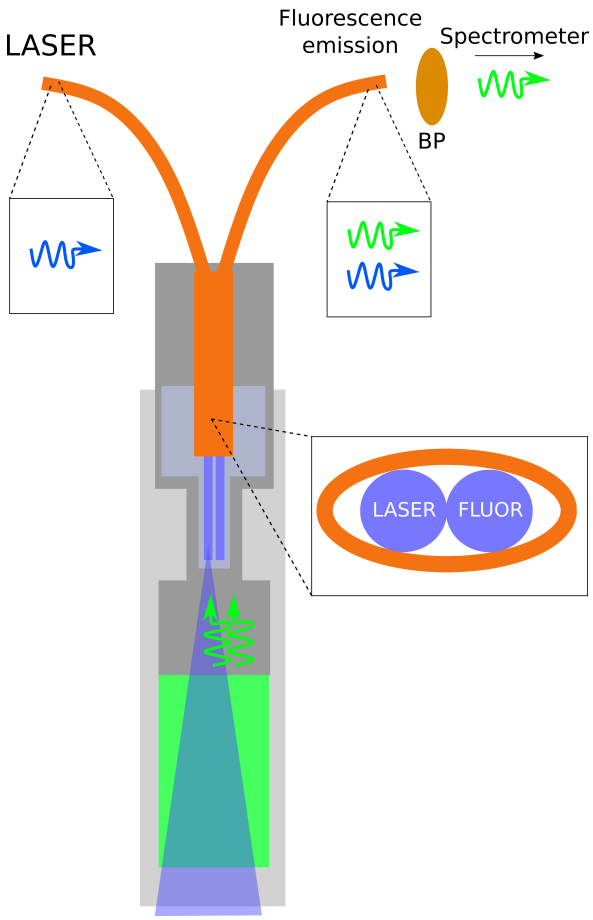


Figure 3.10 A sketch depicting the use of two optical fibers in the NMR Shigemi sample, to enable detection of fluorescence emission from the illuminated sample. Incoming light from the LASER (blue) excites the sample, while outgoing light, including the reflected LASER light (blue) and fluorescence emission from the sample (green), is filtered using a bandpass (BP) filter, before it is detected by the optical spectrometer. The fibers share the protective tubing, in order to make the two-fiber system thin enough to fit in the plunger tube.

3.2 NMR studies of rsFolder "on"- and "off"-states

After optimizing the in-situ illumination set-up, the next steps consisted of the characterization of rsFolder in the "on"- and "off"-state. The next sections describe the NMR approach towards studying rsFolder as a model RSFP protein. During this part of my project not only did we learn new things concerning rsFolder structure and dynamics but we were also able to establish an "NMR-toolbox" consisting of experiments that are relevant to the study of RSFPs.

3.2.1 Sample and spectral quality control

All protein samples were kept at 4°C in-between experiments. For longer waiting times (weeks or months), samples stayed at -80°C. The buffer used in the following sections was HEPES (50mM) at pH=7.5.

An important factor for high signal NMR spectra is the choice of temperature. Considering the overall effect of temperature on global molecular motions (and thus the overall NMR signal for all peaks present in the spectrum) at low temperatures, NMR peaks are broadened, while at higher temperatures, they become narrower. The reason for line broadening upon temperature decrease is due to shorter transverse relaxation times T_2, induced by slower molecular tumbling. Moreover, depending on the local dynamics present, temperature increase can cause NMR lineshape narrowing or broadening, by shifting the timescale of motions present at a local level, from the intermediate to a faster exchange, or from the slow to the intermediate regime, respectively.

Fortunately, rsFolder [61] is a protein which was inspired by a marrying of an RSFP (rsEGFP2) with Superfolder-GFP, a GFP well-known for its stability and endurance to a range of pH and temperatures [169]. Hence, rsFolder has been a very easy protein to work with even at relatively high temperature (40°C) (Figure 3.11).

Nevertheless, a regular checking of sample stability – before and after an experiment, before and after illumination etc – was necessary for reliable data interpretation. Sample quality control was performed by acquiring a 2D ^{1}H-^{15}N BEST-TROSY spectrum of the "on"- (or dark adapted) and the "off"- (or 488nm-illumination-induced) state.

3.2.2 Distinction between "on"- and "off"- states – NMR-observed switching contrast

The spectra for the "on"- and "off"-states were clearly distinct (Figure 3.12). In contrast to crystallography, whereby only small local differences on the side-chains of residues in the vicinity of the CRO are seen between the two states, using NMR we were able to clearly identify two characteristic NMR signatures for the "on"- and the "off"-state. However, at this point, it was not clear whether the differences between the spectra are due to changes in conformation or just an effect of *cis-trans* isomerization of the CRO. More specifically, the presence of an aromatic ring (of the CRO) can have a significant impact on the observed CS, a phenomenon called "ring current shifts". Considering the electronic configuration of aromatic rings, local currents of the overlapped pi-orbitals of a ring can induce a local magnetic field which affects other residues in the vicinity. Nevertheless, distinguishing between the two states from an NMR perspective was important in characterising them independently.

Furthermore, when overlaying the two spectra, it is clear that after 488nm illumination some 3.5% of the protein still remains in the "on"-state. This percentage of remaining

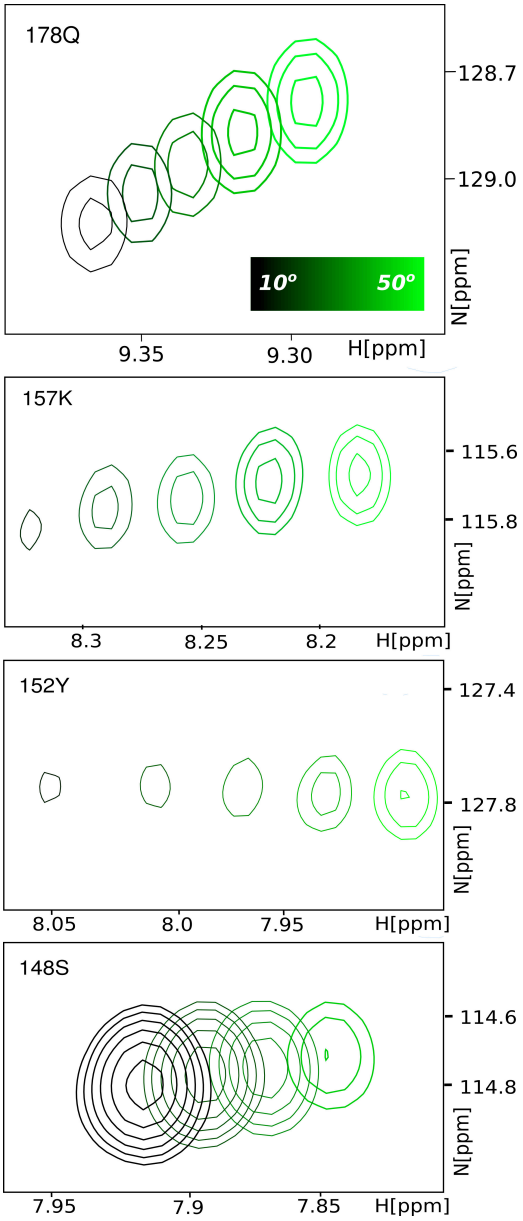


Figure 3.11 Spectral regions of ^{1}H-^{15}N BEST-TROSY spectrum of the rsFolder "on"-state, under different temperatures (10, 20, 30, 40, 50), recorded at 850MHz. Temperature-induced CS changes can be explained by considering the sensitivity of CS to local environment (H-bonding, folding stability and conformational freedom) [211]. From this series of experiments, it was concluded that 40°C should be the default temperature of experiments due to high signal-to-noise for the majority of peaks.

"on"-state is considered to be the NMR-observed switching contrast – assuming that the "on"-state is the only one that can fluoresce upon 488nm light excitation, the presence of 3.5% of "on"-state molecules can contribute to remaining fluorescence after the sample has

reached maximum “off”-state occupancy. Moreover, this value agrees with the expected value from optical studies [61], whereby switching contrast is defined as the remaining fluorescence when the sample has been switched-off using continuous 488nm illumination.

Our results combined with fluorescence studies from literature [61] suggest that the switching contrast of rsFolder originates from the presence of a minor population of “on”-state proteins after long 488nm illumination, while “off”-state molecules do not contribute to fluorescence. This finding is in support of the idea that switching contrast originates from the back-switching of some “off”-state molecules to the fluorescent “on”-state.

Moreover, our NMR experiments indicate that switching contrast is not affected by temperature (in the range of 20-40°C).

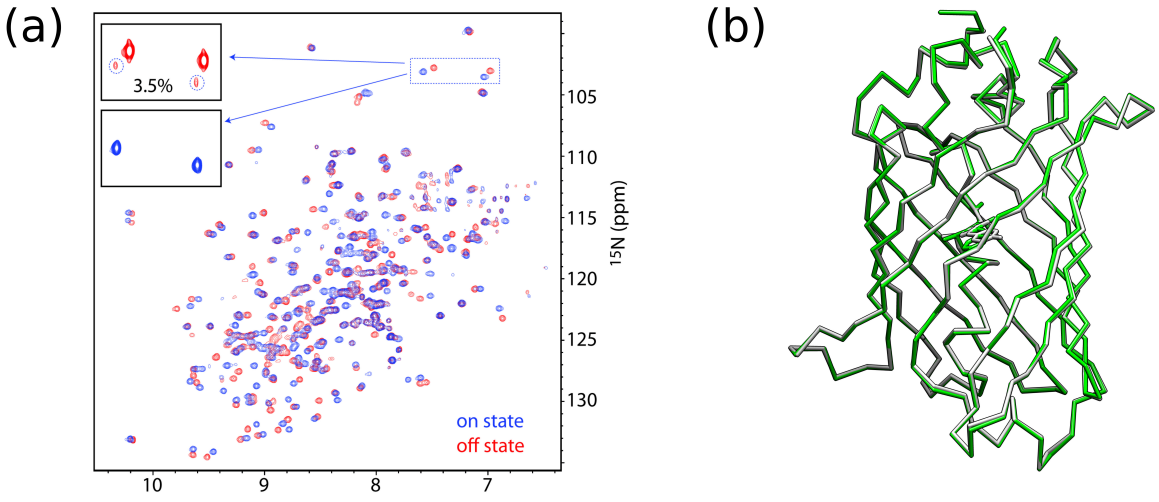


Figure 3.12 (a) Overlaid ^{1}H-^{15}N BEST-TROSY spectra of rsFolder in the “on’- and “off”-state recorded at 850MHz at 40°C. Zooming in a pair of well-resolved peaks, the population of remaining “on”-state detected in the “off”-state spectra is 3.5%, similar to the switching contrast published for rsFolder[61]. The NMR signatures of the two states are very distinct (red and blue, respectively) while, in (b), crystallographic structures show no differences in backbone conformation (green – “on” (PDB:5DTZ), grey – “off” (PDB:5DU0).

3.2.3 NMR-observed photobleaching

The metastable “off”-state of rsFolder was induced by 488nm illumination. While thermal relaxation of rsFolder is very slow, it was nevertheless crucial to maintain the sample in the “off”-state using continuous 488nm illumination. In order to account for the effect of continuous 488nm illumination on the sample, multiple ^{1}H-^{15}N BEST-TROSY spectra were

recorded to assess the quality of the sample over time. Each ^{1}H-^{15}N BEST-TROSY spectrum lasted 8mins and was recorded before and after other NMR experiments (eg inbetween 3D triple-resonance experiments) aimed at studying the "off"-state. The peak positions did not show any change and new peaks did not appear over time – a sign that the protein is not undergoing degradation or unfolding over time due to 488nm illumination. However, peak intensities of the integrated signal across the spectrum plotted against time appeared to decrease. After 20 hours, the overall signal had decreased by 30% and reached a plateau (Figure 3.13). Such light-induced effects were not observed after long periods of 405nm illumination.

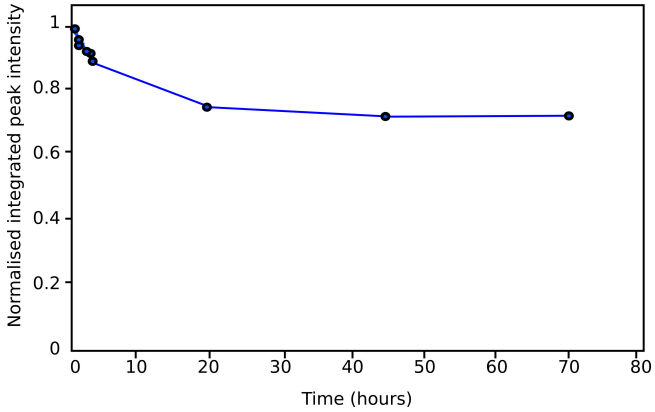


Figure 3.13 Graph showing the integrated peak intensity of "off"-state spectra, after normalisation, against time of 488nm illumination.

Interestingly, in NMR, loss of signal without an obvious "leak" of the sample due to evaporation, is a sign of the formation of bigger sized particles, not detectable by NMR due to slower tumbling, causing faster T_2 relaxation time and thus dramatic line broadening of the NMR peaks. Therefore, we speculated that continuous 488nm illumination causes the formation of oligomers or aggregates which are undetectable by NMR.

The 488nm-illuminated sample was sent to the Analytical Ultra-Centrifugation (AUC) platform at IBS, which used AUC to confirm the existence or not of such big particles. As was expected, they were able to identify the existence of big aggregates accounting for $\approx 25\%$ of the overall sample. After a round of centrifugation at 3000rpm, the sample OD(280) decreased by $\approx 25\%$, indicating that large particles precipitated and did not contribute in the total OD(280).

Photobleaching in FPs has been generally associated to the amount of oxygen species able to enter the barrel and attack the CRO and other residues. According to a study of IrisFP by Duan et al [57], oxygen at low LASER powers oxidizes two residues from the barrel, which seem to "trap" the CRO to a protonated form. At higher LASER powers, oxygen is not involved, but the main photobleaching mechanism includes irreversible re-configuration of residues from the barrel and reduction of the methine bridge of the CRO. The LASER power used in our experiments falls under the low-illumination regime, with respect to Duan's

work. Studies on monomeric red FPs (mCherry and mRFP1) also attribute photobleaching to the weakness of the β-barrel to shield the CRO from outside molecules. Oxygen diffusion pathways have been identified starting from flexible regions of the β-barrel (Chapagain 2011, Campbell 2002). Comparing the effect of 405nm and 488nm illumination, it is clear that rsFolder at 488nm is prone to photobleaching, which causes aggregation of rsFolder to big NMR-unobsevable objects. Potentially, the remaining 3.5$\%$ 488nm-absorbing "on"-state molecules present after continuous 488nm-illumination "consume" the oxygen (or other air substances) present in the sample during photobleaching, until there is no more oxygen left so photobleaching and, thus, aggregation reaches a plateau.

3.2.4 Effect of photocycling and LASER power on NMR observables

Furthermore, to confirm that multiple photocyles do not affect the spectra detected for the two states, we performed a round of 1000 photocycles whereby the two LASERs (405nm, 488nm) were alternating for a duration of 3s and 7s, respectively. Comparing the ^{1}H-^{15}N BEST-TROSY spectra recorded before and after photocycles, we found neither significant changes in CS of peaks in the two spectra, nor the appearance of new peaks. Therefore, we could conclude that the illumination schemes used on the sample do not significantly alter the NMR-observed behaviour (Figure 3.14(a), (b)).

As will be discussed below, our NMR experiments suggested the existence of intermediate exchange dynamics in the light-induced "off"-state state of rsFolder. Such dynamics induced loss of NMR signal for certain residues in the "off"-state. For rsFolder in the "off"-state, there remains a small percentage of molecules that can be found in the "on"-state and can therefore switch upon 488nm illumination. Therefore, such light-induced switching of a minor population from *cis-trans* could potentially lead to the NMR-observed dynamics upon 488nm illumination. In order to ensure that the observed dynamics for the light-induced state report on the intrinsic behaviour of the protein in its "off"-state conformation and are not a result of the LASER power-dependent *cis-trans* isomerization, we acquired spectra of the "off"-state in the presence and absence of 488nm light.

More specifically, we recorded a ^{1}H-^{15}N BEST-TROSY spectrum for the "off"-state with full-power 488nm illumination and 5 mins after the LASER was turned off and the sample remained in the dark – it can be safely assumed that during this short time, thermal relaxation does not have a significant effect on the ratio of *trans* and *cis* populations. Subtracting these two spectra recorded, we could detect no effect of the LASER power on the spectra of the "off"-state (Figure 3.14), suggesting that the observed dynamics are inherent to the "off"-state conformation of rsFolder.

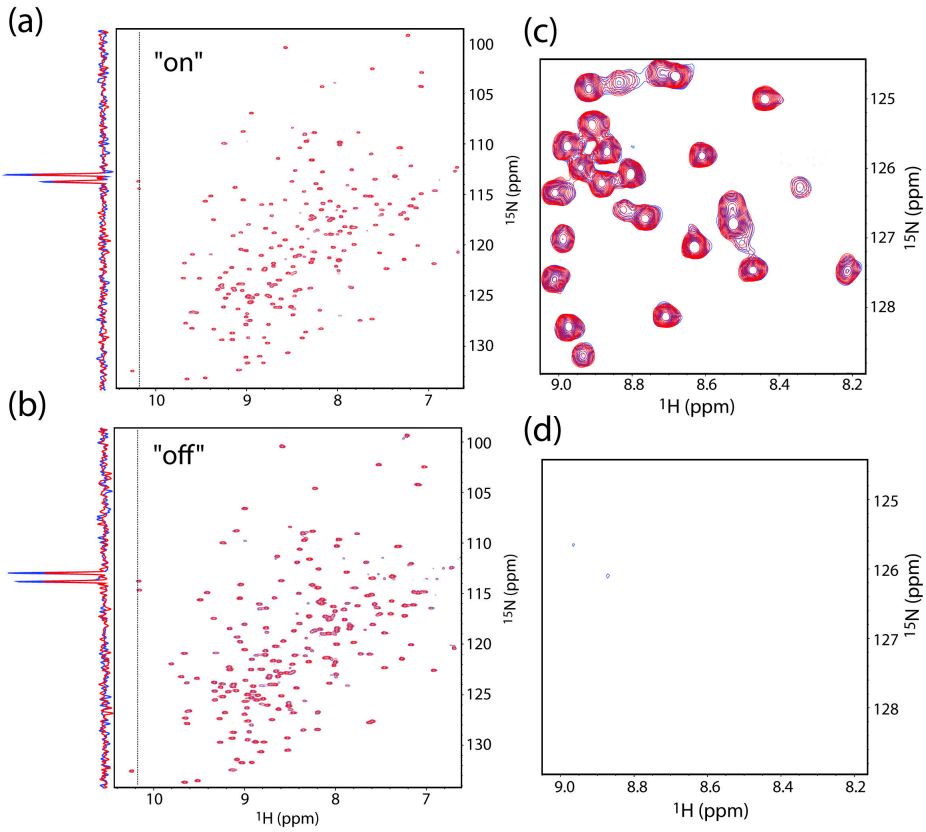


Figure 3.14 (a) Overlaid ^{1}H-^{15}N BEST-TROSY spectra of rsFolder (a)"on'-state and (b) "off-state before and after 1000 photocycles recorded at 850MHz at 40°C. Note that each cycle consisted of a 7s pulse of full-power illumination at 488 nm, followed by 3s full power illumination at 405 nm. On the left of (a) and (b), the overlay of the FT of the Nitrogen dimension for the pair of two peaks in the far left of the amide spectrum. (c) Perfect overlap of peaks of "off"-state spectra acquired with continuous 488nm illumination and in the dark suggest that the "off"-state does not experience changes in its CS or conformational dynamics due to LASER power. (d) Subtracting two "off"-state spectra, one acquired exactly after (in the dark) and during 488nm illumination (at maximum LASER output), there is no resulting signal left. This suggests that rsFolder "off"-state dynamics observed are not a result of "on"-to-"off" population exchange, but rather inherent dynamics of the "off"-state conformation.

3.2.5 NMR-observed thermal relaxation

In an attempt to correlate optical with NMR-observed data, we focused on the thermal relaxation times (ground state "off"-to-"on") of rsFolder, reported to be exceptionally long as compared to other RSFPs [61]. Thermal relaxation was investigated by real-time NMR. Namely, the sample was initially photoswitched to the "off"-state using the 488nm LASER,

after which the LASER was turned off and a series of ^{1}H-^{15}N BEST-TROSY experiments, each one lasting 2.53mins, were recorded at different temperatures. The intensities of peaks originating from the "off"-state, thus, decreased exponentially over time, giving rise to "on"-state peaks instead. When plotting intensities of either state over time, they seem to decrease ("off"-state) and increase ("on"-state) with the same rate, respectively, confirming that the phenomenon observed is "off"-to-"on" thermal relaxation. Furthermore, the rate of thermal relaxation extracted using real-time NMR is comparable to the one reported by optical studies, $k_{thermal} \approx 10^{-4}$ [61]. Minor differences between the two numbers can be explained when considering the effect of temperature on thermal relaxation.

Similarly to the NMR-observed switching contrast, this finding alongside optical studies of thermal relaxation, supports that there is no fluorescence emanating from the "off"-state *trans*-CRO.

By extracting the values of $k_{thermal}$ at different temperatures, we could get insight into the activation energy, E_a, required for the ground state *trans-cis* isomerization of the rsFolder CRO, using equation 2.3, as seen in Figure 3.15. The estimated energy barrier for ground-state isomerization is $E_a = 118 \pm 5kJ/mol$. So far, the only reported values for the ground-state *cis-trans* isomerization available were extracted from NMR studies of a model GFP CRO [95], where they reported a value of approximately $55kJ/mol$. Such discrepancy can be explained by the difference in environmental conditions for the model GFP CRO as compared to the rsFolder CRO burried inside the β-barrel. The model GFP CRO was prepared in ethanol-glass [123], i.e immersed in liquid nitrogen at 77K, in order to assimilate the rigidity of the native CRO inside the GFP barrel. Comparing the two values, the native *trans*-CRO in the rsFolder barrel seems to be more tightly fixed than the model GFP CRO in liquid nitrogen.

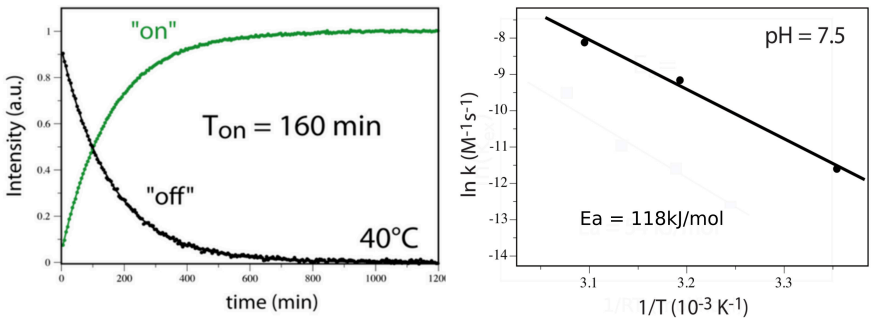


Figure 3.15 Intensity decay (increase) of peak signals corresponding to the "off"-state ("on"-state) as rsFolder thermally relaxes in the dark at 40°C, retrieved from a series of ^{1}H-^{15}N BEST-TROSY experiments over time, recorded at 850MHz at 40°C, on the left. On the right, the Arrhenius plot corresponding to the ground state *trans-cis* isomerization of rsFolder. Data retrieved from thermal relaxation at 10, 20 and 40°C.

3.2.6 Establishing the LASER-Driven EXSY experiment used for backbone assignment of rsFolder "off"-state

As already mentioned in the Materials & Methods part, the characteristic "on"- and "off"-state amide spectra were assigned using conventional 3D triple-resonance experiments for backbone assignment. Nevertheless, the "off"-state amide spectrum assignment proved to be more cumbersome using such 3D experiments, since lower intensity (line broadened) amide peaks observed in the ^{1}H-^{15}N BEST-TROSY spectra of the "off"-state were not detectable in the less sensitive 3D triple-resonance experiments. Therefore, identification of the carbon frequencies and, thus, their assignment was not possible using conventional NMR assignment experiments. This observation is common for residues experiencing intermediate exchange dynamics which, as is discussed further in this chapter, was the case for a number of residues in the "off"-state.

The signal for such amide peaks, despite being low, could still be assigned by making use of the newly established in-situ illumination set-up coupled to the NMR spectrometer. The rationale followed was to bypass the need for a triple-resonance assignment and instead "transfer" the "on"-state amide assignment onto the unknown peaks of the "off"-state amide spectrum.

More specifically, an NMR pulse sequence was created which also included two "LASER channels", so as to photoswitch the protein between the "on"- and the "off"-states. Such an experiment named *LASER-Driven, (LD)-EXSY*, similarly to a SCOTCH experiment [175], triggers the exchange between two conformations of the protein using light during the NMR pulse sequence. In this way, exchange peaks are formed for amides that are experiencing chemical shift changes (as was the case for most line broadened "off"-state peaks) when the protein switches between the two states. Therefore, exchange peaks contain the information about the initial and final CS of an amide pair before and after photoswitching. Using the LD-EXSY experiment, the assignment of the "off"-state increased from 80% to 93% completion, enabling further insight into the protein regions undergoing intermediate exchange dynamics.

In practice, the experiment's signal-to-noise ratio is highly dependent on its ability to photoswitch the sample in a short amount of time. In order for well resolved exchange peaks to be formed, both the "on"- and the "off"-state have to be significantly populated at different moments during the NMR pulse sequence. For this reason, I used the Shigemi tube which has been covered by a reflective material (using a spray) and reduced the concentration to 100μM (sample height 2.5cm, resulting in OD=13.75). The turning on and off of the LASERs was performed digitally, while the total power output was set to 100%. The maximum amount of time of illumination was a compromise between switching efficiency and spin relaxation. Therefore, alongside the tests performed for assessing the performance of the set-up, it was decided to use an illumination time of 2s for the "off"-switching of the protein (i.e 488nm) and 600ms for "on"-switching (i.e 405nm). In this way, the exchange peaks consisted of $\approx 30\%$ of the overall signal.

Exchange peaks arising in the 2D amide spectrum created spectral crowding, which already included both the “on”- and “off”-state peaks. Moreover, correlating exchange peaks to their corresponding “on”- and “off”-state peaks is not straightforward with only the 2D amide spectra at hand. As a consequence, the LD-EXSY was extended to a third dimension, making it possible to discriminate the exchange between the two conformations for each amide peak pair in the 2D spectrum.

Finally, the overall NMR-light pulse sequence used for LD-EXSY comes in two forms; ^{1}H-LD-EXSY and ^{15}N-LD-EXSY associating the hydrogen and nitrogen frequencies of the two states, respectively. Both pulse sequences start with a recycle delay, d1=2s, during which the 488nm LASER switches the majority of the protein sample to the “off”-state, followed by an evolution period t1, during which frequency labelling of the “off”-state hydrogen or nitrogen frequencies, respectively, occurs. Exchange between the two states using the 405nm LASER to switch the proteins to the “on”-state occurs during a mixing time of, d8=600ms, where ^{15}N-magnetization is “stored” along the z-axis (longitudinal spin order) so as to minimize relaxation-induced signal losses. After such mixing step, the pulse sequence continues onto the frequency labelling of the “on”-state amide pair (Figure 3.16(a)). Overall, the resulting 3D spectra (depending on whether it is an ^{1}H-LD-EXSY or an ^{15}N-LD-EXSY) consist of two axis (H, N) corresponding to the “on”-state and one (H or N) originating from the “off”-state CS evolution.

A fast way to plot the data was by creating a projection of the 3D-spectra onto the same type-nuclei dimensions. The latter provides a quick way of distinguishing between residues undergoing CS changes or not after switching, located off-diagonally and diagonally in the spectrum, respectively (Figure 3.16(b), (c)).

3.2.7 Secondary structural elements prediction from NMR CS using the TALOS-N program

Given their sensitivity to local architecture, the CS of backbone ^{1}H, ^{15}N, C_α C_β and *Carbonyl* can be used to calculate backbone torsion angles (ψ and ϕ) which give information regarding the secondary structural elements of a protein. By contrasting them to already published values with experimentally determined structures, the program TALOS-N [196] can predict the secondary structural elements of a protein.

Hence, TALOS-N was used to assess the nature of secondary structural elements of rsFolder in its two ground states in solution at 40°C. The results of such prediction of rsFolder “on”- and “off”-states is shown in Figure 3.17. Firstly, the secondary structural elements in the two states of rsFolder seem to be the same (some information is lacking in the “off”-state due to loss of NMR signal). In its majority, NMR-predicted structural elements are similar to what is expected by crystallography. However, discrepancies are detected in the region of β-7 and the two helices (H2 and H3) inside the barrel which are interrupted by the CRO. In particular, the β-7 strand is extended by 3 residues in solution, which belong to a loop in the crystal structure (146-147-148), the H2 helix is shifted by

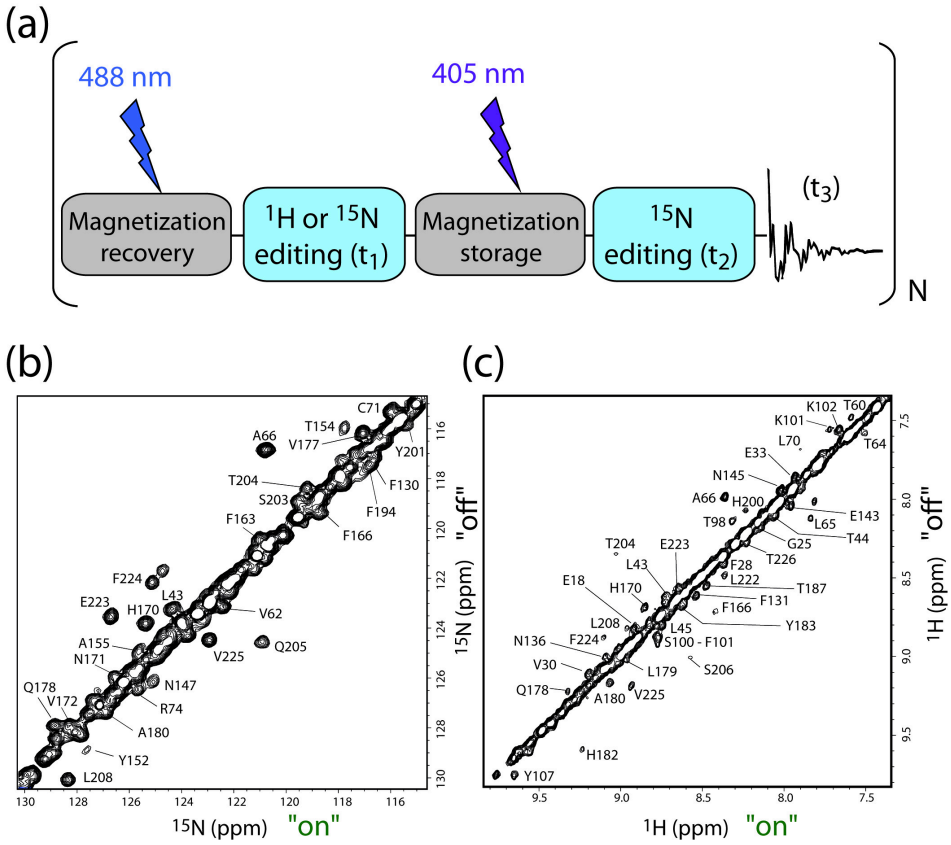


Figure 3.16 (a) Simplified representation of the general principle of LD-EXSY used to assign the backbone amide spectrum of rsFolder "off"-state. The sequence integrates light illumination of the two available channels; a 488nm pulse during the recycle delay lasting 2s, and a 405nm pulse during mixing time whereby magnetization is stored along the for 600 ms. The resulting 3D spectra have 2 "on" state and 1 "off" state chemical shift dimensions. (b) 2D projection of a 3D ^{15}N -edited LD-EXSY spectrum, recorded at 40° at 850MHz. Off-diagonal peaks correspond to rsFolder residues (annotated) which show ^{15}N chemical shift differences between "on"- and "off"-states. (c) Same as (b) but for ^{1}H -edited LD-EXSY

3 residues towards the C-terminal end and H3 helix is not identified as an α-helix using TALOS-N, but rather has a non-canonical shape.

The instability of the β-7 strand has been previously reported in other GFP-type proteins. A crystal structure revealed two conformations for the β-7 strand in ECFP [129], a cyan FP, while Mizuno's work on the Anthozoan Dronpa [146], also revealed increased dynamics in the strands facing the CRO. The shortening of rsFolder's β-7 strand in the crystal structure as compared to the NMR-derived structure can originate from conformational heterogeneity of that region leading to low quality electron densities. Lastly, the atypical shape of H3 in

the crystal structure of rsFolder has been reported in EGFP (an ancestor of rsFolder) as a 3_{10} helix [16].

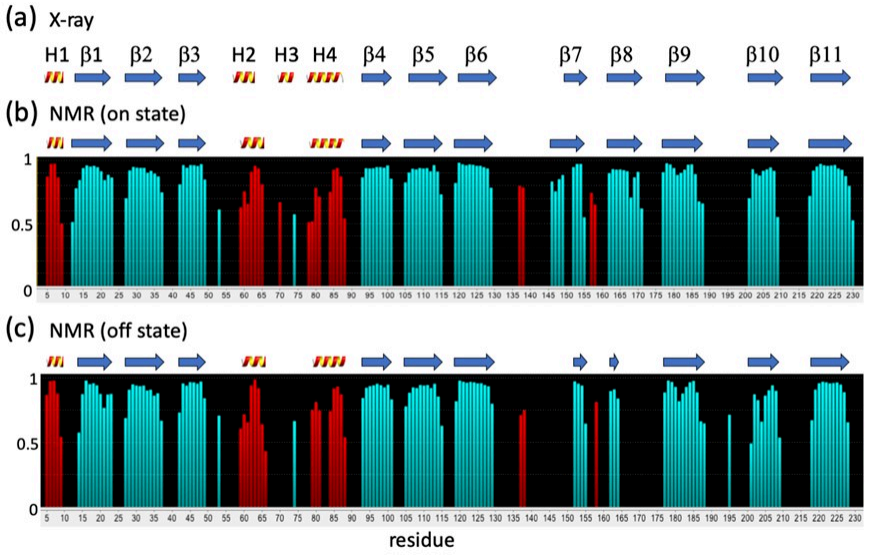


Figure 3.17 Secondary structural elements (α-helix, β-strand) of rsFolder per residue: (a) X-ray structures (PDB: 5DTZ, 5DU0); (b) TALOS-N predictions of the existence of α-helices (red) or β-strand structures (turquoise), calculated from the measured NMR chemical shifts of the "on"-state and at (c) the "off"-state.

3.2.8 Characterization of rsFolder backbone dynamics in the "on"- and "off"-states

With 95% of the assignment completed for rsFolder "on"-state and a total of 93% for the "off"-state, we could proceed to perform NMR relaxation experiments to characterize its dynamics in the dark-adapted and light-induced states.

To start with, a set of ^{15}N-relaxation experiments were performed to obtain values for the global correlation time due to tumbling (i.e τ_c) and to look for regions with significant flexibility. The intensities of only well-resolved amide cross peaks were extracted to estimate the T_1, T_2 and hetNOE values, performing data analysis as explained in the Introduction chapter.

The measured values of the T_1, T_2 and hetNOE are plotted in Figure 3.19(a). From these values, rsFolder appears as a rigid globular protein. Only loops exhibit ps-ns timescale dynamics, while other secondary structural elements are relatively rigid in the context

of the globular motion of the protein. Note that the CRO has not been assigned in this spectrum (since there is no amide pair), therefore its dynamics could not be characterized at this stage.

Using the average values for T_1 and T_2, the τ_c of rsFolder was estimated to be 9.3ns in the "on"- and 9.8ns in the "off"-state, at 40°C (using equation 2.1). Such small increase in τ_c is indicative of the existence of a minority of protein molecules exchanging between a monomeric and an oligomeric state. Alongside the 488nm-induced aggregation observed, this increased tendency for oligomerisation in the "off"-state did not come as a surprise. On the other hand, increased τ_c could be an indication of overall "swelling" of the protein i.e increase of apparent volume, due to dynamics present upon *cis-trans* isomerization.

Nonetheless, while the ^{15}N-relaxation experiments show no evidence of significant dynamics on the ps-ns timescale, there are some important differences to note. These correspond mainly to residues that have been removed from the plots of the "off"-state data due to high error bar values, located at the β-7 and β-8 strand. As evident when comparing the peak intensities of 2D ^{15}N-^{1}H BEST-TROSY spectra of the two states (Figure 3.18), line broadened peaks in the "off"-state do not have sufficient signal quality and, thus, it was not possible to extract T_1, T_2 and hetNOE values accurately (Figure 3.19(b)). Extensive line broadening is, however, an indication of μs-ms timescale dynamics experienced by these residues, which are not present in the "on"-state conformation. Supporting this view, the T2 values of residues C71 and S73, belonging to the non-canonical H2 helix are decreased in the "off"-state, as compared to the "on"-state.

In an attempt to quantify the observed intermediate exchange kinetics of the backbone of the "off"-state, ^{15}N-CPMG-RD was performed. Unfortunately, the data quality was not optimal rendering further quantification of exchange rates and populations impossible. Nevertheless, once comparing the plot of R_2^{eff} values at different CPMG frequencies, ν_{CPMG}, for residue S73 (the only one of reasonable quality) in the two different states, it becomes evident that rsFolder upon 488nm illumination experiences μs-ms dynamics which were not present in the "on"-state (Figure 3.19(c)).

Interestingly, residue S73 lies on one of the helices capping the CRO. It is, therefore, possible that this residue is close enough to "witness" what, at this point, is suspected to be *trans*-CRO-induced dynamics, but not as close as other residues (eg N147) whose NMR signal is almost completely vanished in the "off"-state. Nevertheless, each residue, including S73, experiences a combination of its own and nearby residues' dynamics, including the CRO, making it impossible to decipher where the motions originate from.

Significant chemical shift changes are also observed for residues near the region experiencing intermediate exchange dynamics(Figure 3.18). Hence, the observed lack of assignment, CS changes, line broadening of peaks and lowered T2 values are evidence of changes in dynamics experienced for the two states of rsFolder. These are summarised in the figure below (Figure 3.20). Intermediate exchange dynamics felt by the "off"-state conformation are present mainly in the β-7, 8, 9, 10, 11 strands, all facing the phenol ring of the CRO. On the contrary, the "on"-state shows such dynamics occurring only very locally

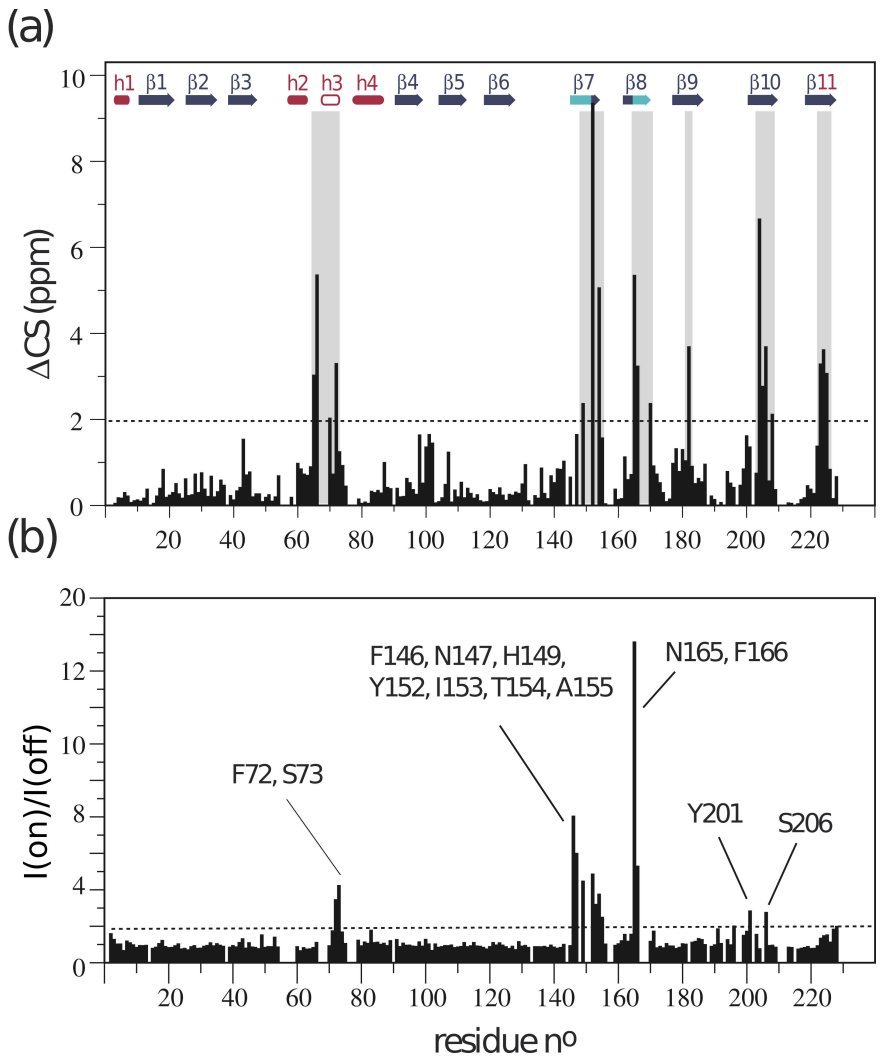


Figure 3.18 Quantitative comparison of 2D ^{1}H-^{15}N correlation spectra for the "on"- and "off"-states of rsFolder (a) The chemical shift difference per residue as extracted from the peak positions of residues. To calculate the ΔCS, the following equation was used: $\Delta CS = \sqrt{(\Delta N)^2 + (10\Delta H)^2}$. On top, the predicted secondary structural elements are depicted (b) The ratio of peak intensities per residue are plotted. Annotated are residues that experience significant line broadening upon *cis-trans* isomerization.

in the β-7, 8 strands.

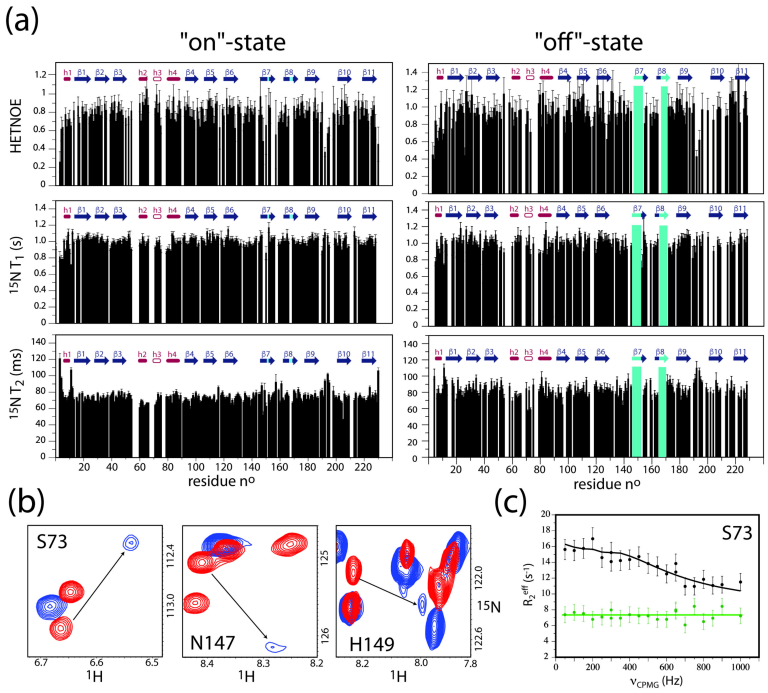


Figure 3.19 ^{15}N relaxation T_1, T_2, hetNOE values for rsFolder "on"- and "off"-states measured at 700MHz at 40°C, with annotated TALOS-N prediction of secondary structural elements on the top. In turquoise, the regions were no peaks are detected in the "off"-state. In (b), examples of severe line-broadening upon *cis-trans* isomerization, (red – "on"-state, blue – "off"-state) as detected by ^1H-^{15}N BEST-TROSY experiments at 850MHz at 40°C. (c) CPMG-RD data recorded for residue S73 at 40° at 850MHz in the "on"- (green) and "off"-(black) states, fitted using the Chemex algorithm. Estimated exchange rate $k_{ex} = 160 \pm 400 s^{-1}$.

3.2.9 The "off"-state shows increased water accessibility and barrel instability

HDX experiments were performed to complement the dynamic view obtained by NMR spin relaxation methods. At neutral or acidic pH, the formation of the H-bond after transient opening is generally much faster than the intrinsic chemical exchange process. In this so-called EX2 exchange regime, the exchange rate constants provide a measure of the exchange-competent state population (p_{open}). Therefore, when p_{open} is increased (directly proportional to the H/D exchange rate observed, k_{obs}), it is either due to a more "flexible" NH bond (due to motions experienced by the residue in question) or due to an increase in water accessibility. In either case, HDX assesses the fold stability of the protein.

For rsFolder, the H/D exchange rates of the "on"- and the "off"-states for a number of residues were different. More specifically, the sample was lyophilised and then immersed in D_2O buffer to initiate the H/D exchange. During the first 5 minutes required for

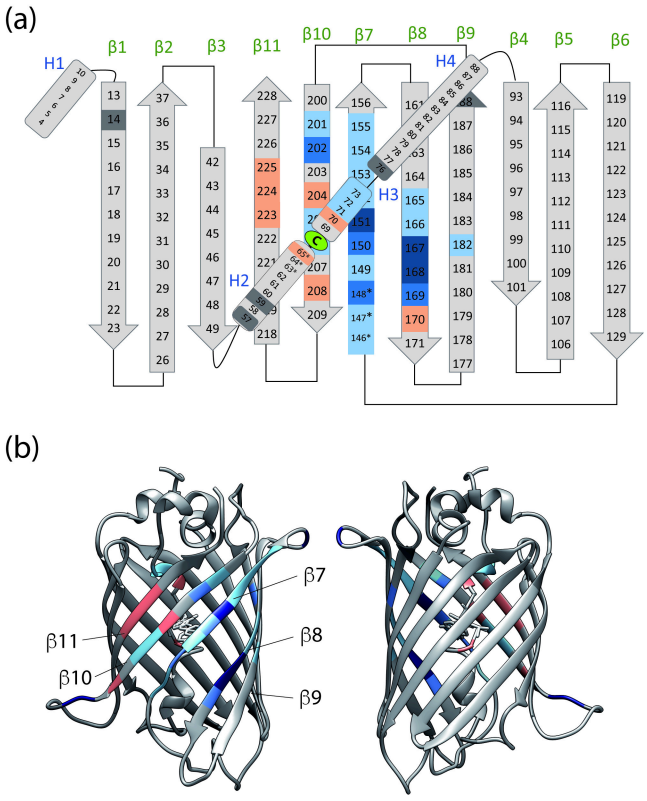


Figure 3.20 Increased conformational dynamics on the μs-ms timescale in the "off" state of rsFolder. The NMR-derived information is color-coded on the (a) open-sheet and (b) 3-dimensional representation of the rsFolder structure as revealed by X-ray crystallopgraphy and NMR (A '*' indicates structural elements only identified by NMR): (i) undetected/unassigned amides in both states (dark blue), (ii) undetected/unassigned amide peaks in the "off"-state only (middle blue), (iii) strongly line-broadened amide NMR signals in the "off"-state (light blue), (iv) amide ^{1}H-^{15}N with large chemical shift changes between "on"- and "off"-states (orange). In addition, proline residues lacking an amide proton are color coded in dark grey.

sample preparation (the sample was either taken from the dark, for the "on"-state, or was illuminated with 488nm-illumination before NMR acquisition, for the "off"-state), insertion in the NMR spectrometer and instrument calibration, some amide hydrogens had been already replaced by deuterium, as evident by the complete disappearance of the corresponding peaks in the first recorded ^{1}H-^{15}N BEST-TROSY of the real-time series. This was the case for the loops and the β-7 and 8 strands of both "on"- and "off"-state. Typically, loops are rapidly exchanging with the solvent, while the increased exchange rate of the β-7 strand can be attributed to its well-known lack of H-bonding with the neighbouring β-8 strand, causing it to present more flexibility [129, 221, 193, 158].

Interestingly, the apparent exchange rates were drastically altered for detectable peaks,

increasing by a factor of 4-100 in the light-induced "off"-state. Helices H2, H3, in particular, showed, on average, a 100-fold increase in solvent accessibility in the "off"-state. The summarised results are shown in Figure 3.21.

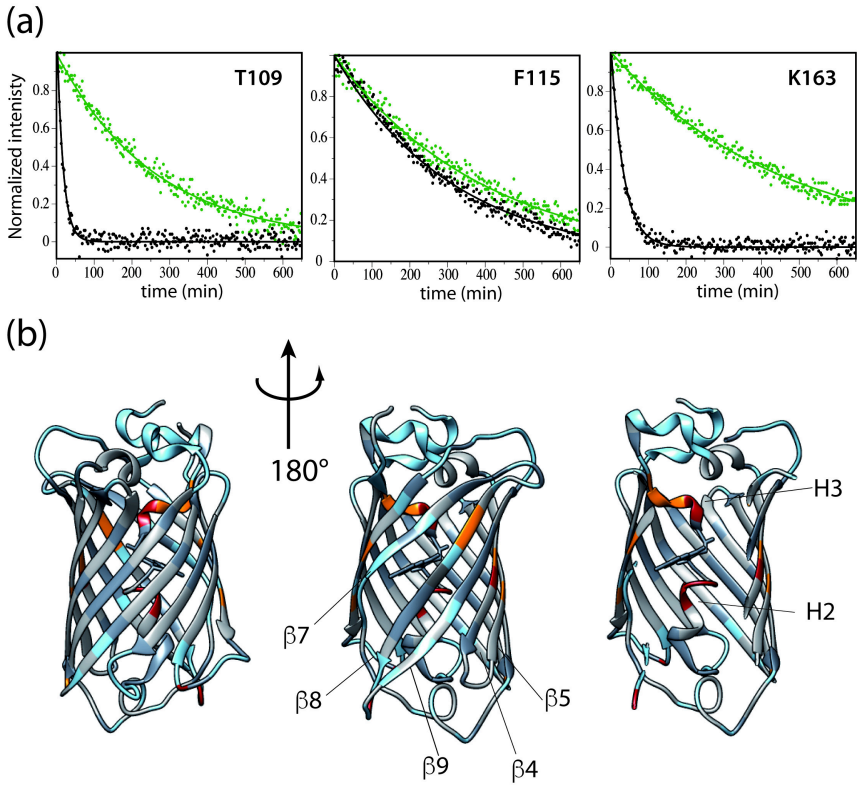


Figure 3.21 HDX NMR data. (a) Representative intensity decay for individual amide hydrogens extracted from real-time 2D NMR data recorded for the "on"- (green) and "off"- (black) states of rsFolder recorded at 40°C at 850MHz. (b) The derived H/D exchange information is color-coded on the crystal structure of rsFolder: (i) residues with amide hydrogen exchange time constants $T_{ex} < 5$ min in both the "on" and off" states (light blue); (ii) residues for which the hydrogen exchange rate is increased by a factor of 4 -20 (orange); (iii) residues for which the hydrogen exchange rate is increased by more than a factor of 20 (red); (iv) residues without a significant change in exchange rate between "on"- and "off"-states (grey); (v) no information due to peak overlap, proline, or unassigned residues in one of the two states (dark grey).

Nevertheless, when investigating the two crystal structures corresponding to the two states, using CAVER (a PyMol plug-in), water channels that could be potentially formed in the protein do not differ significantly (Figure 3.22). Crystal structures can not explain such increase in the *average* H/D exchange rates upon 488nm illumination, when all residues are considered. Therefore, our results using the HDX experiment, could be indicative

of a favouring of a solvent-accessible population in the “off”- as compared to the “on”-state. Interestingly, the fastest exchanging residues belong to the helices docking the CRO, hinting increased dynamics in the *trans*-CRO environment. Additionally, alongside the data concerning “off”-state dynamics mentioned previously, the β-barrel appears to be generally more unstable upon 488nm illumination. Lastly, differences between the dynamics detected in the ^{15}N-relaxation and HDX experiments originate from the sensitivity of each experiment to given processes happening at different timescales ($ps - ms$ or $s - min$).

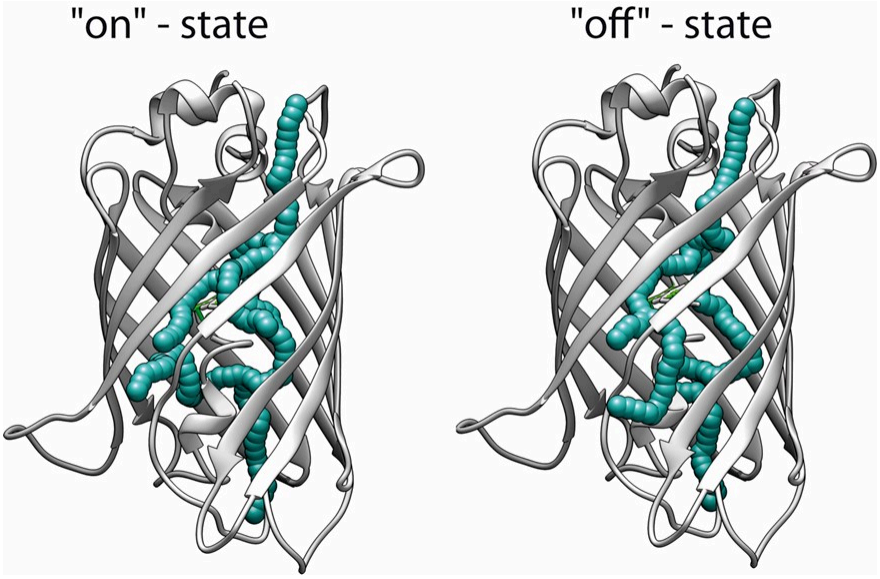


Figure 3.22 3D representations of rsFolder in the “on” and “off” states (crystal structures PDB:5DTZ, 5DU0). Water accessibility channels in the β-barrel structure, as calculated with the PyMol CAVER 3.0 plugin are highlighted by cyan balls.

3.2.10 Characterization of rsFolder side-chain and chromophore dynamics in the “on”- and “off”-states

Methyl side chain assignment and dynamics

In an effort to assign the methyl side chains of rsFolder in its two ground states, a variety of NMR pulse sequences and sample preparation protocols were adapted, as described in the Materials $\&$ Methods section. While incomplete, the assignment of the most “interesting” peaks, i.e peaks undergoing change (CS or line broadening) upon *cis-trans* isomerization, allowed for the characterization of dynamics mainly felt by residues in the vicinity of the CRO.

By overlaying the methyl spectra acquired using the ^{1}H-^{13}C SOFAST-HMQC NMR pulse sequence of the sample in the “on”- and “off”-states, severe line broadening was evident for

residues V62, T63, L70, V151, I168, T204 and V225 upon 488nm illumination (Figure 3.23). When localised in the crystal structure of rsFolder "off"-state, all of these side-chains are in the close vicinity of the phenol ring of the *trans*-CRO. Hence, our data suggest intermediate exchange dynamics present in the vicinity of the *trans*-CRO – CS modulation occurring at the timescale of a few hundreds of μs to ms can originate either from motions of the *trans*-CRO and/or conformational dynamics of the side-chains in the "pocket" of the CRO. Nevertheless, methyl side-chain dynamics further support the view of increased exchange dynamics upon *cis-trans* isomerization. Note that for T204, crystal structures suggest that upon *cis-trans* isomerization, its side chain loses the H-bond with the CRO, potentially due to or being the cause of the NMR-observed increased dynamics.

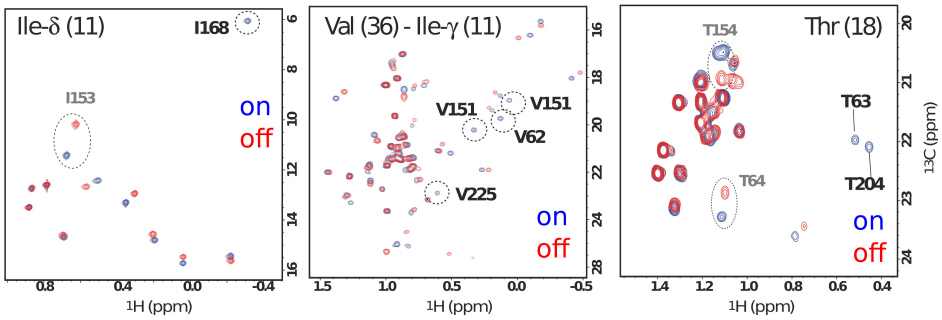


Figure 3.23 Spectral regions of 2D ^{1}H-^{13}C SOFAST-HMQC spectra, recorded at 850MHz at 40°C. On the left, focusing on the Isoleucine-δ region of the "on"- and "off"-state spectra of an rsFolder sample labelled specifically on the Valine-γ and Isoleucine-δ Pro-S carbons. The peak of I168 in the "off"-state is not visible with the chosen counter levels, due to line broadening. In the middle and right, amino-acid edited ^{1}H-^{13}C SOFAST-HMQC spectra of the two states of rsFolder are shown. The peaks of V151, V225, V62, T63, T204 in the "off"-state are not visible with the chosen counter levels.

At this point, it was interesting to further quantify such dynamics using CPMG-RD but this time focusing on $^{13}C_{methyl}$ relaxation rather than ^{15}N as before. More specifically there were two peaks (L70 and I168) identified which could potentially yield some quantitative information concerning chemical exchange. However, their signal-to-noise ratio at 40°C was suboptimal, thus more temperatures were tested, namely 30°C, 50°C and 60°C. Moreover, the sample chosen was specifically labelled in Ile-δ, Val-Leu-γ-ProS and deuterated, thus ensuring higher spectral resolution (less spectral crowding) and signal-to-noise ratio (deuteration suppresses NMR signal absorption from the solvent, originating from solvent hydrogens). Finally, the $^{13}C_{methyl}$-CPMG-RD was chosen at 60°C to run for approximately two days. Unfortunately, the protein aggregated while the experiment was running overnight, thus the experiments could not continue. Nonetheless, it should not come as a surprise that during continuous illumination at 60°C, even rsFolder reaches its limits.

Assignment of aromatic side chain and chromophore NMR spectra

The assignment of aromatic moieties was aided by the development of new Tyrosine and Histidine specific BEST and SOFAST NMR pulse sequences [43]. The motivation behind designing such pulse sequences was mainly the numerous aromatics (in total 33) present in the rsFolder protein sequence but also the optimization of NMR pulse sequences for aromatics assignment.

Firstly, a 2D Aromatic *Constant-Time, CT* ^{1}H-^{13}C HSQC which included the BEST element to ensure signal enhancement was created. By using band-selective pulses, the recycle delay could be set to 300ms (as compared to the conventional 1s delay), thus allowing for more repetitions of the pulse sequence within a given amount of time. Such a spectrum not only provides sharp peaks but also allows for the distinguishing between carbons having an odd or an even number of attached carbons due to the choice of constant time ^{13}C-editing. In this way, for example, Histidine $C_{\epsilon 1}$ having a negative signal could be discerned from Phenylalanines C_{δ} with a positive signal which resonate at similar frequencies.

Particularly in the case of rsFolder, it was important to investigate the frequencies of nuclei in Tyrosine and Histidine rings since the CRO consists of a Tyrosine ring while a Histidine near the CRO (H149) has been reported to play a central role in "rigidifying" the CRO [61]. Hence, Tyrosine and Histidine specific 2D ^{1}H-^{13}C BEST correlation experiments were performed to assign all nuclei present in Tyrosine and Histidine aromatic side chains. Each one of these experiments included magnetization transfer steps via J-couplings.

Consequently, I made use of the newly designed BEST-HDCG, BEST-HDCB, BEST-H(C)N to retrieve the cross-peaks of H_{δ}-C_{γ}, H_{δ}-C_{β} and correlate carbon-bound hydrogens (H_{δ} and H_{ϵ}) to aromatic nitrogens ($N_{\delta 1}$, $N_{\epsilon 2}$) of Histidines, respectively (Figure 3.24).

Furthermore, a long-range SOFAST $H^{C}N$ experiment complemented the spectra acquired using the BEST-H(C)N. More specifically, the SOFAST experiment also enables shorter acquisition times due to lower d_1 interscan delay times but does so not only by the use of the BEST element, but also by adjusting the nutation angle of the excitation pulse to the Ernst angle, thus enabling ^{1}H steady state polarization enhancement [187]. The experiment correlates the ring carbon-bound hydrogens to ring nitrogens 2-3 bonds away. Overall, the BEST H(C)N detects the cross-peaks of protonated ring nitrogens, resonating at higher frequencies (155-185ppm) while the SOFAST $H^{C}N$ is more sensitive to deprotonated nitrogens of Histidines arising in the range of 220-255ppm. Since the BEST-H(C)N experiment transfers magnetization via the ring carbons, a strong J_{CN} coupling of proton-bound nitrogens results in its characteristic sensitivity.

Finally, overlaying the spectra acquired from BEST-H(C)N and SOFAST $H^{C}N$, alongside the spectra acquired using the Histidine-specific 2D BEST correlation sequences and the negative peaks arising in the 2D CT ^{1}H-^{13}C BEST HSQC, all 11 Histidines with their respective ring carbon and nitrogen frequencies could be detected and assigned.

Similarly to Histidines, 2D ^{1}H-^{13}C BEST correlation sequences were written to identify

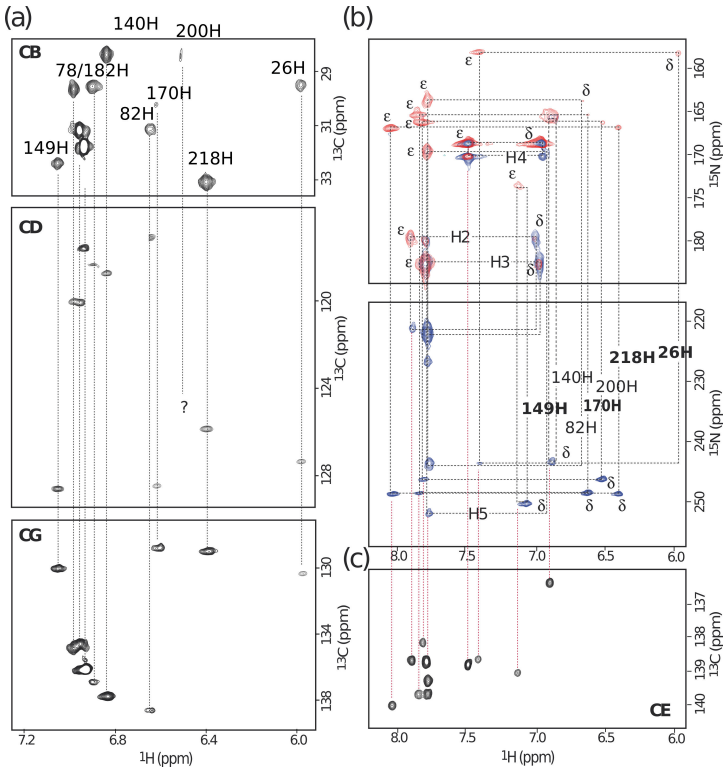


Figure 3.24 Compilation of Histidine-specific 2D correlation spectra, recorded at 850MHz at 40°C, to assign the aromatic side chain hyrogens, carbons and nitrogens of Histidines in rsFolder "on"-state. (a) From top to bottom: BEST-HDCB, negative signals of 2D Arom ^{1}H-^{13}C CT-edited BEST-HSQC and BEST-HDCG. (b) Top: BEST-H(C)N, bottom: SOFAST $H^{C}N$ and (c) 2D Arom ^{1}H-^{13}C CT-edited BEST-HSQC, focusing on the region of H_{ϵ}-$C_{\epsilon 1}$.

the frequencies of ring hydrogens and carbons in Tyrosines. In brief, correlation was possible between H_{δ}-C_{γ}, H_{δ}-C_{β}, H_{ϵ}-C_{δ}, H_{ϵ}-C_{γ} and H_{ϵ}-C_{z} leading to the assignment of most Tyrosines present, including the CRO.(Figure 3.25).

A 2D ^{1}H-^{13}C BEST HSQC without CT-editing gave rise to a new set of low-intensity cross-peaks corresponding to the H_{δ}-C_{δ} and H_{ϵ}-C_{ϵ} of the phenol ring and the H_{β}-C_{β} of the methine bridge of the CRO in the "on"-state i.e its *cis* conformation. These peaks were not visible in the CT version of the pulse sequence, since CT-editing favours resolution but not sensitivity and thus were not part of the assignment shown in Figure 3.25. Using a series of BEST experiments correlating Tyrosine H_{ϵ}-C_{δ}, H_{δ}-C_{ϵ} and H_{ϵ}-C_{z}, alongside the 2D ^{1}H-^{13}C BEST HSQC without CT-editing (Figure 3.26), allowed for the assignment of the ring hydrogen-carbon pairs of the phenol ring of the *cis*-CRO (Table 3.1). The corresponding peaks in the "off"-state were not detectable.

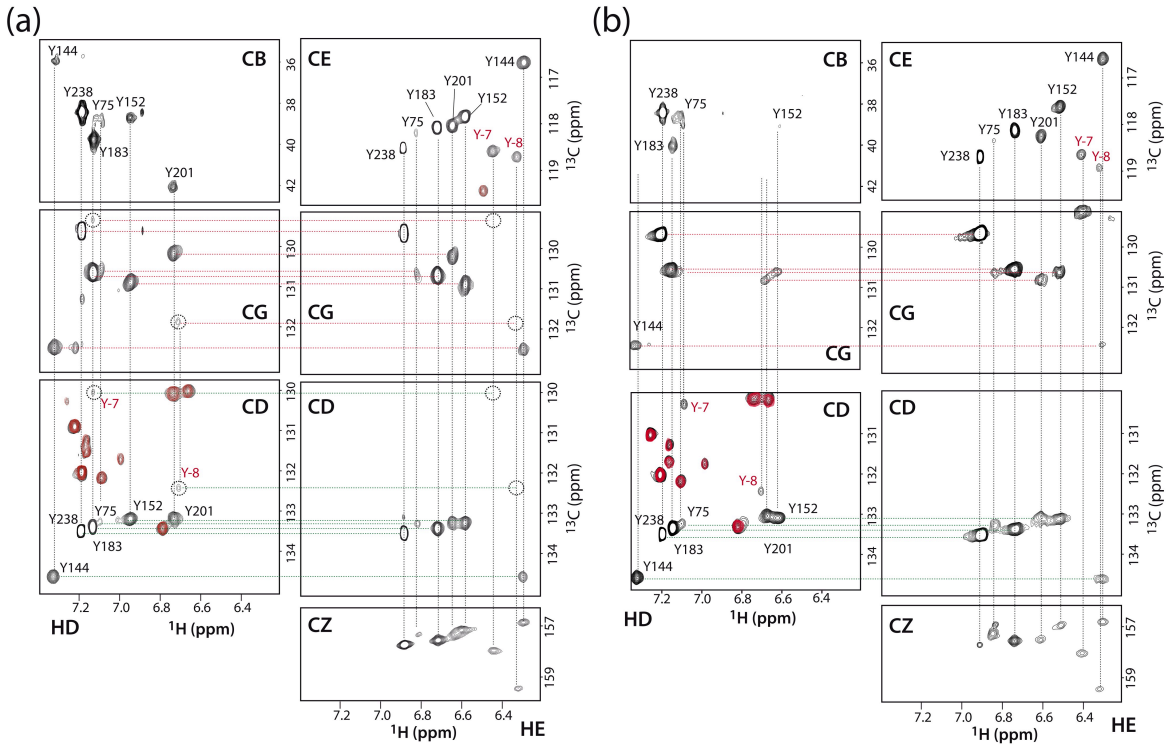


Figure 3.25 Compilation of Tyrosine-specific 2D correlation spectra, recorded at 850MHz at 40°C, to assign the aromatic side chain hyrogens, carbons and nitrogens of Tyrosines in rsFolder "on"- (a) and "off"-state (b). Left column: BEST-H_δ-C_β (on top), BEST-H_δ-C_γ (center), and BEST CT ^1H-^{13}C HSQC – H_δ-C_δ region (bottom). Right column: CT ^1H-^{13}C HSQC – H_ϵ-C_ϵ region (on top), BEST-H_ϵ-C_γ (center), and BEST-H_ϵ-C_δ, BEST-H_ϵ-C_z (bottom). In addition, the CT ^1H-^{13}C HSQC spectral regions recorded for a tyrosine unlabeled sample of rsFolder are plotted on top (red contours). Spectral connectivities are highlighted by dashed lines, and tyrosine ring ^1H-^{13}C pairs are assigned. We could not assign two tyrosines (Y7, Y8), due to the lack of signal in the BEST-H_δ-C_β. Lack of signal is probably due to signal overlap with another peak or lack of sensitivity. Unfortunately, peak intensity in this experiment is inherently inhomogeneous, thus, we cannot conclude on dynamics of the residue affecting its NMR signal quality. Dashed circles highlight the position of peaks that are close to the noise level, and that are fairly or not visible in this figure (chosen contour levels).

Aromatic side chain dynamics

After having secured the assignment, residues experiencing line broadening upon *cis-trans* isomerization were discerned. In particular, H149 shows signs of severe line broadening,

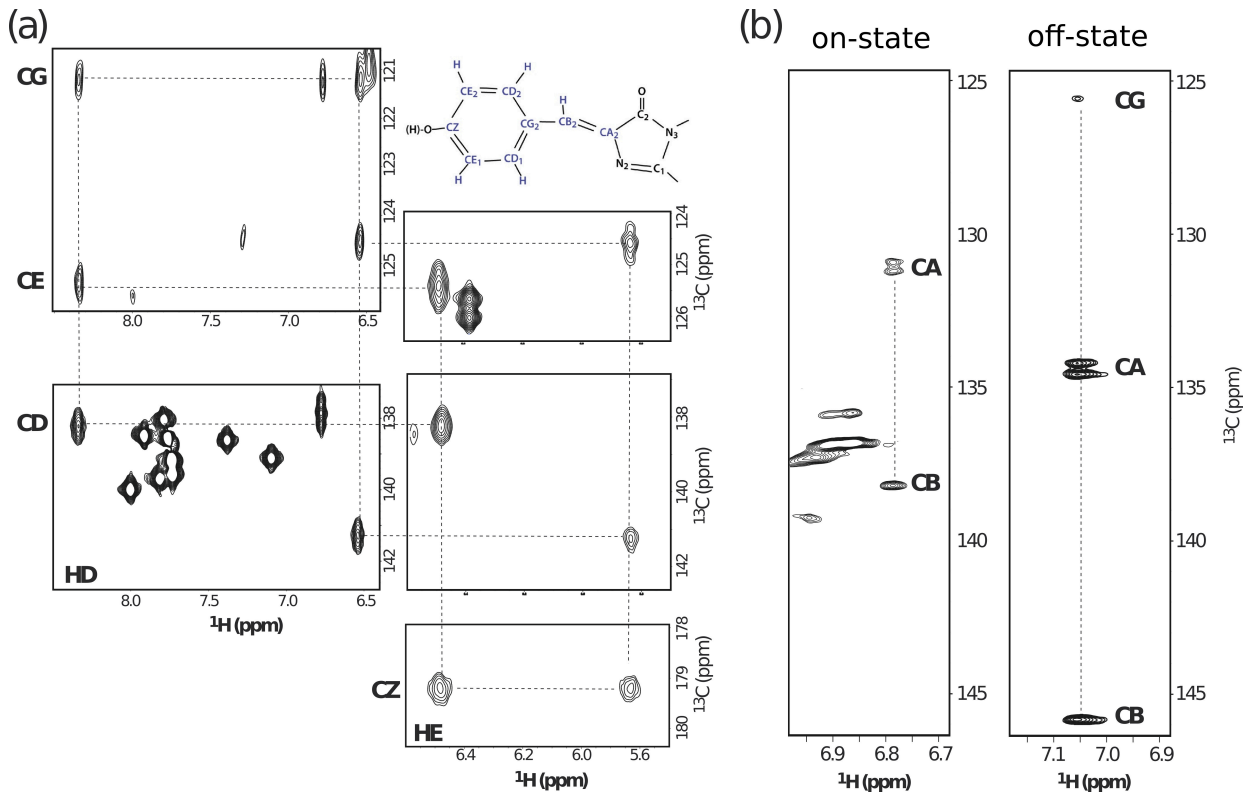


Figure 3.26 (a) Compilation of Tyrosine-specific 2D correlation spectra, recorded at 850MHz at 40°C, to assign the NMR chemical shifts of rsFolder "on"-state CRO's tyrosine ring. The chemical structure of the CRO with the corresponding ring and methine carbons annotated is shown. Spectral connectivities are highlighted by dashed lines. The appearance of two peaks in each spectrum is indicative of motions taking place in the slow exchange regime. (b) 2D HB(CB)CACG spectra recorded for the "on"- (left) and "off"-(right) states of rsFolder, correlating the HB-CB with the neighboring CA and CG of the methine bridge. Note that the correlation peak with CA appears as a doublet, because the neighboring CO is not decoupled. This allows straightforward assignment of the two cross peaks.

while the previously mentioned CRO peaks (H_δ-$C_{\delta 2}$ and H_ϵ-$C_{\epsilon 1}$) are not detectable in the "off"-state (Figure 3.27). Alongside T204, H149 is suggested by crystallography to be one of the H-bonding partners with the CRO in the "on"-state. However, unlike T204, the H-bond of H149 with the CRO is expected to be maintained even in the "off"-state. Our NMR data suggest that H149, similarly to T204, experiences increased dynamics upon *cis-trans* isomerization.

Another important residue is F146. In rsEGFP2, this residue is replaced by a hydrophilic Tyrosine which can readily bind to H149 upon "on"-to-"off" photoswitching. In contrast, in rsFolder, H149 in the "off"-state is expected to be H-bonded with the CRO since it cannot bind to the hydrophobic F146. This H-bond is suggested to be the reason for the

NUCLEUS	"ON"-STATE CS (ppm)	"OFF"-STATE CS (ppm)
C_α	131.1	134.4
C_β(H_β)	137.9(6.79)	145.5(7.05)
C_γ	121.2	125.6
$C_{\delta 1}$($H_{\delta 1}$)	138.2(8.34)	n.d.*
$C_{\delta 2}$($H_{\delta 2}$)	141.3(6.54)	n.d.
$C_{\epsilon 1}$($H_{\epsilon 1}$)	125.5(6.48)	n.d.
$C_{\epsilon 2}$($H_{\epsilon 2}$)	124.6(5.84)	n.d.
C_z	179.2	n.d.

Table 3.1 Chromophore NMR chemical shifts. * not detectable signal

exceptionally high thermal stability of rsFolder's "off"-state, which partly "rigidifies" the *trans*-CRO. As observed by NMR, F146 undergoes a significant CS change, but it maintains its linewidth, indicating that it is very little affected by the dynamics present.

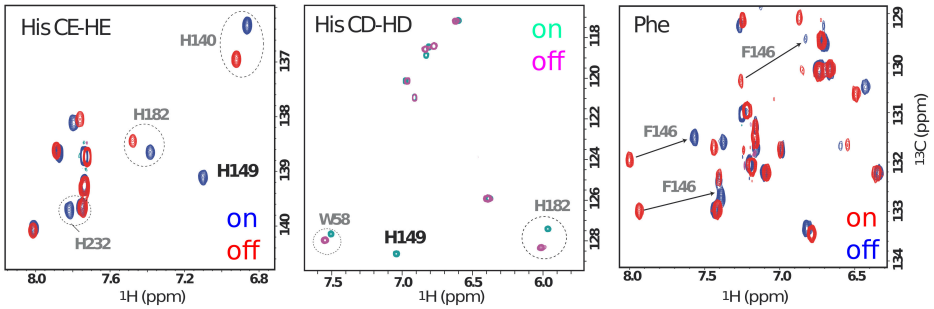


Figure 3.27 Spectral regions of 2D ^1H-^{13}C Aromatic BEST-HSQC spectra, recorded at 850MHz at 40°C. On the left, focusing on the H_ϵ-$C_{\epsilon 1}$ region of the "on"- and "off"-state spectra of rsFolder. The peak of H149 in the "off"-state is not visible with the chosen counter levels, due to line broadening. In the middle, showing the negative NMR peaks of the spectra, focusing on the region of H_δ-$C_{\delta 2}$, where the H149 in the "off"-state is not visible. On the right, showing the CS changes of F146 upon 488nm illumination.

Therefore, the final figure of side chain dynamics of the light-induced "off"-state (including the methyls) is summarised in the figure below (Figure 3.28). As concluded already from the methyl dynamics, motions in the hundreds of μs-ms timescale appear to take place in the pocket of the *trans*-CRO, either due to the CRO itself and/or the nearby side chain residues.

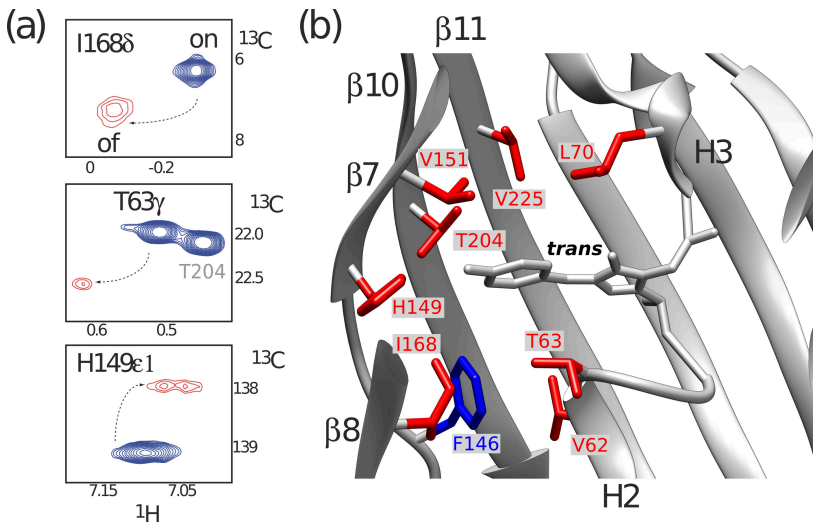


Figure 3.28 rsFolder side chains involved in conformational exchange processes in the "off"state as deduced from extensive line broadening. Examples of peak line broadening observed in the 2D ^{1}H-^{13}C correlation spectra, recorded at 850MHz at 40°C, of rsFolder: "on"-state (blue), "off"-state (red). (b) Zooming in the CRO environment highlighting the side chains which show NMR line broadening in the "off"-state (red). In addition, F146 is plotted as it experiences large chemical shift changes upon light illumination.

Chromophore investigations using NMR

Most importantly, the assignment of the CRO was particularly instructive (Table 3.1). So far, NMR spectroscopists have been able to detect specific frequencies of the CRO, mainly in 1D NMR [158, 1, 218]. Contrary to what is available in the literature, we could achieve an almost complete assignment, including the H_β-C_β of the methine bridge, the H_δ-C_δ, H_ϵ-C_ϵ and C_z of the phenol ring. Consequently, we could gain insight on the configuration and mechanistic properties of the CRO of rsFolder in both ground states.

First of all, the C_z frequency in the "on"-state is that of a deprotonated tyrosine, similarly to the NMR studies on Dronpa [146]. Thus, NMR can confirm the existence of a *deprotonated* CRO in the "on"-state of rsFolder, at pH=7.5. Unfortunately, the C_z frequency of the "off"-state conformation was not visible, most likely due to the existence of intermediate exchange dynamics causing line broadening.

Fortunately, the H_β-C_β were detectable in both states, suggesting that the methine bridge of the CRO remains stable upon *cis-trans* isomerization, despite the motions potentially experienced by the phenol ring. Moreover, the H_β-C_β cross-peaks of the "on"- and "'off"-states are highly distinct. In agreement with other NMR studies as well as DFT simulations [1, 218], these peaks are characteristic of a *cis* and a *trans*-CRO, respectively.

Lastly, the H_δ-C_δ and H_ϵ-C_ϵ peaks detected corresponding to the phenol ring of the

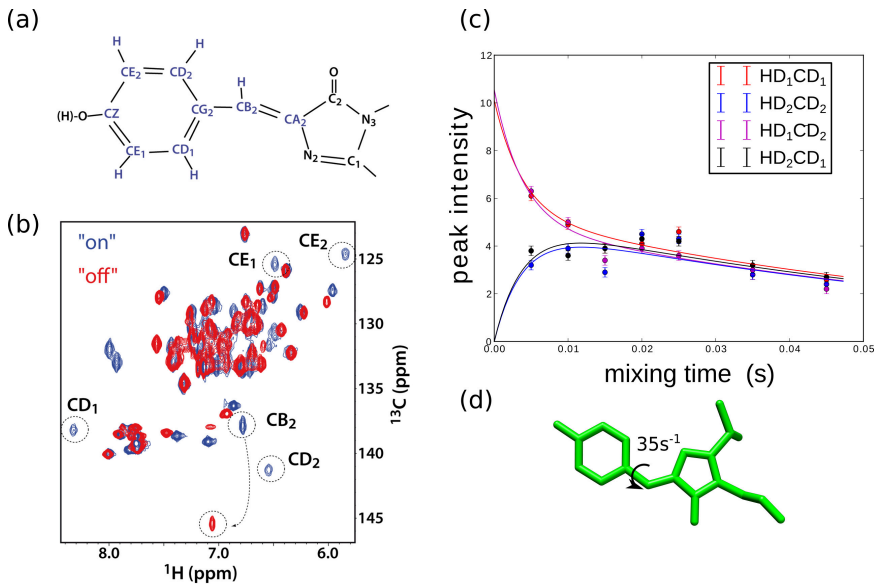


Figure 3.29 (a) Schematic representation of the chemical structure of the CRO. Nuclei giving rise to NMR signals are highlighted by blue letters. (d) Overlay of 2D aromatic ^{1}H-^{13}C BEST-HSQC spectrum of rsFolder, recorded at 850MHz at 40°C without CT-editing to achieve highest experimental sensitivity, in the "on"- (blue) and "off"- (red) state. NMR signals from the tyrosine moiety (Y63) of the CRO are annotated. (c) Fitting of peak intensities against mixing time using the equation of 2.2, taking into account the major and minor peaks formed during the aromatic ^{1}H-^{13}C EXSY experiment in a sample of rsFolder "on"-state, recorded at 40°C at 850MHz. (d) A model of the *cis*-CRO flipping at $71 \pm 3s^{-1}$.

"on"-state *cis*-CRO come in pairs (i.e corresponding to $C_{\delta 1}$, $C_{\delta 2}$ and $C_{\epsilon 1}$, $C_{\epsilon 2}$ frequencies) reflecting slow ring flips (ms-s timescale) which give rise to distinct CS for the two rotamers of the ring. At this point, aromatic ^{1}H-^{13}C EXSY experiments using different exchange times were employed to quantify the CRO ring flip rate, k_{rf}. The resulting intensities of diagonal and cross peaks against mixing times was plotted and fitted using equation 2.2 described in the Materials & Methods, to conclude that the CRO in the "on"-state at 40°C appears to undergo 35 ± 3 flips per second (Figure 3.29 (c),(d)). NMR's ability to quantify ring flips can be useful when assessing the rigidity of the CRO, an intrinsic property defining its fluorescence QY. In the case of the "off"-state, ring flips occurring at μs-ms timescale could be the origin of peak disappearance. However, overall intermediate exchange dynamics experienced by the phenol ring and/or neighbouring residues could also be the cause for "off"-state H_{δ}-C_{δ} and H_{ϵ}-C_{ϵ} peak line broadening. Thus, while at this point it is not possible to clearly identify the origin of line broadening in the NMR signals of the *trans*-CRO, it is clear that the *trans*-CRO is experiencing intermediate exchange dynamics, unlike the *cis*-CRO which appears to be more stable.

Histidine 149 protonation state as seen by NMR

In the crystal structures of rsFolder, H149 appears to form an H-bond with the phenoxy oxygen of the CRO in both the "on"- and the "off"-state. However, crystal packing and cryogenic temperatures might be distorting the proximity and, thus, existence of an H-bond between those two sites. Thus, NMR provided a way to inspect the state of H149 at physiological conditions.

The set of experiments used for assigning Histidines also provided information regarding their protonation state. As described previously, BEST H(C)N acquired spectra overlayed with SOFAST H^CN discriminates between protonated and deprotonated nitrogens. Hence, for nitrogen frequencies of both ring nitrogens of a Histidine to be similar, the Histidine needs to be biprotonated. If not, the Histidine ring is only singly protonated. Furthermore, the ppm difference between C_γ and $C_{\delta 2}$ is sensitive to whether the Histidine is protonated in the $N_{\delta 1}$ or the $N_{\epsilon 2}$, while the $C_{\epsilon 1}$ reports on the charged state (singly or doubly protonated). Therefore, by investigating the frequencies extracted after assignment, we could conclude that H149 in the "on"-state is protonated at the $N_{\delta 1}$ position, suggested by the crystal structure (using the PDB refinement model) to be the nitrogen forming the H-bond with the phenolate ring of the CRO. Unfortunately, due to line broadening in the "off"-state, the protonation state of H149 could not be concluded. Nevertheless, the $C_{\epsilon 1}$ frequency of H149 does not significantly change upon *cis-trans* isomerisation, an indication that H149 remains singly protonated in the "off"-state at pH=7.5.

Furthermore, a 2D ^1H-^{15}N SOFAST HMQC was recorded, aimed at extracting the frequencies of protonated nitrogens of Histidine rings involved in a hydrogen bond. More specifically, the hydrogen involved in a hydrogen bond with the imidazole nitrogen is expected to be protected from solvent exchange. An NMR signal is, therefore, expected to appear in these 2D ^1H-^{15}N SOFAST HMQC spectra as long as the hydrogen bond remains stable over a few hundreds of milliseconds.

The spectrum of the "on"-state at 40°C detected 4 cross-peaks, corresponding to H82, H182, H200 and H218. Nevertheless, the possibility remained that the hydrogen bond between H149 and the CRO is relatively weak at 40°C, thus we decided to re-run the experiment at lower temperatures (15°C and 25°C). In order to assess the strength of the hydrogen bond, the intensity of H182 was taken as a reference, since it did not drastically change upon temperature decrease. Comparing the cross-peak intensities of H149 with H182, we concluded that the hydrogen bond of H149 doubles in strength when decreasing the temperature from 25°C to 15°C. Consequently, our NMR data confirm that the CRO is only weakly bonded with H149 in the "on"-state at physiologically relevant conditions.

No corresponding peak was detected for H149 in the "off"-state, characterized by a protonated CRO. The lack of a H149 peak in the "off"-state does not exclude the possibility of an H-bond existing, potentially between the a deprotonated H149-$N_{\delta 1}$ and the protonated *trans*-CRO. However, the sub-ms dynamics experienced by the *trans*-CRO and/or its neighbouring residues, including H149, suggest that this H-bond is either non-existent or

very weak in solution at pH=7.5.

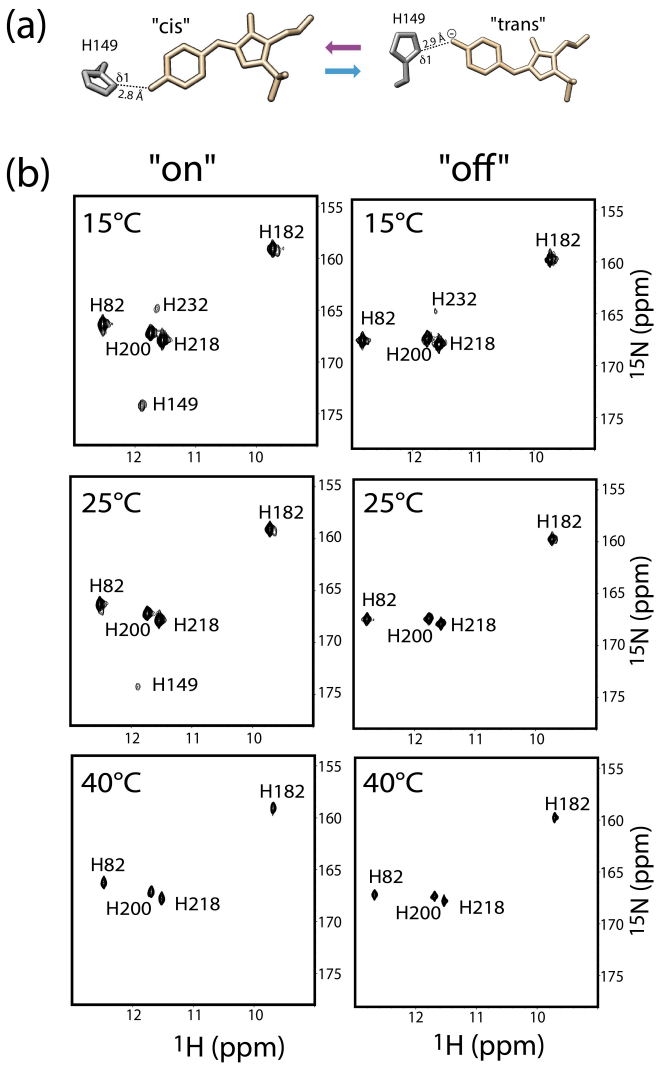


Figure 3.30 (a) Estimation of distances between the phenol ring of the CRO and the $N_{\delta 1}$ of H149 in both the "on"- and the "off"-state crystal structures of rsFolder. (b) 2D ^{1}H-^{15}N SOFAST experiments run at a range of temperatures to discern the existence or not of an H-bond between H149 and the CRO in both the "on"- and the "off"-state, recorded at 850MHz. Peaks detected in all temperatures are annotated corresponding to Histidines engaged in an H-bond.

3.2.11 Interpretation of NMR findings of rsFolder conformation and dynamics upon *cis-trans* isomerization

Using nD solution NMR to characterize rsFolder in its two ground states proved to be very insightful. Not only were we able to complement the static crystallographic structures

by providing information on dynamics happening on various timescales, but also assess rsFolder's mechanistic characteristics (i.e dynamics, H-bonds, protonation states of residues and the CRO) under physiologically more relevant conditions.

Initially, testing out a variety of NMR pulse sequences, the most appropriate experiments were chosen for further studies included in the next parts of the Results section. Consequently, a more streamlined approach could be designed, achieving characterization of future samples of rsFolder and mutants within a few days of NMR experiments.

Assessment of LASER-Driven EXSY experiment

Using the LD-EXSY experiment, the assignment of the "off"-state amide spectra significantly ameliorated. In particular, residues experiencing motions in the intermediate exchange regime could be identified which enabled a more complete view of the dynamic behaviour of rsFolder upon *cis-trans* isomerization. Note that these residues could not have been assigned using conventional 3D triple-resonance experiments. Furthermore, the duration of both ^{1}H- and ^{15}N-LD-EXSY was 24h, a substantial improvement from the one-week long 3D triple-resonance experiments. Not only were these experiments fast, but the assignment process became more straightforward since there was no need to identify carbon frequencies. Hence, the use of LD-EXSY experiments should prove useful when assigning RSFPs, in general, essentially reducing the time needed for assignment of both states to almost a half.

Nevertheless, its ability to provide reliable information highly depends on how fast the sample can be switched between the two states. Therefore, such experiments are useful either for fast RSFPs such as rsFolder or for more powerful and homogeneous illumination set-ups. Having said that, it is to be noted that LD-EXSY experiments tailored for the aromatic region of rsFolder did not yield sufficient signal-to-noise ratio and consequently were not included in the results. Most likely, despite the relatively rapid photoswitching of rsFolder, there was a need to further populate the "off"- and "on"-states during the recycle delay and mixing time, respectively, since the original Arom-HSQC on which the Arom-LD-EXSY was based, is not as sensitive as the BEST-TROSY focusing on amides. Hence, while 30$\%$ of the signal in the amide LD-EXSY originated from exchange peaks, higher signal was required for detection in the aromatic LD-EXSY.

Altogether, the use of LD-EXSY was proven efficient and significantly aided the characterization of the backbone of rsFolder. Its ability to contribute in the assignment of side chains as well will depend on the advances in the in-situ illumination set-up. Overall, it is essential to achieve more uniform illumination within the Shigemi tube, while higher LASER powers should be considered.

Dynamic characterization of the "on"-state of rsFolder

As mentioned in the Introduction, FPs require a relatively rigid CRO in order to fluoresce. In rsFolder in particular, the crystal structures identified three hydrogen bonds (H149, T204, water molecule) which stabilise the *cis*-CRO [61] and thus minimize non-fluorescent de-excitation due to vibrations and motions [139, 154, 157, 212]. Nevertheless, RSFPs require a minimum conformational "freedom" in order for the CRO to be able to alternate between the fluorescent *cis* and the dark *trans* configuration upon illumination [3]. Therefore, we would expect that the "on"- and "off"-states of rsFolder are characterized by different dynamics, accommodating for a rigid deprotonated *cis*-CRO in the "on"-state, while still allowing an amount of flexibility for it to quickly undergo photoisomerization. Moreover, the "off"-state structure of rsFolder, being a metastable state, whereby the H-bonds between the β-barrel and the *trans*-CRO have been reduced, is most likely more dynamic than the "on"-state. Such differences in dynamics could not be discerned from the static crystallographic structures.

From our work, it is evident that rsFolder in its fluorescent "on"-state behaves mostly as a rigid globular protein. Nevertheless, line-broadening leading to loss of NMR signal for residues in the β-7 and 8 strand (V151, K167, I168) can explain the well-known weakness of these strands to form a hydrogen bond between them [129, 221, 193, 158]. This apparent plasticity in this region of the protein has been identified in other GFP-type proteins and is interpreted as "holes" from which water molecules can enter the barrel [200]. Evidently, our HDX data, in agreement with such an interpretation, point towards increased solvent accessibility for residues in these two β strands. Interestingly, in some studies found in the literature concerning GFPs, the β-7 strand is understood as the proton donor which transfers protons from the solvent to the CRO [158]. Consequently, some conformational freedom should be expected to allow for proton exchange. When investigating the potential role of such a defect in the β-barrel of rsFolder, in particular, one could also attribute it to the need for the protein to photoswitch.

Focusing on the H149 residue belonging to the β-7 strand, our NMR data confirm the existence of an H-bond with the CRO's phenolate as suggested by crystallography [61], formed at the $N_{\delta 1}$ of the H149 imidazole ring. However, the cross-peak corresponding to this H-bond only appears at lower temperatures. Hence, it was concluded that H149 only weakly binds to the CRO at physiologically relevant conditions. Potentially, rsFolder would present a higher brightness if this bond was stronger. Numerous studies seem to agree on the need for a strong H-bond network anchoring the CRO and thus improving its fluorescent capabilities [150, 52, 147]. On the other hand, the transient nature of the H149-CRO bond can also be the reason for the fast "on"-"off" photoswitching recorded, ensuring a relatively low energetic barrier between the two states. This interpretation agrees with QM/MM studies done in a Dronpa mutant, whereby the breaking of H-bonds is required for the CRO to undergo *cis-trans* isomerization [150]. Altogether, the unstable H-bond of H149-CRO in the "on"-state can be interpreted as enabling relatively fast photoswitching while maintaining sufficient brightness for rsFolder. Additionally, when compared to other RSFPs, like Dronpa [11, 9, 12], rsFolder having fewer H-bonds (only 3) in the "on"-state

possibly results in its comparatively lower brightness but also fast photoswitching.

After discovering the cross-peaks corresponding to the phenol ring of the CRO, we were able to quantify ring flips. The *cis*-CRO appears to be highly stabilized by the protein environment, flipping at 35 ± 3 flips per second, at 40°C. At this point of the project, this finding was interesting but we could not conclude about its relevance for the photophysical properties of rsFolder. Nevertheless, we could assume its importance in quantifying the "rigidity" of the fluorescent form of the CRO and thus potentially being an NMR observable which could be correlated to the fluorescence quantum yield of rsFolder "on"-state. In addition, CRO's apparent stability is consistent with the overall stability of the rsFolder "on"-state.

rsFolder experiences increased dynamics in the few hundreds of μs-ms timescale upon *cis-trans* isomerization

From the published crystal structures of rsFolder and other green negative RSFPs [12, 61, 53], it is expected that reconfiguration of the CRO upon photoswitching to the *trans* state leads to the breakage of H-bonds. Moreover, the CRO becomes protonated, potentially explaining its inability to fluoresce. When considering the decrease in H-bonds, CROs of RSFPs in their "off"-state conformation probably experience increased dynamics.

During my project, we discovered dynamics present primarily in the rsFolder "off"-state as compared to the "on"-state. More specifically, evidence of line broadening and ^{15}N-CPMG-RD data strongly suggest motions taking place in the few hundreds of μs-ms in the *trans*-CRO and its vicinity. Some residues taking part in such dynamics have already been the focus of literature concerning photoswitching (L70, H149, V151, T204, S206, V225) [61, 87, 47] while others have not been considered so far (V62, T63, I168). HDX data also support an overall destabilization of the β-barrel and the helices flanking the CRO upon 488nm light illumination. Unfortunately, these motions could not be easily quantified due to significant loss of NMR signal of cross-peaks corresponding to backbone, methyl and aromatic side chains of the protein as well as the CRO.

Intermediate exchange motions taking place in the CRO and its vicinity are further supported by the lack of a sufficiently stable H-bond with H149. Taking into account the lack of signal for the phenol ring of the *trans*-CRO, NMR results seem to contradict the published crystallographic data [61]. In particular, a stable H-bond between H149 and the *trans*-CRO contrasts the dynamics in the hundreds of μs-ms detected by NMR (i.e the line-broadened H_ϵ-$N_{\epsilon 1}$ peak of H149 in the "off"-state). Crystal structures most likely represent the lowest energy conformation of the rsFolder "off"-state, whereby due to crystal packing and lowered temperatures, the H149 is 2.9Angstrom away from the phenol ring of the *trans*-CRO.

Our data suggest that the exceptional thermal stability of rsFolder "off"-state is not necessarily the result of CRO "anchoring" to the β-barrel. Overall, 488nm illumination induces destabilisation of the CRO in the "off"-state, which appears to have no stable

H-bonds left to anchor it to the β-barrel. Notably, loss of rigidity can also explain the *trans*-CRO's inability to fluoresce in rsFolder, similarly to Padron0.9, whereby both *cis* and *trans*-CROs are deprotonated, but only the less dynamic CRO (*cis*) is the fluorescent one [28].

Altogether, the NMR-derived information concerning the "off"-state conformation of rsFolder suggests an energetic landscape containing multiple conformations. At cryogenic temperatures, the accessibility to these conformations is limited while at physiologically relevant conditions, the CRO is freely exploring multiple configurations on the hundreds of μs-ms timescale. While the general plasticity of the β-strands opposite to the phenol ring of the CRO comes as no surprise, the information gained regarding the helices flanking the CRO is quite significant. Literature attributes limited importance to these helices – the Q69L [88] (numbered L70 in rsFolder [61]) mutation in EGFP resulted in the photoswitching properties of rsEGFP, while S73 has been included in the number of "hopping" points of GFP [200]. Moreover, XFEL structures of rsEGFP2 reveal an intermediate conformation which becomes transiently populated upon photoswitching, whereby H2 moves downward in an attempt to accommodate for the twisted conformation of the CRO [47]. Our HDX data show up to a 100-fold increase in solvent exchange for the regions H2 and H3 upon *cis-trans* isomerization. Taken together, rsFolder might have a similar photoswitching mechanism where the helices on top and bottom of the CRO need to shift so as to make space for an intermediate twisted conformation. Unfortunately, NMR is not capable of capturing such fast excited states. All in all, this work highlights the importance of attempting H2 and H3 mutations when performing semi-rational engineering on RSFPs.

The insight gained concerning the differences in dynamics of the β strands facing the phenol ring of the CRO as well as the existence (or lack thereof) of the H149 H-bond in the "on"- and "off"-states, could be potentially considered as evidence supporting the "proton wire" model suggested by the group of Noam Agmon in several papers published on GFP [199, 8, 201]. They suggest a role of the β-barrel of GFP as a matrix of "hopping" points for protons. Accordingly, protons originating from the bulk solvent "hop" onto these residues as a way to get transferred to the CRO which continuously "pumps" protons in and out of the barrel. Note that in the case of GFP, only the *cis*-CRO is concerned. In the case of green negative RSFPs, identifying the residues involved in these wires could be instructive in explaining the high pKa of the *trans* as compared to the *cis*-CRO of RSFPs [77] (thus favouring of the protonated form under physiologically relevant conditions) but also identify "problematic" residues which, upon illumination, have a tendency to create free radicals, causing phototoxicity [199]. In the "off"-state of rsFolder, key residues which could be part of an efficient proton wire in the "on"-state (eg H149, N150, T204, S206) [8], show dynamics in the few hundred μs-ms. The difference in these residues' dynamics, alongside the weakening of the H-bond between the CRO and H149 could indicate a proton wire network which in the "off"-state becomes unable to readily protonate and deprotonate the *trans*-CRO. Thus, by identifying residues undergoing increased dynamics upon *cis-trans* isomerization, one could suggest a number of potential "hopping" points. At this point, our NMR data alone are not sufficient to support such a theory, but future work could include simulations to investigate "hopping" points in rsFolder. To conclude, understanding proton

wires in RSFPs can potentially inspire fluorescent protein engineers to create brighter and more photoresistant RSFPs.

In conclusion, this first part of my PhD thesis yielded interesting results for rsFolder. Regions identified which experienced a change in dynamics upon 488nm illumination are interesting to investigate in the future – either through simulations or mutation studies. Most importantly, alongside the studies on Dronpa, this work is significant in the effort to initiate a more active participation of NMR in the field of RSFPs.

The next part of my results section aimed at further probing the dynamics observed, in an effort to identify their origins.

3.3 NMR investigation of ground-state chromophore isomerization and protonation in rsFolder

Identifying appropriate NMR experiments to gain more insight into the CRO of rsFolder was crucial for the work that will be described below. Despite having to work with a 27 kDa protein, approximately one year after the beginning of my thesis I found myself studying a rather small molecule, i.e the CRO, using nD solution NMR. By doing so, I was able to not only better understand ground-state processes leading to the apparent dynamics previously found in the "on"- and "off"-states but also shed light on the energetic landscape of rsFolder.

The primary questions leading to this work concerned the protonation state of the CRO. Protonation influences the optical properties of FPs, since only the anionic (deprotonated) form of the CRO is expected to be fluorescent upon cyan light excitation for green-emitting FPs [213, 167]. Consequently, when one considers the *cis*-CRO in rsFolder, it is important to account for the exchange between its neutral and anionic form. However, other techniques studying RSFPs have limited discriminatory capabilities – X-ray crystallography hardly detects hydrogens and, thus, can only discern between *cis* and *trans* species, while optical studies can reveal the protonation state, but they are only little sensitive to isomerization phenomena. Therefore, so far in the field of RSFPs, it has been a challenge to disentangle *cis-trans* isomerization from protonation/deprotonation processes. One technique that can discriminate between isomerization and protonation states is Raman spectroscopy, as shown in a study on EGFP [90].

Due to the results obtained concerning the distinct NMR signals for the *cis*- and *trans*-CRO configuration, I was optimistic that NMR will be able to overcome such common limitations of other biophysical techniques. In order to further our comprehension of mechanistic processes affecting the CRO and the apparent dynamics in the "on"- and "off"-states, we, thus, had to investigate rsFolder at different pH values.

On the one hand, such pH study would allow us to characterize the various possible CRO configurations. The ability of solution NMR to study proteins in a variety of solvents allowed for not only a pH study but also an evaluation of the effect of the buffer composition on the mechanistic properties of rsFolder. NMR observables were, finally, correlated with optical measurements of rsFolder at different pH values.

On the other hand, by gradually perturbing the population of CRO species present, we could target the main region where dynamics lie in rsFolder. In this way, we were able to conclude on the origins and characteristics of the dynamics previously reported on Section 3.2.

Altogether, this part of my thesis yielded significant results regarding ground-state isomerization and protonation/deprotonation events. Not only could we discern and characterize 4 major CRO configurations – *cis* neutral (N-cis), *cis* anionic (A-cis), *trans* neutral (N-trans) and *trans* anionic (A-trans) –but we could also characterize exchange dynamics between them at a range of pH values and under different buffer conditions.

Interestingly, we discovered pH-induced *cis-trans* isomerization occurring in rsFolder, in the absence of light. Combining this work with optical studies confirmed its importance for microscopy techniques using RSFPs.

3.3.1 NMR deciphers the various chromophore configurations

X-ray crystal structures of rsFolder suggest a *cis*-CRO present in the dark-adapted fluorescent state while after 488nm-light excitation, the CRO adopts a *trans* conformation [61]. In the case of rsFolder at pH=7.5 and 40°C, we were able to retrieve distinct CS values for the C_β nucleus in the methine bridge of the CRO ($\Delta\nu = 7.6ppm$) corresponding to its two isomerization states. Comparing the measurements of the ${}^3J_{HBCO}$ coupling strength for the two H_β-C_β peaks (5Hz and 10Hz) with published values [1, 218], confirmed that they correspond to the *cis* and *trans*-CRO, respectively. NMR pH titration studies revealed a pH-induced CS change of the characteristic H_β-C_β cross-peak of the *cis*-CRO, indicating that the molecule titrates at a range of pH from 4.2 (lower pH leads to unfolding of the protein) to 8. Hence, at low pH the CRO is assumed to be primarily in the N-*cis*, while the A-*cis* is mostly populated at basic pH. With this in mind, the two extreme CS retrieved for the H_β-C_β cross-peak (at pH=4.2 and pH=8) can be assigned to the N-*cis* and the A-*cis* configuration of the CRO. Nevertheless, the NMR signals of the two are averaged according to population ratios into one peak – a sign of fast exchange between the two configurations as will be discussed later in the chapter.

On the contrary, the characteristic *trans*-CRO-peak of H_β-C_β does not change position in the range of pH=4.2 – 8.0. In accordance to similar studies performed in mIrisGFP, the *trans*-CRO has a pKa>10 [77], thus it should not be expected to titrate at the pH range chosen for the *cis*-CRO studies. Consequently, we studied the rsFolder "off"-state at pH=9.5, 10.7, 11.5, where rsFolder NMR spectra confirmed a stable protein. At pH=11.5 under 488nm illumination, a new peak appeared in the aromatic HSQC spectra, slightly shifted from the previously recorded *trans* H_β-C_β peak and was consequently assigned to the A-*trans* configuration. Unlike the *cis* H_β-C_β peak, the two *trans* peaks were both present in the "off"-state spectra at pH=11.5 and neither peak shifted upon pH change, an indication that the *trans*-CRO (de)protonation is in slow exchange (rate of less than $10s^{-1}$), even at high pH values.

Altogether, we could identify and assign 4 major CRO configurations, which are shown in the Figure below (Figure 3.31). The next steps involved characterising their intrinsic dynamics and the inter-conversion between them.

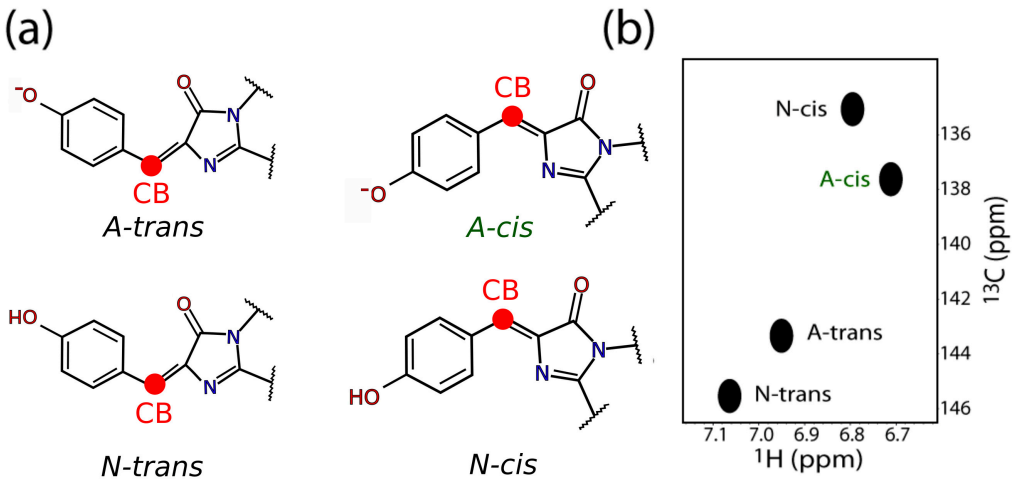


Figure 3.31 (a) Chemical structure of p-HBI chromophore in the 4 main configuration states : A-trans, A-cis, N-trans, and N-cis, with annotated the CB position. (b) Characteristic ^{1}H-^{13}C NMR signature of the CRO $^{13}C_{\beta}$ site in the 4 configurational states. Only the A-*cis* configuration is expected to be fluorescent upon 488nm illumination.

3.3.2 NMR investigation of the *cis*-CRO of rsFolder in the absence of light

Protonation/deprotonation of the *cis*-CRO as observed by NMR

Titration studies were performed for a series of rsFolder samples ranging from pH=4.2 – 8.0. All samples were prepared in McIlvaine buffer which has an exceptional range of buffering from pH=2-8 [141]. The 2D ^{1}H-^{13}C aromatic BEST-HSQC pulse sequence without CT-editing was chosen due to its high sensitivity to observe the low-intensity CRO peaks. Before NMR measurements, samples were kept in the dark for a few days to ensure that the protein is in thermodynamic equilibrium (dark-adapted ground-state).

As mentioned above, the characteristic H_{β}-C_{β} peak corresponding to the *cis*-CRO state, progressively shifted in CS at the pH range chosen. Fitting the CS of C_{β}, in particular, against 10 pH values to a simple bi-molecular reaction as described by the Henderson-Hasselbalch equation, assuming a Hill coefficient of 1, resulted in an apparent pKa value of $pKa_{cis} = 5.5$. A similar pKa value (5.5) has been previously reported, derived from absorption measurements of rsFolder [61]. Moreover, detecting only one peak of the averaged CS of A-*cis* and N-*cis* suggests a fast *cis*-CRO (de)protonation, with exchange rates on the sub-ms timescale ($k_{ex} >> 1000s^{-1}$).

Interestingly, such titration behaviour could be detected for a number of residues in the vicinity of the *cis*-CRO, seen by the series of 2D ^{1}H-^{13}C and ^{1}H-^{15}N spectra recorded.

Extracting the ^{1}H, ^{13}C or ^{15}N CS of these residues at different pH, we could globally fit the same Henderson-Hasselbalch curve with $pKa = 5.45 \pm 0.05$ (Figure 3.32). The almost identical behaviour with the *cis*-CRO titration strongly suggests that all these nuclei specifically report on the protonation/deprotonation of the *cis*-CRO.

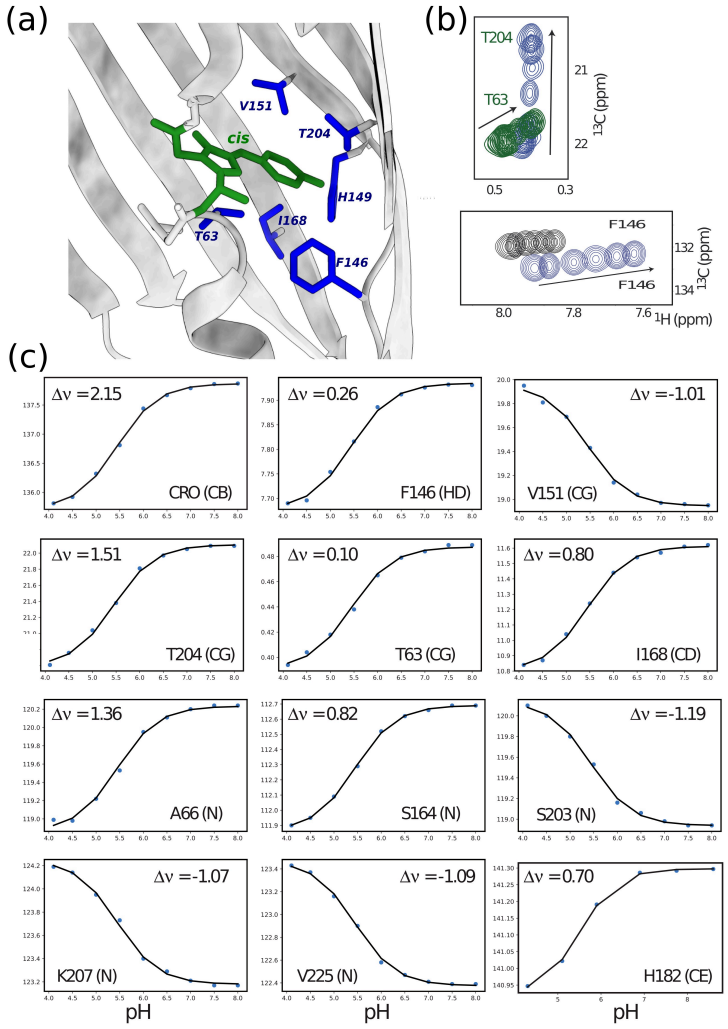


Figure 3.32 (a) Zoom in the crystal structure of rsFolder "on"-state (5DTZ). Highlighted in blue are side chains which appear to titrate with the same pKa as the *cis*-CRO, i.e $pKa = 5.45 \pm 0.05$. (b) Regions of 2D ^{1}H-^{13}C correlation spectra (methyl peaks of T204, T63 and two aromatic peaks of F146), showing examples of CS modulation of peaks upon pH variation, recorded at 850MHz at 40°C. (c) Plotting of the observed CS variations as a result of pH changes for a number of nuclei. Data can be fitted to a bimolecular reaction as described by the Henderson-Hasselbalch equation with $pKa = 5.45 \pm 0.05$.

Effect of buffer composition on the pronotation/deprotonation of the *cis*-CRO

In order to cover the range of pH values desired for this study, we chose to use a series of McIlvaine buffers, consisting of a mixture of disodium phosphate and citric acid. Nonetheless, for most of the studies during my thesis we would use HEPES buffer. Therefore, in an effort to compare the effect of the two buffers, we acquired 2D methyl ^{1}H-^{13}C SOFAST-HMQC and ^{1}H-^{15}N BEST-TROSY spectra of an rsFolder sample in HEPES, pH=6.0. By doing so, we discovered that a number of peaks present in the spectra of the McIlvaine pH=6 sample significantly lost signal (or disappeared) in the spectra of the HEPES pH=6 sample, while their positions remained the same for both buffers. Such buffer-induced line broadening is indicative of a shift of dynamics to the intermediate exchange regime. Given that phosphate ions are known for catalyzing proton exchange [131, 207, 94], we can hypothesize that the McIlvaine buffer increases the rate of protonation/deprotonation of the *cis*-CRO, while HEPES has only a minor effect. Consequently, the linewidths of residues nearby the *cis*-CRO are affected. This idea is supported by the line narrowing of the same residues' peaks at lower pH, i.e at higher availability of hydrogen ions in the solvent. Hence, this comparison between buffers not only supports the hypothesis that peaks titrating with $pKa = 5.45 \pm 0.05$ are "witnesses" of the CRO's protonation/deprotonation, but is also indicative of the effect that the buffer can have on the exchange dynamics of N-*cis* and A-*cis* interconversion.

The observed line broadening opened the possibility of further quantifying the N-*cis*/A-*cis* exchange in HEPES. Consequently, a HEPES, pH=6.2 sample in the dark-adapted state was studied using ^{15}N-CPMG-RD at two magnetic field strengths (corresponding to Larmor frequencies of 70MHz and 85MHz for ^{15}N nuclei) at 40°C. The intensities of 6 amide cross-peaks were chosen for fitting, which also appear to titrate with $pKa = 5.45 \pm 0.05$ in the McIlvaine buffer series (Figure 3.33). More specifically, it was hypothesized that the exchange process leading to line broadening of peaks is the protonation/deprotonation of the *cis*-CRO. Accordingly, for fitting the data recorded on the rsFolder sample at pH=6.2, the ratio was fixed between the major A-*cis* and minor N-*cis* populations to be $p_{Acis} = 0.85$ and $p_{Ncis} = 0.15$, as determined from the Henderson-Hasselbalch curve with pKa=5.45. The resulting apparent exchange rate between A-*cis* and N-*cis* was calculated to be $k_{ex} = 3200s^{-1} \pm 200s^{-1}$. Furthermore, when comparing the CS changes ($\Delta\nu$) predicted from the ^{15}N-CPMG-RD data and measured in the pH titration series, we get a correlation of 0.82 (Figure 3.33). Therefore, ^{15}N-CPMG-RD data support our hypothesis regarding residues reporting on the exchange between N-*cis*/A-*cis*-CRO, with $k_{deproton} = 2720s^{-1}$ and $k_{proton} = 480s^{-1}$, corresponding to the deprotonation and protonation rates of the *cis*-CRO, respectively, at pH=6.2 in HEPES buffer. Note that these amides are probably also experiencing other protonation and/or conformational events, which can explain why the correlation between the observed (from titration) and the expected $\Delta\nu$ assuming only the effect of CRO's (de)protonation (from ^{15}N-CPMG-RD) is not 1.

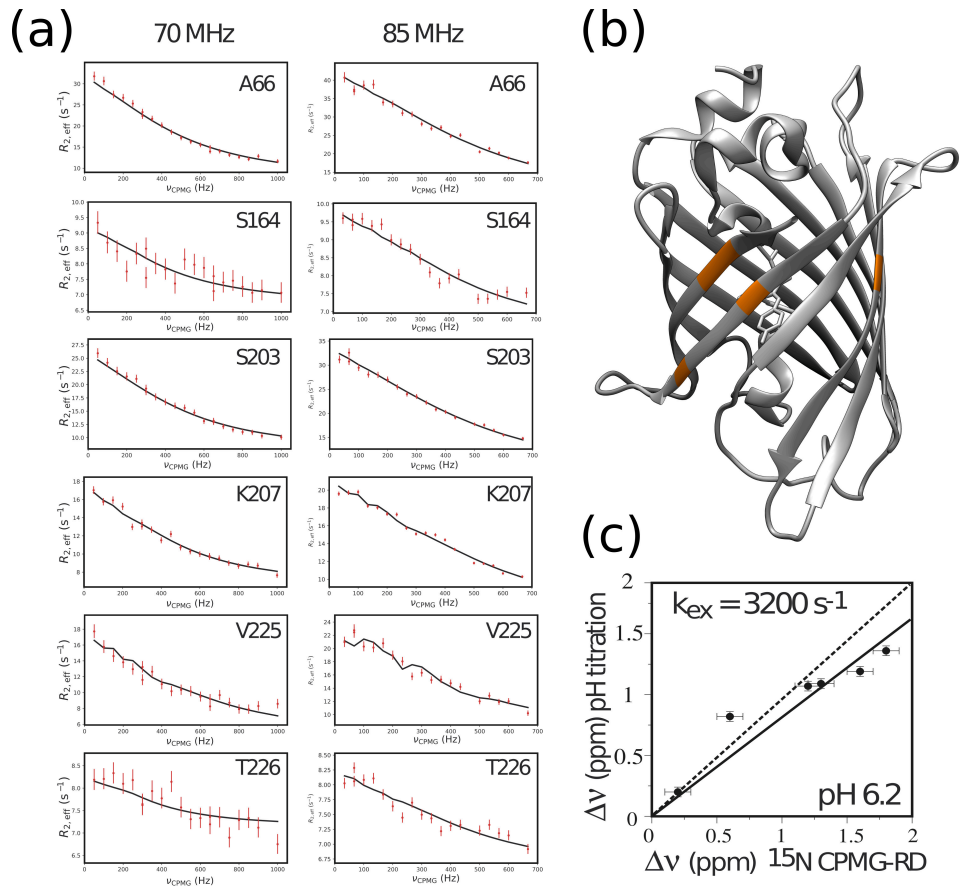


Figure 3.33 (a) The result of fitting using the Chemex algorithm, of ^{15}N-CPMG-RD data recorded at 40°C of 6 amide peaks at two different magnetic field strengths (left and right panel) for a Histidine-unlabelled sample of rsFolder "on"-state. By fixing the population ratio between the major (A) and minor (B) populations to be $p_A = 0.85$ and $p_B = 0.15$, the resulting exchange rate is $k_{ex} = 3200s^{-1}$, with $k_{B->A} = 2720s^{-1}$ and $k_{A->B} = 480s^{-1}$, corresponding to the rates of deprotonation and protonation, respectively. (b) The crystal structure of rsFolder "on"-state, with highlighted in orange, the amide residues chosen for ^{15}N-CPMG-RD data fitting. Note that amides nearer to the CRO (eg N147, T204, V151) could not be fitted due to severe line broadening and histidines were not visible. (c) $\Delta\nu$ values as measured in the pH-titration series and as predicted by the fitting of the ^{15}N-CPMG-RD data for the 6 amides. The straight line corresponds to a slope of m = 0.82, while the dashed line has a slope m = 1.

Conformational dynamics experienced by the protonated and deprotonated configurations of the *cis*-CRO

At this point, the CRO ring flip rate, k_{rf} previously measured at pH=7.5 in HEPES buffer (mentioned in the previous part of the Results section) needed to be reconsidered, keeping

in mind the time averaging between the N-*cis* and A-*cis* conformation. All NMR observables measured for the "on"-state are a population weighted average of the NMR signals corresponding to rsFolder with a protonated and deprotonated *cis*-CRO. Thus the $^{1}H_{\delta}$-$^{13}C_{\delta}$ CRO-pair used to extract the CRO's phenol ring flip rates, should also be considered as a population weighted average of two species of *cis*-CRO present, flipping at different rates, k_{rf}. Hence, having retrieved the pKa_{cis} using NMR, we could deconvolute the contributions on the average CRO-ring flipping rate of the two species of the *cis*-CRO; k_{rf}^{N-cis} and k_{rf}^{A-cis}.

The characteristic NMR signatures of the $C_{\delta 1}$ and $C_{\delta 2}$ of the phenol ring of the *cis*-CRO start to be detected in the aromatic ^{1}H-^{13}C correlation spectra only at high pH values (pH>7.2). Hence, at lower pH the population-weighted average of the observed k_{rf} is increasing, shifting to the μs-ms timescale – the exchange rate between the two phenol rotamers is comparable to the chemical shift difference of the two peaks (at a high-field spectrometer, at 40°C) causing line broadening. This is already a sign that the k_{rf} observed is the population-weighted average of a fast-flipping N-*cis* species behaving similarly to a free tyrosine, which becomes more populated, and the more "rigid" A-*cis*, whose presence decreases.

Initially, running the aromatic ^{1}H-^{13}C-EXSY experiments for values pH=7.2, 7.5, 10.7 we acquired the k_{rf} to be $115 \pm 25s^{-1}$, $35 \pm 2s^{-1}$, $13 \pm 1s^{-1}$, respectively. Nevertheless, the aim was to calculate values for k_{rf} at a range of pH values which would allow the fitting of the k_{rf}-pH data to a Henderson-Hasselbalch curve with a pKa=5.45. Therefore, k_{rf} at lower pH values had to be determined. We could only approximate the value of k_{rf} at pH=6 since the two peaks corresponding to the two rotamer's H_{ϵ}-N_{ϵ} pairs, coalesce into one strongly line broadened peak. The observed coalescence enables us to estimate the k_{rf} at pH=6 to be $k_{ex}/2 = \frac{1}{4\pi\Delta\nu} = 1750s^{-1}$. Extrapolating the k_{rf}-pH Henderson-Hasselbalch curve to the two extremes, we could estimate the values, $k_{rf}^{A-cis} = 10 \pm 1s^{-1}$ and $k_{rf}^{N-cis} = 7500 \pm 1000s^{-1}$, corresponding to an almost rigid A-*cis*-CRO and a freely rotating N-*cis*-CRO, similar to a free CRO in solution.

The role of H149 in protonation of the *cis*-CRO

As mentioned previously, Histidine-specific NMR experiments enabled us to assess the tautomeric state of H149 as well as the possibility of it forming an H-bond with the phenolate of the *cis*-CRO. More specifically, Histidines' $C_{\epsilon 1}$ and C_{γ}, $C_{\delta 2}$ frequencies are sensitive to the total charge ($C_{\epsilon 1}$) and the protonation state (C_{γ}, $C_{\delta 2}$) of the imidazole ring [94]. In the "on"-state rsFolder, pH=7.5, HEPES buffer sample previously recorded, there was strong indication of H149 being protonated at the $N_{\delta 1}$ position of its ring, while ^{1}H-^{15}N SOFAST-HMQC experiments suggested a weak H-bond with the *cis*-CRO's phenolate ring.

In the McIlvaine pH series described at this part of my thesis, no significant CS changes were observed for the $C_{\epsilon 1}$, C_{γ} and $C_{\delta 2}$ frequencies of H149 in the "on"-state, indicating no change of tautomeric state for H149 upon population shift from the A-*cis* to the N-*cis* configuration. Moreover, H149 in the McIlvaine buffer is experiencing line sharpening as

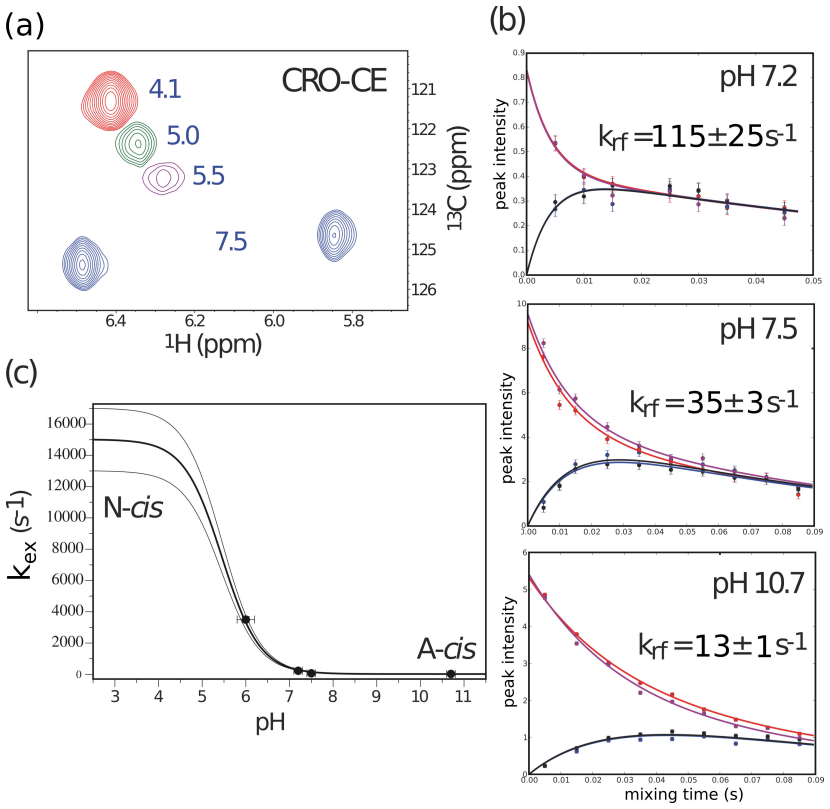


Figure 3.34 (a) Observed coalescence of the H_ϵ-C_ϵ pair of the *cis*-CRO's phenol ring as seen in the aromatic ^1H-^{13}C HSQC spectra of rsFolder "on"-state at different pH values, recorded at 850MHz at 40°C. (b) Results of the aromatic ^1H-^{13}C EXSY experiment for the estimation of ring flipping rates at different pH values, recorded at 850MHz at 40°C. The intensities of peaks considered are the H_δ-C_δ pairs of the *cis*-CRO's phenol ring; in red, the intensity of major peaks, in blue, the intensity of minor (exchange) peaks. (c) Fitting the values of $k_{ex} = 2 * k_{rf}$ extracted for pH>6 to a Henderson-Hasselbalch curve of $pKa = 5.45 \pm 0.05$ and extrapolating for the two extreme pH values, to determine the $k_{rf}^{A-cis} = 10 \pm 1 s^{-1}$ and $k_{rf}^{N-cis} = 7500 \pm 1000 s^{-1}$.

compared to HEPES, suggesting that it experiences faster protonation exchange either of its imidazole ring and/or of the phenol ring of the *cis*-CRO. Altogether, H149 appears as experiencing a shift to faster protonation exchange upon the presence of phosphate ions from the McIlvaine buffer while also maintaining its tautomeric state at different pH values.

If one considers H149 as the proton donor (or acceptor) of the *cis*-CRO, through its $N_{\delta 1}$ nitrogen, in order to maintain a primarily protonated $N_{\delta 1}$, it needs to be part of a larger "proton wire" network, which can rapidly shuffle protons from (or to) the H149 to (or from) the *cis*-CRO. Moreover, this would mean that at lower pH, the increase of hydrogen ions in the solvent forces H149 to continuously "pump" the *cis*-CRO with protons, further populating the protonated *cis* configuration, which forms no H-bond with the H149 (since both N-*cis* and the $N_{\delta 1}$ of H149 are protonated). Most likely, in this scenario, the

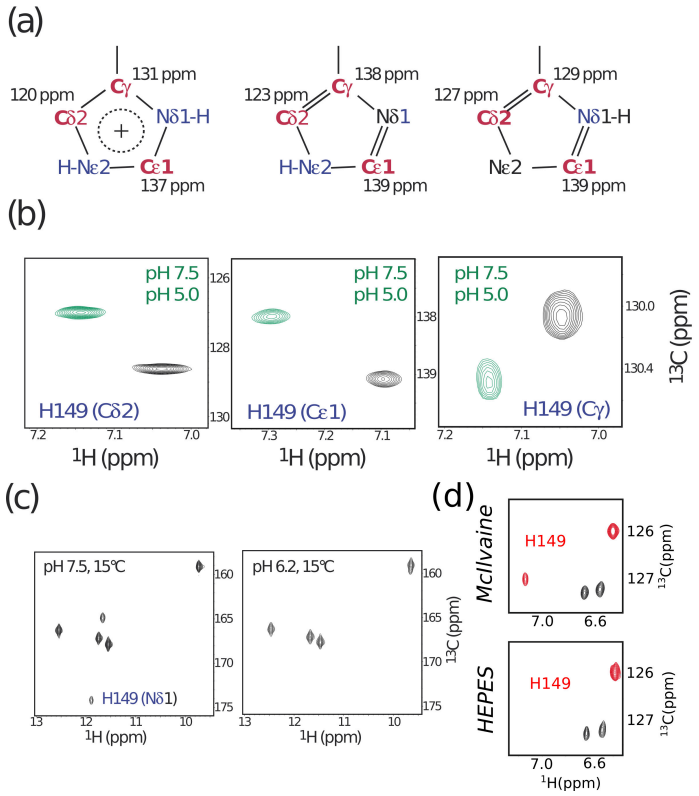


Figure 3.35 (a) Chemical structure of a Histidine imidazole ring in its 3 protonation states, from left to right: biprotonated, $N_{\epsilon 2}$ (τ tautomer) or the $N_{\delta 1}$ (π tautomer) protonated. Characteristic resonant frequencies for the ring carbons $C_\gamma, C_{\epsilon 1}, C_{\delta 2}$ are indicated in ppm. (b) Spectral regions of aromatic ^1H-^{13}C BEST-HSQC spectra, focusing on the cross-peaks corresponding to the $H_\delta - C_{\delta 2}$, $H_\epsilon - C_{\epsilon 1}$, $H_\epsilon - C_{\gamma 1}$ of His149 of rsFolder in the dark-adapted "on" state, at pH=7.5 (black) and pH=5.0(green), recorded at 850MHz at 40°C. The $\Delta\nu$ upon lowering the pH is very small in all three Carbon nuclei. At lower pH, H149 maintains the same tautomeric state as in high pH, being primarily protonated on the $N_{\delta 1}$. (c) Spectra region of ^1H-^{15}N His-SOFAST-HMQC, recorded at 850MHz at 15°C, showing H-bonded Nitrogens of Histidines for pH=7.5 (left) and pH = 5.0 (right) for the dark-adapted "on"-state. The detected cross-peak of His149($N_{\delta 1}$) is annotated. The loss of NMR signal for the H149 peak at lower pH suggests that the H-bond of H149 with the *cis*-CRO is further weakened upon increase of hydrogen ions availability. (d) Spectral regions of aromatic ^1H-^{13}C BEST-HSQC spectra focusing on H149, showing the effect of buffer on H149 NMR signal, at pH=6. The line sharpening of peaks in the McIlvaine buffer indicates that H149 is reporting on a protonation event.

H-bond needs to be constantly broken and reformed. In agreement with this model, a study showed the quick reprotonation of H148 in the bright EGFP, after several proton transfer steps and structural rearrangements of residues neighbouring the *cis*-CRO [2]. The authors highlighted the importance of H149 (H148 in EGFP) as the donor of solvent protons to the phenolate of the CRO. Nevetheless, our NMR data at this point, can only be used to infer the dynamics experienced by H149, and not to discern its role as a proton donor or acceptor of the *cis*-CRO.

Altogether, using our NMR data we were able to put together a kinetic model, concerning the *cis*-CRO. All our conclusions, alongside the corresponding supportive qualitative analysis, are summarized in Figure 3.36.

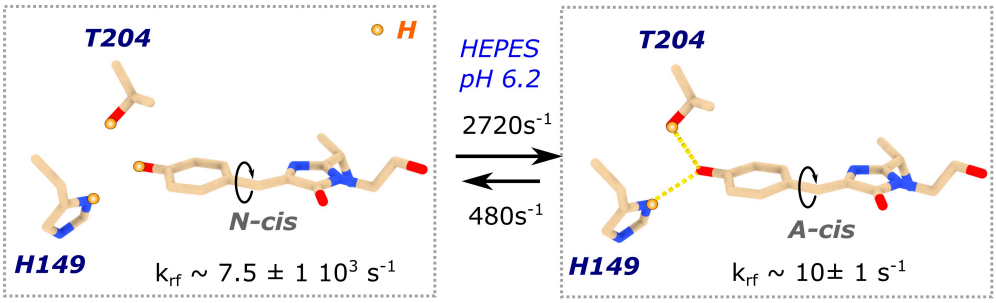


Figure 3.36 Graphical sketch summarizing the major findings on the (de)protonation kinetics, ring-flip dynamics, and H-bond stabilization of the *cis*-CRO in rsFolder.

3.3.3 NMR investigation of the *trans*-CRO of rsFolder

Protonation/deprotonation of the *trans*-CRO as observed by NMR

Similarly to the study of the *cis*-CRO, the *trans*-CRO was also inspected for its (de)protonation pKa value. We know from literature that in similar RSFPs it has been found to be beyond 10 [77]. The durability of rsFolder against "extreme" pH values enabled us to measure the 2D aromatic ^{1}H-^{13}C correlation spectra for samples at very high pH (9.5, 10.7, 11.5) (Figure 3.37). Here, instead of extracting the CS changes of the two *trans* configurations, we rather calculated the relative change of intensity of their corresponding $H_\beta - C_\beta$ cross-peaks. Consequently, we could fit the intensities to a bi-molecular reaction as described by the Henderson-Hasselbalch equation, assuming a Hill coefficient of 1, to estimate the pKa of the *trans*-CRO of rsFolder to be $pKa_{trans} = 12.0 \pm 0.1$.

Furthermore, the two distinct *trans*-CRO peaks detected suggest a slow exchange between the protonated and deprotonated *trans*-CRO, even at high pH values, with a proton exchange rate of $k_{ex} << 500s^{-1}$. While the limited availability of hydrogen ions in the

solvent makes it less probable to quickly protonate and deprotonate the CRO, these data point towards a much less efficient proton "wire" established in the "off"-state conformation of rsFolder. Subsequently, the *trans*-CRO seems to be protected by the barrel in the "off"-state; it is able to "hold-on" to the hydrogen in its phenol ring for longer than the protonated *cis*-CRO does, with a pKa value 2 units higher than an average hydroxyphenyl ring in aqueous solution [174].

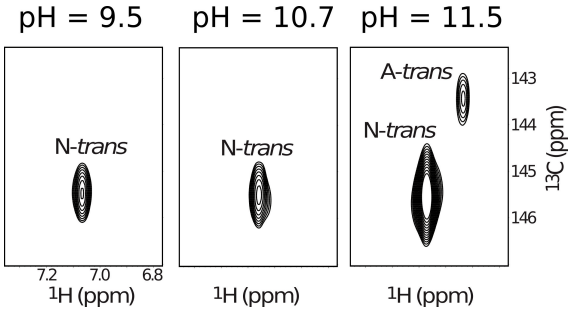


Figure 3.37 Spectral regions of aromatic ^{1}H-^{13}C BEST-HSQC spectra recorded at 850MHz at 40°C, focusing on the H_β-C_β CRO cross-peaks of the 488nm-light-induced state of rsFolder at different pH values. Annotated are the signature peaks of the N-*trans* and A-*trans*-CRO configuration.

Revisiting "off"-state dynamics derived from NMR

Severe line broadening hindered us from acquiring further information regarding the dynamics of the "off"-state of rsFolder in our previous work (Section 3.2). Similar to the "on"-state, replacing HEPES with the phosphate-containing McIlvaine buffer, we could retrieve the signal of lost peaks at pH=7.5 in the 488nm-light-induced "off"-state spectra (backbone and side chain spectra). Such line narrowing, also induced by lowering the pH in the HEPES buffered samples, is an indication of a shift in the timescale of kinetics present and/or the relative population of states (either faster or slower) (Figure 3.38).

Furthermore, 2D methyl ^{1}H-^{13}C SOFAST-HMQC spectra recorded at different temperatures at pH=7.5 (HEPES buffer) of rsFolder "off"-state, suggest line narrowing upon increase of temperature (Figure 3.39). Therefore, since higher temperature speeds up motions, our data place the dynamics experienced by rsFolder in the "off"-state at pH=7.5, at 40°C in the intermediate exchange regime. Upon increase of temperature, hydrogen ions, and/or phosphate ions, the protein dynamics observed are, thus, shifted to the fast rather than slow exchange regime.

In the case of induced faster dynamics, it is possible that the speeding up of proton exchange caused by phosphates and/or increase in hydrogen ion availability is "felt" by residues in the vicinity of the *trans*-CRO. In other words, T63, L70, H149, V151, I168 and T204 are reporting on a protonation event happening near them at a fast to intermediate timescale, depending on the pH, temperature and buffer composition.

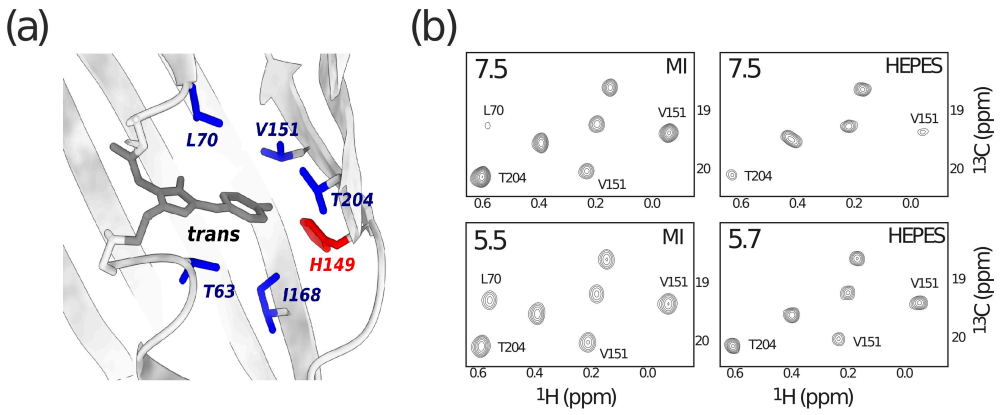


Figure 3.38 (a) Zoom in the crystal structure of the non-fluorescent state of rsFolder in the region of the CRO pocket. Highlighted in red are side chains with nuclei that show pH- and buffer-dependent line narrowing upon increase in hydrogen ion availability. (b) Spectral regions of methyl ^{1}H-^{13}C SOFAST-HMQC spectra recorded at 850MHz at 40°C, of the 488nm-light-induced state of rsFolder, focusing on cross-peaks experiencing line narrowing upon pH-decrease (top to bottom), and buffer change (right to left).

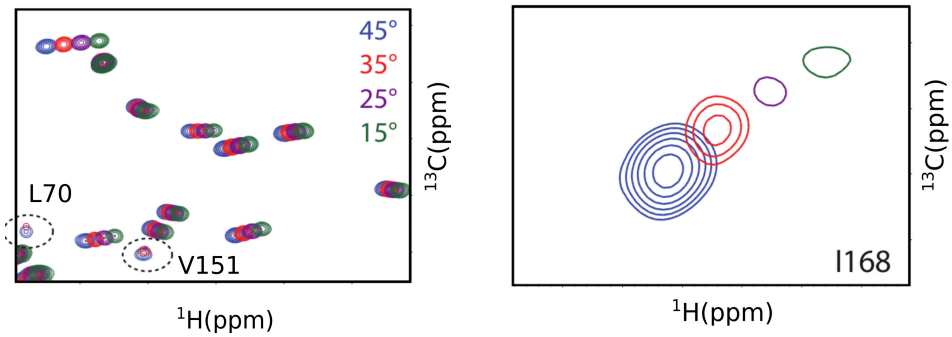


Figure 3.39 Spectral region of 2D methyl ^{1}H-^{13}C SOFAST-HMQC spectra of the "off"-state of rsFolder at pH=7.5, recorded at different temperatures at 850MHz. Focusing on cross-peaks which undergo chemical shift difference upon increase of temperature, significant line narrowing is induced at high temperature in the "off"-state. This behaviour is indicative of residues undergoing intermediate exchange in the "off"-state at low temperatures and fast exchange upon increase of temperature. Left panel shows several methyl residues while the right panel focuses on I168-δ.

The rapidly (de)protonated residue can not be the *trans*-CRO, since its (de)protonation happens at a much slower timescale and higher pH. Hence, the relatively fast protonation of another readily available residue in the *trans*-CRO pocket might be the source of such line broadening. For this scenario to hold, one would expect the peaks of the nuclei

whose CS is affected by the protonation event, to titrate according to pH in the "off"-state spectra recorded, since CS_A and CS_B are expected to be significantly different (A and B configurations have a difference of a proton). So far, peaks experiencing line narrowing were only slightly titrating (Figure 3.40), suggesting that the residue experiencing faster protonation at low pH and/or presence of phosphate ions is characterized by a significantly skewed population (eg p_A»p_B).

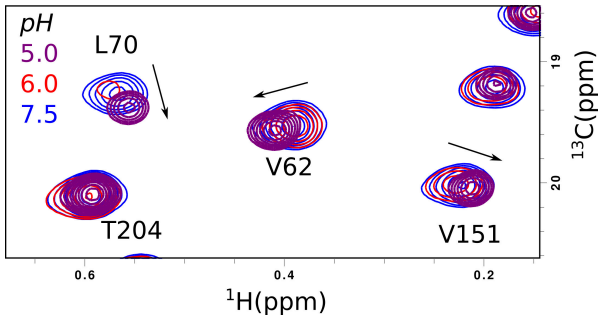


Figure 3.40 Spectral region of 2D methyl ^{1}H-^{13}C SOFAST-HMQC spectra of the "off"-state of rsFolder samples at pH=5.0, 6.0, 7.5, recorded at 40°C and 850MHz. Focusing on cross-peaks which undergo chemical shift difference and line narrowing upon pH change, V62, V151 and L70 appear to slightly titrate, due to the presence of a nearby residue nearby that is characterized by fast exchange between two protonation states.

Inspecting the free energy landscape of the rsFolder "off"-state in the absence of light

In order to explore the pH-induced differences in the energetic landscape of the "off"-state, we decided to study the thermal relaxation times and relative populations of the *cis* and *trans* states of the rsFolder sample at different pH, in the absence of light. Moreover, these experiments allowed us to assess how the observed speeding up of dynamics experienced by the *trans*-CRO pocket affects the *trans*-CRO's stability. Thermal relaxation times measured at different temperatures can be used to estimate activation energies for ground-state *trans-cis* isomerization, E_a, while population ratios of the two states are related to the free energy difference between them, ΔG.

2D amide ^{1}H-^{15}N BEST-TROSY spectra were used to accurately measure populations of *cis* and *trans* species present. More specifically, quantification of the relative *cis* and *trans* populations was done by extracting the relative intensities of peaks which undergo significant CS changes upon *cis-trans* isomerization while their linewidths are not affected by the dynamics of the *trans*-CRO (in $\beta-10$, $\beta-11$ strand, Figure 3.20). This quantification could not be achieved using 2D correlation spectra focusing on methyls or aromatics, since the peaks that undergo CS changes upon *cis-trans* isomerization are mostly line-broadened in the "off"-state. Moreover, the *cis* isomer (present primarily in the "on"-state) is in fast exchange between a protonated and a deprotonated form, which is resulting in a single set

of population-weighted NMR signals, including amide peaks in the spectrum, so we did not consider N-*cis* seperately from A-*cis*. The *trans*-CRO (of the "off"-state) is primarily in its protonated form, so the A-*trans* species were not taken into account in the free energetic landscape. Therefore, assessing the population ratios and thermal relaxation times using 2D amide ^{1}H-^{15}N BEST-TROSY spectra we could derive a simplified free energy landscape characterising the rsFolder CRO at different pH and buffer composition, whereby the CRO can occupy either the *cis* or the *trans* conformation in the ground state (Figure 3.43).

While under 488nm illumination, the majority of the CROs found in the sample adopt the thermodynamically metastable *trans* configuration, in the dark, they are expected to progressively return to the *cis* state. By measuring the rate of thermal relaxation, therefore, we can estimate the difference in activation energies, E_a, needed to undergo ground-state *trans-cis* isomerization at different pH values and buffer composition (using the Arrhenius equation 2.3). To measure NMR-observed thermal relaxation time, we chose to run a series of real-time 2D amide ^{1}H-^{15}N BEST-TROSY experiments on rsFolder samples, at different pH and buffer composition, in the dark after 488nm illumination.

Moreover, differences in free energies ΔG of the "on"- and "off"-states in the ground state can provide insight on the relative "stability" of the *cis*- and the *trans*-CRO at different pH values and buffer composition. To estimate population ratios and, consequently, estimate ΔG, 2D amide ^{1}H-^{15}N BEST-TROSY spectra were recorded for the dark-adapted sample (in the absence of light).

Overall, the thermal relaxation rates measured for a range of pH values in the McIlvaine series reveals an interesting behaviour – as the pH decreases, thermal relaxation significantly increases. Evidently, this is a sign of a higher energetic barrier, E_a needed to overcome for *trans-cis* isomerization in the ground state (Figure 3.41). However, a maximum is reached at pH=5.0, where the thermal relaxation reaches a time constant of 125hours (as compared to 14hours at pH=7.5), before it goes down again at pH<5.0. At pH=4.2, the time constant describing the ground-state *trans-cis* isomerization is back to only 16hours. When extracting the activation energies of ground state *trans-cis* isomerization for rsFolder at pH=5.0 and pH=7.5, we estimated a difference of 5.7kJ/mol (assuming that the transition state remains the same), which is in the range of reported energies required for an H-bond to be formed [161]. Therefore, an H-bond is most likely becoming progressively stronger from high pH to pH=5.0 before it weakens again at pH<5.

The progressive appearance of *trans*-CRO upon lowering pH in the absence of light is a sign of *trans*-CRO stabilisation (Figure 3.42). At pH=5.0, the population of *trans* present in the dark-adapted state is reaching its maximum value 30%, before it diminishes again at lower pH. The free energy differences between the two states for the pH=5.0 and pH=7.5 McIlvaine samples are 2.2kJ/mol and >12kJ/mol respectively (Figure 3.43). Hence the "off"-state (i.e *trans*-CRO) is significantly stabilized at pH=5.0, resulting in pH-induced *cis-trans* isomerization. Note that this pH-induced isomerization could not have been predicted by measuring absorption or emission spectra of the protein since N-*trans* and N-*cis* would not have been easily distinguishable.

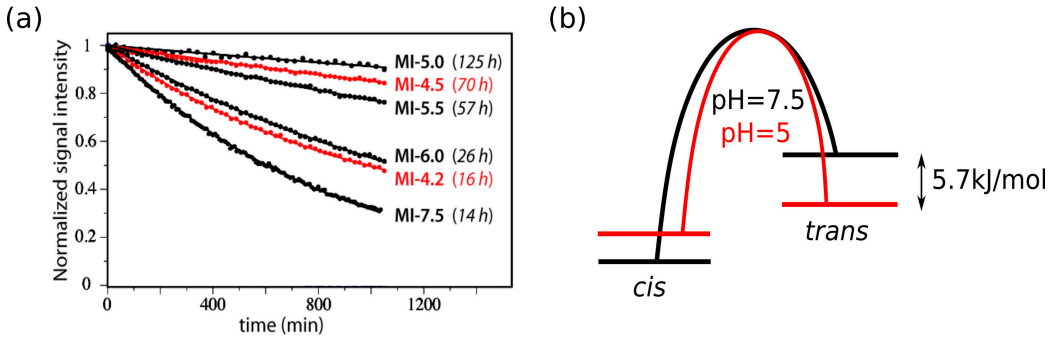


Figure 3.41 (a) Thermal relaxation times of "off"-to-"on" recovery in the absence of light for samples in McIlvaine buffer, pH=4.2 – 8.0, evaluated from real-time NMR spectra recorded at 850MHz at 40°C. (e) Schematic representation of the differences in activation energies, E_a, for the *trans-cis* isomerization in the ground state, assuming the same transition state.

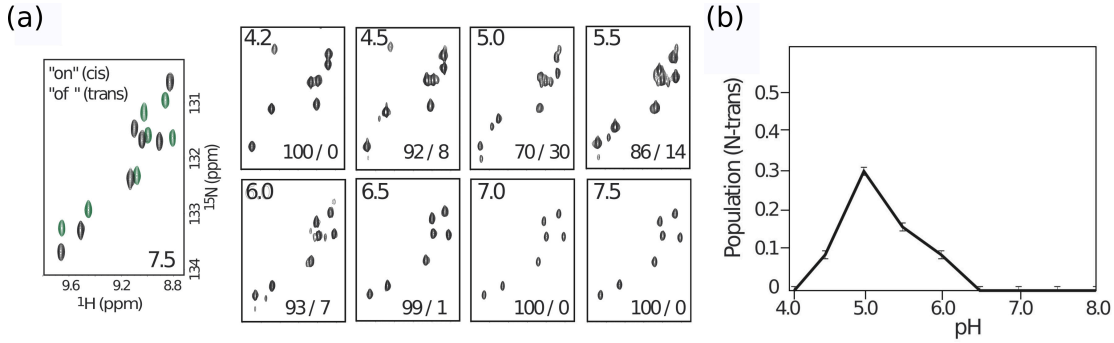


Figure 3.42 (a) On the left, overlay of (color-coded) spectral regions extracted from 2D amide ^{1}H-^{15}N BEST-TROSY spectra recorded at 850MHz, 40°C, for the "on"-(green) and the "off"-state (black), on the right, spectral regions extracted from 2D amide ^{1}H-^{15}N BEST-TROSY spectra recorded at 850MHz at 40°C for rsFolder samples of pH values in the range of 4.2-8.0 in the dark-adapted state. (b) Plotting the percentage intensity of the peaks corresponding to the N-*trans* conformation extracted from the spectra at (a), against pH values.

A potential role of H149 in stabilising the *trans*-CRO

The observed free energy minimum of the *trans*-CRO at pH=5 may potentially be attributed to the formation of an H-bond to the β-barrel causing such structural stabilisation (assuming that thermodynamic stability is due to decreased *trans*-CRO dynamics). One potential H-bond partner is H149, already suggested by the low-temperature crystal structure of the "off"-state of rsFolder [61].

More specifically, the only residue in the vicinity of the *trans*-CRO that could be ex-

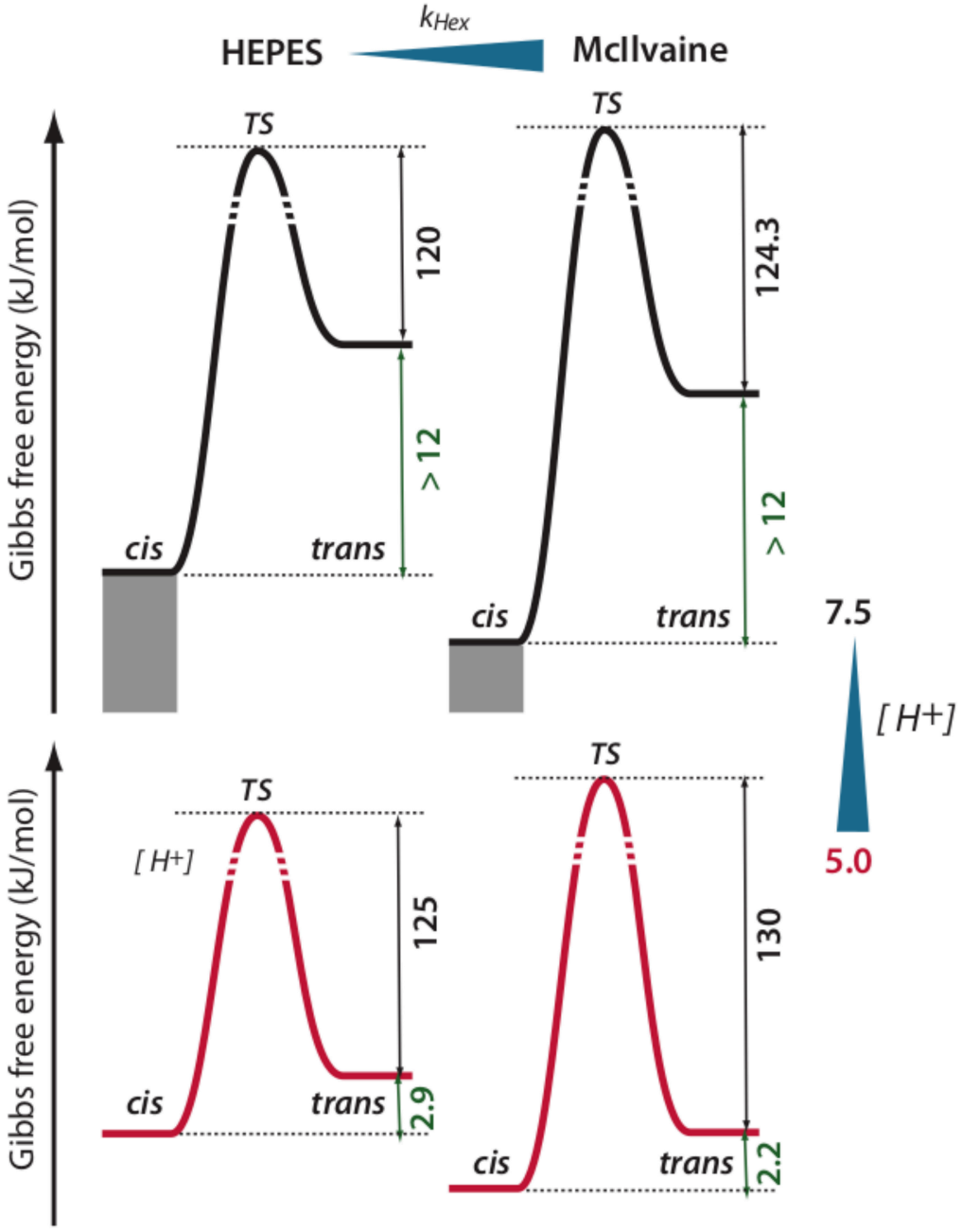


Figure 3.43 Schematic representation of the differences in activation energies, E_a, for the *trans-cis* isomerization and free energies, ΔG, for the *cis*- and *trans*-CRO in the ground state, at pH=5.0 and pH=7.5, for samples in the McIlvaine and HEPES buffer. Values for E_a and ΔG were estimated by measuring the thermal relaxation times and population ratios of *cis* and *trans* species present, respectively, at different pH values and buffer composition.

periencing a faster proton exchange at pH≈5 or in the presence of phosphates is H149. Other residues in the vicinity of the *trans*-CRO have typically pKa values far from 5 (Valines, Isoleucines, Leucines, Threonines). Accordingly, for our hypothesis to be correct, the imidazole ring of H149 would have to exchange protons between its $N_{\epsilon 2}$ and $N_{\delta 1}$ nitrogens. Upon lowering the pH, the most favourable tautomeric state in the "off"-state would need to be the one of a deprotonated $N_{\delta 1}$ (as suggested by crystallography [61]), in order to allow for a transient H-bond with the protonated *trans*-CRO.

In the 2D aromatic ^{1}H-^{13}C BEST-HSQC spectra recorded of the "off"-state at pH=4.2-8, we observed the $C_{\epsilon 1}$ peak of H149 only slightly titrating in the range of pH=5-8 (Figure 3.44). Frequencies of H_{ϵ}-$C_{\epsilon 1}$ are highly similar between the *cis* and the *trans* conformation, indicating no difference in charge state upon isomerization. However, the frequencies of $C_{\delta 2}$ and C_{γ} of H149 in the "off"-state at all pH values tested could not be extracted. Hence, H149 in the "off"-state remains singly protonated at $pH \geq 5$, while the interconversion between the two tautomeric states causes line broadening of the $C_{\delta 2}$ and C_{γ} peaks. Note that the expected differences in CS between the two tautomeric states for $C_{\delta 2}$ and C_{γ} are quite large ($\approx 9ppm$ and $\approx 4ppm$, respectively, shown in Figure 3.35), thus line-broadening is possible if the interconversion between two tautomers is in the fast-intermediate regime. Our data are in support of the hypothesis that H149 is exchanging protons between its ring nitrogens. Although we can not conclude specifically on a favouring of a $N_{\delta 1}$ deprotonated tautomer, we expect that the tautomer interconversion significantly promotes one tautomer (evident from H149 H_{ϵ}-$C_{\epsilon 1}$ weak titration) which enables the H-bonding with the *trans*-CRO and, thus, its apparent stabilization at pH=5.

At pH<5, the H149 H_{ϵ}-$C_{\epsilon 1}$ undergoes significant CS, possibly due to its change in protonation from a singly to a doubly protonated Histidine. This shift would be in agreement with the apparent destabilization of the *trans*-CRO at pH<5, suggesting that the *trans*-CRO's H-bond with H149 is no longer possible once this Histidine is doubly protonated (Figure 3.44). The lack of H-bonds of the *trans*-CRO with the β-barrel results in its reduced presence in the absence of light and its overall destabilization at pH<5.

Altogether, while lack of signal impedes the ability to draw a strong conclusion, our observations point towards the role of H149 in stabilising the *trans*-CRO by forming a transient H-bond at pH=5. We hypothesize that this bond is enabled by a fast exchange of protons between the $N_{\delta 1}$ and $N_{\epsilon 2}$ nitrogens, resulting in a predominantly deprotonated $N_{\delta 1}$ in H149, at pH=5. Additionally, the dynamics previously observed in Section 3.2 experienced by the "off"-state seem to be due to the intermediate-fast protonation event occurring in the pocket of the *trans*-CRO.

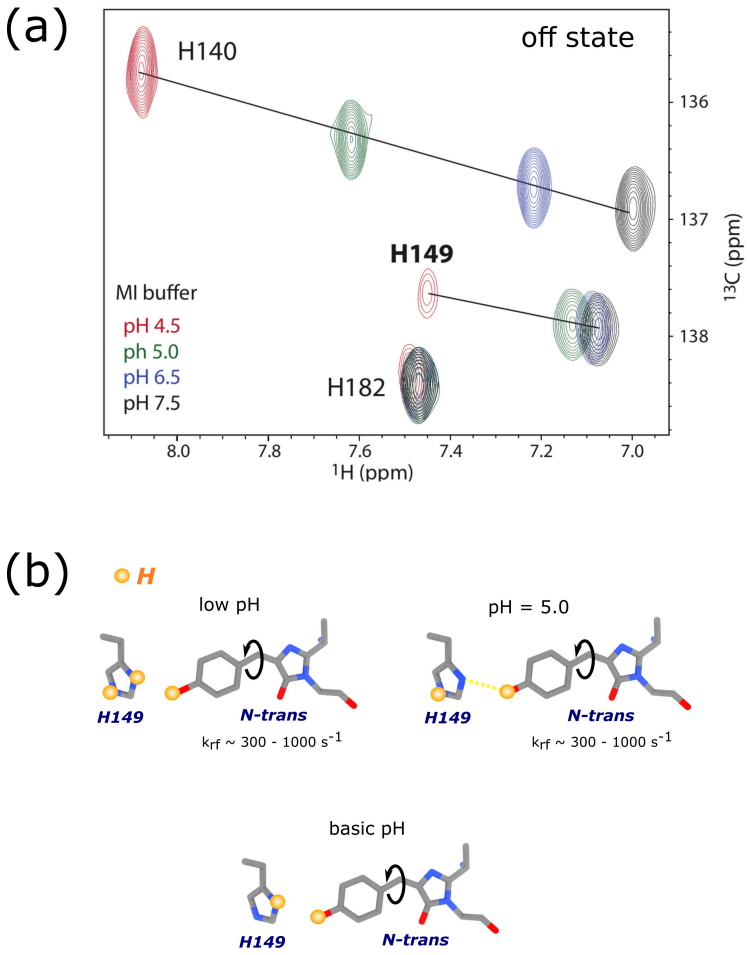


Figure 3.44 (a) Spectral region of 2D aromatic ^{1}H-^{13}C BEST-HSQC spectra, focusing on the cross-peaks corresponding to the $H_\epsilon - C_\epsilon$ of His149 of rsFolder in the 488nm-induced "off"-state at pH=4.5-7.5, recorded at 850MHz at 40°C. Annotated are the peaks corresponding to H140 significantly titrating with pH and H182, not titrating with pH, while H149 is slightly titrating at pH=5-7.5 and significantly shifts at pH<5. (b) Schematic representation of the model proposed for the pH dependence of the H149 – *trans*-CRO bond. At pH<5, H149 is primarily doubly protonated, ubable to form an H-bond with the *trans*-CRO. At pH=5, we hypothesize that H149 is in fast interconversion between its two tautomeric states, with an increased favouring of the tautomer where $N_{\delta 1}$ remains deprotonated, to allow for an H-bond with the *trans*-CRO. At pH>5 H149, there is a favouring of the tautomer where $N_{\delta 1}$ is protonated, not allowing for the formation of an H-bond with the *trans*-CRO.

3.3.4 Impact of NMR observations of pH-induced effects on the optical properties of rsFolder

Preliminary studies on light-excited states using NMR

Typically, in the work presented in parts 3.1 and 3.2, before recording spectra corresponding to the "on"-state of rsFolder, a short but powerful 405nm pulse would be used on the sample to switch the protein "on". When lowering the pH, however, we observed a discrepancy in the populations of CRO species present between 405nm-light-induced or dark-adapted "on"-state samples. More specifically, there is a progressive increase in population of N-*trans*-CRO species present at lower pH, even lower than pH=5, after 405nm illumination. Overall, for pH<7.5, the N-*trans* becomes more favored upon 405nm illumination as compared to a situation where the sample has thermally relaxed to a thermodynamic equilibrium, in the dark. Furthermore, pH-induced population differences are also evident in the 488nm illuminated samples. For pH<6, the *trans* population is decreased, giving rise to an increasing amount of *cis*-CRO species. The fraction of populations present under different illumination schemes and pH values is plotted in Figure 3.45.

Correlating NMR observables with optical measurements

These observations potentially have an impact on the observed photophysical behaviour of rsFolder. To evaluate the correlation between NMR findings and expected photophysics, I performed optical measurements on rsFolder at different pH values. Using the microscopy setup for ensemble measurements, I recorded a series of photocycles of fluorescence for rsFolder samples fixed in PAA gel at pH=5 and pH=8 (Figure 3.46) following protocols described in the Materials $\&$ Methods section.

I fitted the intensity of fluorescence over time using a script based on a kinetic model of exchange between CRO-species with different extinction coefficients for a given wavelength of excitation. More specifically, I estimated the extinction coefficient of the A-*cis*-CRO at 488nm by comparing the absorption spectra of rsFolder samples of same concentration at pH=8.0 ($\epsilon_{488}^{A-cis,pH=8} = 54518 Mol^{-1}s^{-1}$) and pH=5.0 $\epsilon_{488}^{A-cis,pH=5} = 12270 Mol^{-1}s^{-1}$, while the extinction coefficient for the *trans*-CRO at 405nm was evaluated from the expected populations measured in NMR ($\epsilon_{405}^{N-trans,pH=8} = 36137 Mol^{-1}s^{-1}$, $\epsilon_{405}^{N-trans,pH=5} = 32523 Mol^{-1}s^{-1}$).

The fitted photophysical quantities are given in the table below (Table 3.2). We were able to draw three main conclusions: (a) both the "on"-to-"off" and "off"-sto-"on" photoswitching quantum yields seem to decrease at lower pH, (c) the maximum fluorescence emitted in the beginning of each photocycle is significantly decreased at low pH and (c) the switching contrast worsens at acidic pH. To explain the differences observed between the acidic and basic sample, one needs to consider that these values report on either population-weighted average of all CRO-species present (maximum fluorescence, switching contrast)

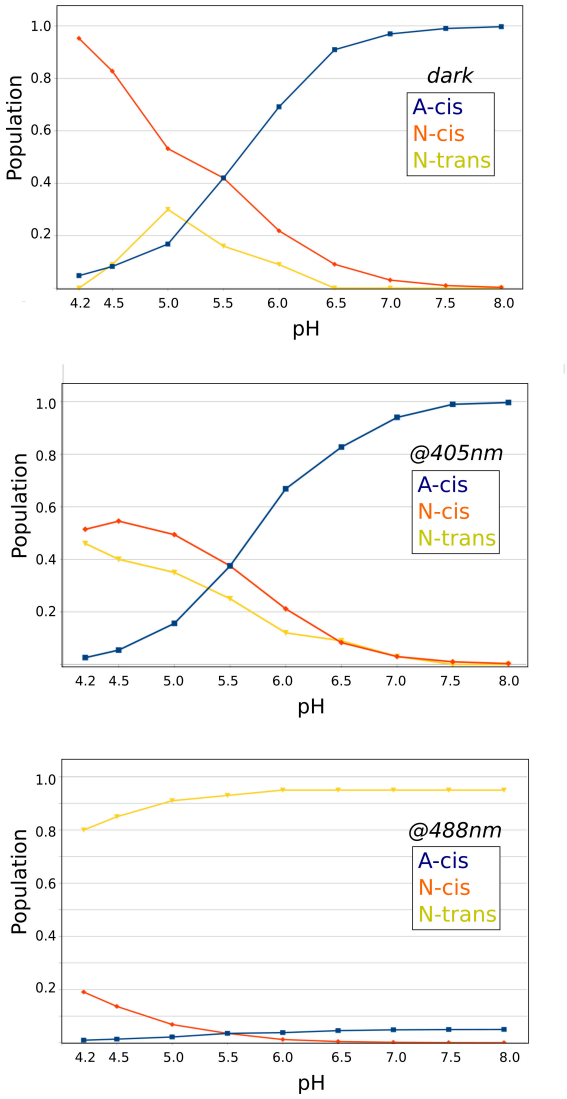


Figure 3.45 Percentage of CRO present at a given pH value and illumination scheme (in the dark, at 405nm and 488nm). Intensities were extracted from the characteristic *cis*-CRO ("on") and *trans*-CRO ("off") signatures detected in the 2D amide ^{1}H-^{15}N BEST-TROSY spectra recorded at 850MHz, 40°C. The total amount of *cis*-CRO species, after taking into account the *trans* species present, is a combination of A-*cis* and N-*cis* species. The corresponding A-*cis* and N-*cis* species for each pH were, therefore, estimated using the $pKa_{cis} = 5.5$.

or population-independent photophysics (switching QY) of each CRO-species.

By using the correct extinction coefficients, we can account for the population differences between the two pH samples, thus allowing for extracting population-independent "on"-to-"off" and "off"-to-"on" switching quantum yields (rather than population-dependent "on"-to-"off" and "off"-to-"on" switching rates). Therefore, our results show that the CRO

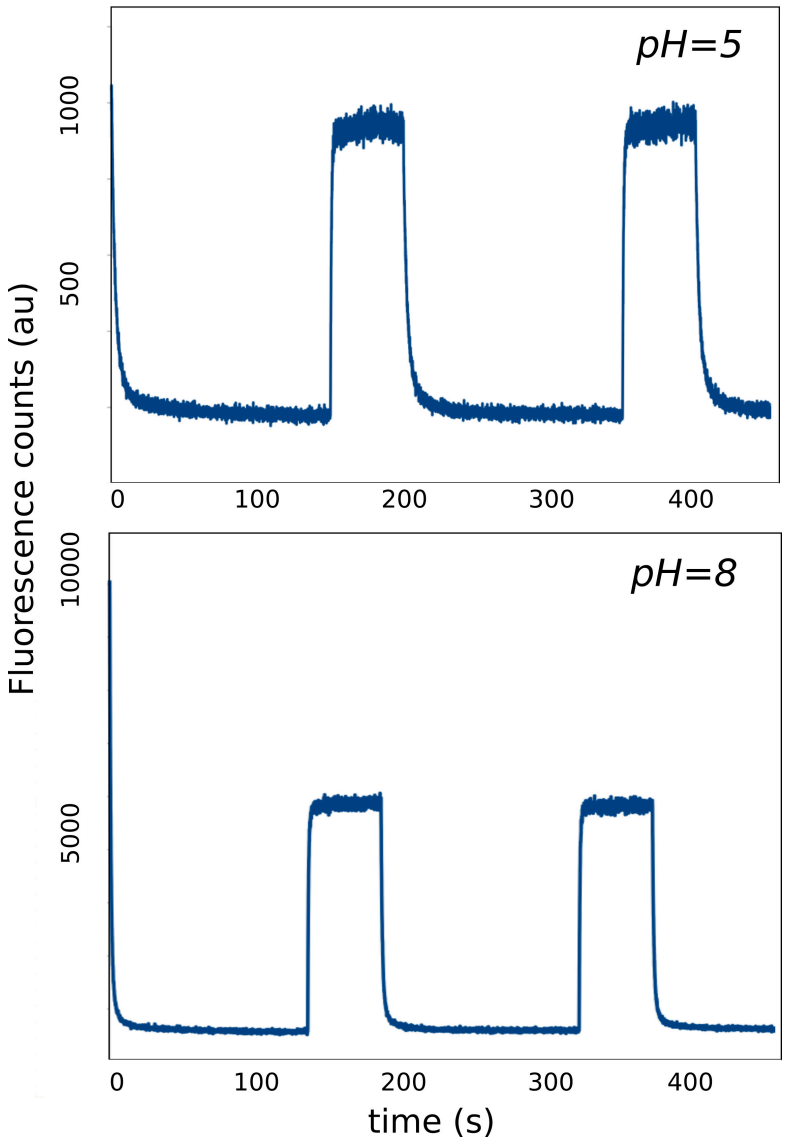


Figure 3.46 Plotted emission intensity over time of the induced photocycles, recorded in the microscope setup of rsFolder samples fixed in PAA gels at pH=5 and pH=8. LASER power, sample concentration and sample thickness were the same between the two samples. The low acidic sample presents decreased fluorescence and switching efficiency as compared to the high pH sample.

photoswitches slower at lower pH, in both "on"-to-"off" and "off"-to-"on" directions. The slowing down of "off"-to-"on" switching would be in agreement with our NMR-derived increase in energetic barriers for *trans-to-cis* isomerization at pH=5 as compared to pH=7.5 (Figure 3.43). However, this is not the case for "on"-to-"off" switching, whereby the energetic barrier as deduced from our NMR data is lower for pH=5 than pH=7.5 and thus was expected to be more efficient at lower pH. Nevertheless, while at both pH values, the

trans-CRO is expected to be primarily protonated, the *cis* state populations of protonated and deprotonated species are different in the different pH values measured. Therefore, the energetic barrier estimated by NMR, reports on the average switching efficiency of both N-*cis* and A-*cis* to switch to the *trans* state. In contrast, optical measurements report on the efficiency of only the fluorescent species (i.e A-*cis*) to isomerize. Most likely, the higher presence of the highly dynamic N-*cis* species at pH=5 is what leads to an apparent decrease in the NMR-derived energetic barrier for isomerization as compared to pH=7.5, indicating that the N-*cis*-CRO requires less energy in order to isomerize than the A-*cis*-CRO, which is primarily populated at basic pH.

However, at this point, the question arises on whether the predicted extinction coefficients used as an input in the kinetic model are accurate. Our NMR-derived model describing the behaviour of the N-*trans* species would suggest that the H-bonding with the H149 is pH-dependent. Therefore, this should have a significant impact on its absorption capabilities at different pH values. Potentially, the more constrained N-*trans*-CRO at pH=5 is more efficient at absorbing 405nm light as compared to the more flexible N-*trans*-CRO at higher pH. In this case, the "off"-to-"on" switching quantum yield would be less than the one estimated.

The maximum fluorescence recorded in the beginning of a photocycle is 8.5 times higher for rsFolder at pH=8 as compared to pH=5. NMR-expected A-*cis* CRO species are 6.4 times higher in the basic sample as compared to the acidic one, while pKa-derived populations would suggest only a 4.2 times higher population. Assuming that the concentration of rsFolder in both samples is similar after the PAA fixation, this result would indicate that at low pH, pH-induced *cis-trans* isomerization significantly hinders the total fluorescence. Thus our NMR data better predict the expected fluorescence, since they report on the amount of individual CRO species present in the sample.

The optically observed switching contrast matches the expected switching contrast from our NMR data. NMR-observed switching contrast in the light-induced regime can be calculated by taking the ratio between A-*cis*-CRO present after 405nm and after 488nm illumination. At pH=5, NMR-observed ratio is 8.2 while at pH=8 it becomes 20. Therefore, our NMR data could predict an observed decrease in experimental switching contrast, due to the alteration of the free energy landscape characterizing the *cis* and *trans* species, upon pH change. NMR's ability to decipher between *cis* and *trans* species allows it to relatively accurately predict the expected switching contrast of rsFolder and potentially other RSFPs.

3.3.5 Interpretation of NMR results on ground-state (de)protonation and isomerization kinetics

Photophysical properties of FPs are mainly influenced by the CRO – its rigidity, planarity and protonation state, influenced by its direct environment. Following the work concerning the generalised dynamics of rsFolder in its two characteristic states, it was crucial to investigate further the intrinsic properties of the CRO. For further CRO inspection, pH titration studies

pH	Max Fluo (Counts)	"on"-"off" QY (%)	"off"-"on" QY (%)	Switching contrast	ϵ_{488}^{A-cis} $(Mol.cm^{-1})$	$\epsilon_{405}^{N-trans}$ $(Mol.cm^{-1})$
5.0	1070 $\pm$ 40	0.30 $\pm$ 0.01	17.1 $\pm$ 0.6	6.1 $\pm$ 0.6	12270	32523
8.0	9130 $\pm$ 80	0.53 $\pm$ 0.03	21.6 $\pm$ 0.6	16.2 $\pm$ 0.3	54518	36137

Table 3.2 Results obtained by fitting the fluorescence photocycles measured for rsFolder samples fixed in PAA gels at pH=5 and pH=8, to a kinetic model including two dark states ("off"-state and photobleaching/blinking). Photocycles were recorded from three different regions of each sample. Correlation between experimental and modelled data is 0.99.

were the most appropriate, given not only the overall sensitivity of NMR to pH-induced phenomena but also the expected change in protonation of the CRO. Consequently, we could derive models explaining the characteristic behaviour of the *cis* and *trans* conformation in its two protonation states in the electronic ground state.

Doing this work, it became clearer how our NMR findings could have an impact on microscopy. This part is, therefore, dedicated in placing this work in the general framework of existing bibliography as well as discussing the impact it may have on the practical use of RSFPs.

The effect of the β-barrel in changing the pKa of the CRO species

The pH of the solvent in which the protein is dissolved seems to affect the protonation state of the CRO in most GFPs [158]. Proton transfer is bi-directional, i.e the CRO gets protonated by receiving a proton from the solvent and gets deprotonated by releasing a proton to the environment. Such proton exchange with the solvent is either direct, when the CRO essentially behaves as a free CRO in solution or via the protein scaffold, when it is protected by the β-barrel.

Generally, free GFP-CROs in solution have a $pKa \approx 8.2$ [210, 191]. In GFPs, the β-barrel surrounding the CRO seems to affect the pKa of its tyrosine-type phenol [191], resulting in *cis*- and *trans*-CROs having a lower (around 6) and higher pKa (around 11), respectively. Notably, depending on the GFP, the pKa of either *cis*- or *trans*-CROs differs, due to differences in the protein sequence.

By deciphering the 4 major CRO configurations by NMR, we were able to further comprehend the protonation/deprotonation of the *cis*- and the *trans*-CRO. Investigating the dynamics of rsFolder in either CRO configuration, we were able to assess how the β-barrel impacts the protonation and isomerization state of the CRO.

rsFolder "on"-state dynamics – how they influence pKa_{cis} values and CRO dynamics

Our previous work (section 3.2) suggested that in the "on"-state at pH=7.5, the *cis*-CRO is well-protected by the β-barrel. One might speculate that, in the conformationally stable "on"-state at pH>pKa_{cis}, the "unwanted" proton is released from the *cis*-CRO via the convenient H-bonds with T204, the water molecule or H149. Nevertheless, rsFolder has a significantly lower pKa_{cis} as compared to its close relative, rsEGFP2 [61], even if in both proteins, their *cis*-CROs are shown in crystallography to be stabilized by the same H-bonds, namely, with H149, T204 and a water molecule. Hence, the choice of H-bond partners does not seem to be the reason for differences in pKa_{cis} values.

Most likely, what defines the pKa_{cis} values is a network of residues around the β-barrel, beyond (but possibly including) the residues involved in H-bonds with the CRO. Many studies suggest the existence of a "proton wire" network in the protein scaffold of GFPs, bringing in and pushing out protons from the CRO, as already discussed in section 3.2 [8, 201, 199]. Therefore, literature suggests that not only the H-bond partners, but the overall β-barrel determines the pKa of the CRO in either isomerization state. The efficiency of this "proton wire" network in GFPs to shuttle protons at different pH values must be, therefore, what defines the CROs rate of (de)protonation. More specifically, if pKa_{cis}<pKa_{free}, the protein scaffold is readily ejecting protons out of the *cis*-CRO in the "on"-state, rendering it deprotonated at pH values whereby a free CRO in solution would be neutral.

Consequently, the difference in pKa_{cis} between rsFolder and rsEGFP2 suggests that rsFolder is characterized by a "proton wire" network that can more efficiently shuttle protons away from the *cis*-CRO as compared to the network in rsEGFP2. The higher efficiency of the rsFolder's "proton wire" network in releasing protons to the solvent has to be related to either a different pattern of hydrogen "hopping" points in the two structures (i.e beyond H149, T204 and the water molecule), or a difference in dynamics that those "hopping" points experience. At this stage of our work, it is not clear which explanation is more appropriate.

Nevertheless, our data regarding rsFolder were instructive in understanding how β-barrel dynamics in the "on"-state influence the pKa and intrinsic properties of either protonation state of the *cis*-CRO. Firstly, NMR results derived from residues nearing the *cis*-CRO allowed for the quantification of the (de)protonation rate in rsFolder, occurring at a fast sub-ms timescale. Secondly, apparent ring flip dynamics were deconvoluted between the intrinsic ring flipping rate of the phenol ring of the N-*cis* and of the A-*cis* species.

Furthermore, the β-barrel must be applying an impressively strong constraining force on the A-*cis*-CRO, essentially achieving a 4-orders of magnitude increase in rigidity ($k_{rf}^{A-cis} = 10 \pm 1s^{-1}$, $k_{rf}^{N-cis} = 7500 \pm 1000s^{-1}$). Such rigidity is enabled by the anchoring of the A-*cis*-CRO to the barrel, via H-bonds to T204, a water molecule and H149, as suggested by crystallography [61]. Moreover, the role of H149 and T204 in stabilizing the A-*cis*-CRO is supported by pre-existing work, namely a structural comparison between the dim wtGFP and its brighter mutant, GFP(S65T). The authors observed an increase in anionic species of

CRO present in GFP as compared to its ancestor, which was attributed to the stabilisation of this species by the H148 and T203 side chains [29].

Below pH=5.5, there is a favouring of the protonated CRO configuration (either N-*cis* or N-*trans*). In the absence of light, we observed increased dynamics, i.e the *cis*-CRO-H149 H-bond is broken leading to a highly dynamic N-*cis*-CRO. Most likely, residues which are part of the "proton wire" network become protonated (such as H149 which probably becomes doubly protonated) and the network is disrupted. Therefore, at low pH, the *cis*-CRO is protonated as a result of the lack of protection from the β-barrel rather than due to an efficient protonation pathway from the solvent to the CRO via the protein scaffold. The N-*cis*-CRO, seems to behave as a free CRO in solution, possibly due to loss of all H-bonds to the β-barrel, further suggesting that the protein scaffold does not protect it from solvent hydrogen ions.

rsFolder "off"-state dynamics – how they influence pKa_{trans} values and CRO dynamics

Upon 488nm illumination, the A-*cis* isomerizes to the N-*trans*. As already concluded from our previous work (section 3.2), the N-*trans* is inherently more dynamic than the A-*cis*-CRO. The increased dynamics of the rsFolder barrel in the "off"-state could be potentially thought of as resulting in a less protected and thus primarily protonated *trans* CRO (as would be the case for a free CRO in solution). However, the *trans* CRO has a much higher pKa than a free CRO and does not behave like a free CRO (experiences intermediate rather than fast dynamics). The extremely high pKa_{trans} suggests that, instead of being less protected, the N-*trans* CRO is still somehow promoted, even at pH values whereby a free CRO would normally be deprotonated (i.e pH>8.2). One reason for the appearance of N-*trans* CRO could therefore be a "proton wire" network of the "off"-state which can quickly reprotonate the *trans*-CRO, even at high pH values. Since the (de)protonation of the *trans*-CRO seems to be very slow, our NMR data do not point towards the existence of such an efficient "proton wire". Instead, we can conclude that the structure and dynamics of the "off"-state β-barrel promote the N-*trans* CRO by partially structurally stabilizing it.

While we could not retrieve conclusive information regarding its H-bonding partners, we hypothesied that H149 is in a transient H-bond with the N-*trans*-CRO, which seems to get stronger upon lowering of pH, achieving maximum N-*trans*-CRO stability at pH=5. The apparent stabilization of the "off"-state CRO at pH=5 can explain the pH-induced isomerization in the absence of light. Hence, the faster proton shuffling causing rapid (de)protonation of nearby residues (most likely H149) allows for the formation of a more stable H-bond between the β-barrel and the CRO. Nevertheless, the stability of the N-*trans*-CRO in the absence of light is diminished once the H149 becomes doubly protonated and cannot accommodate an H-bond with the CRO.

Lastly, no peaks corresponding to the HC pairs (ϵ or δ) of the phenol ring of the *trans*-CRO could be detected, indicating that at all pH values measured, it experiences dynamics (either translational or ring flips) in the few hundreds of μs-ms timescale. Thus the N-*trans*

appears less dynamic than the fast flipping N-*cis*.

CRO species are populated according to β-barrel dynamics

While the (de)protonation efficiency of the β-barrel determines the ratio between A-*cis* and N-*cis*, the H-bond strength of the H149 with the CRO possibly regulates the appearance of N-*trans*-CRO in the absence of light. The progressive appearance of the N-*trans*-CRO, at lower pH values can be potentially explained by a ground state N-*cis*-to-N-*trans* isomerization. More specifically, upon H-bond weakening, the N-*cis*-CRO is undergoing significant conformational changes, including (possibly) hula-twisting. Hence, a part of the population of very mobile N-*cis*-CROs can be partially "trapped" in the N-*trans* conformation possibly via an H-bond which transiently "anchors" it to the β-barrel – allowing for the occurrence of both N-*cis* and N-*trans* configurations.

Altogether, the 3-H-bonded A-*cis*-CRO appears to adopt the most rigid configuration. On the other hand, the N-*trans*-CRO, depending on the pH, forms a transient H-bond with (potentially) H149 which can partially restrict its motions. Lastly, the N-*cis*-CRO behaves like a free CRO, possibly because at low pH, there is no remaining H-bond partner that is readily deprotonated to allow for its "anchoring" with the β-barrel.

Impact on the use of RSFPs in microscopy

The last series of experiments within this section (3.3) aimed at understanding how the predisposition of rsFolder at a particular ground-state energetic landscape might influence the behaviour of its light-excited states. This could be a way to use our NMR data to partially predict the behaviour of rsFolder when used in fluorescence microscopy. Particularly rsFolder which was constructed to endure low pH environments (eg. bacterial periplasm), appears to have significantly altered its photophysical properties in the ensemble level upon lowering the pH.

Typically, in the beginning of the optical measurements recorded in the microscope, a 405nm LASER pulse is used to switch "on" rsFolder, before recording fluorescence using the 488nm LASER for excitation. Our NMR data show important differences between population ratios of *cis* vs *trans* CRO species present upon 405nm illumination at pH=5 as compared to pH=8. Consequently, we retrieved a lower maximum fluorescence point on the first frame recorded of the photocycle of rsFolder at pH=5.

One explanation for the rise of N-*trans* is the tendency of the N-*cis* species to absorb at 405nm causing photoisomerization to the *trans* conformation, as observed in GFP, previously [213, 29]. However, the *trans*-CRO also absorbs at 405nm and isomerizes to the *cis* conformation, thus both N-*cis* and N-*trans* are present at steady-state conditions. Therefore, assuming a pH-independent N-*cis*-to-N-*trans* and *vice versa* switching rate we would expect that the amount of N-*trans* species present at different pH, follows a Henderson-Hasselback

curve of pKa=5.45. In contrast, our data suggest that at pH values around pH=5, there is a higher built-up of N-*trans* species than predicted from such a curve (Figure 3.47). This behaviour is further indication of the pH-dependent stability of the N-*trans*, since its ability to isomerize to the N-*cis* state is pH-dependent.

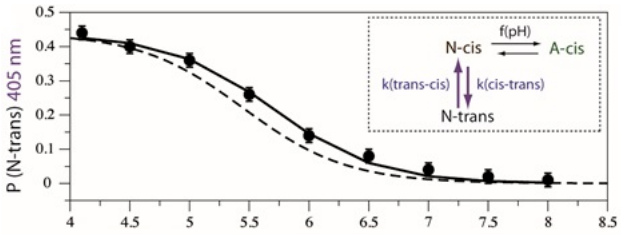


Figure 3.47 Plotted values of NMR-derived N-*trans* species populations at different pH values after 405nm illumination (solid line) and the expected N-*trans* populations assuming a pH-independent N-*cis*-to-N-*trans* switching (dashed line). The amount of N-*trans* species is higher than expected, suggesting pH-dependent switching under 405nm illumination.

After 405nm illumination the relative population of the fluorescent A-*cis* has decreased at low pH, as expected by NMR, due to the built-up of *trans* species from the photoswitched N-*cis* conformer. Note that the built-up of non-fluorescent species (both N-*cis* and N-*trans*) is significantly more than the expected value derived from the pKa_{cis} value, where only N-*cis* is considered. Thus, by taking into account the populations of A-*cis* species present at 405nm illumination, we could predict the decrease in maximum fluorescence of rsFolder at pH=5.

In the microscope, the 488nm LASER is progressively switching "off" the proteins, over periods of around 150s (depending on the sample). A plateau of "minimum fluorescence" is reached, which is related to the switching contrast of the sample. In NMR, after continuous 488nm illumination, upon lowering the pH, N-*trans* becomes less populated, while the total *cis* species is increased (Figure 3.45). This is probably the result of primarily a significant decrease of 488nm-absorbing A-*cis* and a greater occupancy of the 405nm-absorbing N-*cis* species, as well as differences in free energy landscapes for rsFolder in the two pH values. Furthermore, photoswitching rates are LASER-power dependent. Thus, it is expected that in different buffer conditions and 488nm illumination powers, the resulting N-*trans* population at equilibrium after continuous 488nm illumination will vary. Overall, depending on the conditions of the sample, the remaining fluorescent A-*cis* species is expected to vary, resulting in different switching contrast values.

The free energy landscape for isomerization alters with pH and buffer composition, as derived from our NMR experiments. Consequently, the "on"-switching appears slowed-down at pH=5 as compared to pH=8, since the energetic barrier of "off"-to-"on" has increased, and the increased presence of N-*cis* species at low pH, leads to an overall decrease of apparent "on"-to-"off" photoswitching rates. The energetic barrier for the A-*cis*-to-N-*trans*

could not be calculated at this point, since both *cis* species are accounted for in the free energy landscape. Focusing on single molecule properties, the difference in "on"-"off" switching QY would suggest that the A-*cis*-CRO needs to overcome a higher energetic barrier to isomerize at pH=5. Our ground-state NMR data interpret the inherent dynamics and stability of A-*cis* (ΔG_{A-cis}) to be pH-independent, thus, considering that the transition state (TS) for isomerization is little sensitive to pH, we would expect the A-*cis* CRO to require similar activation energy values to switch to the "off"-state at pH=8 and pH=5 ($E_a = TS - \Delta G$), resulting in similar "on"-to-"off" switching QY. The lower switching QY at low pH derived from our fluorescence measurements is, therefore, not in agreement with a pH-independent energetic barrier for A-*cis*-to-N-*trans* isomerization. Hence, the fluorescence data possibly report on excited-state processes that could not have been predicted from the NMR-observed ground-state behaviour of rsFolder.

Generally, with our data we cannot conclude about what are the kinetics of the CRO in the excited state. However, comparing our NMR data of energetic barriers for isomerization with optical measurements of switching quantum yields could potentially suggest that the free energy landscape in the ground state (studied by NMR) partially informs the free energy landscape of the rsFolder CRO in the excited state (studied by fluorescence). Moreover, similarly to the X-ray structures published on rsFolder, investigating mechanistic properties in the ground state can potentially be instructive in understanding the protein's photophysics. This study, thus, showed that NMR with its discriminatory capabilities, its ability to study proteins in different solutions and decipher their dynamics can be of use in better understanding RSFPs optical characteristics in different biological environments.

Future work should include investigating H-bond dynamics present in the excited states of RSFPs which inform their resulting photophysical properties. For such studies, ultrafast spectroscopy or time-resolved kinetic crystallography should be employed.

Chapter Four

Preliminary work and future perspectives

4.1 Correlating NMR observables with photophysical properties of rsFolder mutants

So far in the Results section, I have been describing the efforts of establishing an appropriate setup and identifying the most relevant and useful experiments in the context of NMR investigation of RSFPs. Overall, my PhD thesis has been an attempt to introduce NMR as an active player in the field of RSFPs. Optimizing and streamlining the NMR work on rsFolder proved more cumbersome than expected, despite the significant findings that were obtained. Consequently, this section can be regarded as a proof-of-concept, assessing the correlation between NMR observables and optical properties of rsFolder mutants presenting different photophysical behaviours. Unfortunately, this work could not continue into thorough mechanistic interpretations of the origins of the photophysics observed in the rsFolder mutants studied within the context of my thesis duration. Nevertheless, the preliminary results discussed below point towards an important contribution of NMR in RSFP engineering.

4.1.1 Rationale for choosing rsFolder mutants

As already mentioned, rsFolder is characterized by a sub-optimal switching contrast (reported as 26 in El Khatib's et al publication [61]), despite its impressively fast switching rates and exceptional endurance to "extreme" environments. Fluorescent protein engineers, in an attempt to ameliorate its switching contrast, successfully designed the high-contrast F146Y mutant of rsFolder, rsFolder2 [61]. In reality, the existence of F146 originates from rsFolder's ancestor, Superfolder-GFP, while rsEGFP2 (switching contrast = 57) has a tyrosine at that position. While rsFolder2 structures were not made available in that publication [61], it is safe to assume that they are more similar to rsEGFP2 than rsFolder. Due to its potential for application and proximity to the well-studied rsEGFP2 [47, 226, 232,

61], rsFolder2 was chosen as the central protein for a mutational investigation of selected mutants by NMR.

More specifically, rsFolder2 and 5 of its mutants were chosen to be studied by NMR: H149F, V151A, I168L, T204V and rsFolder3 – undisclosed mutant. All but rsFolder3, were inspired by our NMR conclusions regarding residues affecting or being affected by the CRO's configuration and dynamics. Notably, these mutations were not expected to enhance the properties of rsFolder or rsFolder2, but rather, to provide a range of RSFPs with varying optical properties.

4.1.2 Optical studies on rsFolder2 and mutants

To start with, an explicit optical characterization of rsFolder2 and mutants needed to be performed. With these optical data at hand we could assess whether (a) the proteins are photoswitchable, (b) how efficient is the photoswitching between the two ground states, (c) how close is the excitation wavelength of each protein to the emission wavelength of our LASERs (405, 488nm). Measurements were performed on optical spectrometers and the PALM microscope, at room temperature.

All rsFolder2 mutants are reversibly photoswitchable, as confirmed by the acquired absorption and emission spectra after repeated rounds of 405nm-488nm illumination. Therefore, in our NMR data we should expect to acquire distinct NMR signals corresponding to the "on"- and the "off"-state.

Before performing the ensemble optical measurements in the microscope, important parameters needed to be accounted for, namely the extinction coefficients of rsFolder2 and mutants in regards to absorption at 405nm and 488nm LASER light. More specifically, the data typically acquired in the photocycling experiments in the microscope of fixed proteins in PAA gels, yield the population-weighted values of photoswitching rates and contrast. Hence, a precise evaluation of the extinction coefficients i.e of the relative population of molecules activated upon illumination, is key to achieving a reliable fit of the data into the kinetic model explained in the Materials $\&$ Methods section. Following typical protocols, I could deduce the extinction coefficients at pH=7.5 of rsFolder mutants, as shown in Table 4.1.

Evidently from the absorption and emission spectra, the H149F and T204V mutants at pH=7.5 are a mixture of neutral and anionic species. At this point, the isomerization state of the CRO-species present could not be assessed. Nevertheless, the presence of such a skew in CRO-species populations results in significantly lower measured extinction coefficients, ϵ_{488}^{on} and ϵ_{405}^{off} at pH=7.5 as compared to higher pH.

Depending on the mutant studied, the photocycling regime had to be altered. That was mainly the case for mutants with slower photoswitching than rsFolder or rsFolder2, since more frames needed to be recorded to allow for the sample to reach an equilibrium. Fitting these photocycling curves into a kinetic model enabled the quantification of switching

MUTANT	MAX ABS, λ (nm)	MAX FLUO, λ (nm)	$\epsilon^{cis}_{max\lambda}$ ($Mol.cm^{-1}$)	Φ	BRIGHT (VS EGFP)	"ON"-"OFF" QY (%)	"OFF"-"ON" QY (%)	SWITCHING CONTRAST
WT	482	504	75975	0.34	0.70	0.8	14.1	44.1
Y146F	477	505	71330	0.30	0.50	1.0	22.9	18.8
V151A	483	505	63533	0.22	0.40	1.2	15.0	82.9
rsFolder3	484	505	66955	0.50	0.98	0.3	14.4	22
T204V	503	510	38355	0.41	0.30	0.4	13.3	10.8
H149F	498	505	23056	0.59	0.33	<0.002	<0.05	n/a
I168L	480	505	54394	0.24	0.35	1.0	15	49

Table 4.1 Results from optical measurements of the optical properties of rsFolder2 wild-type(WT) and mutants in HEPES (50mM), pH=7.5. First five columns show the results obtained from absorption and emission measurements of samples in cuvettes, the last 3 columns include the values obtained from fitting the emission profiles over time of photocycles measured in the microscope for fixed samples in PAA gels. H149F mutant emission profiles could not be fitted since the mutant had very slow switching "on"-"off" and "off"-"on" rates.

rates of the mutants. Consequently, the mutants at pH=7.5 can be distinguished into fast, medium and slow switchers. While rsFolder2, rsFolder2 V151A and rsFolder2 I168L present switching kinetics similar to rsFolder, rsFolder3 is slower by an order of magnitude. Particularly the H149F mutant was extremely slow and thus the full photocycle could not be recorded within the timeframe allowed for the experiments. The results are summarised in Table 4.1

Optical measurements also yielded values of the apparent switching contrast, accounting for the remaining fluorescent species by the end of a photocycle. Here again, the mutants showed a range of switching contrast values ranging from $\approx$ 10 to 80.

Lastly, fluorescence quantum yields, Φ, were estimated against fluorescein at room temperature, pH=7.5. For rsFolder3, T204V and H149F, switching quantum yields are higher than the wild-type. Alongside the values for ϵ^{on}_{488}, I could calculate the apparent brightness of each mutant at pH=7.5, some showing an exceptionally high brightness eg rsFolder3.

The table above summarises all the values measured for rsFolder and its mutants by optical means (Table 4.1). After having characterized the mutants in terms of their photophysics, we could start investigating them by NMR.

4.1.3 NMR studies on rsFolder2 and mutants

In order to streamline the NMR "screening" of mutants, we chose a limited number of experiments. Mainly, we were aiming at acquiring quantifiable measurements by NMR which seem to be relevant to specific optical properties as already observed in rsFolder. Altogether, we created the following NMR toolkit:

1. 2D ^{1}H-^{15}N BEST-TROSY experiments of the dark-adapted and 488nm-light-induced states. These experiment aimed at extracting the NMR-observed switching contrast, not only accounting for the remaining fluorescent state upon 488nm illumination but also the remaining non-fluorescent population in the dark-adapted state.

2. Series of 2D ^{1}H-^{15}N BEST-TROSY spectra over time to measure thermal relaxation after 488nm illumination.

3. 2D ^{1}H-^{13}C aromatic EXSY experiments to measure ring flip rates of the mutant's cis-CROs in the "on"-state.

4. Series of 2D ^{1}H-^{13}C SOFAST-HMQC at different pH values of McIlvaine-buffered samples to estimate the pKa of the cis-CRO. (Figure 4.1)

5. ^{15}N-^{1}H TRACT experiments to assess the oligomerisation state, in the dark-adapted and 488nm-light-induced states.

All mutants showed distinct NMR signatures for the "on"- and "off"-states. Moreover, similar CRO C_β and C_δ frequencies with rsFolder allowed for attributing the signals to *cis* and *trans* species present in the mutants. More specifically, the *cis* isomer is primarily populated in the dark-adapted state of all mutants, while the *trans* is induced upon 488nm light.

Following similar procedures and data analysis as described for rsFolder in previous sections, we derived the majority of values of NMR observed switching contrast (for both dark and light-induced states), thermal relaxation times, *cis*-CRO ring flip rates, pKa and oligomerisation states, as summarised in the Table below (Table 4.2). Note that the rsFolder2 I168L mutant was significantly more unstable than the other mutants and only low concentrated samples were achieved, since the protein would readily precipitate. Hence the signal-to-noise ratio of the NMR spectrum recorded for the 488nm-light induced state was significantly reduced, resulting in $\approx$10$\%$ of the total signal originating from noise at 40°C and 850MHz. Therefore, population ratios for the *cis* and *trans* species present in the Lit state of the I168L mutant are not included in this table.

4.1.4 Comparing NMR observables and optical measurements

Given the sensitivity of optical absorption spectra to the protonation state of the CRO, we could deduce that for most mutants, the anionic form, absorbing at around 488nm,

NMR OBSERVABLES	WT	Y146F	V151A	RSFOLDER3	T204V	H149F	I168L
Thermal relaxation "off"-"on" (min)	110	160	500	130	240	3160	440
CRO-cis ring flips (s^{-1} @30°C, pH = 11.4)	14±2	3 ±1	30 ±2	<2	<2	<2	100 ±10
Dark state contrast P_{cis}/P_{trans}	>100	>100	≈9	>100	≈5	≈1.2	n.d
Lit state contrast P_{trans}/P_{cis}	≈55	≈24	>100	≈30	≈10	≈7	n.d
Dark state TRACT (ns)	16.5 ± 0.9	9.9 ± 0.3	10.0 ± 0.3	12.0 ± 0.5	25.8 ± 2.3	13.7 ± 0.7	11.6 ± 0.5
Lit state TRACT (ns)	13.8 ± 0.6	11.6 ± 0.4	11.1 ± 0.4	12.4 ± 0.5	27.3 ± 3.0	13.2 ± 0.6	10.9 ± 0.5
cis-pKa	5.9 ± 0.1	5.5 ± 0.1	6.4 ± 0.1	6.2 ± 0.1	>7.0	>7.0	6.6 ± 0.1
P(A-*cis* (%))	96	99	83	95	<40	<30	73

Table 4.2 Results from NMR measurements of the properties of rsFolder2 wild-type(WT) and mutants. All measurements were recorded pH=7.5 and 40°C, unless stated otherwise. Note that the expected τ_c of the monomeric form of rsFolder2 is in the range of 8.6-9.8ns. pKa values were retrieved for the *cis*-CRO from pH titration studies in the range of pH=4.5-8.0, thus, mutants with pKa>7.0 did not titrate enough to determine their exact pKa value.

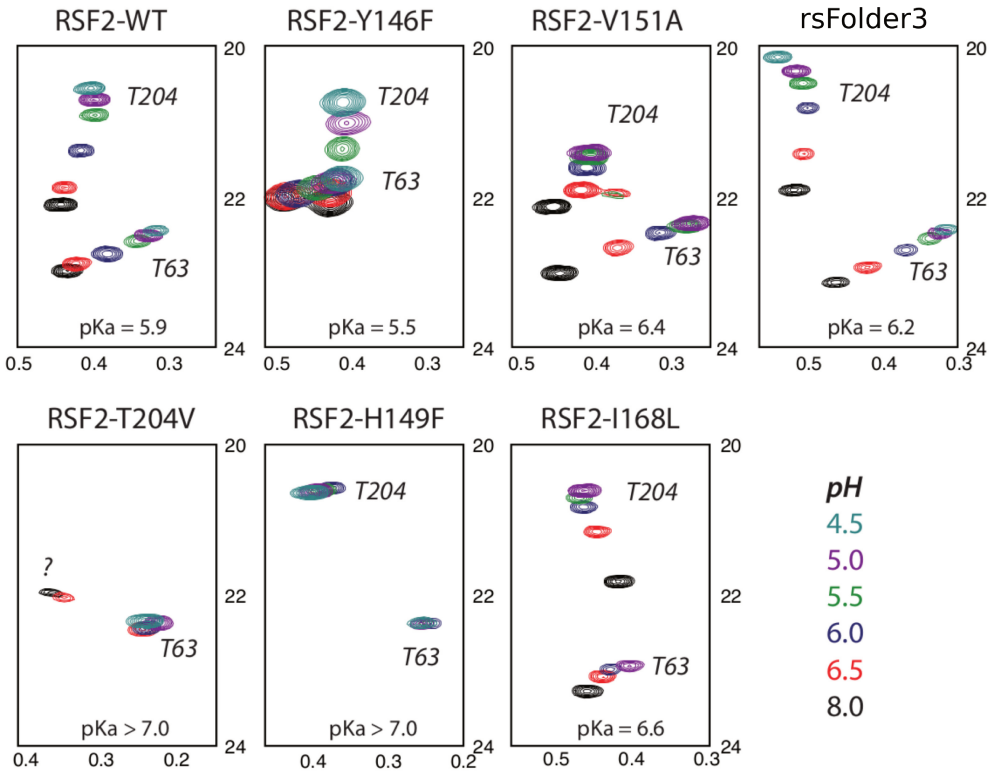


Figure 4.1 Spectral regions of 2D methyl ^{1}H-^{13}C SOFAST-HMQC spectra focusing on T63 and T204 recorded at 850MHz, 40°C for rsFolder2 mutants at different pH values in McIlvaine buffer. CS of T63, T204 and V151 against pH were fitted into a Hasselbalch-Henderson model to retrieve the pKa_{cis} of all mutants.

characterizes the majority of species in the dark-adapted state. The T204V and H149F absorption spectra revealed the existence of a significant population of neutral species in the absence of light. NMR acquired 2D ^{1}H-^{15}N BEST-TROSY spectra can decipher between *cis* and *trans* conformation as described before. When plotting the $\epsilon^{on}_{\lambda_{max}}$ values against A-*cis* population(%) extracted by the 2D ^{1}H-^{15}N BEST-TROSY spectra of the A-*cis* species present in the dark-adapted state (considering the pKa_{cis} value and the existence of *trans* species in the dark-adapted state, at pH=7.5), we could observe a direct correlation. Hence, NMR can potentially approximate the $\epsilon^{on}_{\lambda_{max}}$ of rsFolder2 mutants (and possibly other FPs), given that the $\epsilon^{on}_{\lambda_{max}}$ of a reference protein and its corresponding population ratio of A-*cis* is known.

Furthermore, NMR-observed switching contrast in the 488nm-light-induced state seems to be in good agreement with the values calculated from the photocycling experiments.

Given our previous work regarding pH-induced phenomena on rsFolder, it was decided not to measure the pKa values of mutants by optical means. The main reasoning was that potential pH-induced isomerization of the CRO can result in unreliable measurements of the "real" pKa of the *cis*-CRO. For most mutants, the pKa_{cis} could be retrieved, while mutants

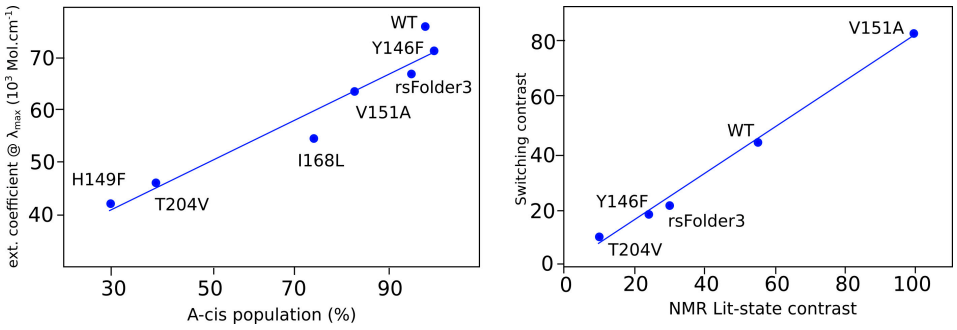


Figure 4.2 Correlating Extinction coefficients at the wavelength of maximum absorption with NMR-observed A-*cis*-CRO populations on the left, and Switching contrast values estimated from fluorescence photocycles with NMR-observed contrast in the 488nm-light induced state, on the right, for rsFolder2 and mutants. Optically measured switching contrast values for H149F are not available and NMR-derived switching contrast for the I168L is unreliable due to low signal-to-noise ratio in the NMR spectrum of the 488nm-light induced state.

with pKa_{cis}>7.0 need to be studied at higher pH range to extrapolate their pKa_{cis} value.

Ring flip rates, k_{rf}, for the *cis*-CRO could potentially be valuable to assess fluorescence QY, Φ. At pH=7.5, k_{rf} does not seem to correlate with the measured Φ for the WT and mutants Y146F, V151A and S206N. However, k_{rf} is the population-weighted average of the ring flips experienced by the fluorescent A-*cis* species, k_{rf}^{Acis} and the non-fluorescent N-*cis*, k_{rf}^{Ncis}. Given the different pKa_{cis} values of each mutant, we decided to measure k_{rf} at high pH (pH=11.4), where we expect the CRO to be at a primarily A-*cis* conformation. The intrinsic ring flip rate of A-*cis* was thus correlated to Φ for the WT and mutants. Ring flip rates k_{rf} deduced for all proteins except rsFolder seem to correlate with Φ, suggesting that A-*cis*-CRO rigidity and corresponding fluorescence efficiency can be evaluated by NMR.

Interestingly, NMR can deduce the oligomeric state of proteins in solution at a given temperature. More specifically, TRACT experiments can give an estimate of the molecular tumbling time, τ_c, of the rsFolder2 mutants in the dark and lit state. Assuming a τ_c of 8.6-9.8ns for the monomeric form, we can conclude that all mutants except the T204V mutant are monomeric in both states, while T204V is most likely dimeric. Moreover, rsFolder2 seems to have a tendency for dimerization with a τ_c measured being between a monomer and a dimer.

Lastly mutants with higher pKa_{cis} possibly undergo similar pH-induced isomerization in the absence of light, as the one observed in rsFolder previously (section 3.3). While 2D methyl ^{1}H-^{13}C SOFAST-HMQC spectra do not show an appearance of two sets of peaks (i.e *cis* and *trans* peaks), 2D amide ^{1}H-^{15}N BEST-TROSY spectra measured at pH=7.5 for mutants of pKa>6.2 include peaks of the “on”- and “off”-state in the absence of light, resulting in decreased apparent dark state contrast (mutants V151A, T204V, H149F, I168L). Typically, 2D methyl spectra are not reliable in quantifying *cis* and *trans* populations since

most methyls that undergo CS change upon *cis-trans* isomerization, also experience change in dynamics in the presence of the *trans*-CRO. As explained previously, NMR-observed switching contrast and thermal relaxation times are more accurately measured by extracting peak intensities from 2D amide spectra. Hence, most likely, the lack of *trans* peaks in the 2D methyl spectra of rsFolder2 and mutants point towards a *trans*-CRO which is undergoing intermediate exchange dynamics, similarly to rsFolder "off"-state at pH=7.5 in HEPES buffer. In order to investigate pH-induced isomerization for rsFolder2 and mutants, we will therefore, in the future, record additional 2D amide ^{1}H-^{15}N BEST-TROSY spectra at different pH, in the absence of light.

4.1.5 Interpretation of NMR and optical studies on rsFolder mutants

Interestingly, the mutations chosen resulted in a versatile pallet of RSFPs. The variability in optical properties enabled us to draw certain conclusions on the potential role that NMR can play as a technique of "screening" for enhanced RSFPs. All mutations were chosen to be in key positions in the vicinity of the CRO, already identified as experiencing changes in dynamics and/or protonation state according to the CRO species present. The effect of certain mutations could not have been predicted (I168L, rsFolder3) while others, were expected to induce certain photophysical behaviour (T204V, H149F).

The correlations between the Dark and Lit-state contrast as measured by NMR with the $\epsilon^{on}_{\lambda_{max}}$ and switching contrast respectively, is further proof that fluorescence originates from the anionic *cis* form of the CRO in all mutants. Moreover, the small differences in absorption wavelength for all A-*cis* species, shows that the absorption properties of the A-*cis*-CRO are only slightly influenced by the residues surrounding it.

H149 and T204 have been identified in both rsFolder and rsEGFP2 to be important for maintaining an H-bond network around the A-*cis*-CRO, alongside a water molecule. Moreover, our NMR findings suggest H149 to be the key residue determining the type of CRO-species present at a given pH. Therefore, replacing it with a hydrophobic phenylalanine, is expected to hinder the formation of (at least) one H-bond with the phenolate of the A-*cis*-CRO. Similarly, the replacement of a T204 with a Valine should hinder the A-*cis*-CRO bond with the threonine. Consequently, both of these mutants are expected to have 2 H-bonds stabilizing the A-*cis*-CRO.

Initially, the H149F and T204V mutants were chosen since their fluorescence capabilities were expected to be worsened as compared to the WT. The rationale behind their design was that by eliminating one of the H-bonds, the A-*cis*-CRO becomes less constrained and thus more likely to undergo radiationless de-excitation. The acquired results in terms of increased Φ as well as the low k_{rf} suggest that the A-*cis*-CRO in both of these mutants is further stabilised, contradicting the expected outcome. The increase in A-*cis*-CRO stability is further supported by the significantly slow "on"-to-"off" switching of these two mutants. As suggested by a review on RSFPs [3], increased Φ tends to correlate to slower switching, due to the increased constraint of the A-*cis*-CRO. This constraining of the CRO is probably a

sign that 2 H-bonds promote more stabilization to its phenolate than 3 H-bonds.

Many studies try to determine the number of H-bonds that a phenolate [31, 166] or, more specifically, a tyrosine's phenolate [192] can have. rsFolder and rsEGFP2 are unique in the fact that their A-*cis*-CRO maintains 3 H-bonds of the phenolate to the β-barrel and a water molecule [61] instead of 2 (eg Dronpa [230] or even 1 (eg Padron [181]) – asFP499 [155], Dronpa M159T [114] mutant and IrisFP [5] maintain 3 H-bonds, 2 of which are with water molecules. In the case where 2 H-bonds provide better anchoring of the A-*cis*-CRO than 3, the existence of 3 candidate H-bond partners in the WT protein potentially caused the phenolate's oxygen to alternate between bonding with any 2 out of the 3 existing H-bond partners. In this scenario, the CRO's rigidity is affected since it constantly breaks and recreates H-bonds with any 2 bonding partners. Thus, the A-*cis*-CRO becomes more stabilized for the H149F and T204V mutants as compared to the WT, possibly since it no longer alternates bonding partners, but rather is preferentially bonding to two specific moieties (T204 or H149 and a water molecule for the H149F and T204V mutants, respectively). X-ray crystal structures for these mutants where the H-bonding partners can be resolved would be instructive in confirming this hypothesis.

On the other hand, in both H149F and T204V mutants, the pKa_{cis} seems to be much higher than in the WT. The increased pKa_{cis} for these mutants is either due to a difference in H-bonding partners, or a result of the CRO behaving like a free CRO in solution, with pKa closer to 8.2. Increased Φ and reduced k_{rf} would suggest that the former explanation is most likely – the proton wire network for these mutants has significantly altered, resulting in different pKa_{cis}.

Moreover, the absorption spectra of the H149F and T204V mutant have a pronounced peak at around 470nm and 490nm, as well as at 405nm. Therefore, at pH=7.5 these mutants can be found in possibly 3 configurations, one neutral, which as shown by our NMR data, corresponds partly to the *trans*-CRO, and two anionic forms. By NMR, we could detect the C_δ peaks of an anionic *cis*-CRO which allowed us to quantify its ring flip rate at pH=11.4. Thus, our NMR signals are either the population-weighted average of the two anionic forms in fast exchange, with each experiencing similar electronic environment (small CS difference between them), or the one anionic form is diminished at pH=11.4. In the future it would be instructive to measure the absorption spectra of H149F and T204V at higher pH to test for the latter hypothesis.

The V151A mutation in rsFolder2 is expected to overcome steric hindrance of the CRO on its way to isomerization. In XFEL studies of rsEGFP2 [47] the CRO passes near V151, thus, replacing the Valine with a shorter residue, namely Alanine, could potentially speed up its switching kinetics. Optical studies suggest that rsFolder2 V151A is switching from "on" to "off" slightly faster than rsFolder2, while "off"-to-"on" switching is not significantly affected. From our NMR study, we would expect that since thermal relaxation which is related to the energy barrier required for "off"-to-"on" switching in the ground-state, is significantly increased for the V151A mutant (500mins) as compared to the WT (110mins), "off"-to-"on" switching quantum yield of rsFolder2 V151A is lower than that of rsFolder2. Most likely, the discrepancies between optical and NMR data suggest that excited- and

ground-state "off"-to-"on" kinetics are not necessarily correlated.

For rsFolder3 and rsFolder2 I168L, it is difficult to interpret the apparent differences in optical properties. The structural re-arrangement that could take place in these mutants leading to differences in switching rates, Φ and switching contrast are not evident, since the residues mutated have not been reported to play a role in the photophysical properties of rsFolder2.

Altogether, while still preliminary, this work shows the potential advantages and limitations of NMR in providing mechanistic interpretations and predictions of the photophysical properties of RSFPs. It further supports the model suggested in section 3.3, according to which, the photophysical properties of RSFPs are influenced by the dynamics of the hydrogen-bond network and proton wire of the barrel, while the fluorescent CRO species maintains its intrinsic properties (absorption, isomerization and protonation state). In order for this work to be completed, it would be instructive to have X-ray crystal structures available that could allow for more structurally-relevant interpretation of NMR-observables.

4.2 Preliminary NMR and optical studies of rsFolder in different sample conditions relevant to microscopy

Towards the end of my PhD, I attempted to study rsFolder in sample conditions relevant to microscopy. For many super-resolution microscopy applications, synthetic tags are still used whose properties can be manipulated with the addition of specific buffers [97, 214]. Particularly in dSTORM microscopy, fluorescent tags, such as Alexa Fluor 647, are used to label the sample, which upon addition of the so-called "STORM buffer" experiences fluorescence quenching. The stochastic "on" switching of the tag once the buffer has been consumed is used to achieve nanometer scale localisation of the tag. Notably, there is an interest in combining fluorescent tags with fluorescent proteins in labelling samples for super-resolution microscopy [63]. However, fluorescent proteins seem to be affected by the buffer conditions, with many of them photoswitching to the "off"-state irreversibly [63].

In order to investigate the effect of super-resolution buffers on the behaviour of rsFolder, I introduced β-mercaptoethylamine (MEA), a reducing agent used in typical STORM buffers [97, 214], in rsFolder samples and attempted to study them by optical means and NMR. The fluorescence recorded in the microscope setup was completely diminished for the sample with 100mM MEA (the chosen concentration for NMR studies), therefore, smaller quantities were used (20mM). The characteristic fluorescence photocycles recorded for the rsFolder sample with and without MEA show some interesting differences (Figure 4.3). Plotting the normalised fluorescence intensity against time, we could observe decreased rates for rsFolder "on"-to-"off" switching, while the "off"-to-"on" switching is slightly faster. Moreover, the photobleaching induced is significant in the MEA sample and the switching contrast appears to have increased.

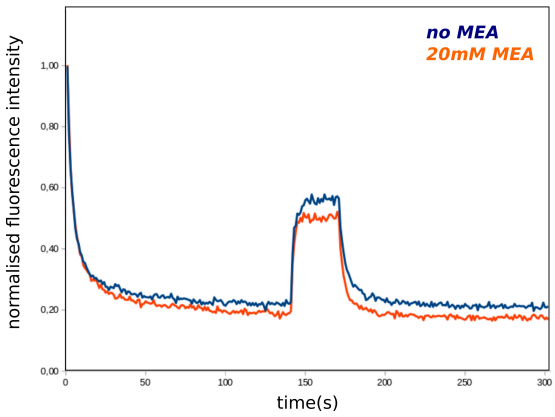


Figure 4.3 Plotting the normalised fluorescence intensity of the photocycles recorded in the microscope for rsFolder with and without MEA. The "on"-to-"off" photoswitching appears slowed down upon addition of 20mM of MEA, while the remaining fluorescence after long 488nm illumination has decreased as compared to the sample without MEA. The switching "off"-to-"on" appears slightly faster in the sample with MEA.

The characteristic "on"-state 2D amide spectrum changes significantly upon addition of MEA. Hence, there is potentially some atomistic information that NMR can provide regarding the effect of MEA on the rsFolder structure. Nevertheless, as already noted previously, the 2D amide spectra are very sensitive to changes in the CRO conformation. Therefore, I run some 2D methyl spectra focusing on Threonines, to assess whether the main effect of MEA is on the CRO and/or the β-barrel and α-helices around the CRO. Threonines shift only slightly with the addition of MEA, suggesting that the CS differences observed in the amide region probably originate from a re-configuration of the CRO, upon addition of MEA (Figure 4.4). Note that the spectrum of rsFolder with 100mM MEA did not overlap with any of the spectra recorded for the "on"-state of rsFolder at different pH values, thus, the effect of MEA is not only the result of a difference in pH of the sample.

Overall, this preliminary study shows that NMR can be used to investigate the mechanistic effect of buffers relevant to microscopy on RSFPs's properties. For this work to continue, further NMR experiments should be included to study the aromatic region, protein dynamics and NMR-observed thermal relaxation and photobleaching rates.

4.3 Future perspectives for studying rsFolder to complement and/or prove NMR conclusions

NMR investigations allowed for the formulation of certain models that could explain the properties of rsFolder in its "on"- and "off"-states, at different temperatures, pH values,

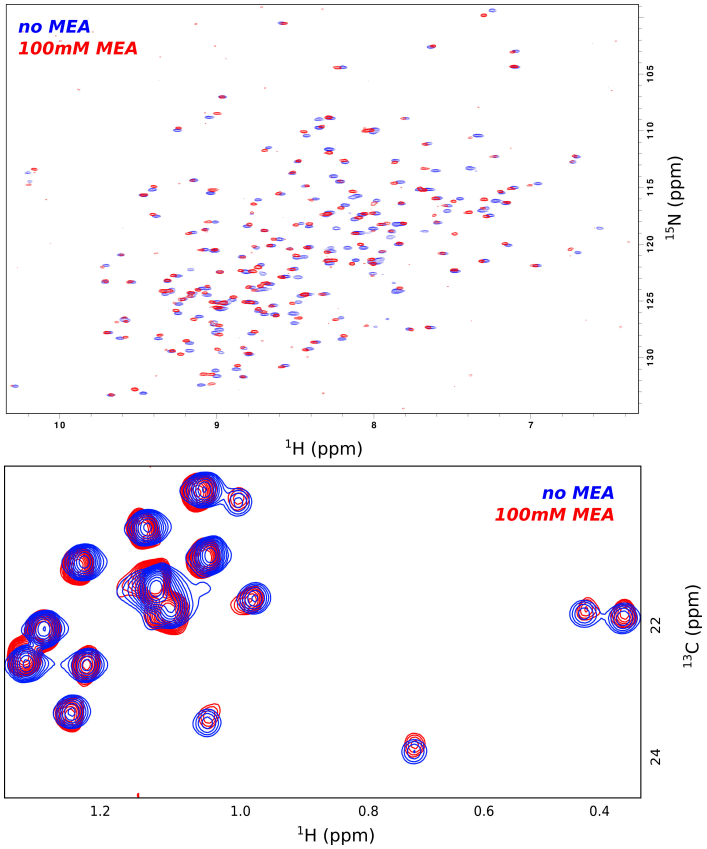


Figure 4.4 On the top, 2D amide ^{1}H-^{15}N BEST-TROSY spectra and on the bottom, 2D methyl ^{1}H-^{13}C SOFAST-HMQC spectra recorded at 40°C at 700MHz, of rsFolder "on"-state, without MEA and with 100mM MEA. Significant CS are observed across the 2D amide spectrum, while only slight differences are seen in the methyl region, indicating that MEA has an effect primarily of the CRO, rather than the protein scaffold.

buffer compositions and illumination schemes. Nevertheless, in order to further solidify or complement the models suggested by our NMR findings, it would be interesting to combine this work with results from X-ray crystallography and/or MDs (and potentially other techiques). Below, I suggest potential points that, while raised by our NMR data, could be further investigated using other techniques.

Histidine 149 role in stabilising the CRO in the "on"- and "off"-states

So far, NMR results could point towards the role of H149 in stabilising the rsFolder CRO in its "on'- and "off"-states at different pH values and temperatures. We were able to investigate the protonation state of H149 and its dynamics and this allowed us to create models concerning the existence or not of an H-bond with the CRO. Nevertheless, NMR

cannot directly conclude on the existence of H-bonds. Hence, it would be interesting to investigate H-bond formation using other techniques.

One such technique could be X-ray crystallography performed at room temperature. While cryogenic temperatures tend to result in better resolved crystal structures, room temperature measurements might allow for the observation of rsFolder conformations at physiologically more relevant conditions. Thus, the proof of existence of an H-bond between the CRO and H149 in the two states at higher temperatures would complement the models suggested by our NMR studies.

In case the data is not of sufficient quality, H-bonds can also be studied in cryogenic temperature structures of rsFolder at different pH values. By changing the pH, the protonation of residues and, thus, their ability to form an H-bond should alter. Consequently, this should result in differences in the conformation of residues, even at cryogenic temperatures.

Another technique that could be instructive to support the formation of a H149-CRO H-bond is MD simulations. Such simulations could assess the lifetime and, therefore, strength of this H-bond at physiologically more relevant conditions than crystallography.

Investigating the origin and effect of dynamics experienced by the CRO to neighbouring residues

The first part of our work suggests that the "off"-state of rsFolder is experiencing dynamics in the few hundreds of μs-ms, particularly in the region of and including the *trans*-CRO, at pH=7.5. After pH titration studies, we could conclude that there is dynamics present in both the "on"- and "off"-states. More specifically, such dynamics seem to be related to a pH-dependent interconversion between the neutral and anionic forms in the case of the *cis*-CRO in the "on"-state and (most likely) the two tautomers of H149 in the case of the *trans*-CRO in the "off"-state. From our data we can conclude that, when the *cis*-CRO is in its anionic state, it can form an H-bond with the protonated $N_{\delta 1}$ of the "on"-state barrel which significantly restricts its motions. Upon protonation of its phenol ring, its dynamics are increased, most likely due to steric hindrance with the protonated H149 $N_{\delta 1}$ and the loss of H-bonds. Similarly, depending on the tautomeric state of H149 in the "off"-state, the protonated *trans*-CRO potentially forms an H-bond or experiences steric hindrance with H149, which could result in an increase of the motions it experiences. Hence, while we can investigate dynamics present, we cannot clearly decipher the origin of such dynamics, in terms of which residues induce it and whether dynamics include motions and/or protonation events.

For our models to be further supported, it would be useful to have X-ray crystal structures of the protonated *cis*-CRO and its relevant configuration in relation to the β-barrel, as well as crystal structures of the "off"-state at pH=5. Moreover, MD simulations could aid in deconvoluting between motions experienced by the CRO and/or the residues in its vicinity.

"Proton wires" responsible for the (de)protonation of the *cis*-CRO

The work on "proton wire" networks primarily includes MD simulations. In the case of rsFolder, simulating the protein at different pH conditions could be useful in further understanding the residues responsible for shuttling protons out of the *cis*-CRO, allowing for such low pKa_{cis}. Our work already suggests the importance of H149 in stabilizing the A-*cis*-CRO but it cannot conclude on its role as a proton donor/acceptor of the CRO. Investigating such "proton wires" could be instructive in understanding the change in pKa_{cis} of rsFolder mutants.

Chapter Five

Conclusion

In conclusion, during my PhD thesis I was able to establish an in-situ illumination set-up that could be used to illuminate my samples inside the NMR spectrometer and, consequently, study rsFolder in its "on"- and "off"-states under different environmental conditions. Choosing the most appropriate NMR experiments to run, I was able to examine rsFolder under physiologically relevant conditions. I could decipher differences in the dynamics experienced by the CRO and the protein scaffold under different illumination schemes, pH values and buffer composition.

The results of my thesis indicate the presence of sub-ms dynamics in the "off"-state of rsFolder under physiologically relevant conditions. In contrast, the "on"-state is characterized by a relatively rigid protein scaffold which potentially enables the A-*cis*-CRO to minimize non-radiative de-excitation pathways upon the absorption of a cyan photon. Using NMR, residues that are affected by CRO dynamics have been identified, which could influence the photophysical properties of rsFolder. Moreover, the residues in the vicinity of the CRO (mainly H149) seem to regulate the presence of different CRO species (neutral or anionic, *cis* or *trans*), depending on the pH, buffer composition and temperature. The NMR-observed ground-state free energy landscape characterising *cis-trans* isomerization at different pH values can partially explain the observed changes in photophysics of rsFolder at acidic as compared to basic environments.

Overall, my PhD work points towards the importance of using NMR to characterize RSFPs. While still preliminary, the NMR study of rsFolder mutants can potentially show how NMR can be used to understand the differences observed in photophysical properties of RSFPs. The ability of NMR to study proteins in solution can also enable scientists to investigate the effect of different buffers on the mechanistic properties of RSFPs and how those can modulate their photophysical behaviour.

REFERENCES

1. Abbandonato, G., Signore, G., Nifosì, R., Voliani, V., Bizzarri, R. & Beltram, F. Cis-trans photoisomerization properties of GFP chromophore analogs. *European Biophysics Journal* **40,** 1205–1214. ISSN: 01757571 (2011).

2. Abbruzzetti, S., Grandi, E., Viappiani, C., Bologna, S., Campanini, B., Raboni, S., Bettati, S. & Mozzarelli, A. Kinetics of acid-induced spectral changes in the GFP-mut2 chromophore. *Journal of the American Chemical Society* **127,** 626–635. ISSN: 00027863 (2005).

3. Acharya, A., Bogdanov, A. M., Grigorenko, B. L., Bravaya, K. B., Nemukhin, A. V., Lukyanov, K. A. & Krylov, A. I. Photoinduced chemistry in fluorescent proteins: Curse or blessing? *Chemical Reviews* **117,** 758–795. ISSN: 15206890 (2017).

4. Adam, V., Berardozzi, R., Byrdin, M. & Bourgeois, D. Phototransformable fluorescent proteins: Future challenges. *Current Opinion in Chemical Biology* **20,** 92–102. ISSN: 18790402 (2014).

5. Adam, V., Lelimousin, M., Boehme, S., Desfonds, G., Nienhaus, K., Field, M. J., Wiedenmann, J., McSweeney, S., Nienhaus, G. U. & Bourgeois, D. Structural characterization of IrisFP, an optical highlighter undergoing multiple photo-induced transformations. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 18343–18348. ISSN: 00278424 (2008).

6. Adam, V., Mizuno, H., Grichine, A., ichi Hotta, J., Yamagata, Y., Moeyaert, B., Nienhaus, G. U., Miyawaki, A., Bourgeois, D. & Hofkens, J. Data storage based on photochromic and photoconvertible fluorescent proteins. *Journal of Biotechnology* **149,** 289–298. ISSN: 01681656 (2010).

7. Adam, V., Nienhaus, K., Bourgeois, D. & Nienhaus, G. U. Structural basis of enhanced photoconversion yield in green fluorescent protein-like protein Dendra2. *Biochemistry* **48,** 4905–4915. ISSN: 00062960 (2009).

8. Agmon, N. Proton pathways in green fluorescence protein. *Biophysical Journal* **88,** 2452–2461. ISSN: 00063495 (2005).

9. Ando, R., Flors, C., Mizuno, H., Hofkens, J. & Miyawaki, A. Highlighted generation of fluorescence signals using simultaneous two-color irradiation on Dronpa mutants. *Biophysical Journal* **92,** L97–L99. ISSN: 00063495 (2007).

10. Ando, R., Hama, H., Yamamoto-Hino, M., Mizuno, H. & Miyawaki, A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 12651–12656. ISSN: 00278424 (2002).

11. Ando, R., Mizuno, H. & Miyawaki, A. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* **306,** 1370–1373. ISSN: 00368075 (2004).

12. Andresen, M., Stiel, A. C., Trowitzsch, S., Weber, G., Eggeling, C., Wahl, M. C., Hell, S. W. & Jakobs, S. Structural basis for reversible photoswitching in Dronpa. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 13005–13009. ISSN: 00278424 (2007).

13. Andresen, M., Wahl, M. C., Stiel, A. C., Gräter, F., Schäfer, L. V., Trowitzsch, S., Weber, G., Eggeling, C., Grubmüller, H., Hell, S. W. & Jakobs, S. Structure and mechanism of the reversible photoswitch of a fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 13070–13074. ISSN: 00278424 (2005).

14. Anthis, N. J. & Clore, G. M. Visualizing transient dark states by NMR spectroscopy. *Quarterly Reviews of Biophysics* **48,** 35–116 (2015).

15. Arai, Y., Takauchi, H., Ogami, Y., Fujiwara, S., Nakano, M., Matsuda, T. & Nagai, T. Spontaneously Blinking Fluorescent Protein for Simple Single Laser Super-Resolution Live Cell Imaging. *ACS Chemical Biology* **13,** 1938–1943. ISSN: 15548937 (2018).

16. Arpino, J. A., Rizkallah, P. J. & Jones, D. D. Crystal Structure of Enhanced Green Fluorescent Protein to 1.35 Å Resolution Reveals Alternative Conformations for Glu222. *PLoS ONE* **7.** ISSN: 19326203. doi:10.1371/journal.pone.0047132 (2012).

17. Barondeau, D. P., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Understanding GFP posttranslational chemistry: Structures of designed variants that achieve backbone fragmentation, hydrolysis, and decarboxylation. *Journal of the American Chemical Society* **128,** 4685–4693. ISSN: 00027863 (2006).

18. Barondeau, D. P., Putnam, C. D., Kassmann, C. J., Tainer, J. A. & Getzoff, E. D. Mechanism and energetics of green fluorescent protein chromophore synthesis revealed by trapped intermediate structures. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 12111–12116. ISSN: 00278424 (2003).

19. Bartkiewicz, M., Kazazić, S., Krasowska, J., Clark, P. L., Wielgus-Kutrowska, B. & Bzowska, A. Non–fluorescent mutant of green fluorescent protein sheds light on

the mechanism of chromophore formation. *FEBS Letters* **592,** 1516–1523. ISSN: 18733468 (2018).

20. Barut, A. O. & Salamin, Y. I. The einstein a-coefficient of spontaneous emission: a relativistic calculation in the Heisenberg representation. *Zeitschrift für Physik D Atoms, Molecules and Clusters* **21,** 1–5. ISSN: 01787683 (1991).

21. Bax, A., Sparks, S. W. & Torchia, D. A. Detection of Insensitive Nuclei. *Methods in Enzymology* **176,** 134–150 (1989).

22. Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J. & Hess, H. F. Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313,** 1642–1645. ISSN: 00368075 (2006).

23. Bodenhausen, G. & Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chemical Physics Letters* **69,** 185–189. ISSN: 00092614 (1980).

24. Bogusky, M. J., Leighton, P., Schiksnis, R. A., Khoury, A., Lu, P. & Opella, S. J. Nitrogen-15 NMR spectroscopy of proteins in solution. *Journal of Magnetic Resonance (1969)* **86,** 11–29. ISSN: 00222364 (1990).

25. Bourgeois, D. Deciphering structural photophysics of fluorescent proteins by kinetic crystallography. *International Journal of Molecular Sciences* **18.** ISSN: 14220067. doi:10.3390/ijms18061187 (2017).

26. Bourgeois, D. & Adam, V. Reversible photoswitching in fluorescent proteins: A mechanistic view. *IUBMB Life* **64,** 482–491. ISSN: 15216543 (2012).

27. Brakemann, T., Stiel, A. C., Weber, G., Andresen, M., Testa, I., Grotjohann, T., Leutenegger, M., Plessmann, U., Urlaub, H., Eggeling, C., Wahl, M. C., Hell, S. W. & Jakobs, S. A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching. *Nature Biotechnology* **29,** 942–950. ISSN: 10870156 (2011).

28. Brakemann, T., Weber, G., Andresen, M., Groenhof, G., Stiel, A. C., Trowitzsch, S., Eggeling, C., Grubmüller, H., Hell, S. W., Wahl, M. C. & Jakobs, S. Molecular basis of the light-driven switching of the photochromic fluorescent protein padron. *Journal of Biological Chemistry* **285,** 14603–14609. ISSN: 00219258 (2010).

29. Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormö, M. & Remington, S. J. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 2306–2311. ISSN: 00278424 (1997).

30. Brutscher, B. & Solyom, Z. *Polarization-enhanced Fast-pulsing Techniques* **11.** ISBN: 9781782628361 (2017).

31. Buytendyk, A. M., Graham, J. D., Collins, K. D., Bowen, K. H., Wu, C. H. & Wu, J. I. The hydrogen bond strength of the phenol-phenolate anionic complex: A computational and photoelectron spectroscopic study. *Physical Chemistry Chemical Physics* **17,** 25109–25113. ISSN: 14639076 (2015).

32. Cavanagh, J., Fairbrother, W. J., Palmer, A. G. I., Rance, M. & Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice* ISBN: 978-0-12-164491-8 (2007).

33. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasherf, D. C. Green Fluorescent Protein as a Marker for Gene Expression. **263** (1994).

34. Chang, J., Romei, M. G. & Boxer, S. G. Structural Evidence of Photoisomerization Pathways in Fluorescent Proteins. *Journal of the American Chemical Society* **141,** 15504–15508. ISSN: 15205126 (2019).

35. Chao, F.-A. & Byrd, R. A. Protein dynamics revealed by NMR relaxation methods. *Emerging Topics in Life Sciences* **2,** 93–105. ISSN: 2397-8554 (2018).

36. Chattoraj, M., King, B. A., Bublitz, G. U. & Boxer, S. G. Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 8362–8367. ISSN: 00278424 (1996).

37. Chen, Y., Müller, J. D., Ruan, Q. & Gratton, E. Molecular brightness characterization of EGFP in vivo by fluorescence fluctuation spectroscopy. *Biophysical Journal* **82,** 133–144. ISSN: 00063495 (2002).

38. Chiragwandi, Z. G., Gillespie, K., Zhao, Q. X., Willander, M. & Panas, I. Ultraviolet driven negative current and rectifier effect in self-assembled green fluorescent protein device. *Applied Physics Letters* **89,** 2004–2007. ISSN: 00036951 (2006).

39. Chirgwandi, Z. G., Panas, I., Johansson, L. G., Nordén, B., Willander, M., Winkler, D. & Ågren, H. Properties of a biophotovoltaic nanodevice. *Journal of Physical Chemistry C* **112,** 18717–18721. ISSN: 19327447 (2008).

40. Chmyrov, A., Keller, J., Grotjohann, T., Ratz, M., D'Este, E., Jakobs, S., Eggeling, C. & Hell, S. W. Nanoscopy with more than 100,000 'doughnuts'. *Nature Methods* **10,** 737–740. ISSN: 15487091 (2013).

41. Chmyrov, A., Leutenegger, M., Grotjohann, T., Schönle, A., Keller-Findeisen, J., Kastrup, L., Jakobs, S., Donnert, G., Sahl, S. J. & Hell, S. W. Achromatic light patterning and improved image reconstruction for parallelized RESOLFT nanoscopy. *Scientific Reports* **7,** 1–9. ISSN: 20452322 (2017).

42. Christou, N. E., Ayala, I., Giandoreggio-Barranco, K., Byrdin, M., Adam, V., Bourgeois, D. & Brutscher, B. NMR Reveals Light-Induced Changes in the Dynamics of a Photoswitchable Fluorescent Protein. *Biophysical Journal* **117,** 2087–2100. ISSN: 15420086 (2019).

43. Christou, N. E. & Brutscher, B. BEST and SOFAST experiments for resonance assignment of histidine and tyrosine side chains in 13C/15N labeled proteins. *Journal of Biomolecular NMR* **72,** 115–124. ISSN: 15735001 (2018).

44. Chudakov, D. M., Matz, M. V., Lukyanov, S. & Lukyanov, K. A. Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. *Physiology Review* **90,** 1103–1163 (2010).

45. Chudakov, D. M., Verkhusha, V. V., Staroverov, D. B., Souslova, E. A., Lukyanov, S. & Lukyanov, K. A. Photoswitchable cyan fluorescent protein for protein tracking. *Nature Biotechnology* **22,** 1435–1439. ISSN: 10870156 (2004).

46. Colletier, J. P., Sliwa, M., Gallat, F. X., Sugahara, M., Guillon, V., Schirò, G., Coquelle, N., Woodhouse, J., Roux, L., Gotthard, G., Royant, A., Uriarte, L. M., Ruckebusch, C., Joti, Y., Byrdin, M., Mizohata, E., Nango, E., Tanaka, T., Tono, K., Yabashi, M., Adam, V., Cammarata, M., Schlichting, I., Bourgeois, D. & Weik, M. Serial Femtosecond Crystallography and Ultrafast Absorption Spectroscopy of the Photoswitchable Fluorescent Protein IrisFP. *Journal of Physical Chemistry Letters* **7,** 882–887. ISSN: 19487185 (2016).

47. Coquelle, N., Sliwa, M., Woodhouse, J., Schirò, G., Adam, V., Aquila, A., Barends, T. R., Boutet, S., Byrdin, M., Carbajo, S., Mora, E. D. L., Doak, R. B., Feliks, M., Fieschi, F., Foucar, L., Guillon, V., Hilpert, M., Hunter, M. S., Jakobs, S., Koglin, J. E., Kovacsova, G., Lane, T. J., Lévy, B., Liang, M., Nass, K., Ridard, J., Robinson, J. S., Roome, C. M., Ruckebusch, C., Seaberg, M., Thepaut, M., Cammarata, M., Demachy, I., Field, M., Shoeman, R. L., Bourgeois, D., Colletier, J. P., Schlichting, I. & Weik, M. Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nature Chemistry* **10,** 31–37. ISSN: 17554349 (2018).

48. Craggs, T. D. Green fluorescent protein: Structure, folding and chromophore maturation. *Chemical Society Reviews* **38,** 2865–2875. ISSN: 03060012 (2009).

49. Croce, A. C. & Bottiroli, G. Autofluorescence spectroscopy and imaging: A tool for biomedical research and diagnosis. *European Journal of Histochemistry* **58,** 320–337. ISSN: 20388306 (2014).

50. Croke, R. L., Patil, S. M., Quevreaux, J., Kendall, D. A. & Alexandrescu, A. T. NMR determination of pKa values in α-synuclein. *Protein Science* **20,** 256–269. ISSN: 09618368 (2011).

51. Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, L. A. & Tsien, R. Y. Understanding, improving and using green fluorescent proteins. *Trends in Biochemical Sciences* **20,** 448–455. ISSN: 09680004 (1995).

52. Cui, G., Lan, Z. & Thiel, W. Intramolecular hydrogen bonding plays a crucial role in the photophysics and photochemistry of the GFP chromophore. *Journal of the American Chemical Society* **134,** 1662–1672. ISSN: 00027863 (2012).

53. De Zitter, E., Ridard, J., Thédié, D., Adam, V., Lévy, B., Byrdin, M., Gotthard, G., Van Meervelt, L., Dedecker, P., Demachy, I. & Bourgeois, D. Mechanistic Investigations of Green mEos4b Reveal a Dynamic Long-Lived Dark State. *Journal of the American Chemical Society* **142,** 10978–10988. ISSN: 15205126 (2020).

54. De Zitter, E., Thédié, D., Mönkemöller, V., Hugelier, S., Beaudouin, J., Adam, V., Byrdin, M., Van Meervelt, L., Dedecker, P. & Bourgeois, D. Mechanistic investigation of mEos4b reveals a strategy to reduce track interruptions in sptPALM. *Nature Methods* **16,** 707–710. ISSN: 15487105 (Aug. 2019).

55. Dertingera, Colyera, R., Iyera, G., Weissa, S. & Enderlein, J. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *PNAS* **106,** 22287–92 (2009).

56. Dixit, R. & Cyr, R. Cell damage and reactive oxygen species production induced by fluorescence microscopy: Effect on mitosis and guidelines for non-invasive fluorescence microscopy. *Plant Journal* **36,** 280–290. ISSN: 09607412 (2003).

57. Duan, C., Adam, V., Byrdin, M., Ridard, J., Kieffer-Jaquinod, S., Morlot, C., Arcizet, D., Demachy, I. & Bourgeois, D. Structural evidence for a two-regime photobleaching mechanism in a reversibly switchable fluorescent protein. *Journal of the American Chemical Society* **135,** 15841–15850. ISSN: 15205126 (2013).

58. Duan, C., Byrdin, M., El Khatib, M., Henry, X., Adam, V. & Bourgeois, D. Rational design of enhanced photoresistance in a photoswitchable fluorescent protein. *Methods and Applications in Fluorescence* **3,** 014004. ISSN: 2050-6120 (2015).

59. Duwé, S., De Zitter, E., Gielen, V., Moeyaert, B., Vandenberg, W., Grotjohann, T., Clays, K., Jakobs, S., Van Meervelt, L. & Dedecker, P. Expression-Enhanced Fluorescent Proteins Based on Enhanced Green Fluorescent Protein for Super-resolution Microscopy. *ACS Nano* **9,** 9528–9541. ISSN: 1936086X (2015).

60. Eggeling, C., Willig, K. I., Sahl, S. J. & Hell, S. W. Lens-based fluorescence nanoscopy. *Quarterly Reviews of Biophysics* **48,** 178–243. ISSN: 14698994 (2015).

61. El Khatib, M., Martins, A., Bourgeois, D., Colletier, J.-P. & Adam, V. Rational design of ultrastable and reversibly photoswitchable fluorescent proteins for super-resolution imaging of the bacterial periplasm. *Scientific Reports* **6,** 18459. ISSN: 2045-2322 (2016).

62. Eletsky, A., Atreya, H. S., Liu, G. & Szyperski, T. Probing structure and functional dynamics of (large) proteins with aromatic rings: L-GFT-TROSY (4,3)D HCCH NMR spectroscopy. *Journal of the American Chemical Society* **127,** 14578–14579. ISSN: 00027863 (2005).

63. Endesfelder, U., Malkusch, S., Flottmann, B., Mondry, J., Liguzinski, P., Verveer, P. J. & Heilemann, M. Chemically induced photoswitching of fluorescent probes-A general concept for super-resolution microscopy. *Molecules* **16,** 3106–3118. ISSN: 14203049 (2011).

64. Ernst, R. R. Nuclear Magnetic Resonance Fourier Transform Spectroscopy (Nobel Lecture). *Angewandte Chemie (International ed. in English)* **31,** 805–930 (1992).

65. Eyal, E., Gerzon, S., Potapov, V., Edelman, M. & Sobolev, V. The limit of accuracy of protein modeling: Influence of crystal packing on protein structure. *Journal of Molecular Biology* **351,** 431–442. ISSN: 00222836 (2005).

66. Farber, P. J. & Mittermaier, A. Relaxation dispersion NMR spectroscopy for the study of protein allostery. *Biophysical Reviews* **7,** 191–200. ISSN: 18672469 (2015).

67. Farjon, J., Boisbouvier, J., Schanda, P., Pardi, A., Simorre, J.-P. & Brutscher, B. Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. *Journal of American Chemical Society* **131,** 8571–8577 (2009).

68. Faustino, A. F., Barbosa, G. M., Silva, M., Castanho, M. A., Da Poian, A. T., Cabrita, E. J., Santos, N. C., Almeida, F. C. & Martins, I. C. Fast NMR method to probe solvent accessibility and disordered regions in proteins. *Scientific Reports* **9,** 1–13. ISSN: 20452322 (2019).

69. Flynn, P. F., Mattiello, D. L., Hill, H. D. & Wand, A. J. Optimal use of cryogenic probe technology in NMR studies of proteins. *Journal of the American Chemical Society* **122,** 4823–4824. ISSN: 00027863 (2000).

70. Fölling, J., Bossi, M., Bock, H., Medda, R., Wurm, C. A., Hein, B., Jakobs, S., Eggeling, C. & Hell, S. W. Fluorescence nanoscopy by ground-state depletion and single-molecule return. *Nature Methods* **5,** 943–945. ISSN: 15487091 (2008).

71. Fron, E., Flors, C., Schweitzer, G., Habuchi, S., Mizuno, H., Ando, R., De Schryver, F. C., Miyawaki, A. & Hofkens, J. Ultrafast excited-state dynamics of the photoswitchable protein dronpa. *Journal of the American Chemical Society* **129,** 4870–4871. ISSN: 00027863 (2007).

72. Fuchs, J., Böhme, S., Oswald, F., Hedde, P. N., Krause, M., Wiedenmann, J. & Nienhaus, G. U. A photoactivatable marker protein for pulse-chase imaging with superresolution. *Nature Methods* **7,** 627–630. ISSN: 15487091 (2010).

73. Fushman, D. in *Protein NMR Techniques* 485–511 (2011). doi:10.1007/978-1-61779-480-3_24.

74. Galbraith, C. G. & Galbraith, J. A. Super-resolution microscopy at a glance. *Journal of Cell Science* **124,** 1607–1611. ISSN: 00219533 (2011).

75. Gautier, A., Juillerat, A., Heinis, C., Corrêa, I. R., Kindermann, M., Beaufils, F. & Johnsson, K. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chemistry and Biology* **15,** 128–136. ISSN: 10745521 (2008).

76. Gautier, A. & Tebo, A. G. Fluorogenic Protein-Based Strategies for Detection, Actuation, and Sensing. *BioEssays* **40,** 1–10. ISSN: 15211878 (2018).

77. Gayda, S., Nienhaus, K. & Nienhaus, G. U. Mechanistic insights into reversible photoactivation in proteins of the GFP family. *Biophysical Journal* **103,** 2521–2531. ISSN: 00063495 (2012).

78. Gebregiworgis, T., Marshall, C. B., Nishikawa, T., Radulovich, N., Sandí, M. J., Fang, Z., Rottapel, R., Tsao, M. S. & Ikura, M. Multiplexed Real-Time NMR GTPase Assay for Simultaneous Monitoring of Multiple Guanine Nucleotide Exchange Factor Activities from Human Cancer Cells and Organoids. *Journal of the American Chemical Society* **140,** 4473–4476. ISSN: 15205126 (2018).

79. Gharbia, Y. A., Milton, G. & Katupitiya, J. The effect of optical fiber endface surface roughness on light coupling. *Optical Fabrication, Testing, and Metrology* **5252,** 201. ISSN: 0277786X (2004).

80. Gill, M. L., Byrd, R. A. & Palmer, A. G. Dynamics of GCN4 facilitate DNA interaction: A model-free analysis of an intrinsically disordered region. *Physical Chemistry Chemical Physics* **18,** 5839–5849. ISSN: 14639076 (2016).

81. Gobeil, S. M., Ebert, M. C., Park, J., Gagné, D., Doucet, N., Berghuis, A. M., Pleiss, J. & Pelletier, J. N. The Structural Dynamics of Engineered β-Lactamases Vary Broadly on Three Timescales yet Sustain Native Function. *Scientific Reports* **9,** 1–12. ISSN: 20452322 (2019).

82. Greenfield, N. J. Analysis of the kinetics of folding of proteins and peptides using circular dichroism. *Nature Protocols* **1,** 2891–2899 (2006).

83. Grigorenko, B. L., Krylov, A. I. & Nemukhin, A. V. Molecular Modeling Clarifies the Mechanism of Chromophore Maturation in the Green Fluorescent Protein. *Journal of the American Chemical Society* **139,** 10239–10249. ISSN: 15205126 (2017).

84. Grigorenko, B., Savitsky, A., Topol, I., Burt, S. & Nemukhin, A. trans and cis Chromophore structures in the kindling fluorescent protein asFP595. *Chemical Physics Letters* **424,** 184–188. ISSN: 00092614 (2006).

85. Grimsley, G. R., Scholtz, J. M. & Pace, C. N. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Science* **18,** 247–251. ISSN: 09618368 (2009).

86. Groot, M. L., Van Wilderen, L. J. & Di Donato, M. Time-resolved methods in biophysics. 5. Femtosecond time-resolved and dispersed infrared spectroscopy on proteins. *Photochemical and Photobiological Sciences* **6,** 501–507. ISSN: 1474905X (2007).

87. Grotjohann, T., Testa, I., Leutenegger, M., Bock, H., Urban, N. T., Lavoie-Cardinal, F., Willig, K. I., Eggeling, C., Jakobs, S. & Hell, S. W. Diffraction-unlimited all-optical imaging and writing with a photochromic GFP. *Nature* **478,** 204–208. ISSN: 00280836 (2011).

88. Grotjohann, T., Testa, I., Reuss, M., Brakemann, T., Eggeling, C., Hell, S. W. & Jakobs, S. rsEGFP2 enables fast RESOLFT nanoscopy of living cells. *eLife* **2012,** 1–14. ISSN: 2050084X (2012).

89. Habuchi, S., Ando, R., Dedecker, P., Verheijen, W., Mizuno, H., Miyawaki, A. & Hofkens, J. Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 9511–9516. ISSN: 00278424 (2005).

90. Habuchi, S., Cotlet, M., Gronheid, R., Dirix, G., Michiels, J., Vanderleyden, J., De Schryver, F. C. & Hofkens, J. Single-Molecule Surface Enhanced Resonance Raman Spectroscopy of the Enhanced Green Fluorescent Protein. *Journal of the American Chemical Society* **125.** PMID: 12848545, 8446–8447 (2003).

91. Hajdu, S. I. A Note from History : The Discovery of Blood Cells. **33,** 237–238 (2003).

92. Halle, B. Biomolecular cryocrystallography: Structural changes during flash-cooling. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 4793–4798. ISSN: 00278424 (2004).

93. Harper, S. M., Neil, L. C., Day, I. J., Hore, P. J. & Gardner, K. H. Conformational Changes in a Photosensory LOV Domain Monitored by Time-Resolved NMR Spectroscopy. *Journal of the American Chemical Society* **126,** 3390–3391. ISSN: 00027863 (2004).

94. Hass, M. A., Hansen, D. F., Christensen, H. E., Led, J. J. & Kay, L. E. Characterization of conformational exchange of a histidine side chain: Protonation, rotamerization, and tautomerization of His61 in plastocyanin from Anabaena variabilis. *Journal of the American Chemical Society* **130,** 8460–8470. ISSN: 00027863 (2008).

95. He, X., Bell, A. F. & Tonge, P. J. Ground state isomerization of a model green fluorescent protein chromophore. *FEBS Letters* **549,** 35–38. ISSN: 00145793 (2003).

96. He, X., Bell, A. F. & Tonge, P. J. Isotopic labeling and normal-mode analysis of a model green fluorescent protein chromophore. *Journal of Physical Chemistry B* **106,** 6056–6066. ISSN: 10895647 (2002).

97. Heilemann, M., Van De Linde, S., Schüttpelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., Tinnefeld, P. & Sauer, M. Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angewandte Chemie - International Edition* **47,** 6172–6176. ISSN: 14337851 (2008).

98. Heim, R., Cubitt, A. B. & Tsien, R. Y. Improved green fluorescence. *Nature* **373.** doi:10.1038/373663b0 (1995).

99. Hell, S. Far-field optical nanoscopy. *Science* **316,** 1153–1159 (2007).

100. Hell, S. W., Dyba, M. & Jakobs, S. Concepts for nanoscale resolution in fluorescence microscopy. *Current Opinion in Neurobiology* **14,** 599–609. ISSN: 09594388 (2004).

101. Henriques, R., Griffiths, C., Rego, E. H. & Mhlanga, M. M. PALM and STORM: Unlocking live-cell super-resolution. *Biopolymers* **95,** 322–331. ISSN: 00063525 (2011).

102. Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M. & Kern, D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450,** 913–916. ISSN: 14764687 (2007).

103. Heras, B. & Martin, J. L. Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallographica Section D: Biological Crystallography* **61,** 1173–1180. ISSN: 09074449 (2005).

104. Hess, S. T., Girirajan, T. P. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal* **91,** 4258–4272. ISSN: 00063495 (2006).

105. Hofmann, M., Eggeling, C., Jakobs, S. & Hell, S. W. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 17565–17569. ISSN: 00278424 (2005).

106. Hooft, R. W., Sander, C. & Vriend, G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins: Structure, Function and Genetics* **26,** 363–376. ISSN: 08873585 (1996).

107. Hutchison, C. D. M., Cordon-preciado, V., Morgan, R. M. L., Nakane, T., Ferreira, J., Dorlhiac, G., Sanchez-gonzalez, A., Johnson, A. S., Fitzpatrick, A., Fare, C., Marangos, J. P., Hong, C., Id, Y., Hunter, M. S., Deponte, D. P., Boutet, S., Owada, S., Tanaka, R., Tono, K., Iwata, S. & Thor, J. J. V. X-ray Free Electron Laser Determination of Crystal

Structures of Dark and Light States of a Reversibly Photoswitching Fluorescent Protein at Room Temperature. *Int. J. Mol. Sci.* **18,** 1918. ISSN: 1422-0067 (2017).

108. Hwang, T. L., Mori, S., Shaka, A. J. & Van Zijl, P. C. Application of phase-modulated CLEAN chemical EXchange spectroscopy (CLEANEX-PM) to detect water - Protein proton exchange and intermolecular NOEs. *Journal of the American Chemical Society* **119,** 6203–6204. ISSN: 00027863 (1997).

109. Hwang, T. L., Van Zijl, P. C. & Mori, S. Accurate quantitation of water-amide proton exchange rates using the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *Journal of Biomolecular NMR* **11,** 221–226. ISSN: 09252738 (1998).

110. Ilgen, P., Grotjohann, T., Jans, D. C., Kilisch, M., Hell, S. W. & Jakobs, S. RESOLFT Nanoscopy of Fixed Cells Using a Z-Domain Based Fusion Protein for Labelling. *PLOS ONE* **10,** 1–11 (Sept. 2015).

111. Jensen, M. R., Ruigrok, R. W. & Blackledge, M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Current Opinion in Structural Biology* **23,** 426–435. ISSN: 0959440X (2013).

112. Jeon, S., Kim, J., Lee, D., Baik, J. W. & Kim, C. Review on practical photoacoustic microscopy. *Photoacoustics* **15,** 100141. ISSN: 22135979 (2019).

113. Joosten, R. P., Salzemann, J., Bloch, V., Stockinger, H., Berglund, A. C., Blanchet, C., Bongcam-Rudloff, E., Combet, C., Da Costa, A. L., Deleage, G., Diarena, M., Fabbretti, R., Fettahi, G., Flegel, V., Gisel, A., Kasam, V., Kervinen, T., Korpelainen, E., Mattila, K., Pagni, M., Reichstadt, M., Breton, V., Tickle, I. J. & Vriend, G. PDB-REDO: Automated re-refinement of X-ray structure models in the PDB. *Journal of Applied Crystallography* **42,** 376–384. ISSN: 16005767 (2009).

114. Kaucikas, M., Fitzpatrick, A., Bryan, E., Struve, A., Henning, R., Kosheleva, I., Srajer, V., Groenhof, G. & Van Thor, J. J. Room temperature crystal structure of the fast switching M159T mutant of the fluorescent protein dronpa. *Proteins: Structure, Function and Bioinformatics* **83,** 397–402. ISSN: 10970134 (2015).

115. Kay, L. E., Torchia, D. A. & Bax, A. Backbone Dynamics of Proteins As Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **28,** 8972–8979. ISSN: 15204995 (1989).

116. Keedy, D. A., Kenner, L. R., Warkentin, M., Woldeyes, R. A., Hopkins, J. B., Thompson, M. C., Brewster, A. S., Van Benschoten, A. H., Baxter, E. L., Uervirojnangkoorn, M., McPhillips, S. E., Song, J., Alonso-Mori, R., Holton, J. M., Weis, W. I., Brunger, A. T., Soltis, S. M., Lemke, H., Gonzalez, A., Sauter, N. K., Cohen, A. E., van den Bedem, H., Thorne, R. E. & Fraser, J. S. Mapping the conformational landscape of a dynamic enzyme by multitemperature and XFEL crystallography. *eLife* **4,** 1–26. ISSN: 2050-084X (2015).

117. Keedy, D. A., Van Den Bedem, H., Sivak, D. A., Petsko, G. A., Ringe, D., Wilson, M. A. & Fraser, J. S. Crystal cryocooling distorts conformational heterogeneity in a model michaelis complex of DHFR. *Structure* **22,** 899–910. ISSN: 18784186 (2014).

118. Keeler, J. *Understanding NMR spectroscopy* ISBN: 978-0470746080 (Wiley, 2005).

119. Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H. & Johnsson, K. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nature Biotechnology* **21,** 86–89. ISSN: 10870156 (2003).

120. Kerfah, R., Plevin, M. J., Pessey, O., Hamelin, O., Gans, P. & Boisbouvier, J. Scrambling free combinatorial labeling of alanine-β, isoleucine-δ1, leucine-proS and valine-proS methyl groups for the detection of long range NOEs. *Journal of Biomolecular NMR* **61,** 73–82. ISSN: 15735001 (2015).

121. Kihara, M., Yajima, Y. & Watanabe, H. Analyzing Deterioration in Optical Performance of Fiber Connections with Refractive Index Matching Material Using Incorrectly Cleaved Fiber Ends. *IEICE Transactions on Communications* **E96.B,** 2206–2212 (Sept. 2013).

122. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1814,** 942–968. ISSN: 15709639 (2011).

123. Kojima, S., Ohkawa, H., Hirano, T., Maki, S., Niwa, H., Ohashi, M., Inouye, S. & Tsuji, F. I. Fluorescent properties of model chromophores of tyrosine-66 substituted mutants of Aequorea green fluorescent protein (GFP). *Tetrahedron Letters* **39,** 5239–5242. ISSN: 00404039 (1998).

124. Kovacs, H., Moskau, D. & Spraul, M. Cryogenically cooled probes—a leap in NMR technology. *Progress in Nuclear Magnetic Resonance Spectroscopy* **46,** 131–155 (2005).

125. Krishna Rao, D. & Bhuyan, A. K. Complexity of aromatic ring-flip motions in proteins: Y97 ring dynamics in cytochrome c observed by cross-relaxation suppressed exchange NMR spectroscopy. *Journal of Biomolecular NMR* **39,** 187–196. ISSN: 09252738 (2007).

126. Krishnan, V. & Rupp, B. Macromolecular Structure Determination: Comparison of X-ray Crystallography and NMR Spectroscopy. *eLS.* doi:10.1002/9780470015902.a0002716.pub2 (2012).

127. Kühn, T. & Schwalbe, H. Monitoring the kinetics of ion-dependent protein folding by time-resolved NMR spectroscopy at atomic resolution. *Journal of the American Chemical Society* **122,** 6169–6174. ISSN: 00027863 (2000).

128. Kuprov, I. & Hore, P. J. Uniform illumination of optically dense NMR samples. *Journal of Magnetic Resonance* **171,** 171–175. ISSN: 10907807 (2004).

129. Lelimousin, M., Noirclerc-Savoye, M., Lazareno-Saez, C., Paetzold, B., Le Vot, S., Chazal, R., Macheboeuf, P., Field, M. J., Bourgeois, D. & Royant, A. Intrinsic dynamics in ECFP and cerulean control fluorescence quantum yield. *Biochemistry* **48,** 10038–10046. ISSN: 00062960 (2009).

130. Li, Y., Forbrich, A., Wu, J., Shao, P., Campbell, R. E. & Zemp, R. Engineering Dark Chromoprotein Reporters for Photoacoustic Microscopy and FRET Imaging. *Scientific Reports* **6,** 1–12. ISSN: 20452322 (2016).

131. Lindskog, S. & Coleman, J. E. The catalytic mechanism of carbonic anhydrase. *Proceedings of the National Academy of Sciences of the United States of America* **70,** 2505–2508. ISSN: 00278424 (1973).

132. Lipari, G. & Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *Journal of the American Chemical Society* **104,** 4546–4559. ISSN: 15205126 (1982).

133. Liu, X., Wong, T. T. W., Shi, J., Ma, J., Yang, Q. & Wang, L. V. Label-free cell nuclear imaging by Grüneisen relaxation photoacoustic microscopy. *Optics Letters* **43,** 947. ISSN: 0146-9592 (2018).

134. Longa, D., Bouvigniesa, G. & Kaya, L. E. Measuring hydrogen exchange rates in invisible protein excited states. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 8820–8825. ISSN: 10916490 (2014).

135. Los, G. V., Encell, L. P., McDougall, M. G., Hartzell, D. D., Karassina, N., Zimprich, C., Wood, M. G., Learish, R., Ohana, R. F., Urh, M., Simpson, D., Mendez, J., Zimmerman, K., Otto, P., Vidugiris, G., Zhu, J., Darzins, A., Klaubert, D. H., Bulleit, R. F. & Wood, K. V. HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chemical Biology* **3,** 373–382. ISSN: 15548929 (2008).

136. Lukacs, A., Haigney, A., Brust, R., Addison, K., Towrie, M., Greetham, G. M., Jones, G. A., Miyawaki, A., Tonge, P. J. & Meech, S. R. Protein photochromism observed by ultrafast vibrational spectroscopy. *Journal of Physical Chemistry B* **117,** 11954–11959. ISSN: 15206106 (2013).

137. Lukyanov, K. A., Chudakov, D. M., Lukyanov, S. & Verkhusha, V. V. Photoactivatable fluorescent proteins. **6,** 885–891 (2005).

138. Ma, Y., Sun, Q. & Smith, S. C. The mechanism of oxidation in chromophore maturation of wild-type green fluorescent protein: A theoretical study. *Physical Chemistry Chemical Physics* **19,** 12942–12952. ISSN: 14639076 (2017).

139. Maddalo, S. L. & Zimmer, M. The Role of the Protein Matrix in Green Fluorescent Protein Fluorescence. *Photochemistry and Photobiology* **82,** 367. ISSN: 0031-8655 (2006).

140. McGlinchey, M. J. Symmetry breaking in NMR spectroscopy: The elucidation of hidden molecular rearrangement processes. *Symmetry* **6,** 622–654. ISSN: 20738994 (2014).

141. McIlvaine, T. C. A buffer solution for colorimetric comparison. *Journal of Biological Chemistry* **49,** 183–186. ISSN: 1098-6596 (1921).

142. Mei, Z., Treado, J. D., Grigas, A. T., Levine, Z. A., Regan, L. & O'Hern, C. S. Analyses of protein cores reveal fundamental differences between solution and crystal structures, 1–8. ISSN: 1097-0134 (2019).

143. Melckebeke, V., Simorre, J.-p. & Brutscher, B. Amino Acid-Type Edited NMR Experiments for Methyl - Methyl Distance Measurement in 13 C-Labeled Proteins powerful method for structure determination of proteins . The, 9584–9591 (2004).

144. Miyawaki, A., Shcherbakova, D. M. & Verkhusha, V. V. Red fluorescent proteins: Chromophore formation and cellular applications. *Current Opinion in Structural Biology* **22,** 679–688. ISSN: 0959440X (2012).

145. Mizuno, H., Mal, T. K., Tong, K. I., Ando, R., Furuta, T., Ikura, M. & Miyawaki, A. Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein. *Molecular Cell* **12,** 1051–1058. ISSN: 10972765 (2003).

146. Mizuno, H., Mal, T. K., Wälchli, M., Fukano, T., Ikura, M. & Miyawaki, A. Molecular basis of photochromism of a fluorescent protein revealed by direct 13C detection under laser illumination. *Journal of Biomolecular NMR* **48,** 237–246. ISSN: 09252738 (2010).

147. Mizuno, H., Mal, T. K., Wälchli, M., Kikuchi, A., Fukano, T., Ando, R., Jeyakanthan, J., Taka, J., Shiro, Y., Ikura, M. & Miyawaki, A. Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 9227–9232. ISSN: 00278424 (2008).

148. Mondal, P. P. Multiple excitation nano-spot generation and confocal detection for far-field microscopy. *Nanoscale* **2,** 381–384. ISSN: 20403364 (2010).

149. Montelione, G. T., Lyons, B. A., Emerson, D. S. & Tashiro, M. An efficient triple resonance experiment using carbon-13 isotopic mixing for determining sequence-specific resonance assingments of isotopically-enriched proteins. *Journal of American Chemical Society,* 10974–10975 (1992).

150. Morozov, D. & Groenhof, G. Hydrogen Bond Fluctuations Control Photochromism in a Reversibly Photo-Switchable Fluorescent Protein. *Angewandte Chemie - International Edition* **55,** 576–578. ISSN: 15213773 (2016).

151. Morris, G. A. & Freeman, R. Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer. *Journal of the American Chemical Society* **101,** 760–762. ISSN: 15205126 (1979).

152. Muñoz, V. & Cerminara, M. When fast is better: Protein folding fundamentals and mechanisms from ultrafast approaches. *Biochemical Journal* **473,** 2545–2559. ISSN: 14708728 (2016).

153. Nienhaus, K. & Nienhaus, G. U. Chromophore photophysics and dynamics in fluorescent proteins of the GFP family. *Journal of Physics Condensed Matter* **28,** 443001. ISSN: 1361648X (2016).

154. Nienhaus, K. & Nienhaus, G. U. Fluorescent proteins for live-cell imaging with super-resolution. *Chemical Society Reviews* **43,** 1088–1106 (2014).

155. Nienhaus, K., Renzi, F., Vallone, B., Wiedenmann, J. & Nienhaus, G. U. Chromophore-protein interactions in the anthozoan green fluorescent protein asFP499. *Biophysical Journal* **91,** 4210–4220. ISSN: 00063495 (2006).

156. Nitschke, P., Lokesh, N. & Gschwind, R. M. Combination of illumination and high resolution NMR spectroscopy: Key features and practical aspects, photochemical applications, and new concepts. *Progress in Nuclear Magnetic Resonance Spectroscopy* **114-115,** 86–134. ISSN: 00796565 (2019).

157. Olsen, S., Lamothe, K. & Martínez, T. J. Protonic gating of excited-state twisting and charge localization in GFP chromophores: A mechanistic hypothesis for reversible photoswitching. *Journal of the American Chemical Society* **132,** 1192–1193. ISSN: 15205126 (2010).

158. Oltrogge, L. M., Wang, Q. & Boxer, S. G. Ground-State Proton Transfer Kinetics in Green Fluorescent Protein. *Biochemistry* **53,** 5947–5957. ISSN: 15204995 (2014).

159. Oregioni, A., Stieglitz, B., Kelly, G., Rittinger, K. & Frenkiel, T. Determination of the pKa of the N-terminal amino group of ubiquitin by NMR. *Scientific Reports* **7,** 1–8. ISSN: 20452322 (2017).

160. Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remingtont, S. J. Crystal Structure of the Aequorea victoria Green Fluorescent Protein. **273,** 392–396 (1996).

161. Pace, N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M., Takano, K., Hebert, E. J., Shirley, B. A. & Grimsley, G. R. Contribution of hydrogen bonds to protein stability. *Protein Science* **23.** doi:10.1002/pro. (2014).

162. Pakhomov, A. A. & Martynov, V. I. GFP Family: Structural Insights into Spectral Tuning. *Chemistry and Biology* **15,** 755–764. ISSN: 10745521 (2008).

163. Palmer, A. G. Chemical exchange in biomacromolecules: Past, present, and future. *Journal of Magnetic Resonance* **241,** 3–17 (2014).

164. Palmer, A. G. NMR characterization of the dynamics of biomacromolecules. *Chemical Reviews* **104,** 3623–3640. ISSN: 00092665 (2004).

165. Palmer, A. G., Kroenke, C. & Loria, J. P. Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules. *Methods in Enzymology* **339.** ISSN: 00406376 (2001).

166. Parthasarath, R., Subramanian, V. & Sathyamurthy, N. Hydrogen bonding in phenol, water, and phenol-water clusters. *Journal of Physical Chemistry A* **109,** 843–850. ISSN: 10895639 (2005).

167. Patterson, G. H., Knobel, S. M., Sharif, W. D., Kain, S. R. & Piston, D. W. Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *Biophysical Journal* **73,** 2782–2790. ISSN: 00063495 (1997).

168. Patterson, G. H. & Lippincott-Schwartz, J. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science* **297,** 1873–1877. ISSN: 00368075 (2002).

169. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* **24,** 79–88. ISSN: 10870156 (2006).

170. Pelton, J. G., Torchia, D. A., Meadow, N. D. & Roseman, S. Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signal-transducing protein from escherichia coli, using two-dimensional heteronuclear NMR techniques. *Protein Science* **2,** 543–558. ISSN: 1469896X (1993).

171. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 12366–12371. ISSN: 00278424 (1997).

172. Petzold, D., Nitschke, P., Brandl, F., Scheidler, V., Dick, B., Gschwind, R. M. & König, B. Visible-Light-Mediated Liberation and In Situ Conversion of Fluorophosgene. *Chemistry - A European Journal* **25,** 361–366. ISSN: 15213765 (2019).

173. Piprek, J. in *Semiconductor Optoelectronic Devices* 83–120 (2003). doi:10.1016/b978-0-08-046978-2.50029-6.

174. Platzer, G., Okon, M. & McIntosh, L. P. PH-dependent random coil 1H, 13C, and 15N chemical shifts of the ionizable amino acids: A guide for protein pK a measurements. *Journal of Biomolecular NMR* **60,** 109–129. ISSN: 15735001 (2014).

175. Pouwels, P. J. & Kaptein, R. Diagonal-Peak Suppression in the SCOTCH Experiment, a 2D NMR Technique for Studying Photochemical Reactions. *Journal of Magnetic Resonance* **101,** 337–341 (1993).

176. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. Primary structure of the Aequorea victoria green-fluorescent protein. **111,** 229–233 (1992).

177. Querard, J., Gautier, A., Le Saux, T. & Jullien, L. Expanding discriminative dimensions for analysis and imaging. *Chemical Science* **6,** 2968–2978. ISSN: 20416539 (2015).

178. Quérard, J., Zhang, R., Kelemen, Z., Plamont, M. A., Xie, X., Chouket, R., Roemgens, I., Korepina, Y., Albright, S., Ipendey, E., Volovitch, M., Sladitschek, H. L., Neveu, P., Gissot, L., Gautier, A., Faure, J. D., Croquette, V., Le Saux, T. & Jullien, L. Resonant out-of-phase fluorescence microscopy and remote imaging overcome spectral limitations. *Nature Communications* **8,** 1–8. ISSN: 20411723 (2017).

179. Quillin, M. L., Anstrom, D. M., Shu, X., O'Leary, S., Kallio, K., Chudakov, D. M. & Remington, S. J. Kindling fluorescent protein from Anemonia sulcata: Dark-state structure at 1.38 Å resolution. *Biochemistry* **44,** 5774–5787. ISSN: 00062960 (2005).

180. Rasia, R. M., Brutscher, B. & Plevin, M. J. Selective Isotopic Unlabeling of Proteins Using Metabolic Precursors: Application to NMR Assignment of Intrinsically Disordered Proteins. *ChemBioChem* **13,** 732–739. ISSN: 14394227 (2012).

181. Regis Faro, A., Carpentier, P., Jonasson, G., Pompidor, G., Arcizet, D., Demachy, I. & Bourgeois, D. Low-temperature chromophore isomerization reveals the photoswitching mechanism of the fluorescent protein Padron. *Journal of the American Chemical Society* **133,** 16362–16365. ISSN: 00027863 (2011).

182. Renz, M. Fluorescence Microscopy - A Historical and Technical Perspective, 767–779 (2013).

183. 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. ISSN: 00062960 (2004).

184. Sahl, S. J., Hell, S. W. & Jakobs, S. *Fluorescence nanoscopy in cell biology* Sept. 2017. doi:10.1038/nrm.2017.71.

185. Salvi, N., Abyzov, A. & Blackledge, M. Atomic resolution conformational dynamics of intrinsically disordered proteins from NMR spin relaxation. *Progress in Nuclear Magnetic Resonance Spectroscopy* **102-103,** 43–60. ISSN: 00796565 (2017).

186. Sanderson, M. J., Smith, I., Parker, I. & Bootman, M. D. Fluorescence Microscopy. doi:10.1101/pdb.top071795.Fluorescence (2016).

187. Schanda, P. & Brutscher, B. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society* **127,** 8014–8015. ISSN: 00027863 (2005).

188. Schanda, P., Forge, V. & Brutscher, B. Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 11257–11262. ISSN: 00278424 (2007).

189. Schanda, P., Kupĉe, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional deteronuclear correlation spectra of proteins within a few seconds. *Journal of Biomolecular NMR* **33,** 199–211. ISSN: 09252738 (2005).

190. Schanda, P., Van Melckebeke, H. & Brutscher, B. Speeding up three-dimensional protein NMR experiments to a few minutes. *Journal of the American Chemical Society* **128,** 9042–9043. ISSN: 00027863 (2006).

191. Scharnagl, C. & Raupp-Kossmann, R. A. Solution pKa values of the green fluorescent protein chromophore from hybrid quantum-classical calculations. *Journal of Physical Chemistry B* **108,** 477–489. ISSN: 15206106 (2004).

192. Scheiner, S., Kar, T. & Pattanayak, J. Comparison of various types of hydrogen bonds involving aromatic amino acids. *Journal of the American Chemical Society* **124,** 13257–13264. ISSN: 00027863 (2002).

193. Seifert, M. H. J., Georgescu, J., Ksiazek, D., Smialowski, P., Rehm, T., Steipe, B. & Holak, T. A. Backbone dynamics of green fluorescent protein and the effect of Histidine 148 substitution. *Biochemistry* **42,** 2500–2512. ISSN: 00062960 (2003).

194. Seifert, M. H., Ksiazek, D., Azim, M. K., Smialowski, P., Budisa, N. & Holak, T. A. Slow exchange in the chromophore of a green fluorescent protein variant. *Journal of the American Chemical Society* **124,** 7932–7942. ISSN: 00027863 (2002).

195. Shaka, A. J., Barker, P. B. & Freeman, R. Computer-optimized decoupling scheme for wideband applications and low-level operation. *Journal of Magnetic Resonance (1969)* **64,** 547–552. ISSN: 00222364 (1985).

196. Shen, Y. & Bax, A. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular NMR* **56,** 227–241. ISSN: 09252738 (2013).

197. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *Journal of Biomolecular NMR* **44,** 213–223 (2009).

198. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent. *Journal of cellular and comparative physiology* **59,** 223–239. ISSN: 00959898 (1962).

199. Shinobu, A. & Agmon, N. Proton wire dynamics in the green fluorescent protein. *Journal of Chemical Theory and Computation* **13,** 353–369. ISSN: 15499626 (2017).

200. Shinobu, A. & Agmon, N. The hole in the barrel: Water exchange at the GFP chromophore. *Journal of Physical Chemistry B* **119,** 3464–3478. ISSN: 15205207 (2015).

201. Shinobu, A., Palm, G. J., Schierbeek, A. J. & Agmon, N. Visualizing proton antenna in a high-resolution green fluorescent protein structure. *Journal of the American Chemical Society* **132,** 11093–11102. ISSN: 00027863 (2010).

202. Sikic, K., Tomic, S. & Carugo, O. Systematic Comparison of Crystal and NMR Protein Structures Deposited in the Protein Data Bank. *The Open Biochemistry Journal* **4,** 83–95. ISSN: 1874-091X (2010).

203. Smith, M. J., Marshall, C. B., Theillet, F. X., Binolfi, A., Selenko, P. & Ikura, M. Real-time NMR monitoring of biological activities in complex physiological environments. *Current Opinion in Structural Biology* **32,** 39–47. ISSN: 1879033X (2015).

204. Smyrnova, D., Moeyaert, B., Michielssens, S., Hofkens, J., Dedecker, P. & Ceulemans, A. Molecular Dynamic Indicators of the Photoswitching Properties of Green Fluorescent Proteins. *Journal of Physical Chemistry B* **119,** 12007–12016. ISSN: 15205207 (2015).

205. Solyom, Z., Schwarten, M., Geist, L., Konrat, R., Willbold, D. & Brutscher, B. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *Journal of Biomolecular NMR* **55,** 311–321. ISSN: 09252738 (2013).

206. Su, X., Zheng, Y., Yang, J., Li, Q. & Li, J. Method for end-face treatment of specialty optical fiber with thick-walled silica glass tubes. *Applied Optics* **55,** 8271. ISSN: 0003-6935 (2016).

207. Sudmeier, J., Evelhoch, J. & Jonsson, N.-H. Dependence of NMR lineshape analysis upon chemical rates and mechanisms: Implications for enzyme histidine titrations. *Journal of Magnetic Resonance (1969)* **40,** 377–390. ISSN: 00222364 (1980).

208. Syu, Y. S., Wu, C. Y. & Lee, Y. C. Double-sided freeform lens for light collimation of light emitting diodes. *Applied Sciences (Switzerland)* **9.** ISSN: 20763417. doi:10.3390/app9245452 (2019).

209. Takaba, K., Tai, Y., Eki, H., Dao, H. A., Hanazono, Y., Hasegawa, K., Miki, K. & Takeda, K. Subatomic resolution X-ray structures of green fluorescent protein. *IUCrJ* **6,** 387–400. ISSN: 20522525 (2019).

210. Toniolo, A., Olsen, S., Manohar, L. & Martinez, T. J. Ultrafast excited state dynamics in the green fluorescent protein chromophore. *Femtochemistry and Femtobiology: Ultrafast Events in Molecular Science* **31,** 425–432 (2004).

211. Trainor, K., Palumbo, J. A., MacKenzie, D. W. & Meiering, E. M. Temperature dependence of NMR chemical shifts: Tracking and statistical analysis. *Protein Science* **29,** 306–314. ISSN: 1469896X (2019).

212. Tsai, M. S., Ou, C. L., Tsai, C. J., Huang, Y. C., Cheng, Y. C., Sun, S. S. & Yang, J. S. Fluorescence Enhancement of Unconstrained GFP Chromophore Analogues Based on the Push-Pull Substituent Effect. *Journal of Organic Chemistry* **82,** 8031–8039. ISSN: 15206904 (2017).

213. Tsien, R. Y. The Green Fluorescent Protein (GFP). *Annual Reviews Biochemistry* **67,** 509–544. ISSN: 22289674 (1996).

214. Van De Linde, S., Endesfelder, U., Mukherjee, A., Schüttpelz, M., Wiebusch, G., Wolter, S., Heilemann, M. & Sauer, M. Multicolor photoswitching microscopy for subdiffraction-resolution fluorescence imaging. *Photochemical and Photobiological Sciences* **8,** 465–469. ISSN: 14749092 (2009).

215. Van Thor, J. J., Gensch, T., Hellingwerf, K. J. & Johnson, L. N. Phototransformation of green fluorescent protein with UV and visible light leads to decarboxylation of glutamate 222. *Nature Structural Biology* **9,** 37–41. ISSN: 10728368 (2002).

216. Van Thor, J. J., Pierik, A. J., Nugteren-Roodzant, I., Xie, A. & Hellingwerf, K. J. Characterization of the photoconversion of green fluorescent protein with FTIR spectroscopy. *Biochemistry* **37,** 16915–16921. ISSN: 00062960 (1998).

217. Vetschera, P., Mishra, K., Fuenzalida-Werner, J. P., Chmyrov, A., Ntziachristos, V. & Stiel, A. C. Characterization of Reversibly Switchable Fluorescent Proteins in Optoacoustic Imaging. *Analytical Chemistry* **90,** 10527–10535. ISSN: 15206882 (2018).

218. Voliani, V., Bizzarri, R., Nifosì, R., Abbruzzetti, S., Grandi, E., Viappiani, C. & Beltram, F. Cis-trans photoisomerization of fluorescent-protein chromophores. *Journal of Physical Chemistry B* **112,** 10714–10722. ISSN: 15206106 (2008).

219. Von Stetten, D., Giraud, T., Carpentier, P., Sever, F., Terrien, M., Dobias, F., Juers, D. H., Flot, D., Mueller-Dieckmann, C., Leonard, G. A., De Sanctis, D. & Royant, A. In crystallo optical spectroscopy (icOS) as a complementary tool on the macromolecular crystallography beamlines of the ESRF. *Acta Crystallographica Section D: Biological Crystallography* **71,** 15–26. ISSN: 13990047 (2015).

220. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. & Laue, E. D. The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics* **59,** 687–696. ISSN: 08873585 (2005).

221. Wachter, R. M., Elsliger, M. A., Kallio, K., Hanson, G. T. & Remington, S. J. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure* **6,** 1267–1277. ISSN: 09692126 (1998).

222. Wallace, M., Adams, D. J. & Iggo, J. A. Titrations without the Additions: The Efficient Determination of p K a Values Using NMR Imaging Techniques. *Analytical Chemistry* **90,** 4160–4166. ISSN: 15206882 (2018).

223. Wan, L., Nagata, T., Morishita, R., Takaori-Kondo, A. & Katahira, M. Observation by Real-Time NMR and Interpretation of Length- and Location-Dependent Deamination Activity of APOBEC3B. *ACS Chemical Biology* **12,** 2704–2708. ISSN: 15548937 (2017).

224. Wang, S., Ding, M., Chen, X., Chang, L. & Sun, Y. Development of bimolecular fluorescence complementation using rsEGFP2 for detection and super-resolution imaging of protein-protein interactions in live cells. *Biomedical Optics Express* **8,** 3119. ISSN: 2156-7085 (2017).

225. Ward, W. *In Bioluminescence and Chemiluminescence: Basic Chemistry and Analytical* (New York: Academic Press, 1981).

226. Weik, M., Coquelle, N., Sliwa, M., Woodhouse, J., Schiro, G., Adam, V., Aquila, A., Barends, T., Boutet, S., Byrdin, M., Doak, B., Feliks, M., Fieschi, F., Foucar, L., Guillon, V., Hilpert, M., Hunter, M., Jakobs, S., Koglin, J., Kovacsova, G., Levy, B., Liang, M., Nass, K., Ridard, J., Robinson, J., Roome, C., Ruckebusch, C., Thepaut, M., Cammarata, M., Demachy, I., Field, M., Shoeman, R., Bourgeois, D., Colletier, J.-P., Schlichting, I. & Weik, M. Time-resolved serial femtosecond crystallography on photoswitchable fluorescent proteins. *Acta Crystallographica Section A Foundations and Advances* **72,** s39–s39. ISSN: 2053-2733 (2016).

227. Weininger, U., Brath, U., Modig, K., Teilum, K. & Akke, M. Off-resonance rotating-frame relaxation dispersion experiment for 13C in aromatic side chains using L-optimized TROSY-selection. *Journal of Biomolecular NMR* **59,** 23–29. ISSN: 15735001 (2014).

228. Wiedenmann, J., Ivanchenko, S., Oswald, F., Schmitt, F., Röcker, C., Salih, A., Spindler, K. D. & Nienhaus, G. U. EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 15905–15910. ISSN: 00278424 (2004).

229. Wilmann, P. G., Petersen, J., Devenish, R. J., Prescott, M. & Rossjohn, J. Variations on the GFP chromophore: A polypeptide fragmentation within the chromophore revealed in the 2.1-Å crystal structure of a nonfluorescent chromoprotein from Anemonia sulcata. *Journal of Biological Chemistry* **280,** 2401–2404. ISSN: 00219258 (2005).

230. Wilmann, P. G., Turcic, K., Battad, J. M., Wilce, M. C., Devenish, R. J., Prescott, M. & Rossjohn, J. The 1.7 Å Crystal Structure of Dronpa: A Photoswitchable Green

Fluorescent Protein. *Journal of Molecular Biology* **364,** 213–224. ISSN: 00222836 (2006).

231. Wolff, C., Kind, J., Schenderlein, H., Bartling, H., Feldmeier, C., Gschwind, R. M., Biesalski, M. & Thiele, C. M. Studies of a photochromic model system using NMR with ex-situ and in-situ irradiation devices. *Magnetic Resonance in Chemistry* **54,** 485–491. ISSN: 1097458X (2016).

232. Woodhouse, J., Nass Kovacs, G., Coquelle, N., Uriarte, L. M., Adam, V., Barends, T. R., Byrdin, M., de la Mora, E., Bruce Doak, R., Feliks, M., Field, M., Fieschi, F., Guillon, V., Jakobs, S., Joti, Y., Macheboeuf, P., Motomura, K., Nass, K., Owada, S., Roome, C. M., Ruckebusch, C., Schirò, G., Shoeman, R. L., Thepaut, M., Togashi, T., Tono, K., Yabashi, M., Cammarata, M., Foucar, L., Bourgeois, D., Sliwa, M., Colletier, J. P., Schlichting, I. & Weik, M. Photoswitching mechanism of a fluorescent protein revealed by time-resolved crystallography and transient absorption spectroscopy. *Nature Communications* **11,** 1–11. ISSN: 20411723 (2020).

233. Wüthrich, K. *NMR of proteins and nucleic acids* ISBN: 9781782620525. doi:10.139/9781782622758-00348 (1986).

234. Xu, S., Ni, S. & Kennedy, M. A. NMR Analysis of Amide Hydrogen Exchange Rates in a Pentapeptide-Repeat Protein from A. thaliana. *Biophysical Journal* **112,** 2075–2088. ISSN: 15420086 (2017).

235. Yadav, D., Lacombat, F., Dozova, N., Rappaport, F., Plaza, P. & Espagne, A. Real-time monitoring of chromophore isomerization and deprotonation during the photoactivation of the fluorescent protein Dronpa. *Journal of Physical Chemistry B* **119,** 2404–2414. ISSN: 15205207 (2015).

236. Yao, J., Kaberniuk, A. A., Li, L., Shcherbakova, D. M., Zhang, R., Wang, L., Li, G., Verkhusha, V. V. & Wang, L. V. Multiscale photoacoustic tomography using reversibly switchable bacterial phytochrome as a near-infrared photochromic probe. *Nature Methods* **13,** 67–73. ISSN: 15487105 (2015).

237. Yuwen, T., Bah, A., Brady, J. P., Ferrage, F., Bouvignies, G. & Kay, L. E. Measuring Solvent Hydrogen Exchange Rates by Multifrequency Excitation 15 N CEST: Application to Protein Phase Separation. *Journal of Physical Chemistry B* **122,** 11206–11217. ISSN: 15205207 (2018).

238. Zhang, L., Patel, H. N., Lappe, J. W. & Wachter, R. M. Reaction progress of chromophore biogenesis in green fluorescent protein. *Journal of the American Chemical Society* **128,** 4766–4772. ISSN: 00027863 (2006).

239. Zhang, W., Niu, X., Ding, J., Hu, Y. & Jin, C. Intra- and inter-protein couplings of backbone motions underlie protein thiol-disulfide exchange cascade. *Scientific Reports* **8,** 1–14. ISSN: 20452322 (2018).

240. Zhang, X., Chen, X., Zeng, Z., Zhang, M., Sun, Y., Xi, P., Peng, J. & Xu, P. Development of a reversibly switchable fluorescent protein for super-resolution optical fluctuation imaging (SOFI). *ACS Nano* **9,** 2659–2667. ISSN: 1936086X (2015).

APPENDIX

Appendix A

Protein sequence of rsFolder:

MVSKGEELFT GVVPILVELD GDVNGHKFSV RGEGEGDATN GKLTLKFICT TGKLPVPWPT LVTTLAYGVL CFSRYPDHMK RHDFFKSAMP EGYVQERTIS FKDDGTYKTR AEVKFEGDTL VNRIELKGID FKEDGNILGH KLEYNFNSHN VYITADKQKN GIKSNFKIRH NVEDGSVQLA DHYQQNTPIG DGPVLLPDNH YLSTQSKLSK DPNEKRDHMV LLEFVTAAGI THGMDELYK

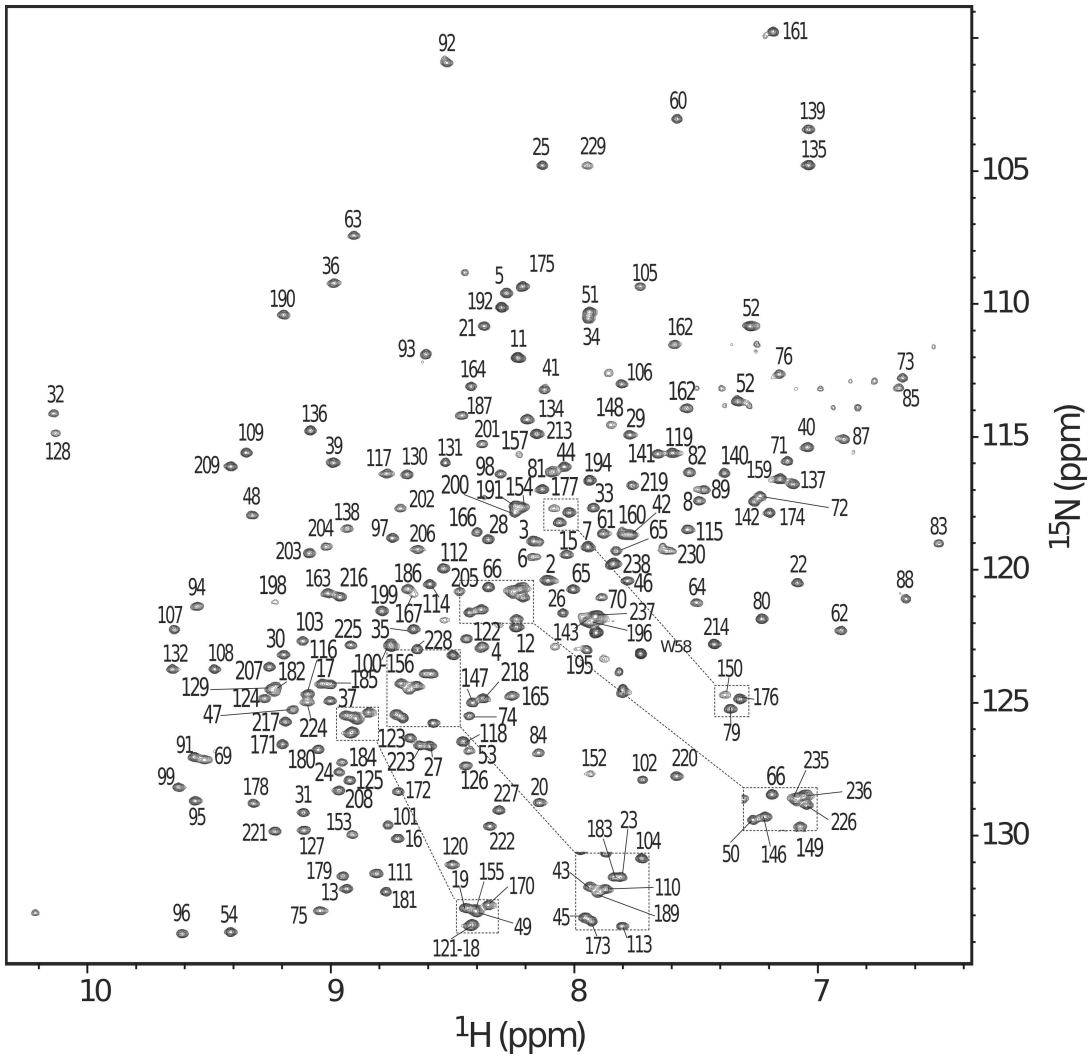


Figure A.1 2D amide ^{1}H-^{15}N BEST-TROSY spectra of rsFolder “on”-state, recorded at 40° at 850MHz. All assigned cross-peaks are annotated with residue number.

Journal of Biomolecular NMR
https://doi.org/10.1007/s10858-018-0216-z

COMMUNICATION

BEST and SOFAST experiments for resonance assignment of histidine and tyrosine side chains in $^{13}C/^{15}N$ labeled proteins

Nina Eleni Christou[1] · **Bernhard Brutscher**[1]

Received: 1 October 2018 / Accepted: 16 November 2018

Abstract
Aromatic amino-acid side chains are essential components for the structure and function of proteins. We present herein a set of NMR experiments for time-efficient resonance assignment of histidine and tyrosine side chains in uniformly $^{13}C/^{15}N$-labeled proteins. The use of band-selective ^{13}C pulses allows to deal with linear chains of coupled spins, thus avoiding signal loss that occurs in branched spin systems during coherence transfer. Furthermore, our pulse schemes make use of longitudinal ^{1}H relaxation enhancement, Ernst-angle excitation, and simultaneous detection of ^{1}H and ^{13}C steady-state polarization to achieve significant signal enhancements.

Keywords BEST · Aromatics · NMR assignment · Longitudinal relaxation enhancement · Histidine tautomerization

Introduction

Aromatic amino-acid side chains are essential components for the structure and function of proteins. Aromatic rings are generally buried within the hydrophobic protein core, representing a significant volume fraction of the protein core, and stabilizing the protein fold via hydrophobic contacts, hydrogen bonding, and π-stacking interactions. If aromatic side chains are on a proteins' surface they are often involved in inter-molecular interactions with other proteins or nucleic acids (Baker and Grant 2007), contributing significantly to the binding energy of such complexes (Bogan and Thorn 1998). They are also frequently found as part of active sites in enzymes contributing to the protein's catalytic activity or stabilizing the substrate. In particular, histidine and tyrosine play prominent roles in enzyme catalysis (Bartlett et al. 2002; Holliday et al. 2009) and metal ion coordination (Yamashita et al. 1990). The chemical properties of histidine play also important roles in ion channels (Hu et al. 2006) and proton transfer proteins (Shimahara et al. 2007). Nuclear spins in the aromatic side chains are essential probes for the protein's structure, dynamics, and functional properties. For NMR studies of these proteins, it is therefore important to have available a set of experiments for time-efficient resonance assignment of aromatic side chains, and in the case of histidine, for the determination of their protonation and tautomeric states. Despite the availability of a number of NMR triple-resonance (H–C–N) pulse experiments that have been proposed in the past for this purpose, mainly in the 1990s (Yamazaki et al. 1993; Löhr and Rüterjans 1996; Prompers et al. 1998), aromatic side chain assignment of uniformly $^{13}C/^{15}N$ labeled proteins remains a challenge, and still mainly relies on ambiguous ^{1}H–^{1}H through-space contacts as revealed by NOESY-type experiments. This observation holds especially true for larger proteins studied at relatively low sample concentration, typically a few hundred micro-molar solutions. Major problems of these heteronuclear correlation experiments are relaxation-induced signal losses during the various coherence transfer steps, as well as large variations in peak intensities as a consequence of differences in local side chain mobility, and eventually strong coupling effects due to almost identical chemical shift values for neighboring ^{13}C spins (Jacob et al. 2002). Here we focus on histidine and tyrosine side chains, and propose a set of sensitivity-improved 2D ^{1}H–^{13}C and ^{1}H–^{15}N pulse sequences that correlate aromatic ring protons with ^{13}C and ^{15}N nuclei in the same aromatic ring, or with the adjacent CB of the aliphatic chain. Each experiment is optimized for a particular side chain, either histidine or tyrosine, by properly adjusting the selectivity of the shaped ^{13}C pulses and transfer delays. All pulse sequences are of the out-and-back type, and

✉ Bernhard Brutscher
bernhard.brutscher@ibs.fr

[1] Univ. Grenoble Alpes, CEA, CNRS, IBS, 38000 Grenoble, France

Published online: 21 November 2018

they all make use of proton longitudinal relaxation enhancement (Pervushin et al. 2002) to improve the experiments' sensitivity resulting in significantly reduced NMR data acquisition times. The performance of these experiments is demonstrated for a 200 µM $^{13}C/^{15}N$-labelled sample of the 250-residue protein rsfolder (El Khatib et al. 2016), a photoswitchable variant of the green fluorescent protein (GFP) studied in our laboratory, that is rich in aromatic residues.

BEST optimization of aromatic $^{1}H–^{13}C$ correlation experiments

Figure 1a shows a constant-time (CT) $^{1}H–^{13}C$ HSQC pulse sequence used for correlating aromatic ^{1}H and ^{13}C via one-bond J-couplings. The pulse sequence is a standard INEPT-based HSQC, except that our pulse scheme uses band-selective shaped ^{1}H pulses for selective manipulation of aromatic (and amide) protons. Such experiments have been termed BEST, an acronym for *Band-selective Excitation Short-Transient* (Schanda et al. 2006; Lescop et al. 2007). BEST experiments achieve enhanced steady-state ^{1}H polarization of the excited spins at short recycle delays, by leaving other proton spins that are in close spatial proximity at thermodynamic equilibrium. This allows for faster repetition rates of the pulse sequence, and thus higher sensitivity and reduced overall experimental time requirements to reach a target signal-to-noise ratio (SNR) (Brutscher and Solyom 2017). Longitudinal relaxation optimized NMR pulse sequences for aromatic moieties have been proposed before for HCCH correlation experiments (Eletsky et al. 2005) and ^{13}C relaxation measurements (Weininger et al. 2012, 2014). Contrary to our BEST-type sequences, in this earlier work band-selective flip-back pulses are used to restore the polarization of non-aromatic ^{1}H spins. The differences between these two approaches are twofold: (i) frequency-selective pulses will excite all ^{1}H (aromatic and amides) resonating within the chosen band width, while flip-back schemes achieve selectivity of aromatic ^{1}H only due to their particular ^{13} C chemical shifts. (ii) In BEST-type experiments, the passive spins, acting as a "thermal spin bath" remain unperturbed during the whole experiment, while in flip-back experiments they follow a complex coherence-transfer trajectory, and their steady-state polarization becomes affected by spin relaxation, pulse imperfections, and spin-coupling evolution. In our hands, BEST-type experiments typically outperform flip-back approaches in uniformly labeled proteins, as long as the majority of protons, and especially the very mobile side chain protons (e.g. methyls) resonate outside the chosen band width.

In the following, we will use the CT $^{1}H–^{13}C$ HSQC experiment of Fig. 1a to demonstrate the improvement achieved by the BEST approach over conventional pulse schemes, based on hard-power ^{1}H pulses, for aromatic spin moieties in the protein rsfolder. The β-barrel structure of rsfolder is shown in Fig. 1b, highlighting the positions of the 33 aromatic side chains present in this protein. Note that also the chromophore that is responsible for the optical properties of this fluorescent protein is made up by a tyrosine side chain as one major component. The aromatic $^{1}H–^{13}C$ spectrum of this protein recorded with the pulse sequence of Fig. 1a is shown in Fig. 1c. Due to the constant-time ^{13}C editing, carbons with an odd number of directly attached carbons give rise to a negative peak intensity, while those with an even number of neighbors are showing up positive. Although the spectrum is well resolved, a particularly unpleasant feature of this spectrum is the extreme heterogeneity of detected peak intensities with the strongest peaks being 2 orders of magnitude more intense than the weakest ones, due to differences in dynamics (ring flips, side chain mobility), as well as the presence of strong $^{13}C–^{13}C$ scalar couplings between adjacent carbons with (almost) identical chemical shifts. As stated above, this presents a major challenge for subsequent NMR correlation experiments required for sequential resonance assignment of these aromatic spin moieties.

In order to estimate the sensitivity gain achieved by the BEST implementation in the pulse sequence of Fig. 1a under moderately fast pulsing conditions, we have recorded the same spectrum using a similar pulse sequence, where the band-selective ^{1}H pulses have been replaced by hard pulses, except for the last 180° pulse that has been applied band-selectively for water suppression purposes. A histogram of the measured intensity ratios (BEST versus conventional hard-pulse version) in these two spectra, recorded in both cases with an inter-scan delay of 300 ms, is shown in Fig. 1d. The intensity gain achieved by the BEST implementation varies from a factor close to 1 up to 2.5 for individual sites, with an average gain of about 75%. As expected, very mobile aromatic side chains, giving rise to the most intense cross peaks, are only little effected by the BEST implementation, resulting in a more uniform intensity distribution in the BEST spectrum, as reported before for amide $^{1}H–^{15}N$ moieties (Favier and Brutscher 2011). These data clearly reflect the significantly reduced longitudinal relaxation times of aromatic ^{1}H in BEST-type experiments. It should be mentioned, however, that the sensitivity improvement will be less important if both experiments are performed with individually optimized inter-scan delays.

An important feature of this experiment is the correct handling of ^{13}C steady state polarization. It is well known that aromatic ring carbons have a large chemical shift anisotropy (CSA) (Veeman 1984), leading to strong relaxation interference of the ^{13}C CSA and the $^{13}C–^{1}H$ dipolar interaction at high magnetic field strengths (Pervushin et al. 1998). As a consequence of the effect of CSA-dipolar cross-correlated relaxation (Brutscher 2000), part of the

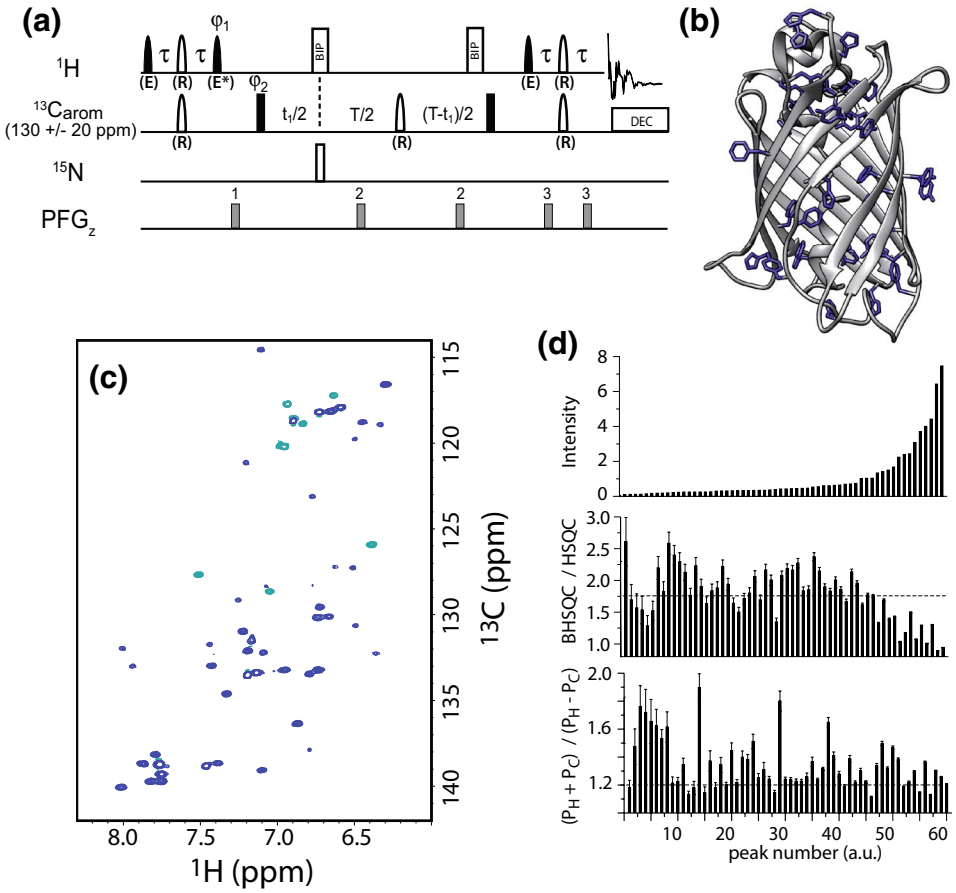
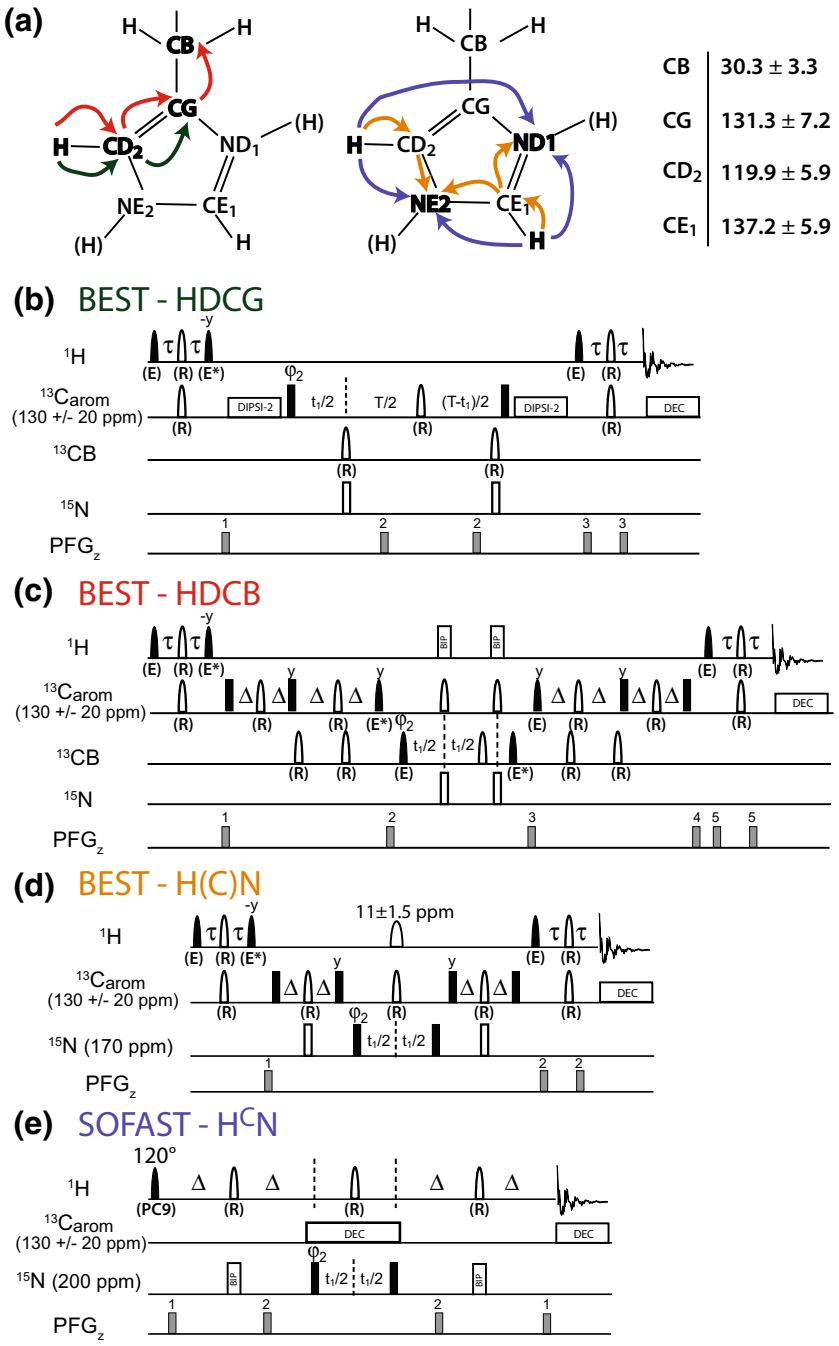


Fig. 1 **a** Pulse sequence of BEST CT ^{1}H–^{13}C HSQC experiment for aromatic protein side chains. Filled and open pulse symbols indicate 90° and 180° rf pulses, respectively. Unless indicated, all pulses are applied with phase x. The following pulse shapes are used for band-selective ^{1}H pulses: E-BURP2 (E), REBURP (R), and time-inversed E-BURP2 (E*) (Geen and Freeman 1991). In addition, a pair of broadband ^{1}H inversion pulses (BIP) (Smith et al. 2001) results in an effective 360° rotation on ^{1}H polarization. Selective ^{1}H pulses are typically centered at 7.8 ppm, covering a bandwidth of 4.0 ppm (aromatic region). Band-selective ^{13}C 180° pulses are applied with a REBURP shape centered at 130 ppm (C_{arom}) and 35 ppm (CB) covering a band width of 40 ppm. ^{15}N pulses are applied at a carrier frequency of 200 ppm. The ^{1}H–^{13}C transfer delay is set to $\tau = 1/(2J_{CH}) - 0.5\delta_1$ with $J_{CH} \approx 180$ Hz and δ_1 the length of the ^{1}H REBURP pulse. The CT delay is set to $T = 1/J_{CC} = 14$ ms. As discussed in the text, for optimal sensitivity φ_1 is set to $\varphi_1 = y$. A 2-step phase cycle is applied with $\varphi_2 = x, -x$ and $\varphi_{rec} = x, -x$. Quadrature detection in t_1 is achieved by time-proportional phase incrementation of φ_2 according to STATES-TPPI. ^{13}C decoupling during detection makes use of a WALTZ65 sequence covering a bandwidth of 35 ppm. Pulsed field gradients are applied along the z-axis (PFGz) with durations of 300 µs to 500 µs and field strengths ranging from 6 to 50 G/cm. The same number indicates identical gradient pulses. **b** Ribbon plot of the structure of rsfolder (PDB entry 5DTZ). Aromatic side chains are highlighted in blue color. All spectra shown in this paper were recorded at 40 °C on a 200 µM sample of rsfolder in a 5 mm Shigemi tube (300 µl) on a Bruker Avance IIIHD 850 MHz spectrometer equipped with a cryogenically cooled HCN probe. **c** ^{1}H–^{13}C correlation spectra recorded with the BEST CT ^{1}H–^{13}C HSQC pulse sequence. The recycle delay was set to 300 ms, and eight scans were recorded per t_1 increment, leading to a total acquisition time of about 30 min. Green and blue contours are representative of negative and positive peak intensities. **d** Comparison of peak intensities extracted from a ^{1}H–^{13}C correlation spectrum recorded with a conventional hard-pulse sequence (HSQC) or with a BEST-type sequence (BHSQC) using a recycle delay of 300 ms in both cases. Intensities extracted from the HSQC spectrum (top), and ordered for increasing peak intensity. Peak intensity ratios of BHSQC over HSQC (center). Peak intensity ratios measured in ^{1}H–^{13}C correlation spectra recorded with the BEST CT ^{1}H–^{13}C HSQC pulse sequence, and setting the phase φ_1 either to $+y$ ($P_H + P_C$) or $-y$ ($P_H - P_C$)

^{13}C steady-state polarization (C_z) becomes transferred into 2-spin order ($2C_zH_z$) during the CT editing period of length T:

$$C_z \overset{90^\circ}{\rightarrow} C_x \overset{T}{\rightarrow} 2C_xH_z \overset{90^\circ}{\rightarrow} 2C_zH_z \quad (1)$$

This 2-spin order is then transferred into detectable ^{1}H single-quantum coherence by the subsequent INEPT step. The detected signal is therefore a combination of two contributions, arising from ^{1}H ($S_H(T)$)and ^{13}C ($S_C(T)$) steady-state polarizations, and given by the following expression (for $t_1 = 0$):



$$S(T) \propto S_H(T) \pm S_C(T) = P_H\gamma_H \cosh\left(\Gamma_{CSA,DD}T\right) \pm P_C\gamma_C \sinh\left(\Gamma_{CSA,DD}T\right) \quad (2)$$

with γ_H and γ_C the gyromagnetic ratios of ^{1}H and ^{13}C, P_H and P_C the respective steady state polarizations, and $\Gamma_{CSA,DD}$ the cross-correlated relaxation rate constant. Whether the two contributions add or subtract depends on the choice of the φ_1 phase value (y or −y) of the final ^{1}H 90° pulse in the initial INEPT transfer step. We can estimate the contribution from ^{13}C steady-state polarization to the detected signal by recording separate data sets with $\varphi_1 = y$ and $\varphi_1 = -y$, respectively. The observed peak intensity ratios (Fig. 1e) vary from 1.1 to 1.6, demonstrating the importance of properly adjusting this pulse phase value ($\varphi_1 = -y$ in the present case), and indicating that up to about 20% of the observed signal originates from ^{13}C steady-state polarization in our experimental conditions. The issue of proper handling ^{13}C (or ^{15}N) polarization in heteronuclear correlation experiments has been well documented for TROSY-type experiments (Pervushin et al. 1998; Brutscher et al. 1998), but it is often

◂**Fig. 2** BEST/SOFAST pulse sequences optimized for histidine side chains. **a** Coherence-transfer pathways (color-coded) exploited in the ^{1}H–^{13}C (left) and ^{1}H–^{15}N (center) correlation experiments. Average and standard deviation of the expected ^{13}C chemical shifts, as extracted from the BMRB data base, are indicated on the right. The symbols and nomenclatures used in these pulse schemes are identical to those introduced in Fig. 1a. Selective ^{1}H pulses are centered at 7.8 ppm, covering a bandwidth of 4.0 ppm. Band-selective ^{13}C 180° pulses are applied with a REBURP shape centered at 130 ppm (C_{arom}) and 35 ppm (CB) covering a band width of 40 ppm. ^{15}N pulses are applied at a carrier frequency of 200 ppm. Note that a phase $\varphi_1 = -y$ has to be used in these sequences to add the contributions originating from ^{1}H and ^{13}C steady-state polarizations. **b** BEST-HDCG pulse sequence. The CD-CG transfer is realized by an isotropic mixing sequence, consisting in two DIPSI-2 (Shaka et al. 1988) cycles of ~6 ms total length, and the CT delay is set to T = 13.3 ms. **c** BEST-HDCB pulse sequence. The transfer delay Δ is set to 3.1 ms. **d** BEST-H(C)N pulse sequence. The transfer delay Δ is set to 13.3 ms. The REBURP ^{1}H pulse applied during t_1 is shifted to 11 ppm, covering a band width of 3 ppm, in order to decouple ^{15}N-bound protons without affecting aromatic protons. As this experiment is only efficient for correlations with proton-bound nitrogen spins, the ^{15}N carrier frequency should be placed at 170 ppm for optimal performance. **e** SOFAST-H^{C}N pulse sequence. The band-selective ^{1}H excitation pulse is applied with a PC9 shape (Kupce and Freeman 1993) and a flip angle of 120°. The transfer delays are set to Δ = 11 ms – $0.5\delta_1$, and Δ' = 11 ms – $0.5\delta_1$– $0.5\delta_2$, with δ_1 and δ_2 the length of the REBURP and PC9 pulses. The ^{15}N 180° pulses are applied with a BIP shape and a $\gamma B_1/2\pi$ field of about 7 kHz

neglected for HSQC-type experiments, where its effect on peak intensities may become similarly important.

The observations and conclusions made here for the case of a simple CT ^{1}H–^{13}C correlation experiment remain valid for all further experiments discussed in this paper, although the actual enhancement numbers will vary from one pulse sequence to the other. All pulse sequences presented in this paper have been included in the NMRlib pulse sequence library for Bruker spectrometers that can be freely downloaded from the IBS web site at http://www.ibs.fr/research/scientific-output/software/pulse-sequence-tools.

Histidine experiments

A set of four sensitivity-enhanced two-dimensional (2D) heteronuclear correlation experiments, specifically tailored and optimized for the assignment of histidine side chain nuclei (^{1}H, ^{13}C, ^{15}N) is displayed in Fig. 2. The first two pulse sequences (Fig. 2b, c) correlate the HD_2 ring proton with either the quaternary CG ring carbon (BEST-HDCG), or the aliphatic CB (BEST-HDCB). Both experiments exploit large one-bond J-couplings for coherence transfer, with the transfer delays adjusted to about 1/(2J). In BEST-HDCG, ^{13}C editing is achieved in a CT manner to avoid line splitting due to the large one-bond CG–CD couplings, while in BEST-HDCB no CT ^{13}C editing is used for sensitivity reasons. This leaves the aliphatic CB–CA coupling active during t_1, leading to line broadening (or line splitting) in the ^{13}C dimension. The two latter pulse sequences (Fig. 2d, e) were designed to correlate the carbon-bound ring protons (HD_2 and HE_1) with the two nitrogens (ND_1, NE_2) of the imidazole ring. The BEST-H(C)N experiment performs a relayed transfer via one bond couplings from H to C to N (Sudmeier et al. 1996; Schmidt et al. 2008), while the SOFAST-H^{C}N uses small two-bond (and three-bond) $^{2}J_{HN}$ ($^{3}J_{HN}$) couplings to transfer directly from H to N (Pelton et al. 1993). The SOFAST pulse scheme (Schanda and Brutscher 2005; Schanda et al. 2005) allows further enhancement of the ^{1}H steady-state polarization by adjusting the nutation angle of the excitation pulse to the Ernst angle. More experimental details about these experiments are provided in the figure caption.

In Fig. 3 are plotted 2D correlation spectra recorded with these pulse sequences on a sample of 200 µM $^{13}C/^{15}N$ labeled rsfolder at 850 MHz ^{1}H frequency. The four data sets were collected overnight with individual data acquisition times of 1 h (Best-HDCG and BEST-H(C)N), 3 h (Best-HDCB), and 10 h (SOFAST-H^{C}N). All expected 11 histidines are observed in these spectra, although with very different peak intensities. Interestingly, the two ^{1}H–^{15}N correlation spectra (Fig. 3b) provide very complementary information. While BEST-H(C)N is almost only giving rise to correlation peaks with protonated nitrogens resonating at higher field (^{15}N frequency range 155–185 ppm), the long-range SOFAST-H^{C}N experiment is more sensitive for correlations with deprotonated nitrogens (^{15}N frequency range 220–255 ppm). The latter observation is due to significantly stronger $^{1}J_{CN}$ coupling constants for proton-bound nitrogens. The differences observed in the SOFAST-H^{C}N spectrum are not due to the off-resonance effects of the ^{15}N pulses, but we rather believe that it is somehow related to chemical exchange of the ^{15}N-bound ^{1}H. Fortunately, taking both spectra together, identification of the expected peak patterns allows straightforward and unambiguous assignment of the 11 imidazole spin systems, correlating the two carbon-bound protons (HD_2 and HE_1), with the two nitrogens (ND_1 and NE_2) in a characteristic 4-peak pattern. Connecting the identified histidine ring spin systems to a particular position in the protein sequence is then achieved via the HD–CB correlation information, as CB chemical shifts are typically available after the sequential protein backbone assignment step. In case of CB frequency degeneracy, the assignment will of course be ambiguous, and additional NMR information, e.g. from NOESY-type spectra, is required to solve these remaining ambiguities. The HD–CG correlation spectrum adds additional chemical shift information on the CG ring carbon that is of importance for the determination of the tautomeric state of the uncharged (neutral) histidine side chains, as will be discussed in more detail in the following.

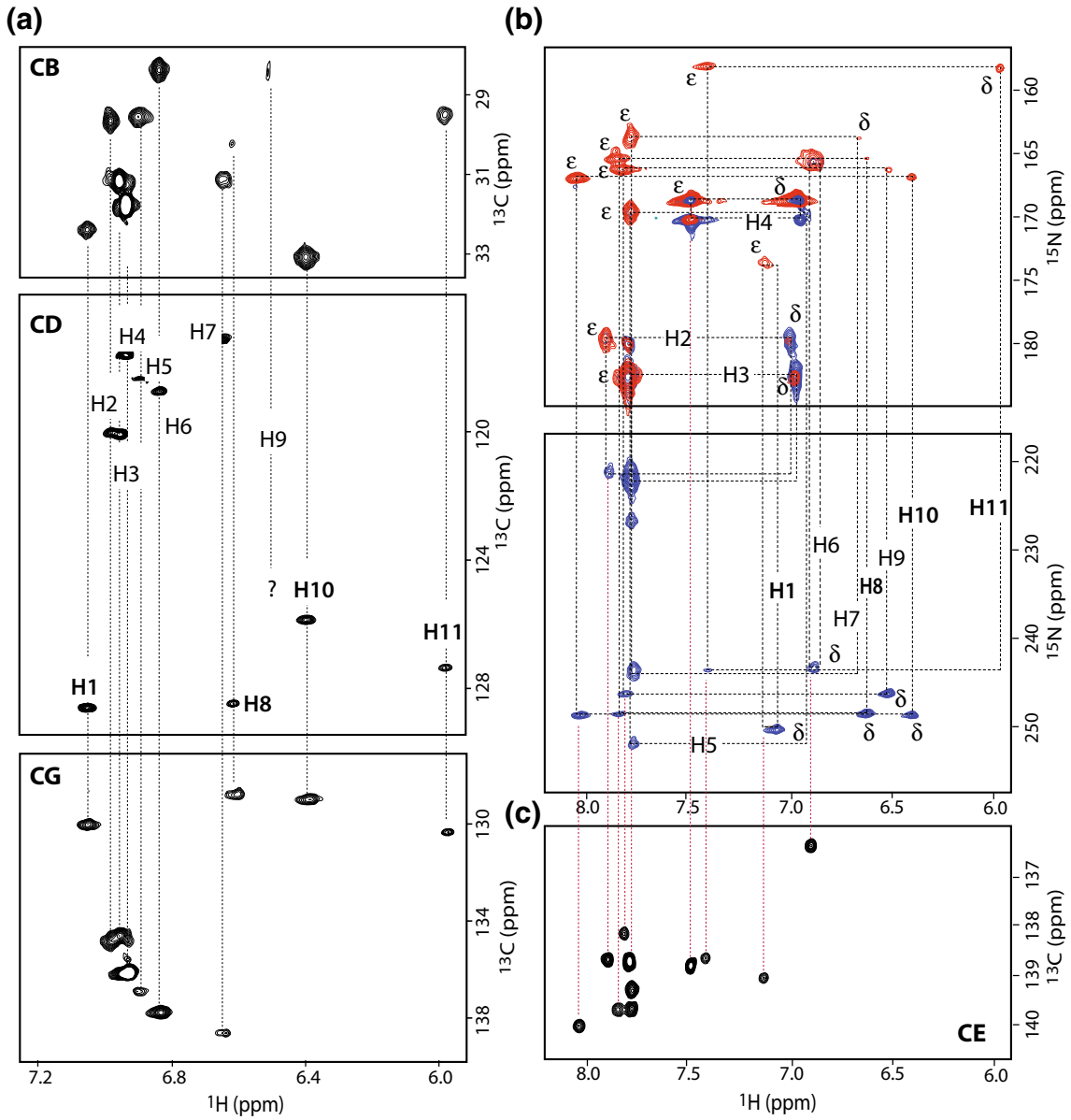


Fig. 3 Histidine 2D correlation spectra recorded with the pulse sequences of Fig. 2b–e (and Fig. 1a) on a sample of 200 µM rsfolder (40 °C, 850 MHz). All experiments were performed with an inter-scan delay of 300 ms. **a** ^{1}H–^{13}C spectra correlating the HD_2 with CB achieved by BEST-HDCB (top) and CG achieved by BEST-HDCG (bottom). For completeness, in the central part of the plot the spectral region of the BEST CT ^{1}H–^{13}C HSQC spectrum showing the HD_2–CD_2 correlations is plotted. In the CD and CG spectra, only negative contours are plotted. The observed 11 histidine spin systems are annotated. **b** ^{1}H–^{15}N spectra correlating the HD_2 and HE_1 with ND_1 and NE_2 using either the BEST-H(C)N experiment (red contours) or the SOFAST-H^{C}N pulse sequence (blue contours). Total data acquisition times were 1 h for BEST-HDCG and BEST-H(C)N, 3 h for BEST-HDCB, and 10 h for SOFAST-H^{C}N. The identified peak patterns for each histidine side chain are highlighted by dashed lines and annotated by arbitrary numbers (H1, H2,…,H11). For completeness, the HE_1–CE_1 correlation region of the BEST CT ^{1}H–^{13}C HSQC spectrum is shown in **c**

Histidine side chains in a protein at physiological pH can populate different protonation states: a bi-protonated state where both imidazole ring nitrogens are protonated, and the histidine has a net positive charge, as well as two neutral tautomers with only either the ND_1 (π tautomer) or the NE_2 (τ tautomer) protonated. NMR is a particular powerful technique to obtain valuable information about the relative populations of these states, and to measure the local pKa values of individual histidines, that can provide crucial insights into the protein's function. It has been shown that the ring ^{13}C and ^{15}N chemical shifts, as well as one-bond and multiple-bond heteronuclear coupling constants are sensitive to the protonation state of the histidine ring nitrogens (Blomberg et al. 1976; Pelton et al. 1993; Sudmeier et al.

2003; Li and Hong 2011; Hansen and Kay 2014), and a number of such NMR parameters have been used in the past for the study of histidine protonation states. In the following, we will demonstrate how this information can be extracted in a straightforward way from the correlation spectra of Fig. 3. A bi-protonated state can easily be recognized from the spectra in Fig. 3b as a peak pattern in the upper part of the spectrum with the two nitrogens having similar ^{15}N chemical shifts. This is clearly the case only for histidine **H4** in our rsfolder construct at pH 7.5. For the discrimination of the two neutral tautomers, the most efficient way is the exploitation of ^{13}C chemical shifts. Experimental as well as theoretical studies have shown that the ^{13}CG (^{13}CD) shifts in the π tautomer are downfield (upfield) shifted with respect to the shifts observed in the τ tautomer. The chemical shift difference $\Delta_C = \delta(CG) - \delta(CD)$ therefore provides a particularly robust and simple way to identify the tautomeric state of neutral histidine conformers. For rsfolder we can easily distinguish 4 histidines (**H1, H8, H10**, and **H11**—highlighted by bold letters in Fig. 3a) that predominantly adopt a π tautomeric state with $\Delta_C < 4$ ppm, while all others have much larger Δ_C values, typically around 20 ppm, indicative of a neutral τ tautomer (or bi-protonated state). We would like to emphasize here that the tautomeric information would have been much more difficult, or even impossible, to derive from the ^{1}H–^{15}N spectra only, as often reported in the literature. In particular, the H(C)N correlation experiment gives rise to strong cross peaks between the HE_1 and the protonated nitrogen (estimated $^{1}J_{CN} > 10$ Hz), a much weaker peak for the HD_2–N(H) correlation (estimated $^{1}J_{CN} < 5$ Hz), and no peak with the deprotonated nitrogen (estimated $^{1}J_{CN} < 2$ Hz), whatever the tautomeric state. This makes it impossible to distinguish between π and τ tautomeric states from such a spectrum, contrary to what has been reported (Sudmeier et al. 1996; Schmidt et al. 2008).

Tyrosine experiments

A set of 2D ^{1}H–^{13}C BEST correlation experiments specifically tailored for the assignment of tyrosine side chains is displayed in Fig. 4. The experiments exploit the well-separate chemical shift ranges of CZ, CE, and (CD, CG) aromatic ring carbons to avoid branched spin-systems (except for the CG → CB transfer step), and achieve highest possible coherence transfer efficiencies (Fig. 4a). The first two pulse sequences correlate the HD with either the CG (Fig. 4b) or CB (Fig. 4c), while the latter two start from HE and correlate with either CD (Fig. 4d) or CG (Fig. 4e). As for the histidine experiments, aromatic ^{13}C chemical shifts are edited in a CT fashion to avoid line splitting due to the large $^{1}J_{CC}$ in aromatic rings. The most critical (and inefficient) part in these sequences is the CG → CB transfer step in the BEST-HDCB experiment (Fig. 4c), as the CG nuclear spins cannot be decoupled from CD during this transfer step, resulting in the evolution of spin coherence under the influence of three one-bond ^{13}C–^{13}C couplings ($^{1}J_{CGCD1}$, $^{1}J_{CGCD2}$, and $^{1}J_{CGCB}$). Therefore, the transfer amplitude is given by the following relation (with $^{1}J_{CGCD1} = {}^{1}J_{CGCD2}$):

$$A(\delta) = \left\{ \frac{1}{2} sin\left(2\pi J_{CGCD} 2\delta\right) \times sin\left(\pi J_{CGCB} 2(\delta - \epsilon)\right) \times exp\left(-T_2^{eff} 2\delta\right) \right\} \quad (3)$$

with T_2^{eff} an effective relaxation time constant. Assuming coupling constants of $J_{CGCD} = 70$ Hz and $J_{CGCB} = 45$ Hz, and neglecting spin relaxation, the transfer efficiency is maximal for delays set to $2\delta = 12.5$ ms and $2\varepsilon = 1.5$ ms, reduced by the 180° ^{13}C pulse length as explained in the figure caption.

2D correlation spectra recorded with these pulse sequences are shown in Fig. 5 for the HD-detected (left) and HE-detected (right) experiments. Again, the whole set of spectra could be recorded overnight with individual data acquisition times of 2 h for BEST-HDCG, BEST-HECD, and BEST-HECG, and 7 h for BEST-HDCB. From these spectra, 8 tyrosine spin systems are identified, and the ^{1}H and ^{13}C nuclei of each side chain (aromatic ring) can be unambiguously connected despite the huge intensity variations observed in these spectra. The connectivity information benefits from the redundancy in using both the CD and CG chemical shifts to correlate the HD and HE of the same aromatic spin system. This allows to overcome some of the ^{13}C frequency degeneracy, e.g. the similar CD shifts of Y6 and Y7, and to improve the confidence in distinguishing small correlation peaks from spectral noise or artefacts. The assignment of tyrosine HD–CD and HE–CE correlation peaks has been confirmed a posteriori by a BEST CT ^{1}H–^{13}C HSQC recorded on a rsfolder sample that has been produced with unlabeled (^{12}C) tyrosine in the bacterial expression medium. Only five out of the eight tyrosine spin systems, however, can be connected to the aliphatic CB carbon and thus sequentially assigned. Note that even in a 4-times longer BEST-HDCB experiment we were not able to detect additional cross peaks for these tyrosine spin systems (Y4, Y5, and Y8), giving rise to only weak peak intensities even in the most sensitive ^{1}H–^{13}C correlation experiments, eventually due to an unfavorable kinetic exchange regime of ring flips in these aromatic side chains.

Conclusions

We have presented herein a set of two-dimensional aromatic-proton-detected correlation experiments for NMR assignment of histidine and tyrosine side chains in uniformly

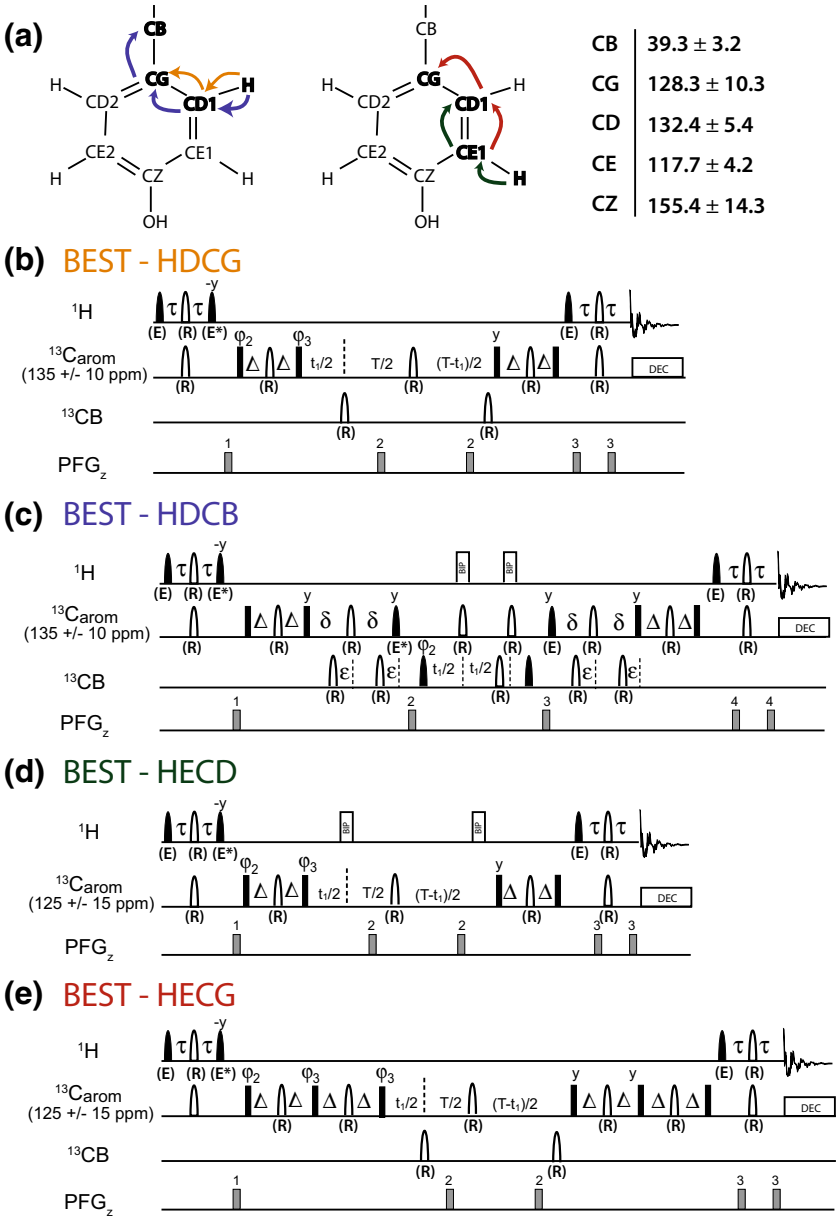


Fig. 4 BEST pulse sequences optimized for tyrosine side chains. **a** Coherence-transfer pathways (color-coded) exploited in these "out-and-back" ^{1}H–^{13}C correlation experiments starting either from HD (left) or HE (center). Average and standard deviation of the expected ^{13}C chemical shifts, as extracted from the BMRB data base, are indicated on the right. The symbols and nomenclatures used in these pulse schemes are identical to those introduced in Fig. 1a. Selective ^{1}H pulses are centered at 7.8 ppm, covering a bandwidth of 4.0 ppm. Band-selective ^{13}CB pulses are centered at 35 ppm covering a band width of 40 ppm. Note that a phase $\varphi_1 = -y$ has to be used in all these sequences to add the contributions originating from ^{1}H and ^{13}C steady-state polarizations. The transfer delays are set to $\Delta = 3.6$ ms, $\delta = 6.25$ ms $- 0.5\delta_3$, and $\varepsilon = 750$ µs $- 0.5\delta_3$, with δ_3 the length of the ^{13}C REBURP pulse; the CT delay is set to $T = 14.2$ ms. The BEST-HDCG (**b**) and BEST-HDCB (**c**) pulse sequences start and detect HD spins. In these sequences, the $^{13}C_{arom}$ pulses are applied with band-selective pulse shapes centered at 135 ppm, and covering a band width of 20 ppm. This ensures unidirectional coherence transfer as the ^{13}CE resonates outside this frequency band. The BEST-HECD (**d**) and BEST-HECG (**e**) pulse sequences start and detect HE spins. Here, shaped $^{13}C_{arom}$ pulses cover a larger band width of 30 ppm (centered at 125 ppm), thus now covering also the CE chemical shifts, but still decoupling CE from CZ

$^{13}C/^{15}N$ labeled proteins. The improved sensitivity brought about by the BEST/SOFAST pulse schemes considerably reduces the required experimental time to record a full set of NMR data. This is especially important for aromatic spin systems giving rise to very weak NMR signals that require long experimental times to achieve sufficient SNR. We have demonstrated the performance of these experiments for a 250-residue protein containing 11 histidine and 8 tyrosine

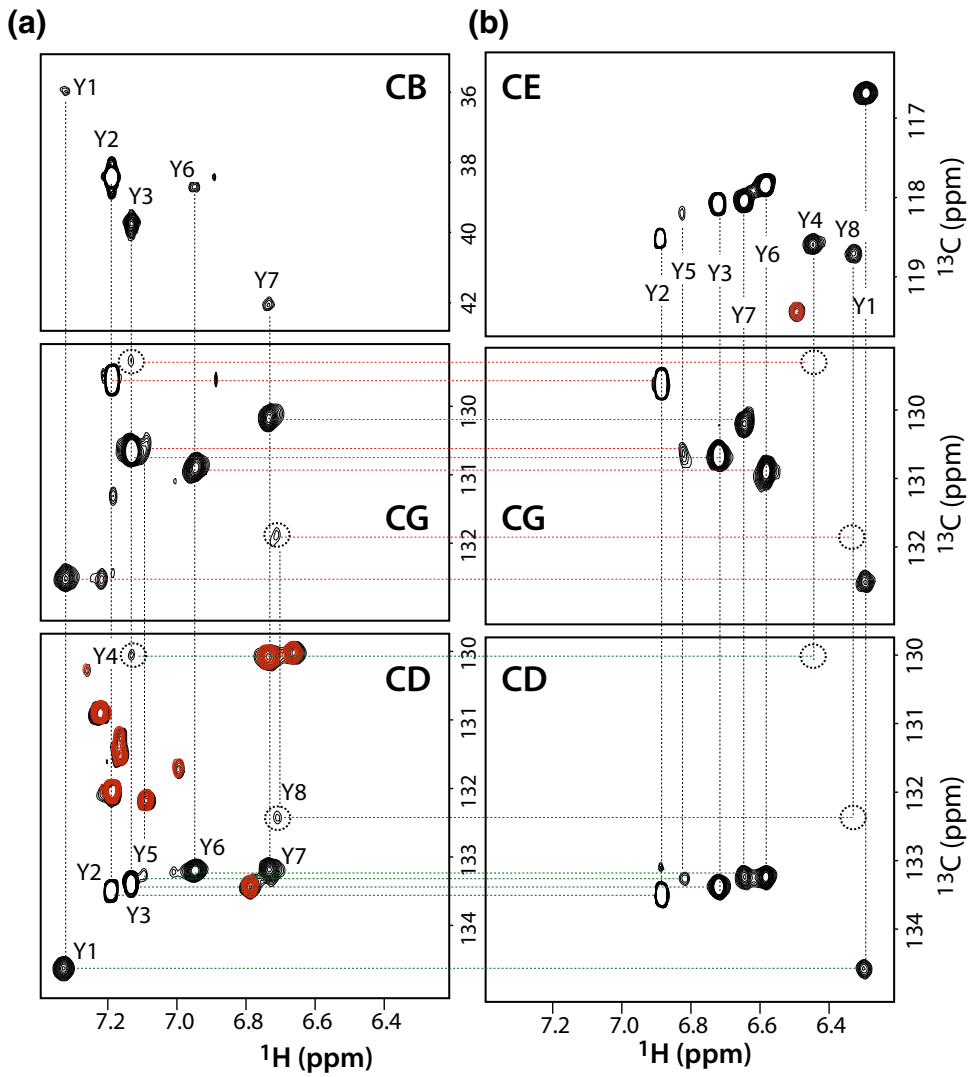


Fig. 5 Tyrosine 2D $^{1}H–^{13}C$ correlation spectra recorded with the pulse sequences of Fig. 4b–e (and Fig. 1a) on a sample of 200 µM rsfolder (40 °C, 850 MHz). All experiments were performed with an inter-scan delay of 300 ms. Only the interesting spectral regions are plotted. **a** BEST-HDCB (on top), BEST-HDCG (center), and BEST CT $^{1}H–^{13}C$ HSQC – HDCD region (bottom). **b** CT $^{1}H–^{13}C$ HSQC – HECE region (on top), BEST-HECG (center), and BEST-HECD (bottom). Total data acquisition times were 2 h for BEST-HDCG, BEST-HECD, and BEST-HECG, and 7 h for BEST-HDCB. In addition, the CT $^{1}H–^{13}C$ HSQC spectral regions recorded for a tyrosine unlabeled sample of rsfolder are plotted on top (red contours). Spectral connectivities are highlighted by dashed lines, and the tyrosine spin systems are annotated by arbitrary numbers (Y1, Y2,…,Y8). Dashed circles highlight the position of peaks that are close to the noise level, and that are fairly or not visible in this figure (chosen contour levels). Despite the low SNR of these peaks, the spin systems Y4 and Y8 can be connected with high fidelity due to the redundant information obtained from the CG and CD connectivity information

(+chromophore) residues. Despite a huge variation in peak intensities, already observed in a simple CT $^{1}H–^{13}C$ correlation spectrum, all expected His and Tyr spin systems could be unambiguously identified, and most of them connected to the sequential backbone assignment via correlation with the CB carbon. In the case of histidine, the set of experiments also allowed straightforward identification of the histidine's protonation state and tautomeric form. The proposed pulse sequences yield excellent water suppression and can thus be recorded on protein samples in H_2O. We therefore believe that they present a useful addition to the NMR toolbox for the atomic-resolution characterization of proteins in solution. Furthermore, they can be easily implemented on any modern Bruker NMR spectrometer by using the python setup scripts available from our laboratory within the NMRlib sequence library. Most of the proposed pulse schemes can be easily extended to 3D versions by additional CT frequency editing during the transfer delays if frequency degeneracy prevents the proper connection of the expected peak patterns. Furthermore, the H(C)C-type pulse schemes of Figs. 2 and 4 may also be used for Trp and Phe side chains, by properly adjusting the excitation bands of the ^{13}C pulses and transfer delays to the relevant chemical shift ranges and coupling topologies.

Acknowledgements We thank Dominique Bourgeois and Virgile Adam for making the rsfolder plasmid available, and Isabel Ayala and Karine Giandoreggio for protein sample preparation. We acknowledge support from the CNRS (Défis Instrumentation 2018). This work used the NMR and isotope labeling platforms of the Grenoble Instruct center (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

References

Baker CM, Grant GH (2007) Role of aromatic amino acids in protein–nucleic acid recognition. Biopolymers 85:456–470

Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) Analysis of catalytic residues in enzyme active sites. J Mol Biol 324:105–121

Blomberg F, Maurer W, Rüterjans H (1976) Nuclear magnetic resonance investigation of 15N-labeled histidine in aqueous solution. J Am Chem Soc 2932:8149–8159

Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280:1–9

Brutscher B (2000) Principles and applications of cross-correlated relaxation in biomolecules. Concepts Magn Reson 12:207–229

Brutscher B, Solyom Z (2017) Polarization-enhanced Fast-pulsing Techniques. In: Fast NMR data acquisition. pp 1–48

Brutscher B, Boisbouvier J, Pardi A et al (1998) Improved Sensitivity and resolution in 1 H–13 C NMR experiments of RNA. J Am Chem Soc 120:11845–11851

El Khatib M, Martins A, Bourgeois D et al (2016) Rational design of ultrastable and reversibly photoswitchable fluorescent proteins for super-resolution imaging of the bacterial periplasm. Sci Rep 6:18459

Eletsky A, Atreya HS, Liu G, Szyperski T (2005) Probing structure and functional dynamics of (large) proteins with aromatic rings: L-GFT-TROSY (4,3)D HCCH NMR spectroscopy. J Am Chem Soc 127:14578–14579

Favier A, Brutscher B (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR. J Biomol NMR 49:9–15

Geen H, Freeman R (1991) Band-Selective Radiofrequency Pulses. J Magn Reson 93:93–141

Hansen AL, Kay LE (2014) Measurement of histidine pKa values and tautomer populations in invisible protein states. Proc Natl Acad Sci 111:E1705–E1712

Holliday GL, Mitchell JBO, Thornton JM (2009) Understanding the functional roles of amino acid residues in enzyme catalysis. J Mol Biol 390:560–577

Hu J, Fu R, Nishimura K et al (2006) Histidines, heart of the hydrogen ion channel from influenza A virus: toward an understanding of conductance and proton selectivity. Proc Natl Acad Sci 103:6865–6870

Jacob J, Louis JM, Nesheiwat I, Torchia DA (2002) Biosynthetically directed fractional13C labeling facilitates identification of Phe and Tyr aromatic signals in proteins. J Biomol NMR 24:231–235

Kupce E, Freeman R (1993) Polychromatic selective pulses. J Magn Reson A 102:122–126

Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. J Magn Reson 187:163–169

Li S, Hong M (2011) Protonation, tautomerization, and rotameric structure of histidine: A comprehensive study by magic-angle-spinning solid-state NMR. J Am Chem Soc 133:1534–1544

Löhr F, Rüterjans H (1996) Novel pulse sequences for the resonance assignment of aromatic side chains in 13C-labeled proteins. J Magn Reson 112:259–268

Pelton JG, Torchia DA, Meadow ND, Roseman S (1993) Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated I11 Glc, a signal-transducing protein from *Escherichia coli*, using two-dimensional heteronuclear NMR techniques. Protein Sci 2:543–558

Pervushin K, Riek R, Wider G, Wüthrich K (1998) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in 13C-labeled proteins. J Am Chem Soc 120:6394–6400

Pervushin K, Vo B, Eletsky A (2002) Longitudinal 1H relaxation optimization in TROSY NMR spectroscopy. J Am Chem Soc 124:12898–12902

Prompers JJ, Groenewegen A, Hilbers CW, Pepermans HAM (1998) Two-dimensional NMR experiments for the assignment of aromatic side chains in 13 C-labeled proteins. J Biomol NMR 75:68–75

Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc 127:8014–8015

Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. J Biomol NMR 33:199–211

Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043

Schmidt H, Himmel S, Walter KFA et al (2008) Transverse relaxation-optimized HCN experiment for tautomeric states of histidine side-chains. J Korean Magn Reson Soc 12:89–95

Shaka AJ, Lee CJ, Pines A (1988) Iterative schemes for bilinear operators; application to spin decoupling. J Magn Reson 77:274–293

Shimahara H, Yoshida T, Shibata Y et al (2007) Tautomerism of histidine 64 associated with proton transfer in catalysis of carbonic anhydrase. J Biol Chem 282:9646–9656

Smith MA, Hu H, Shaka AJ (2001) Improved broadband inversion performance for NMR in liquids. J Magn Reson 151:269–283

Sudmeier JL, Ash EL, Günther UL et al (1996) HCN, a triple-resonance NMR technique for selective observation of histidine and tryptophan side chains in13C/15N-labeled proteins. J Magn Reson 113:236–247

Sudmeier JL, Bradshaw EM, Coffman Haddad KE et al (2003) Identification of histidine tautomers in proteins by 2D 1H/13Cδ2 one-bond correlated NMR. J Am Chem Soc 125:8430–8431

Veeman WS (1984) Carbon-13 chemical shift anisotropy. Prog Nucl Magn Reson Spectrosc 16:193–235

Weininger U, Respondek M, Akke M (2012) Conformational exchange of aromatic side chains characterized by L-optimized TROSY-selected ^{13}C CPMG relaxation dispersion. J Biomol NMR 54:9–14

Weininger U, Brath U, Modig K et al (2014) Off-resonance rotating-frame relaxation dispersion experiment for13C in aromatic side chains using L-optimized TROSY-selection. J Biomol NMR 59:23–29

Yamashita MM, Wesson L, Eisenman G, Eisenberg D (1990) Where metal ions bind in proteins. Proc Natl Acad Sci 87:5648–5652

Yamazaki T, Forman-Kay JD, Kay LE (1993) Two-dimensional NMR experiments for correlating 13CB and 1HD chemical shifts of aromatic residues in 13C-labeled proteins via scalar couplings. J Am Chem Soc 115:11054–11055

Biophysical Journal
Article

Biophysical Society

NMR Reveals Light-Induced Changes in the Dynamics of a Photoswitchable Fluorescent Protein

Nina-Eleni Christou,[1] Isabel Ayala,[1] Karine Giandoreggio-Barranco,[1] Martin Byrdin,[1] Virgile Adam,[1] Dominique Bourgeois,[1] and Bernhard Brutscher[1,*]
[1]Institut de Biologie Structurale, Université Grenoble Alpes, CEA, CNRS, Grenoble, France

ABSTRACT The availability of fluorescent proteins with distinct phototransformation properties is crucial for a wide range of applications in advanced fluorescence microscopy and biotechnology. To rationally design new variants optimized for specific applications, a detailed understanding of the mechanistic features underlying phototransformation is essential. At present, little is known about the conformational dynamics of fluorescent proteins at physiological temperature and how these dynamics contribute to the observed phototransformation properties. Here, we apply high-resolution NMR spectroscopy in solution combined with in situ sample illumination at different wavelengths to investigate the conformational dynamics of rsFolder, a GFP-derived protein that can be reversibly switched between a green fluorescent state and a nonfluorescent state. Our results add a dynamic view to the static structures obtained by x-ray crystallography. Including a custom-tailored NMR toolbox in fluorescent protein research provides new opportunities for investigating the effect of mutations or changes in the environmental conditions on the conformational dynamics of phototransformable fluorescent proteins and their correlation with the observed photochemical and photophysical properties.

SIGNIFICANCE Phototransformable fluorescent proteins (PTFPs) are essential tools for superresolution microscopy. In practical applications, however, researchers often encounter problems when using PTFPs in a particular cellular context because the environmental conditions (pH, temperature, redox potential, oxygen level, viscosity, etc.) affect their brightness, photostability, and phototransformation kinetics. Rational fluorescent protein engineering exploits the mechanistic information available from structural studies, mainly x-ray crystallography, to design new PTFP variants with improved properties for particular applications. Here, we apply NMR spectroscopy in solution to investigate the light-induced changes in conformational dynamics of rsFolder, a reversibly switchable fluorescent protein. The dynamic view offered by NMR highlights protein regions that comprise potentially interesting mutation points for future mutagenesis campaigns.

INTRODUCTION

The discovery and engineering of phototransformable fluorescent proteins (PTFPs), derived either from *Aequorea victoria* green fluorescent protein (GFP) or other fluorescent proteins from Anthozoa species (corals and anemones), has triggered exciting new modalities in fluorescence microscopy (1) and biotechnology (2). In particular, PTFPs have become indispensable tools for modern superresolution fluorescence microscopy techniques that achieve resolution beyond the diffraction limit of light (3,4). At present, a large number of PTFPs have been discovered or engineered that differ in their specific response to light (photoactivation, photoconversion, or reversible photoswitching), in the kinetics of this response, in their photostability and general fluorescence characteristics (brightness, emission color, Stokes shift, etc.), as well as in their oligomerization state, folding stability, and maturation efficiency. Furthermore, these properties are generally dependent on environmental conditions (5). The requirements for an optimal PTFP marker are different for specific applications, which call for further expansion and improvements of the PTFP toolbox.

Understanding at near-atomic resolution the mechanistic features responsible for the observed photochemical and photophysical properties of PTFPs is a key step to envision the rational design of new variants optimized for a particular application or experimental context. Most mechanistic studies on PTFPs reported in the literature have used x-ray

Submitted July 3, 2019, and accepted for publication October 25, 2019.
*Correspondence: bernhard.brutscher@ibs.fr
Editor: Michael Sattler.
https://doi.org/10.1016/j.bpj.2019.10.035

crystallography as the method of choice to determine structures of their different ground-state conformations. Such structures are essentially static and, because data are typically collected at cryogenic temperature, they are representative of frozen states. Kinetic crystallography has also been used in combination with optical spectroscopy and/or molecular dynamics simulations to decipher the structures of short-lived ground-state species such as intermittent dark states (6,7). Recently, ultrafast time-resolved crystallography using an x-ray free-electron laser source has been able to image an electronically excited short-lived intermediate state in rsEGFP2 that becomes populated during photoswitching between a nonfluorescent ("off") and a fluorescent ("on") state (8), shedding some light on the switching dynamics of a PTFP at the picosecond timescale.

Although fluorescent proteins are generally quite resistant to their environment, an increasing number of studies have shown that PTFP photophysics are affected by external factors such as pH, redox potential, oxygen level, viscosity, temperature, etc. (5,9–13). These environmental effects are difficult to study by x-ray crystallography in which structural data are collected in conditions imposed by crystallogenesis that may differ substantially from the conditions found during in situ fluorescence imaging. Moreover, little is known about the conformational dynamics of PTFPs at physiological temperatures and their relevance for the observed fluorescence and phototransformation properties. To address these issues, multidimensional solution NMR spectroscopy may provide a complementary tool to investigate at atomic resolution the conformational dynamics of PTFPs in different states and over a wide range of temperatures and buffer conditions. However, quite surprisingly, so far only little information on PTFPs has been obtained from NMR (11,14,15), whereas a few more studies on chromophore dynamics in conventional GFPs can be found in the literature (16–18).

Here, we introduce a toolbox of multidimensional and multinuclear (^{1}H, ^{13}C, ^{15}N) NMR experiments specifically tailored to investigate the structural dynamics of PTFPs. We apply this toolbox to probe the conformational dynamics of rsFolder, a PTFP that can be reversibly switched from a stable fluorescent "on" state to a metastable nonfluorescent "off" state, and vice versa, upon illumination with cyan or violet light, respectively. Such reversibly photoswitchable fluorescent proteins (RSFPs) are required for a variety of applications, ranging from superresolution microscopy (RESOLFT, NL-SIM, pcSOFI, and others) (19–21), optical lock-in detection (OLID) (22), multiplexed imaging (OPIOM) (23), up to the development of high-density optical storage media (2). rsFolder has been genetically engineered as a chimera between rsEGFP2 (24) and superfolder GFP (25) to facilitate proper folding in highly reducing environments such as the bacterial periplasm (26). The photophysical properties of rsFolder have been characterized by optical spectroscopy at room temperature, and high-resolution crystallographic structures are available for both the "on" and "off" states (26). These structures reveal the typical 11-stranded β-barrel fold, enclosing an endogenous *p*-HBI chromophore, formed by maturation of an Ala-Tyr-Gly tripeptide. Upon light illumination at 488 nm (405 nm), the chromophore undergoes a *cis* to *trans* (*trans* to *cis*) isomerization and protonation (deprotonation) at the ζ position of the phenol ring. Although the β-barrel structures for the "on" and "off" states almost perfectly superpose (C^{α} rmsd = 0.21 Å), slight structural rearrangements are observed for a few side chains located in the close vicinity of the chromophore (Fig. 1). Our NMR data complement this static picture by adding information on the changes in local conformational sampling after photoswitching, providing some new insight toward the long-term goal of better understanding how the photophysical properties of PTFPs are influenced by different types of local molecular flexibility. This work opens the door to further studies aiming at evaluating how the conformational dynamics of PTFPs respond to mutational or environmental changes.

MATERIALS AND METHODS

Protein samples

Purification of rsFolder was performed as described previously (26). The full-length rsFolder complementary DNA was cloned in the ampicillin-resistant expression plasmid pET15b (Addgene, Teddington, UK), and constructs were designed to bear a six-residue N-terminal His-tag linked to the protein coding region via a thrombin cleavage sequence. For initial NMR experiments detecting aromatic side-chain protons, the His-tag was cleaved using a thrombin protease. Otherwise, all experiments were recorded

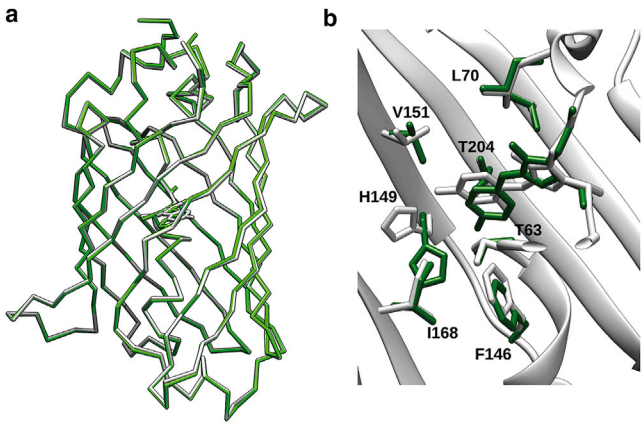


FIGURE 1 Crystallographic structures of rsFolder in the fluorescent "on" state (*green*) and the nonfluorescent "off" state (*gray*). (*a*) Superposition of the main chain coordinates highlighting that the overall β-barrel structure of the protein is identical in the two states. (*b*) Zoom on the chromophore region of rsFolder shows the change in chromophore geometry from *cis* to *trans*, as well as some side-chain rearrangements in the vicinity of the chromophore. The most pronounced structural changes upon light illumination are observed for F146, H149, and V151. To see this figure in color, go online.

without cleavage of the His-tag because those samples showed a higher time stability. *Escherichia coli* BL21(DE3) cells were transformed with the plasmid described above for protein expression. An isotopically enriched minimal medium, containing $^{15}NH_4Cl$ (1 g/L) and ^{13}C-glucose (2 g/L), was used to grow the cells until the cultures reached an OD at 600 nm between 0.6 and 0.8, at which point protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside. For uniformly $^{15}N/^{13}C$ labeled samples, 1 mM isopropyl β-D-1-thiogalactopyranoside was added and induction was performed overnight at 20°C. The cells were harvested by centrifugation and subsequently lysed by sonication in lysis buffer (50 mM HEPES, 100 mM NaCl (pH 7.5)). The lysate was centrifuged at 4°C for 30 min at 46,000 × *g*. The protein present in the supernatant was purified using a Ni-NTA column (Qiagen, Hilden, Germany). The protein was eluted by increasing the concentration of imidazole. The resulting protein was further purified using size-exclusion chromatography with a Superdex 75 column (GE Healthcare, Chicago, IL) equilibrated with the NMR buffer (50 mM HEPES (pH 7.5)).

In situ laser illumination inside the NMR magnet

An in situ laser illumination device similar to that described by Mizuno et al. (15) was built. It comprises an Oxxius laser box (wavelength combiner L4Cc; Oxxius, Lannion, France) hosting two lasers of wavelengths 405 nm (LBX-405-180-CSB-OE) and 488 nm (LBX-488-200-CSB-OE) and an optical fiber (FG105UCA, 0.22 NA, 105 μm core diameter, 2 mm diameter protective cover, 5 m length; Thorlabs, Newton, NJ). The fiber end was left naked and routinely cleaved when necessary, using a ceramic fiber scribe (CSW12-5; Thorlabs). The fiber was inserted into the plunger of a 5 mm diameter Shigemi tube (Shigemi, Tokyo, Japan). Importantly, the Shigemi plunger was filled with water before inserting the fiber to avoid a sudden change in the refractive index at the air-glass interface, thus reducing the divergence of the laser beam. The laser power was set and monitored using the Oxxius software from a laptop connected to the laser combiner, and the laser output was controlled either manually using a shutter or by digital on/off modulation from the NMR console connected to the laser box via transistor-transistor logic (TTL) lines. The difference between nominal laser power as displayed by the Oxxius software and the measured output power at the fiber tip was ~30%, suggesting a coupling efficiency of ~70%. For short illumination times (less than 1 min), the full laser power was used (200 mW nominal), whereas for longer exposure times (>>1 min), the output was set to 30% of the maximal power to limit the effect of photobleaching. To allow smooth insertion of the fiber-coupled Shigemi tube into the NMR magnet, the use of two Bruker sample holders put on top of each other proved to be helpful (Fig. S1 *d*).

NMR measurements

All NMR experiments were performed on Bruker Avance III HD spectrometers (Billerica, MA), at the high-field NMR facility of Institut de Biologie Structurale, operating at magnetic field strengths of 600-MHz (H/D exchange), 700-MHz (^{15}N relaxation, NMR assignment), and 850-MHz (LD-EXSY, ^{15}N CPMG-RD, H/D exchange, methyl side chain, aromatics, and chromophore NMR experiments) and equipped with cryogenically cooled triple-resonance probes and pulsed z-field gradients. NMR measurements were performed at 40°C on NMR samples containing 100–200 μM [U-^{13}C, ^{15}N]-enriched rsFolder in 50 mM HEPES (pH 7.5) and 5% (v/v) D_2O. The samples were stored at 4°C between data recordings, and the sample quality was routinely checked for possible aggregation and/or degradation by recording ^{1}H-^{15}N correlation spectra. "On"-state data were recorded after short (5 s) illumination at 405 nm in the absence of laser illumination, whereas off-state data were acquired after a short (10 s) high-power illumination at 488 nm, under continuous illumination (488 nm) using reduced (30%) laser power to prevent on-state recovery by thermal relaxation. Most NMR experiments used in this study are implemented in the NMRlib pulse sequence library for Bruker spectrometers (27) that can be freely downloaded from our web site (http://www.ibs.fr/research/scientific-output/software/pulse-sequence-tools). NMR data processing was performed using TopSpin 3.5 (Bruker BioSpin), whereas data analysis was done using the CCPNMR software (28).

For sequential resonance assignment of the "on" state, a set of five triple-resonance three-dimensional (3D) BEST-TROSY experiments (29,30), HNCO, HNcoCA, HNCA, HNcoCACB, and HNCACB, was recorded. For the "off" state, two LD-EXSY data sets, (^{15}N, ^{15}N, ^{1}H) and (^{1}H, ^{15}N, ^{1}H), were recorded to transfer the "on"-state ^{1}H, ^{15}N assignments to the "off" state and complemented by 3D BEST-TROSY HNcaCO, HNCACB to extend the assignment to CO, CA, and CB. Pulse sequences of the LD-EXSY experiment are provided in Fig. S4. Additional side-chain assignments of aromatic tyrosine and histidine residues, including the tyrosine moiety of the *p*-HBI chromophore, was achieved using a series of specifically tailored two-dimensional (2D) correlation experiments as described recently (31). Partial methyl (^{1}H, ^{13}C) assignments were obtained by a combination of 3D CcoNH-TOCSY and HccoNH-TOCSY (32,33) experiments, complemented by 3D methyl-edited HCC-COSY and amino-acid-edited 2D SOFAST ^{1}H-^{13}C correlation spectra (34,35). NMR assignments of rsFolder have been deposited with the Biological Magnetic Resonance Bank under accession numbers 50010 for the fluorescent "on" state, and 50011 for the nonfluorescent "off" state.

^{15}N relaxation experiments (T_1, T_2, HETNOE) were performed using standard pulse schemes (36) implemented in NMRlib. For T_1 (T_2) measurements, 11 (10) time points were recorded with relaxation delays varying from 0 to 1.8 s (0–240 ms). For the HETNOE, reference and saturation spectra were recorded in an interleaved manner with the recycle delay set to 5 s and the saturation time set to 3 s. Relaxation data for 181 on- state and 162 off-state backbone amide sites could be obtained at 40°C from crosspeaks that were well resolved and of sufficient signal to noise in the recorded 2D ^{1}H–^{15}N spectra, whereas all other (overlapping or weak) peaks were excluded from further analysis. The overall correlation time has been estimated from the average T_1 and T_2 values (37) measured for residues in the β-barrel as $\tau_c = \sqrt{(6T_1/T_2) - 7}/4\pi\nu_N$. Additional ^{15}N CPMG relaxation-dispersion (RD) experiments were performed using a BEST-TROSY CMPG pulse sequence (38) with the relaxation delay set to 40 ms and the CPMG frequency ranging from 50 to 1000 s^{-1}.

For H/D exchange measurements, lyophilized samples of protonated $^{13}C/^{15}N$-labeled rsFolder were dissolved in 300 μL D_2O (sample concentration of 200 μM), transferred to a 5 mm Shigemi tube with the optical fiber inserted, and quickly put into the NMR spectrometer for data recording at 40°C. The decay of amide ^{1}H NMR signals, because of the exchange with deuterons, was then monitored for ~20 h by a series of ^{1}H-^{15}N BEST-TROSY spectra with a time resolution of 2.5 min per experiment. Because of sample preparation and spectrometer setup (locking, shimming, etc.), the dead time of these kinetic experiments was ~5 min. To estimate the accuracy and reproducibility of the measurements, two data sets were recorded for both "on" and "off" states (at 600 and 850 MHz). To obtain more quantitative data for amide sites that are slowly exchanging in the "on" state but much faster in the "off" state, we performed an additional H/D exchange experiment. In this experiment, the exchange measurement starts with the "on" state, and the sample is switched after a few minutes to the "off" state using high-power laser (488 nm) illumination.

RESULTS

Laser-driven photoswitching

To perform a set of time-consuming multidimensional (2D and 3D) NMR experiments for the metastable "off" state of rsFolder, we have adapted a setup similar to the one proposed by Mizuno et al. (15) that allows one to illuminate the sample in the NMR magnet during data collection with 488

or 405 nm laser light via an optical fiber introduced into the glass insert of a standard Shigemi tube (Fig. 2). This illumination device is portable, can be used with any high-field NMR spectrometer, and can be synchronized with an NMR pulse sequence to switch the sample between the "on" and "off" states on demand (Fig. S1).

To characterize the efficiency of our illumination device to switch rsFolder between the "on" and "off" states, we have performed different types of NMR experiments. Real-time one-dimensional ^{1}H NMR spectra recorded under continuous laser illumination (488 nm) are shown in Fig. S1 for different laser powers. The two NMR resonances visible in this spectral window correspond to a methyl ^{1}H that is spectroscopically resolved in both states and thus presents a convenient probe for such one-dimensional NMR-detected switching kinetics. As expected, the switching kinetics depend on the chosen laser power, with a measured "on"-off switching half time of a few seconds at the highest available laser output. In a second type of experiment, shown in Fig. 2 *d*, we have recorded 2D ^{1}H-^{15}N readout spectra after a variable illumination time. The spectra in Fig. 2 *b* illustrate the decay (buildup) of the NMR peak intensities when switching either from "on" to "off" or from "off" to on. Our data show that the "off"-to-"on" switching under 405 nm illumination is much faster than what is observed for the reverse process under 488 nm illumination. We would like to stress here that these measurements only provide an estimate of the switching efficiency that may vary by up to a factor of 2 because of a number of experimental factors (fiber end aging, water evaporation in Shigemi plunger, exact positioning of fiber, etc.).

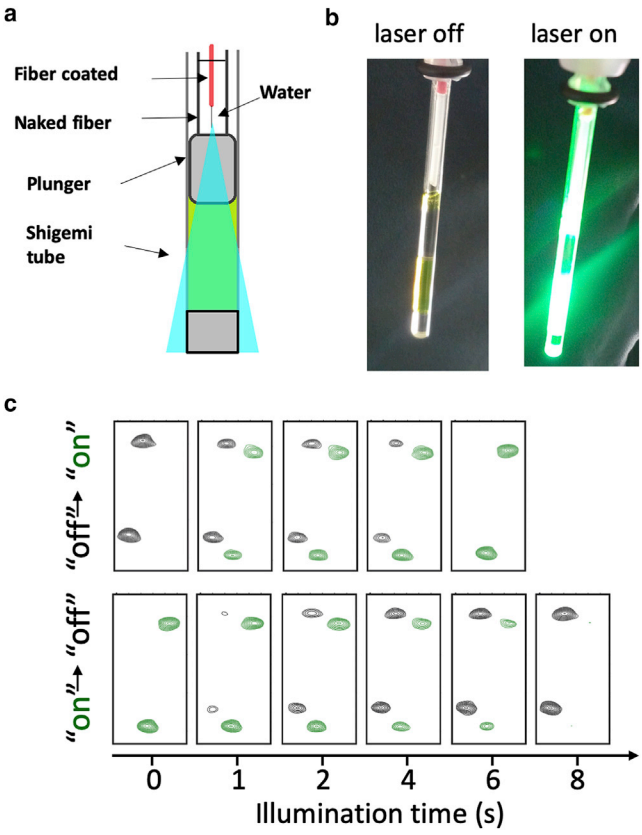


FIGURE 2 In situ NMR sample illumination device. (*a*) A schematic representation of the insertion of the optical fiber into the water-filled plunger of a Shigemi NMR tube. The NMR sample is illuminated from the top in a conic geometry. (*b*) A sample of rsFolder: picture with and without light illumination. (*c*) Switching kinetics measured for rsFolder using variable illumination times and a BEST-TROSY ^{1}H-^{15}N readout sequence. Only a small spectral region is shown, and the on-state peaks have been artificially color coded in green. The laser power (measured at the fiber end) was 140 mW at 488 nm ("on"-to-"off" switching) and 120 mW at 405 nm ("on"-to-"off" switching). To see this figure in color, go online.

A superposition of the ^{1}H-^{15}N fingerprint spectra recorded in the absence ("on" state) and presence of 488 nm light illumination ("off" state) is plotted in Fig. S2. Many residues (amide groups) show significant peak shifts in these spectra, indicating that the change in chromophore geometry induced by photon absorption has a widespread effect on the amide ^{1}H and ^{15}N chemical shifts in the β-barrel (see below). The laser-induced peak shifts are reversible as illustrated in Fig. S3, *a* and *b* by the superposition of "on"-state and "off"-state spectra recorded before and after 1000 switching cycles using alternate sample illumination by cyan and violet light. We found that photobleaching resulted in a progressive loss of peak intensity (~25% after 1000 cycles under the chosen experimental conditions) but did not induce any noticeable peak shifts in the spectra, indicating that the NMR-observable conformational "on" and "off" states are not affected by photobleaching. The absence of additional peaks in our NMR spectra developing over time suggests that photo-damaged rsFolder may form large aggregates not observable by solution NMR.

The observed peak shifts between rsFolder "on" and "off" states offer the possibility to quantify a "switching efficiency" that can be related to the fluorescence switching contrast of importance in microscopy applications. Quantification of the residual peak intensity of "on"-state peaks in the "off"-state spectrum provides a measure of the relative state populations. For rsFolder, we have measured this residual on-state population to be ~3.5% (Fig. S2), independent of the sample temperature in the range of 20–40°C. This NMR-derived value closely matches results from fluorescence spectroscopy (26), where a residual fluorescence intensity of ~4% in the "off" state at room temperature was obtained, in line with the notion that the rsFolder "off" state is fully nonfluorescent. Under steady-state conditions and in the presence of continuous 488-nm light illumination, rsFolder is mainly populating the "off" state (~96.5%) but still constantly switching between "off" and "on" states with an exchange rate that depends on the light power density in the NMR sample. To evaluate experimentally whether this exchange process is a source of NMR line broadening in the "off" state, we have recorded ^{1}H-^{15}N correlation spectra with and without light illumination (Fig. S3, *c* and *d*). The absence of any significant change

in peak intensities observed in these spectra provides a clear evidence that such effects can be safely neglected. NMR also enables the measurement of the "off"-to-"on" back-switching efficiency, which in the case of rsFolder turns out be close to 100% (Fig. S2).

Laser-aided backbone resonance assignments and secondary structure

$^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, ^{13}CO, and ^{15}N resonance assignment of rsFolder at 40°C in the "on" state was obtained from a series of 3D BEST-TROSY HNC correlation experiments (30). Almost complete (95%) assignment was obtained for residues 2–230, corresponding to the β-barrel structure. Only 12 nonproline residues (W58, V69, D77, H78, E133, Y144, V151, K167, I168, K210, D211, and K215) could not be assigned from these spectra. Of those, V69, V151, and K167 are in direct contact with the chromophore based on crystallographic data. The ^{1}H-^{15}N correlation spectrum, annotated with the assigned residue number, is shown in Fig. S5.

Recording and analyzing such a set of 3D assignment spectra is a time-consuming task (typically 1 week of data recording and another week of data analysis). To speed up the assignment process for the "off" state, we have adapted a different strategy that makes use of our laser-induced in situ switching capabilities. The "on"-state assignment can be directly transferred to the "off" state by performing an LD-EXSY, conceptually similar to SCOTCH experiments (39), proposed in the past to monitor light-induced chemical reactions. The basic scheme of an LD-EXSY experiment is shown in Fig. 3 *a*, whereas a more detailed drawing of the pulse sequence is provided in Fig. S4. During the magnetization recovery time in the beginning of the pulse sequence (typically a few seconds), a laser illumination (488 nm) switches the proteins in the sample to the "off" state, allowing one to edit either the ^{1}H or ^{15}N chemical shift for this state. After chemical shift editing, the magnetization is stored as longitudinal spin order (N_z), and a second laser pulse, this time at 405 nm, converts the proteins back to the "on" state during a mixing period (exchange time) of typically a few hundred milliseconds. The remainder of the sequence allows one to edit the ^{1}H and ^{15}N frequencies in the "on" state. The resulting 3D data set thus comprises one "off"-state (^{1}H or ^{15}N) and two "on"-state (^{1}H and ^{15}N) chemical shift dimensions. Examples of 2D ^{15}N-^{15}N correlation planes extracted for two residues of rsFolder undergoing large chemical shift changes upon light illumination from a 3D LD-EXSY spectrum are shown in Fig. 3 *b*. Fig. 3 *c* shows annotated 2D projections of two 3D data sets recorded for rsFolder, with either the ^{1}H or ^{15}N chemical shift labeled in the "off"-state spectral dimension. Off-diagonal peaks detected in these spectra originate from protein residues with amide chemical shifts (^{1}H or ^{15}N) that differ between the two states, whereas diagonal peaks are either due to incomplete switching or chemical shifts that are not affected by photoswitching. These LD-EXSY spectra thus provide a straightforward and

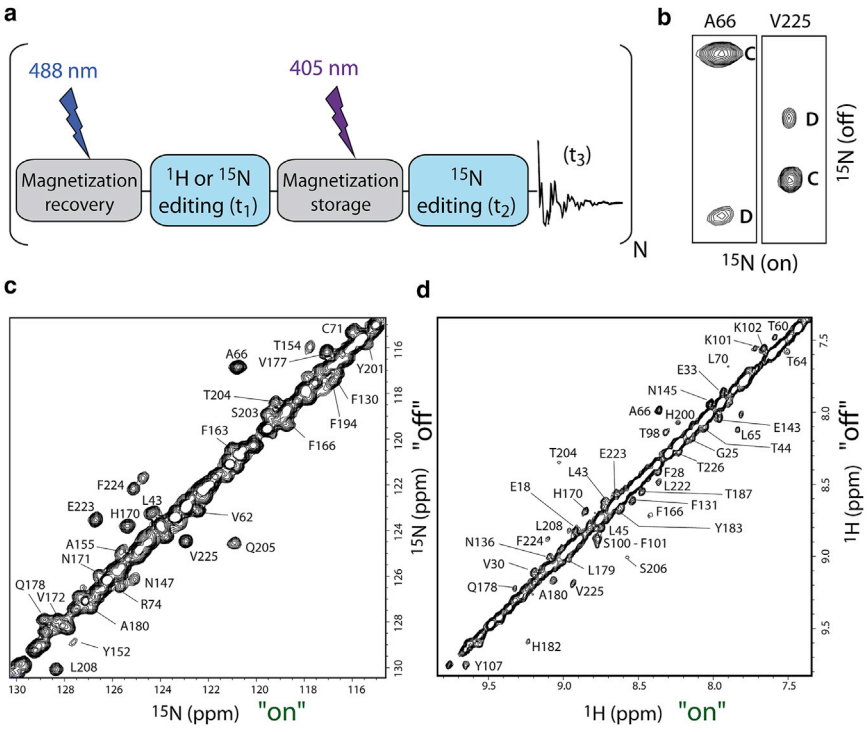


FIGURE 3 Laser-driven exchange spectroscopy (LD-EXSY) experiment. (*a*) Schematic drawing of LD-EXSY pulse sequence including light illumination at two different wavelengths: at 405 nm during the recycle delay (magnetization recovery, 2 s) and at 488 nm during the EXSY mixing time (magnetization storage, 600 ms). The laser sources are coupled to the NMR console via TTL lines and switched on and off by specific commands in the NMR pulse sequence code. The resulting 3D spectra have two "on"-state and one "off"-state chemical shift dimensions. The full pulse sequence is shown in Fig. S4. (*b*) 2D slices extracted from the ^{15}N-edited LD-EXSY, corresponding to residues A66 and V225, both showing significant ^{15}N chemical shift changes between the "on" and "off" states. The relative intensities observed for the diagonal (D) and crosspeaks (C) provide a measure of the switching performance of our illumination device under the chosen experimental conditions. (*c*) A 2D projection of a 3D ^{15}N-edited LD-EXSY spectrum. Off-diagonal peaks correspond to rsFolder residues (annotated), which show ^{15}N chemical shift differences between "on" and "off" states. (*d*) A 2D projection of a 3D ^{1}H-edited LD-EXSY. To see this figure in color, go online.

time-efficient way for amide ^{1}H, ^{15}N assignments in the "off" state. Additional ^{13}C chemical shift information in the "off" state was then extracted from 3D BT-HNcaCO and BT-HNCACB spectra. This resulted in an assignment completeness of 93% (residues 2–230), with a total of 16 unassigned nonproline residues (W58, E133, Y144, S148, N150, V151, K157, Q158, N165, K167, I168, R169, L202, K210, D211, and K215), of which S148, V151, and K167 are in direct contact with the chromophore based on crystallographic data. Note that in the case of severely line-broadened amide resonances (see below), no cross-peaks were observed in the triple-resonance data, and assignment was only possible with the LD-EXSY experiment.

The assigned NMR frequencies, and in particular the ^{13}C chemical shifts, provide useful information on the secondary structural elements (α-helix, β-strand) present in rsFolder under our experimental conditions. In Fig. S6, the secondary structural features derived from our measured chemical shifts using the TALOS-N program are plotted (40). A comparison with the x-ray structures of rsFolder (Protein Data Bank, PDB: 5DTZ and 5DU0) shows that the β-barrel structure in solution is almost identical to that observed in the crystal at cryogenic temperature. Nevertheless, a few slight differences are observed for strand β7 and the two helices (H2 and H3) inside the barrel to which the chromophore is connected. The β7 strand in solution is extended by three residues at the N-terminal end, helix H2 is shifted by three residues toward the C-terminus, and helix H3 has a noncanonical backbone geometry that is not identified as an α-helix by TALOS-N. Interestingly, it has been concluded from the hydrogen bonding patterns observed in the high-resolution crystal structure of EGFP (41), an "ancestor" protein of rsFolder, that helix H2 is composed of a 3_{10} helix (P57–L61 in rsFolder) followed by a short regular α-helix (L61–L65 in rsFolder), in agreement with our NMR data.

Backbone dynamics in the "on" and "off" states

To obtain information on the molecular tumbling correlation time and local backbone dynamics of rsFolder, we have performed ^{15}N relaxation measurements in both the "on" and "off" states at 40°C. The measured T_1, T_2, and HETNOE data are plotted in Fig. 4 *a* as a function of the protein sequence, with the secondary structural elements as identified from TALOS-N plotted on top of each graph. These data indicate an overall rigid monomeric globular protein. The T_1 and T_2 time constants measured for residues involved in the β-barrel structure allow one to estimate a tumbling correlation time (at 40°C) of 9.3 ns for rsFolder in the "on" state and 9.8 ns in the off state. The slight increase in the apparent correlation time (molecular size) can be explained by the presence of a low population of dimeric (or higher oligomeric) species that are in fast exchange with the rsFolder monomer. This finding further

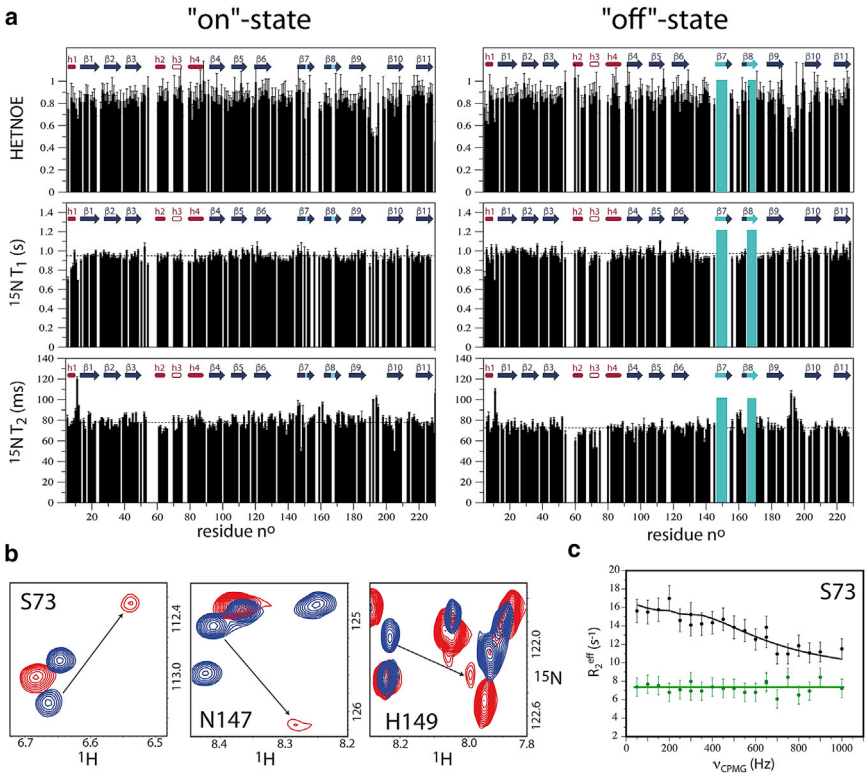


FIGURE 4 Backbone dynamics of rsFolder in "on" and "off" states. (*a*) ^{15}N relaxation data (T_1, T_2, and HETNOE) measured at 700 MHz and 40°C on the same $^{13}C/^{15}N$ labeled rsFolder sample, plotted as a function of protein sequence. Secondary structural elements as identified from the ^{13}C chemical shifts by the program TALOS-N are plotted on top of each graph. Cyan bars indicate peptide regions for which no (or very weak) peaks are detected in the "off" state. Dashed lines indicate average T_1 and T_2 values calculated for residues in the β-strand regions (on state: T_1^{av} = 950 ms and T_2^{av} = 78 ms; off state: T_1^{av} = 980 ms and T_2^{av} = 72 ms) and (*b*) ^{1}H-^{15}N BEST-TROSY spectral regions (*red*: "on" state; *blue*: "off" state) zooming on line-broadened "off"-state peaks. (*c*) CPMG-RD data recorded for residue S73 at 850 MHz in the "on" (*green*) and "off" (*black*) states. The straight line corresponds to a fit of the data using the program ChemEx (50). To see this figure in color, go online.

supports our observation of rsFolder being more aggregation prone in the "off" state under light illumination than in the thermodynamically stable "on" state. Increased flexibility on the picosecond to nanosecond timescale (longer T_2 and reduced HETNOE values) is detected in some loop regions, especially the loops connecting the first helix (H1) and strand β1 and the loop between β-strands 9 and 10. Most interestingly, there is no significant difference in the fast timescale dynamics, as seen from these ^{15}N relaxation data, between the "on" and "off" states. However, some differences are observed that can be ascribed to increased conformational dynamics on the microsecond to millisecond timescale in the "off" state of rsFolder. A first observation is the lack of peak intensity for some amides in β-strands 7 and 8 in the ^{15}N relaxation spectra (Fig. 4 *b*, *turquoise bars*) as well as reduced T_2 values measured for residues C71 and S73 in the noncanonical helix H3. The reduction in peak intensities for some residues in the "off" state can also be observed in standard 2D ^{1}H-^{15}N BEST-TROSY spectra, as illustrated in Fig. 4 *b*. To further characterize these conformational dynamics, we have performed CPMG-type RD experiments. Unfortunately, the signal/noise ratio obtained for the above-mentioned residues was too low to perform a quantitative analysis. Only S73 yielded data of reasonable quality (Fig. 4 *c*) that confirm the presence of a conformational exchange process in the "off" state that leads to a modulation of the chemical shift during the relaxation period, whereas no relaxation dispersion is observed for this residue in the "on" state. Because of the limited experimental sensitivity, we did not attempt to extract kinetic (k_{ex}) or thermodynamic (state populations) information from these dispersion curves.

Fig. 5 summarizes the main changes in NMR observables after light illumination. The following NMR data have been color coded on the structure of rsFolder in an open-sheet and a 3D representation: 1) unobserved amide resonances in both the "on" and "off" states (*dark blue*), 2) amide resonances undetected in the "off" state only (*blue*), 3) amide resonances with significantly reduced peak intensity in the "off" state as extracted from the graph of Fig. S7 *b* and/or decreased ^{15}N T_2 values (*light blue*), and 4) significant ^{1}H-^{15}N chemical shift changes ($\Delta_{CS} > 2$ ppm) as extracted from the graph of Fig. S7 *a* (*orange*). Note that the reported chemical shift changes are not a direct evidence of changes in structure or conformational dynamics because part of it may be due to long-range ring-current-induced chemical shift effects that follow from the change in orientation of the chromophore. Interestingly, all of these NMR data point toward significant changes in conformational dynamics in a particular β-barrel region. The missing "on"-state peaks for a few residues in β7 (V151) and β 8 (K167, I168) indicate some possible dynamics already present in the fluorescent state, highlighting a well-known defect in the β-barrel structure and a hotspot for backbone fluctuations (16,42,43). After the *cis-trans* isomerization of the chromophore, the

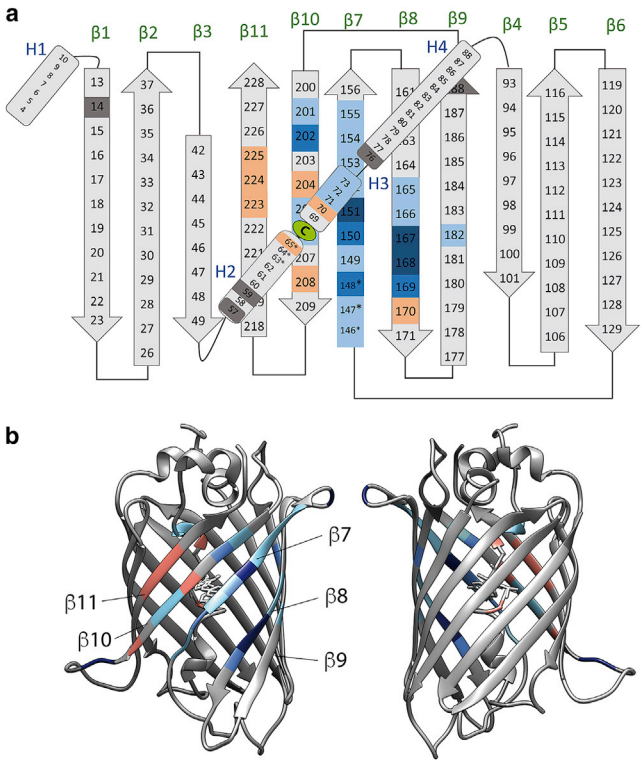


FIGURE 5 Increased conformational dynamics on the microsecond-millisecond timescale in the "off" state of rsFolder. The NMR-derived information is color coded on the (*a*) open-sheet and (*b*) three-dimensional (3D) representation of the rsFolder structure as revealed by x-ray crystallography and NMR (* indicates structural elements only identified by NMR): 1) undetected/unassigned amides in both states (*dark blue*), 2) undetected/unassigned amide peaks in the "off" state only (*middle blue*), 3) strongly line-broadened amide NMR signals in the "off" state (*light blue*), and 4) amide ^{1}H-^{15}N with large chemical shift changes between "on" and "off" states (*orange*). The color coding was performed in a hierarchical order, that is, from 1 to 4 according to the importance of these data to inform about the presence of conformational dynamics. In addition, proline residues lacking an amide proton are color coded in dark gray. To see this figure in color, go online.

conformational dynamics occurring on a microsecond to millisecond timescale are expanded to a larger region close to the phenol ring of the chromophore, comprising strands 7, 8, 10, and eventually 11, as well as helix H3.

Fold stability and water accessibility

Complementary information on the conformational dynamics of backbone amides in the "on" and "off" states of rsFolder has been obtained from H/D exchange measurements. For amide hydrogens engaged in an intramolecular H-bond, H/D exchange is assumed to proceed via transient opening of the hydrogen bond followed by interaction with a solvent exchange catalyst (44). Therefore, NMR-derived H/D exchange rates provide an indirect measure of the dynamic fluctuations of the protein backbone leading to transient breakage of hydrogen bonds. At neutral or acidic

pH, the formation of the H-bond after transient opening is generally much faster than the intrinsic chemical exchange process. In this so-called EX2 exchange regime, the exchange rate constants provide a measure of the exchange-competent state population. A second structural parameter that needs to be considered is water accessibility. Typically, an amide hydrogen surrounded by water has a higher exchange probability than the same amide hydrogen buried in the interior of a globular protein. For rsFolder, the available crystal structures for the "on" and "off" states do not indicate any major change in water-accessible cavities inside the β-barrel as illustrated in Fig. S8. Therefore, we consider that changes in the observed H/D exchange rates are mainly due to a population increase (or decrease) of exchange-competent, unprotected conformational substates.

H/D exchange kinetics for hydrogen-bonded amides of rsFolder were measured by time-resolved 2D ^{1}H-^{15}N NMR. The exchange process was initiated by dissolving the lyophilized protein in D_2O outside the NMR instrument. Typically, unprotected solvent-accessible amide sites were fully exchanged during the experimental dead time of ~5 min. The exchange process was then monitored for ~20 h, allowing one to quantify exchange processes with time constants $T_{ex} = 1/k_{ex}$ spanning a range of three orders of magnitude (5 min $< T_{ex} <$ 5000 min). The measured exchange time constants are color coded (logarithmic scale) on the secondary structure (open-sheet representation) of rsFolder in Fig. S9 for the "on" and "off" states. Examples of intensity decays extracted from the real-time NMR data for individual amide sites in the "on" and "off" states are shown in Fig. 6 *a*. A first observation from these H/D exchange data is that, although overall the amide protons in the β-barrel of the "on" state are well protected from solvent exchange, strand β7 is much less protected because of the lack in hydrogen bonding with the adjacent β8 strand. In the "off" state, a significant increase in the measured solvent exchange rates is detected for amides in the neighboring strands β8 and β10, further emphasizing the increased dynamics in this part of the β-barrel structure upon light illumination, leading to a destabilization of the hydrogen bonding network. Some local destabilization is also observed for amide hydrogens in strands β4 and β5, whereas no significant changes are observed for the remainder of the β-barrel (β1–3, β6, β11) as illustrated in Figs. 6 *b* and S10. The most pronounced changes in protection from solvent exchange, however, are observed for the two helices (H2 and H3) flanking the chromophore in the interior of the β-barrel, with a decrease in solvent protection ranging from a factor of 5 up to a factor of 100. These data clearly indicate the presence of conformational dynamics affecting these helical elements induced by the chromophore *cis-trans* isomerization. Interestingly, although ^{15}N relaxation and line broadening data already identified increased dynamics in helix H3, they did not point toward changes in helix H2 dynamics. Our findings highlight the complementarity of NMR relaxation and H/D exchange data in providing information on local protein dynamics, the latter being sensitive to much longer timescales and smaller chemical shift changes.

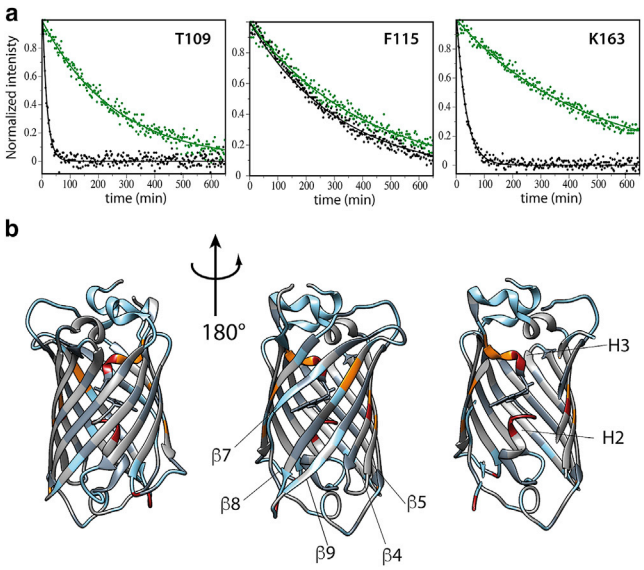


FIGURE 6 H/D exchange NMR data. (*a*) Representative intensity decay for individual amide hydrogens extracted from real-time 2D NMR data recorded for the "on" (*green*) and "off" (*black*) states of rsFolder at 40°C and 850 MHz ^{1}H frequency. (*b*) The derived H/D exchange information is color coded on the crystal structure of rsFolder: 1) residues with amide hydrogen exchange time constants $T_{ex} < 5$ min in both the "on" and off states (*light blue*), 2) residues for which the hydrogen exchange rate is increased by a factor of 4–20 (*orange*), 3) residues for which the hydrogen exchange rate is increased by more than a factor of 20 (*red*), 4) residues without a significant change in exchange rate between "on" and "off" states (*gray*), and 5) no information due to peak overlap, proline, or unassigned residues in one of the two states (*dark gray*). The same data are plotted on an open-sheet representation of rsFolder in Fig. S10. To see this figure in color, go online.

Side-chain and chromophore dynamics in the "on" and "off" states

Line broadening effects similar to those already discussed in the context of backbone dynamics (Fig. 4 *b*) are also observed in the ^{1}H-^{13}C correlation spectra of rsFolder in the "off" state, indicating that a number of protein side chains undergo conformational exchange. Examples of methyl groups (I168 and T63) and an aromatic side chain (H149) that become severely line broadened in the "off" state are shown in Fig. 7 *a*. To identify the protein regions that are involved in conformational side-chain dynamics, we have performed ^{1}H-^{13}C resonance assignment of methyl and aromatic side-chain moieties in the "on" and "off" states of rsFolder as further explained in the Materials and Methods. Fig. S11 highlights ^{1}H-^{13}C spectral regions displaying NMR signals that experience either large chemical shift changes or severe line broadening in the "off" state.

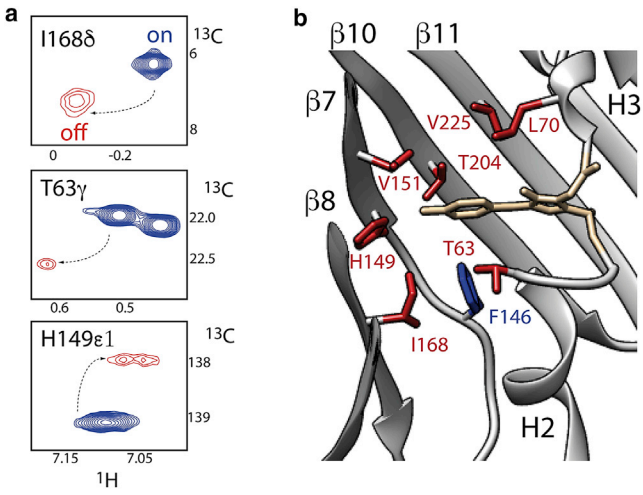


FIGURE 7 rsFolder side chains involved in conformational exchange processes in the "off" state as deduced from extensive line broadening. (*a*) Examples of peak line broadening observed in the ^{1}H-^{13}C CT-BEST-HSQC correlation spectra of rsFolder: "on" state (*blue*) and "off" state (*red*). (*b*) Zoom on the chromophore environment highlighting the side chains, which show NMR line broadening in the "off" state (*red*). In addition, F146 is plotted as it experiences large chemical shift changes upon light illumination. To see this figure in color, go online.

The side chains involved in conformational dynamics are all located in close proximity to the chromophore (Fig. 7 *b*), attached to β-strands 7, 8, 10, and 11 or helices 2 and 3, the same backbone regions that have been shown to experience conformational exchange in the "off" state. These data further support the conclusion that the chromophore itself and/or the surrounding environment experience significant conformational exchange dynamics in the "off" state leading to chemical shift modulation on the microsecond to millisecond timescale.

NMR also provides information on the dynamics of the chromophore and its stabilization in the β-barrel by hydrogen bonding. The exceptional thermal stability of rsFolder in the "off" state was tentatively attributed to the H-bond that H149 maintains with the phenol ring of the chromophore in both "on" and "off" states (26), as illustrated in Fig. 8 *a*. To validate or contradict this hypothesis, we first performed NMR experiments that reveal the protonation state of H149. In the "on" state, the imidazole ring of H149 is predominantly in a neutral π-tautomeric state, protonated at the Nδ_1 position (31). In the "off" state, as discussed in the previous paragraph, the ^{1}H-^{13}C correlation peaks of H149 (Hδ_1-Cδ_1 and Hε_2-Cε_2) are severely line broadened, and no correlation peaks are detected in aromatic HCC-type experiments. Therefore, no conclusions can be drawn on the protonation state of H149 in the "off" state. Further information on the involvement of H149 in hydrogen bonding interactions was obtained by a ^{1}H-^{15}N SOFAST-HMQC experiment (34) tailored to the chemical shifts of histidine side chains. The presence of a hydrogen bond protects the "labile" imidazole hydrogen(s) from solvent exchange. Therefore, a correlation peak for the protonated histidine ring nitrogens is expected if the lifetime of the hydrogen is at least a few hundred milliseconds. Fig. 8 *b* shows such correlation spectra recorded for the "on" and "off" states at different temperatures (15, 25, and 40°C). At low temperature (15°C), a total of 6 out of the 10 histidines of rsFolder give rise to a crosspeak in the "on" state indicating their involvement in a hydrogen bond. Whereas four of them (H82, H182, H200, and H218) remain visible also at higher temperature (40°C), no correlation peak is detected for H149 and H232 at this temperature. Taking H218 as an internal reference, the relative H-bond stability (peak intensity) decreases by a factor of 2 when increasing the temperature from 15 to 25°C. Our NMR data validate the crystal-structure-derived hypothesis of H-bond formation between the Nδ_1 of H149 and the phenol ring of the chromophore. However, we can also conclude that this H-bond is relatively unstable as compared to other H-bonds involving histidines H82, H182, H200, and H218. In the "off" state, the same peak patterns and T-dependence is observed, except for in H149, which does not give rise to a correlation peak at any temperature. This observation, however, does not allow one to conclude the absence of an H-bond between H149 and the chromophore in the "off" state because the protonated phenol ring has to interact with an unprotonated imidazole ring nitrogen to form an H-bond, which is undetectable in our NMR spectra. Nevertheless, the observed line broadening of the nonexchangeable ring hydrogens of H149, resulting from conformational exchange dynamics, makes the hypothesis of a stable H-bond in the "off" state very unlikely.

We also attempted to obtain some information on the NMR signature of the chromophore itself. Fig. 8 *d* shows the aromatic region of the ^{1}H-^{13}C correlation spectra of rsFolder in the "on" state (*blue spectrum*). The chromophore contains five C-H pairs (Fig. 8 *c*) in the tyrosine part of the chromophore that is made up from the A66-Y67-G68 tripeptide in the rsFolder sequence. The five corresponding correlation peaks detected in the spectrum of Fig. 8 *d* were identified and unambiguously assigned by a set of 2D correlation spectra shown in Fig. S12, recorded with pulse sequences that were specifically tailored for tyrosine side chains (31). These NMR data allowed us to assign all ^{1}H and ^{13}C resonances (except for CO) of the tyrosine moiety of the chromophore (Table 1). The CZ chemical shift of 179.2 ppm is characteristic of an anionic phenol group as expected for a deprotonated chromophore in the fluorescent "on" state. The observation of distinct correlation peaks for the symmetric CD and CE sites of the phenol ring indicates that the chromophore is structurally stabilized in the β-barrel with slow ring flip dynamics ($k_{ex} < 500\ s^{-1}$) at 40°C. These slow ring flips are in contrast to all other tyrosine and phenylalanine side chains of rsFolder for which a single correlation peak was detected for the symmetric δ1 and δ2 (ε1 and ε2) sites, indicative of fast ring flip dynamics

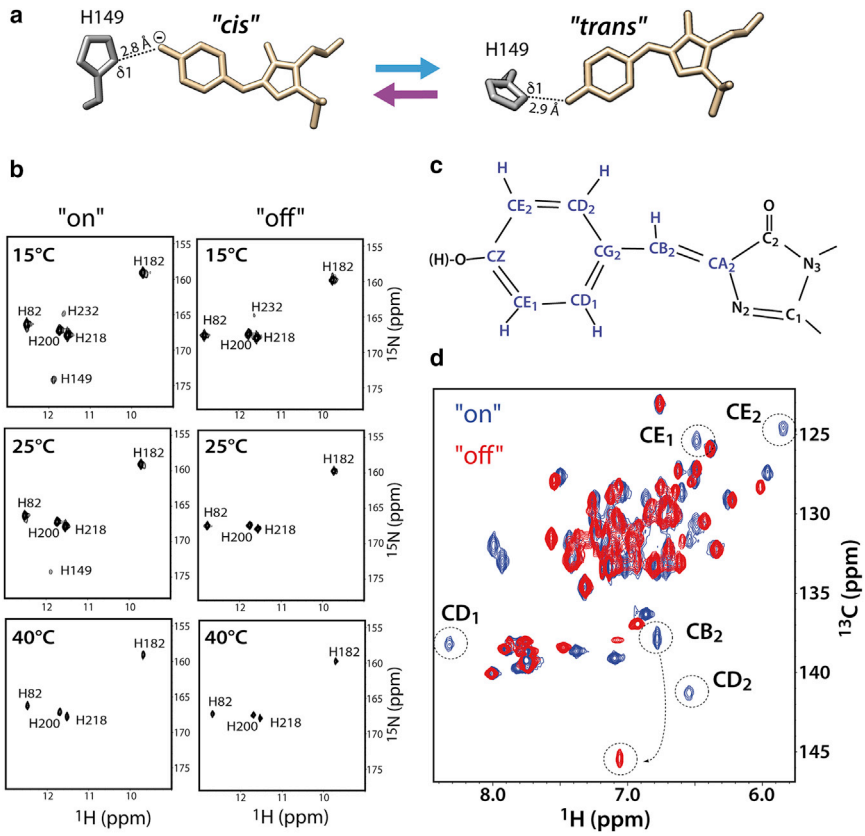


FIGURE 8 NMR information of the chromophore. (*a*) Chromophore and His149 side-chain conformations extracted from the available crystal structures of rsFolder in the "on" (*cis* configuration) and "off" (*trans* configuration) states. (*b*) A temperature series of ^{1}H-^{15}N SOFAST-HMQC spectra of histidine side chains in the "on" and the "off" states recorded at 850 MHz ^{1}H frequency. The detected crosspeaks are annotated by the residue number. (*c*) A schematic representation of the chemical structure of the chromophore. Nuclei giving rise to NMR signals are highlighted by blue letters, and their NMR frequencies are provided in Table 1. (*d*) Overlay of 2D aromatic ^{1}H-^{13}C BEST-HSQC spectra (31) of rsFolder, recorded without CT ^{13}C editing to achieve highest experimental sensitivity, in the "on" (*blue*) and "off" (*red*) state. NMR signals from the tyrosine moiety (Y67) of the chromophore are annotated. To see this figure in color, go online.

leading to chemical shift averaging. In the nonfluorescent "off" state (Fig. 8 *d*, *red spectrum*), the four correlation peaks of the chromophore's phenol group disappear, indicative of a significant change in chromophore dynamics. Only the signal of the HB-CB in the methine bridge of the chromophore remains detectable in the dark state, with the CB resonance experiencing an impressive 7.6 ppm shift after *cis-trans* isomerization of the chromophore. The NMR assignment of the chromophore in the "off" state could be further extended to the CA and CG (Fig. S12 *b*) that also shift by 3.3 and 4.4 ppm, respectively (see Table 1). Our NMR data indicate a destabilization of the chromophore in the "off" state leading to motion of the phenol ring on the millisecond timescale. However, whether these motions correspond to increased ring flip dynamics, local fluctuations around the methine bridge, or both cannot be distinguished from these data.

DISCUSSION

In situ NMR illumination and time-efficient NMR experiments

Our portable laser illumination device allowed us to perform a large set of 2D and 3D correlation experiments on a single rsFolder sample. Because of its portability, the setup is compatible with any high-field NMR instrument and probe, provided that the optical fiber is long enough for the sample to reach its position in the magnet. Illumination from the top (fiber placed in the Shigemi plunger) ensures minimal perturbation of the magnetic field homogeneity and thus results in close-to-optimal NMR line shapes. However, relatively high laser power is required to ensure fast "on"-to-"off" (or "off"-to-"on") switching in particular because of nonuniform illumination and high optical density of the sample. The LD-EXSY experiment that makes use of fast switching capabilities provides a convenient and time-efficient way to transfer NMR assignments of fast-switching PTFPs from one conformational state to another. This is of importance because NMR resonance assignment is the most time-consuming step for investigating the dynamic properties of PTFPs. Once a sufficient level of resonance assignment has been achieved, NMR provides a unique tool to access local dynamics in the protein backbone, side chains, and chromophore on a broad range of timescales, as will be discussed further in the following paragraph(s).

Conformational dynamics in the fluorescent state of rsFolder

The β-barrel scaffold of fluorescent proteins stabilizes the anionic *cis* conformation of the chromophore via hydrogen

TABLE 1 Chromophore ^{1}H and ^{13}C Chemical Shifts

Y67 NMR Shifts (ppm)	On State	Off State
CA	131.1	134.4
CB (HB)	137.9 (6.79)	145.5 (7.05)
CG	121.2	125.6
CD1 (HD1)	138.2 (8.34)	n.d.
CD2 (HD2)	141.3 (6.54)	n.d.
CE1 (HE1)	125.5 (6.48)	n.d.
CE2 (HE2)	124.6 (5.84)	n.d.
CZ	179.2	n.d.

n.d., not detected.

bonding, π-stacking, and cation-π interactions, as well as steric hindrance. This restricts the chromophore's internal motion observed in free solution and prevents radiationless decay upon light excitation, a prerequisite for a high-fluorescence quantum yield. However, this conformational rigidity is somehow in contradiction with the capability of RSFPs to switch efficiently between distinct conformations, implying local structural rearrangements, as well as changes in the hydrogen bonding network upon light excitation. Therefore, a detailed experimental characterization of the structural heterogeneity and conformational flexibility of the chromophore and its protein environment is important for understanding and improving the photochromic properties of RSFPs. Such dynamic information is typically not available from crystallographic structures. In the case of rsFolder, the available x-ray structures show a well-defined β-barrel fold and a single chromophore conformation in both the "on" and "off" states. Here, we have applied solution NMR spectroscopy to complement this static picture. Information on the local dynamics over timescales ranging from picoseconds to seconds has been obtained from a combination of ^{15}N spin relaxation measurements, NMR line width analysis, and H/D exchange data. Our NMR data point toward a well-known defect in the β-barrel structure of GFP-derived fluorescent proteins around strand β7 (16,18,42,45) where the H-bonds formed with the neighboring strands, typical of an antiparallel β-sheet conformation, are weakened by longer distances and other structural irregularities. This leads to backbone fluctuations and transient opening of the β-barrel in the fluorescent "on" state of rsFolder as evidenced by a few unobservable amide resonances (V151, K167, and I168) and fast H/D exchange rates measured for amide hydrogens in strand β7. Similar observations have been made previously for conventional GFPs (16,18). The dynamic "holes" in the β-barrel between β7 and the neighboring strands have also been identified by molecular dynamics simulations as major entry points of water molecules into the interior of the β-barrel (43). Although certainly not a sufficient condition, such conformational plasticity is likely required for efficient photoswitching.

The data show clear evidence of an anionic chromophore, with the phenolate oxygen hydrogen bonded to the Nδ1 of H149. This finding is in agreement with the crystal structure (26) that indicates that the chromophore's phenolate oxygen is involved in three hydrogen bond interactions with H149, T204, and a water molecule. However, our NMR data show that the H-bond between the chromophore and H149 is thermally unstable. At physiological temperature, the bond is constantly broken and reformed on a submillisecond timescale. It has been shown by QM/MM simulations of the RSFP Dronpa (46) that two or three H-bonds stabilizing the chromophore are required for high-fluorescence quantum yields, whereas less hydrogen bonding favors photoswitching of the chromophore by a hula-twist mechanism. We speculate that the weak H-bond observed with H149 in rsFolder is a characteristic property of RSFPs that have been engineered to obtain reasonable fluorescence quantum yields combined with efficient switching capability.

Light-induced conformational dynamics in the chromophore and its environment

As for a majority of RSFPs, on-off photoswitching in rsFolder involves a *cis-trans* isomerization of the chromophore coupled to deprotonation at the hydroxy group of the phenol moiety. In our NMR spectra, this photoinduced isomerization results in large ^{13}C shifts of carbons in the methine bridge (CA, CB, and CG) of the chromophore. Because ^{13}C chemical shifts are particularly sensitive to the local geometry (torsion angles) of the chemical bonds (47), the methine bridge ^{13}C chemical shifts can be considered a spectral signature of *p*-HBI being in *cis* or *trans* configuration. In principle, the protonation state at the chromophore's phenol moiety can be inferred from the $^{13}C_Z$ chemical shift, which is expected to be 175–180 ppm for a phenolate and 155–160 ppm for a phenol. However, for rsFolder in the "off" state, no NMR signals could be detected for the phenol moiety in the aromatic ^{1}H-^{13}C spectra, indicative of a chromophore undergoing conformational exchange dynamics on the microsecond to millisecond timescale. Similarly, NMR signals of nuclei in the protein side chains H149 and T204, forming hydrogen bonds with the phenolate oxygen in the "on" state, become strongly line broadened in the "off" state. These NMR observations clearly indicate a destabilization of the chromophore and the absence of any stable hydrogen bonding interaction with the β-barrel. In the crystal structures of rsFolder, the observed distances between the histidine δ1 nitrogen and the phenol oxygen are only little altered (from 2.8 to 2.9 Å) upon photoisomerization of the chromophore (Fig. 7). Our findings thus contradict the conclusion drawn from these structures that H149 remains hydrogen bonded to the hydroxyl group of the phenol in the "off" state. Most likely, the crystal structure observed at cryogenic temperatures represents the lowest energy conformation in the "off" state, which at higher temperature is in exchange with alternate conformations that are separated by low energy barriers. A loss of hydrogen bonding of the

chromophore phenolate with the β-barrel scaffold has also been observed for other RSFPs (e.g., Dronpa, rsEGFP2, and rsGreen) (8,46,48,49). Of note, this has also been observed recently in the case of a long-lived dark state in the photoconvertible fluorescent protein mEos4b (7).

The chromophore *cis-trans* isomerization also induces increased structural dynamics at the β-barrel side (β7–β10) facing the chromophore's phenol group, the two helical regions (H2 and H3) connecting the chromophore with the rest of the protein, and a number of side chains attached to these peptide regions, and pointing toward the chromophore. In particular, all side chains that undergo a structural rearrangement in the "off" state (T63, F146, H149, V151, and I168), as seen by x-ray crystallography (Fig. 1), are heavily line broadened in our NMR data (Fig. 6), except for F146, which shows large chemical shift changes. These observations further consolidate the hypothesis that photoisomerization in rsFolder leads to a polymorphic and highly dynamic chromophore environment with multiple substates (conformations) of comparable free energy that interconvert on the micro-to-millisecond timescale. Most likely, the structural fluctuations of the chromophore and surrounding side chains also affect the backbone regions to which they are attached, thus explaining the observed amide ^{1}H line broadening and increased solvent exchange rates. These data are fully in line with the notion of a fully nonfluorescent dark state unable to undergo radiative deexcitation upon photon absorption. Similar findings on changes in backbone dynamics upon photoswitching have previously been reported for the anthozoan-derived RSFP Dronpa (14).

Concerning the helical region H3, we know that L70 is an important residue for the photoswitching properties of rsEGFP2 (24) and rsFolder (26), whereas S73 has been suggested as one of the "hopping" points of protons when entering from the solvent onto their way to the chromophore (43). In our NMR data, the H3 region of rsFolder shows increased conformational dynamics in the "off" state, as evidenced by line broadening (C71, F72, S73) and faster H/D exchange (L70, C71). It has also been reported recently using time-resolved serial crystallography at an x-ray free-electron laser that during photoswitching in rsEGFP2, an intermediate state becomes populated, in which the chromophore is twisted, and helix H2 is moved downward along the β-barrel axis (8). Here, we have shown that water exchange for several amide hydrogens (V62, T63, and T64) in H2 is increased by up to two orders of magnitude. We may therefore speculate that conformations similar to those seen in the excited state may be also transiently occupied under equilibrium conditions at ambient temperature, possibly involved in the thermally induced relaxation mechanism of rsFolder from the "off" state to the "on" state. So far, the chromophore connecting helical elements, H2 and H3, have found only little attention in the field of GFP engineering. In the future, it will be interesting to investigate the effect of point mutations in H2 or H3 on the photophysical properties.

CONCLUSIONS

In this work, we have used multidimensional solution NMR spectroscopy coupled with in situ sample illumination as a complementary high-resolution tool to study light-induced changes in the conformational dynamics of the PTFP rsFolder. Our results add a dynamic dimension to the static view provided by x-ray crystallographic protein structures. In particular, we identified dynamic hotspots in the chromophore environment and the β-barrel structure that may suggest future mutation sites to be probed by, for example, saturation mutagenesis to alter the photophysical properties of RSFPs. Furthermore, NMR spectroscopy provides a unique tool to investigate the influence of environmental conditions on the conformational dynamics of PTFPs and correlate them with their photophysical and photochemical properties. Therefore, including NMR spectroscopy in the toolbox used for rational fluorescent protein engineering will provide new opportunities for further improving advanced fluorescent markers for a wide range of microscopy and biotechnological applications.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2019.10.035.

AUTHOR CONTRIBUTIONS

D.B. and B.B. conceived and directed the research. I.A. and K.G.-B. expressed the protein and prepared the samples. N.-E.C. and B.B. performed the NMR experiments and analyzed the data. V.A. and M.B. contributed plasmids and helped with the experimental setup. N.-E.C., D.B., and B.B. wrote the article. All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Dr. Paul Schanda for many stimulating discussions. This work used the NMR and isotope labeling platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) within the Grenoble Partnership for Structural Biology. Institut de Biologie Structurale acknowledges integration into the Interdisciplinary Research Institute of Grenoble (CEA).

Financial support from the CNRS (Défis Instrumentation 2018) is acknowledged. Platform access was supported by FRISBI (ANR-10-INBS-05-02) and GRAL, a project of the Université Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003).

REFERENCES

1. Schermelleh, L., A. Ferrand, …, G. P. C. Drummen. 2019. Super-resolution microscopy demystified. *Nat. Cell Biol.* 21:72–84.

2. Adam, V., H. Mizuno, …, J. Hofkens. 2010. Data storage based on photochromic and photoconvertible fluorescent proteins. *J. Biotechnol.* 149:289–298.
3. Betzig, E., G. H. Patterson, …, H. F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science.* 313:1642–1645.
4. Hess, S. T., T. P. Girirajan, and M. D. Mason. 2006. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* 91:4258–4272.
5. Endesfelder, U., S. Malkusch, …, M. Heilemann. 2011. Chemically induced photoswitching of fluorescent probes–a general concept for super-resolution microscopy. *Molecules.* 16:3106–3118.
6. Adam, V., P. Carpentier, …, D. Bourgeois. 2009. Structural basis of X-ray-induced transient photobleaching in a photoactivatable green fluorescent protein. *J. Am. Chem. Soc.* 131:18063–18065.
7. De Zitter, E., D. Thédié, …, D. Bourgeois. 2019. Mechanistic investigation of mEos4b reveals a strategy to reduce track interruptions in sptPALM. *Nat. Methods.* 16:707–710.
8. Coquelle, N., M. Sliwa, …, M. Weik. 2018. Chromophore twisting in the excited state of a photoswitchable fluorescent protein captured by time-resolved serial femtosecond crystallography. *Nat. Chem.* 10:31–37.
9. Duan, C., V. Adam, …, D. Bourgeois. 2014. Structural basis of photoswitching in fluorescent proteins. *In* Photoswitching Proteins: Methods and Protocols, Methods in Molecular Biology. S. Cambridge, ed. Humana Press, pp. 177–202.
10. Avilov, S., R. Berardozzi, …, D. Bourgeois. 2014. In cellulo evaluation of phototransformation quantum yields in fluorescent proteins used as markers for single-molecule localization microscopy. *PLoS One.* 9:e98362.
11. Cloin, B. M. C., E. De Zitter, …, L. C. Kapitein. 2017. Efficient switching of mCherry fluorescence using chemical caging. *Proc. Natl. Acad. Sci. USA.* 114:7013–7018.
12. Duwé, S., and P. Dedecker. 2019. Optimizing the fluorescent protein toolbox and its use. *Curr. Opin. Biotechnol.* 58:183–191.
13. Kao, Y. T., X. Zhu, and W. Min. 2012. Protein-flexibility mediated coupling between photoswitching kinetics and surrounding viscosity of a photochromic fluorescent protein. *Proc. Natl. Acad. Sci. USA.* 109:3220–3225.
14. Mizuno, H., T. K. Mal, …, A. Miyawaki. 2008. Light-dependent regulation of structural flexibility in a photochromic fluorescent protein. *Proc. Natl. Acad. Sci. USA.* 105:9227–9232.
15. Mizuno, H., T. K. Mal, …, A. Miyawaki. 2010. Molecular basis of photochromism of a fluorescent protein revealed by direct 13C detection under laser illumination. *J. Biomol. NMR.* 48:237–246.
16. Seifert, M. H., J. Georgescu, …, T. A. Holak. 2003. Backbone dynamics of green fluorescent protein and the effect of histidine 148 substitution. *Biochemistry.* 42:2500–2512.
17. Seifert, M. H., D. Ksiazek, …, T. A. Holak. 2002. Slow exchange in the chromophore of a green fluorescent protein variant. *J. Am. Chem. Soc.* 124:7932–7942.
18. Oltrogge, L. M., Q. Wang, and S. G. Boxer. 2014. Ground-state proton transfer kinetics in green fluorescent protein. *Biochemistry.* 53:5947–5957.
19. Grotjohann, T., I. Testa, …, S. W. Hell. 2011. Diffraction-unlimited all-optical imaging and writing with a photochromic GFP. *Nature.* 478:204–208.
20. Rego, E. H., L. Shao, …, M. G. Gustafsson. 2012. Nonlinear structured-illumination microscopy with a photoswitchable protein reveals cellular structures at 50-nm resolution. *Proc. Natl. Acad. Sci. USA.* 109:E135–E143.
21. Dedecker, P., G. C. Mo, …, J. Zhang. 2012. Widely accessible method for superresolution fluorescence imaging of living systems. *Proc. Natl. Acad. Sci. USA.* 109:10909–10914.
22. Marriott, G., S. Mao, …, Y. Yan. 2008. Optical lock-in detection imaging microscopy for contrast-enhanced imaging in living cells. *Proc. Natl. Acad. Sci. USA.* 105:17789–17794.
23. Querard, J., T. Z. Markus, …, L. Jullien. 2015. Photoswitching kinetics and phase-sensitive detection add discriminative dimensions for selective fluorescence imaging. *Angew. Chem. Int.Engl.* 54:2633–2637.
24. Grotjohann, T., I. Testa, …, S. Jakobs. 2012. rsEGFP2 enables fast RESOLFT nanoscopy of living cells. *eLife.* 1:e00248.
25. Pédelacq, J. D., S. Cabantous, …, G. S. Waldo. 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* 24:79–88.
26. El Khatib, M., A. Martins, …, V. Adam. 2016. Rational design of ultra-stable and reversibly photoswitchable fluorescent proteins for super-resolution imaging of the bacterial periplasm. *Sci. Rep.* 6:18459.
27. Favier, A., and B. Brutscher. 2019. NMRlib: user-friendly pulse sequence tools for Bruker NMR spectrometers. *J. Biomol. NMR.* 73:199–211.
28. Vranken, W. F., W. Boucher, …, E. D. Laue. 2005. The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins.* 59:687–696.
29. Favier, A., and B. Brutscher. 2011. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR.* 49:9–15.
30. Solyom, Z., M. Schwarten, …, B. Brutscher. 2013. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J. Biomol. NMR.* 55:311–321.
31. Christou, N. E., and B. Brutscher. 2018. BEST and SOFAST experiments for resonance assignment of histidine and tyrosine side chains in $^{13}C/^{15}N$ labeled proteins. *J. Biomol. NMR.* 72:115–124.
32. Montelione, G. T., B. A. Lyons, …, M. Tashiro. 1992. An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. *J. Am. Chem. Soc.* 114:10974–10975.
33. Grzesiek, S., J. Anglister, and A. Bax. 1993. Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. *J. Magn. Reson. Ser. B.* 101:114–119.
34. Schanda, P., E. Kupce, and B. Brutscher. 2005. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR.* 33:199–211.
35. Van Melckebeke, H., J. P. Simorre, and B. Brutscher. 2004. Amino acid-type edited NMR experiments for methyl-methyl distance measurement in ^{13}C-labeled proteins. *J. Am. Chem. Soc.* 126:9584–9591.
36. Farrow, N. A., R. Muhandiram, …, L. E. Kay. 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry.* 33:5984–6003.
37. Fushman, D., R. Weisemann, …, H. Rüterjans. 1994. Backbone dynamics of ribonuclease T1 and its complex with 2'GMP studied by two-dimensional heteronuclear NMR spectroscopy. *J. Biomol. NMR.* 4:61–78.
38. Franco, R., S. Gil-Caballero, …, B. Brutscher. 2017. Probing conformational exchange dynamics in a short-lived protein folding intermediate by real-time relaxation-dispersion NMR. *J. Am. Chem. Soc.* 139:1065–1068.
39. Pouwels, P. J., and R. Kaptein. 1993. Diagonal-peak suppression in the SCOTCH experiment, a 2D NMR technique for studying photochemical reactions. *J. Magn. Reson. Ser. A.* 101:337–341.
40. Shen, Y., and A. Bax. 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR.* 56:227–241.
41. Arpino, J. A., P. J. Rizkallah, and D. D. Jones. 2012. Crystal structure of enhanced green fluorescent protein to 1.35 Å resolution reveals alternative conformations for Glu222. *PLoS One.* 7:e47132.
42. Lelimousin, M., M. Noirclerc-Savoye, …, A. Royant. 2009. Intrinsic dynamics in ECFP and Cerulean control fluorescence quantum yield. *Biochemistry.* 48:10038–10046.
43. Shinobu, A., and N. Agmon. 2015. The hole in the barrel: water exchange at the GFP chromophore. *J. Phys. Chem. B.* 119:3464–3478.
44. Dempsey, C. E. 2001. Hydrogen exchange in peptides and proteins using NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* 39:135–170.

45. Wachter, R. M., M. A. Elsliger, …, S. J. Remington. 1998. Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. *Structure.* 6:1267–1277.

46. Morozov, D., and G. Groenhof. 2016. Hydrogen bond fluctuations control photochromism in a reversibly photo-switchable fluorescent protein. *Angew. Chem. Int.Engl.* 55:576–578.

47. de Dios, A. C., J. G. Pearson, and E. Oldfield. 1993. Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach. *Science.* 260:1491–1496.

48. Duwé, S., E. De Zitter, …, P. Dedecker. 2015. Expression-enhanced fluorescent proteins based on enhanced green fluorescent protein for super-resolution microscopy. *ACS Nano.* 9:9528–9541.

49. Andresen, M., A. C. Stiel, …, S. Jakobs. 2007. Structural basis for reversible photoswitching in Dronpa. *Proc. Natl. Acad. Sci. USA.* 104:13005–13009.

50. Vallurupalli, P., G. Bouvignies, and L. E. Kay. 2012. Studying "invisible" excited protein states in slow exchange with a major state conformation. *J. Am. Chem. Soc.* 134:8148–8161.