Conformational, hydration, and protonation dynamics of cytochrome *c* oxidase and a tracking-free analysis method for diffusion in fluorescence microscopy

A dissertation

Submitted in Partial Fulfilment of the

Requirements for the Degree of Dr. rer. nat.

to the Department of Physics

of Freie Universität Berlin

by

Alexander Wolf

Berlin, 2019

Supervisor: Prof. Dr. Ulrike Alexiev (Fachbereich Physik, Freie Universität Berlin)

Second examiner: Prof. Dr. Karsten Heyne (Fachbereich Physik, Freie Universität Berlin)

Date of defense: August 17, 2020

Abstract

As the terminal enzyme in the respiratory chain, cytochrome *c* oxidase (C*c*O) facilitates the maintaining of a transmembrane proton gradient by converting molecular oxygen to water and pumping protons across the membrane. This proton gradient is used by the enzyme ATP synthase to produce the energy storage molecule ATP. Among the most pressing open questions regarding C*c*O is how the redox state of its catalytic center controls proton uptake through one of its proton channels, the K-channel, and the location of proton release on the opposite side of the membrane.

To elucidate these questions, an expression system was developed in Part I of this thesis enabling the background-free fluorescence labeling of *Paracoccus denitrificans* C*c*O, which is homologous to the human enzyme. Subsequent conjugation of C*c*O with environmentally sensitive fluorophores allowed for the investigation of nanoenvironmental parameters on localized sites on the protein surface in different enzymatic states. Steady-state and time-resolved spectroscopy techniques such as time-correlated single-photon counting (TCSPC) allowed for the detection of conformational changes, rearrangements of the local hydrogen bonding network, and changes in the hydration shell around the K-channel entrance, depending on the redox state of the enzyme. The combined results yield a model, in which the catalytic center of C*c*O exhibits nanoenvironmental control over the proton uptake surface above the K-channel. There, locally defined increases in conformational flexibility in water dynamics facilitate conditions putatively beneficial for the uptake of protons, which are subsequently translocated through the K-channel to the enzyme's core.

In the second part of this thesis, a novel tracking-free analysis method was developed for the determination of diffusive modes in single-molecule video microscopy. This method distinguishes itself from single-particle tracking (SPT) based techniques in its generation of step length distributions (SLDs), and thus diffusivities, from particle positions without relying on assumptions about the identity of particles as the movie progresses, are a major source of error in SPT. The herein developed method, Diffusion Analysis of Nanoscopic Ensembles (DANAE), however, generates correct SLDs even when faced with high diffusivities, high particle densities, long exposure times, as well as changes in the detectability of particles due to their photo-state, and changes imposed by ageing effects during the course of the microscopic movie, as shown through Monte Carlo Simulations (MCSs). Finally, a two-color derivation of the method (2cDANAE) is presented, which produces SLDs exclusively in case of co-diffusion between the microscopic species in each channel and can thus be used to determine the interaction of biomolecules with spectrally different fluorescence labels in multichannel single-molecule microscopy experiments.

Zusammenfassung

Die Cytochrom *c* Oxidase (C*c*O) ist das terminale Enzym der Atmungskette, als welches sie den transmembranen Protonengradienten aufrechterhält durch die Umwandlung von molekularem Sauerstoff zu Wasser und durch das Pumpen von Protonen durch die Membran. Dieser Protonengradient wird vom Enzym ATP Synthase dazu verwendet den chemischen Energieträger ATP zu produzieren. Zu den ungeklärtesten Aspekten der C*c*O gehört der Mechanismus über den der Redox-Zustand des katalytischen Zentrums die Aufnahme von Protonen über den K-Kanal, einen der Protonenkanäle der C*c*O, steuert, sowie der Ort des Protonenaustritts.

Zu diesem Zweck wurde in Teil I dieser Arbeit ein Expressionssystem etabliert, das die hintergrundfreie Fluoreszenz-markierung der C*c*O des Bakteriums *Paracoccus denitrificans* ermöglicht. Die dabei eingesetzten umgebungssensitiven Fluorophore ermöglichten die Untersuchung der Nanoumgebung an der N-Seite des K-Kanals in Abhängigkeit vom Redox-Zustand des katalytischen Zentrums der C*c*O. Als Messmethoden dienten dabei fluoreszenzbasierte spektroskopische Verfahren wie die zeitkorrelierte Einzelphotonenzählung. In dem resultierenden Modell übt das katalytische Zentrum der C*c*O Einfluss auf die Nanoumgebung an der Oberfläche des K-Kanals aus, wo eine Erhöhung der Proteinflexibilität und Wassermobilität die Bedingungen für eine effiziente Protonenaufnahme befördern.

Im zweiten Teil dieser Arbeit wurde eine neue Analysemethode für die Bestimmung von Diffusionsmoden in Einzelmolekül-Mikroskopievideos etabliert. Die Besonderheit dieser neuen Methode, Diffusion Analysis of Nanoscopic Ensembles (DANAE), im Vergleich zu Single Particle Tracking (SPT)-basierten Methoden liegt darin, dass die Bestimmung von Schrittlängenverteilungen (engl. step length distributions, SLDs) und damit der Diffusionsparameter der Moleküle gänzlich ohne die Zuweisungen von Teilchenidentitäten erfolgt, einer großen Fehlerquelle bei SPT. Anhand von Monte-Carlo-Simulationen (MCS) wurde demonstriert, dass mit DANAE korrekte SLDs auch bei sehr hohen Diffusionskonstanten, sehr hoher Teilchendichte, geringer zeitlicher Auflösung, sowie Veränderungen in der Beobachtbarkeit der Teilchen durch variablen Photozustand oder Alterungseffekte während des Aufnahmezeitraums generiert werden. Im letzten Abschnitt wird eine Zweikanal-Variante des Algorithmus (2cDANAE) vorgestellt, welche SLDs ausschließlich in Anwesenheit von Kodiffusion der sich in den Mikroskopieaufnahmen der beiden Farbkanäle bewegenden Teilchen generiert. In spektral getrennten Fluoreszenzmikroskopie-Experimenten mit unterschiedlich gelabelten Biomolekülen kann somit Aufschluss über deren potentielle Interaktionen gewonnen werden.

Table of contents

List of important abbreviations

AD	Alzheimer's disease
ADP/ATP	adenosine diphosphate/triphosphate
β-DM	n-dodecyl β-D-maltoside
BADAN	6-bromoacetyl-2-demithylaminonaphtalene
BNC	binuclear center
C*c*O	cytochrome *c* oxidase
DANAE	Diffusion Analysis of Nanoscopic Ensembles
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
ETC	electron transport chain
FAD/$FADH_2$	oxidized/reduced flavin adenine dinucleotide
FMA	fluorescein-5-maleimide
FCS	Fluorescence Correlation Spectroscopy
FOV	field of view
$HICT_m$/$HICT_i$	mobile/immobile hydrogen-bonded ICT state
IAF	5-iodoacetamido-fluorescein
ISC	intersystem crossing
ICT	intramolecular charge transfer
JSA/JDA	jump step/distance analysis
LD	label dynamics
LDAO	lauryldimethylamine oxide
MCS	Monte Carlo simulation
MD	Molecular Dynamics
MSD	mean squared displacement
NAD/NADH	oxidized/reduced nicotinamide adenine dinucleotide
OXPHOS	oxidative phosphorylation
PICS	Particle Image Correlation Spectroscopy
PLS	proton loading site
SLD	step length distribution
SPT	single-particle tracking
SR	solvent relaxation
TAE	Tris-acetate-EDTA buffer
TCSPC	time-correlated single photon counting

TDFS	time-dependent fluorescence shift
TMH	transmembrane helix

Aim of the work

The first part of this thesis focusses on C*c*O, the terminal enzyme of the respiratory chain. In humans, mitochondrial C*c*O deficiency is connected to several severe diseases.[2, 3] While many aspects of this enzyme have been established over the past decades, the mechanism by which the enzyme facilitates selective activation of the proton channel in the K-pathway and the proton exit pathway(s) remain obstacles to completing the model of C*c*O's function.

While numerous computational and theoretical studies have proposed solutions to the selective activation of the K-channel (see e.g. [4-6]), experimental corroboration has been hindered by the absence of large scale conformational changes during the O→R transition[7] and the difficulty of obtaining experimental data conclusively localized in and around the K-channel.[8, 9] One strategy to overcome this obstacle and to obtain experimental insight into the redox-dependent nanoenvironmental conditions in terms of conformational and water dynamics changes around the K-channel entrance is site-directed mutagenesis in conjunction with selective labeling with environmentally sensitive fluorophores as a means to probe localized changes at the proton uptake surface around the K-channel entrance. The same holds true for revealing the proton exit pathway(s). This strategy requires the availability of an expression system for single-cysteine reactive C*c*O mutants. The aim of this work was to establish a homologous expression system allowing for mutagenesis of C*c*O subunits 1-3 to achieve labeling on both the N- and P-side, to characterize the mutants, establish site-specific labeling with environmentally sensitive fluorophores, and finally to measure the redox state-dependent changes at the labeling sites with steady-state and time-resolved spectroscopic techniques.

The second part of the thesis establishes and tests an algorithm that generates step length distributions from single-molecule video microscopy experiments. Several methods exist to extract the diffusive properties of moving particles, many of which are based on SPT[10], i.e. tracing the trajectories of the particles over the span of the movie. While this is possible for sufficiently well separated and slowly moving particles, it becomes increasingly harder the more tracjectories overlap per captured frame of the video. This is due to the fact that SPT based methods need to establish the identity of individual particles in each frame to construct a trace, thus requiring the evaluation of particle features like inter-frame proximity or intensity[11]. These features, however, become increasingly less predictive the more other particles exhibit similar values for them. This puts many constrains on the experimental design (number of particles, localization accuracy vs. temporal resolution), which needs to be optimized in order to reduce the inaccuracies contained in the signal from the assignment of particle identities alone. The aim of the second part of this thesis is therefore to establish an algorithm that does not rely on assigning particle identities and thus circumvents the problems associated with SPT, to test it on simulated microscopy data, and lastly to provide solutions to problems such as boundary effects.

PART I: K-channel dynamics of *Paracoccus denitrificans* cytochrome *c* oxidase

1. Significance and mechanism of cytochrome *c* oxidase

Cytochrome *c* oxidase (C*c*O) is an essential enzyme in the oxidative energy conversion system of eukaryotes, including humans, and many aerobic prokaryotes. The oxygen enrichment of the Earth's atmosphere through photosynthetic oxygenation by oceanic cyanobacteria, which probably evolved during the Archean-Proterozoic transition, correlates with—and likely enabled—the transition from prokaryotic to eukaryotic life[12]. As a part of the electron transport chain (ETC), cytochrome *c* oxidase partakes in the buildup of a transmembrane proton gradient. ATP synthase uses this gradient to produce ATP as a chemical energy storage (section 1.1). In eukaryotes, C*c*O is so essential that even slight abnormalities in its function or assembly can lead to a variety of severe metabolic diseases and death (section 1.2). C*c*O is present in the inner mitochondrial membrane of eukaryotes and has high structural and functional homology to bacterial C*c*O, where it resides in the inner membrane (section 1.3). Due to the large degree of homology throughout the domains of life and its less complex structure, bacterial C*c*O is well suited to study the function and catalysis of its human homologue (section 1.4).

1.1. Oxidative phosphorylation and the electron transport chain

Oxidative phosphorylation (OXPHOS) is a highly efficient metabolic pathway converting energy by oxidizing donor molecules to phosphorylate adenosine diphosphate (ADP) to its triphosphate form (ATP):[13]

$$\mathrm{ADP} + \mathrm{P_i} + \mathrm{H^+} \rightarrow \mathrm{ATP} + \mathrm{H_2O} \qquad \Delta G^0 = +30.5\ \mathrm{kJ\ mol^{-1}}\ .$$

Enzymes can later use the chemical energy stored in ATP to catalyze a multitude of endergonic reactions and thereby facilitate cellular homeostasis and proliferation. As can be seen in the reaction scheme, the synthesis of ATP requires energy itself. This energy is provided by the proton motive force (PMF), a transmembrane proton gradient generated in the electron transport chain (ETC). The PMF is the sum of both the chemical gradient (ΔpH) and the electrical gradient ($\Delta\Psi$). This

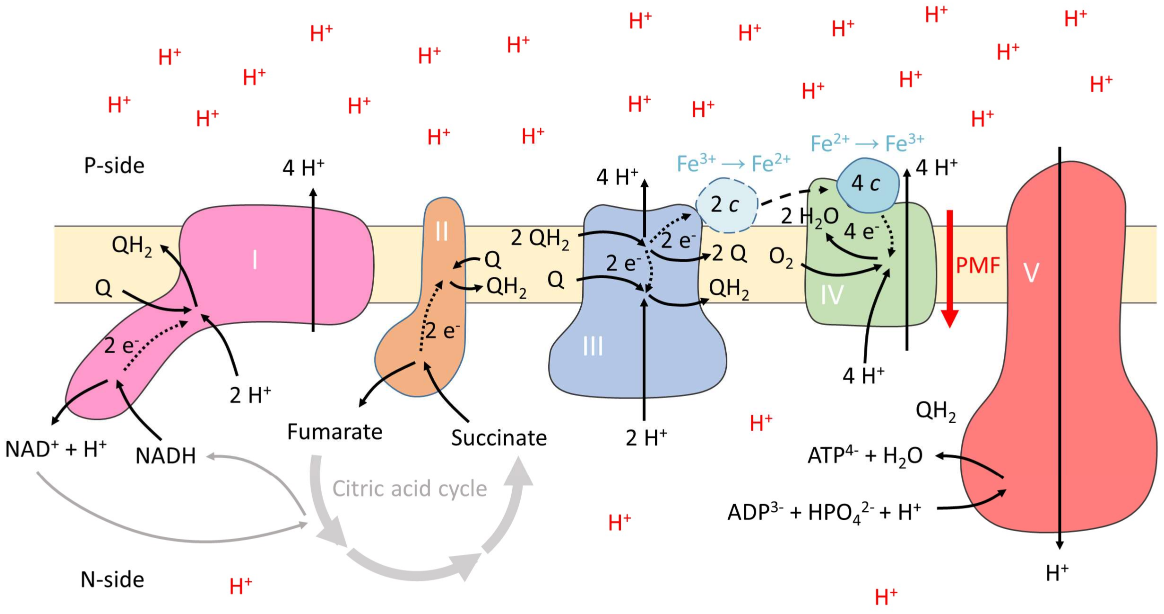


Figure 1: The electron transport chain (ETC) with respiratory complexes I-IV and ATP-synthetase as complex V. The insulating membrane is shown in light yellow. Complex II partakes in both the ETC and citric acid cycle (intermediate FAD reduction and $FADH_2$ oxidation steps were intentionally omitted to improve clarity).

concept of a chemiosmotic coupling between the redox reactions in the ETC and ATP synthesis was highly revolutionary when Peter Mitchell introduced[14] it in 1961.

The ETC consists of four membrane protein complexes, numbered consecutively in Roman numerals. To illustrate the flow of electrons and protons, ATP synthase is sometimes referred to as complex V. The ETC couples electron transport to proton transport across the selectively insulating membrane from the N-side (negative charge, high pH) to the P-side (positive charge, low pH). For the ETC to function there need to be electron donors. The citric acid cycle continually supplies a steady pool of these molecules, reduced nicotinamide adenine dinucleotide (NADH) and succinate, from the oxidation of acetyl-coenzyme A. The ETC transports electrons from these donors via the electron carriers coenzyme Q and cytochrome *c* to their final destination, molecular oxygen in complex IV. Figure 1 illustrates the entire chain in a simplified schematic.

One entry for electron donor molecules in the ETC is NADH-Q-oxidoreductase (complex I)[15], a large membrane protein with a minimal number of 11 subunits in bacteria and at least 40 in eukaryotes.[16] Complex I accepts two electrons from NADH to a flavin mononucleotide cofactor. Both electrons then move along a chain of iron-sulfur-clusters to their target, coenzyme Q. Due to the abundance of coenzyme Q, it also carries the name ubiquinone in the oxidized form and ubiquinol when it is twice reduced. Q is reduced to QH_2 with the two protons being taken up from the N-side. The large membrane-bound part of the enzyme further pumps four protons from the N to the P side. The exact mechanism of this pumping, however, is still elusive. As it converts NADH to $NAD^+ + H^+$, complex I is also called NADH-dehydrogenase. Complex I has the potential to also reduce molecular oxygen to superoxide, especially under conditions of reverse electron transfer due to high ΔpH and a highly reduced Q pool, and when the $NADH/NAD^+$ ratio is high.[17] The superoxide ion is both itself a reactive oxygen species (ROS) and a precursor to other ROS species that can have both regulatory as well as highly damaging effects, which is why they have emerged as a research focus in the treatment of cancer and metabolic diseases like Alzheimer's disease[18] (AD) and the study of ageing.

Succinate-Q-oxidoreductase (complex II)[19] marks the second entry into the ETC. It is the only enzyme partaking in both the citric acid cycle and the ETC. It consists of two small membrane-bound subunits and two larger hydrophilic subunits on the N-side. Complex II catalyzes the oxidation of succinate to fumarate while reducing a covalently bound cofactor, flavin adenine dinucleotide FAD, to $FADH_2$. This reaction consists of two electron and two proton transfers. The electrons are then transported through three iron-sulfur clusters to coenzyme Q in the hydrophobic part of the enzyme. The electron transfers together with deprotonation thus restore FAD. There is also a heme cofactor close to the bound Q. Its possible role in the Q reduction is, however, not clear yet. Together with two close by protons, the two electrons finally reduce Q to QH_2, which then exits from the enzyme. Complex II is the only enzyme in the ECT that does not actively transport protons across the membrane.

Both complexes II and I produce one QH_2 per reaction cycle that then diffuses through the membrane as a hydrophobic electron carrier. The next step involves the oxidation of the two-electron carrying ubiquinol in conjunction with the reduction of the soluble one-electron carrying cytochrome *c* by complex III, Q-cytochrome-*c*-oxidoreductase.[20] Its reaction cycle is termed the Q-cycle. In it, the directed transport of protons across the membrane is not the result of active proton pumping, but a consequence of the directed supply and release of protons from the reduction of ubiquinol and oxidation of ubiquinone. In the first part of the Q-cycle, a bound QH_2 donates one electron to an oxidized cytochrome *c* at the P-side of the membrane and another electron to a second bound Q, which also sits in the transmembrane part of the enzyme, however closer to the N-side. The electron for cytochrome *c* reduction moves through an iron-sulfur cluster and a heme group, while two heme groups facilitate electron transport for the reduction of the second Q. The now reduced cytochrome *c* diffuses away from the enzyme and the oxidized Q reenters the membrane, while two of its protons pass into the P-side proton reservoir. The one-electron reduced semiquinone radical $Q^{\bullet -}$, however, remains in the enzyme. Another QH_2 then enters complex III from the membrane and the process is repeated, another electron thus reduces cytochrome *c* and two more protons enter the P-side. The semiquinone radical receives the second electron and the two protons needed for its conversion into QH_2 enter the enzyme from the N-side. The overall cycle therefore takes up two protons from the N-side and uses them to regenerate QH_2 while releasing four protons into the P-side from the Q oxidation.

The small cytochrome *c* molecule has itself a heme cofactor with either a ferrous (Fe^{2+}) or ferric (Fe^{3+}) iron and is both structurally and genetically very homologous even across different domains. This is why every eukaryotic cytochrome *c* tested so far was able to interact with every complex IV enzyme of any species, including bacteria.[21] Upon its reduction, cytochrome *c* interacts with the terminal enzyme of the respiratory chain, cytochrome *c* oxidase (complex IV).[22] Cytochrome *c* donates the electron to C*c*O through a copper cofactor (Cu_A) and diffuses away. From there, electrons move through a first heme *a* cofactor to a second one (heme a_3) that is further coordinated with another copper cofactor (Cu_B). Together, heme a_3 and Cu_B form the binuclear center (BNC), where oxygen binds as the last electron acceptor of the electron transport chain to both the heme a_3 iron and Cu_B and is subsequently reduced to two water molecules. This process needs four electrons and four protons. This means that for a total cycle C*c*O has to reduce four cytochrome *c* proteins. The four protons needed for the oxygen reduction enter the enzyme from the N-side through two different channels. At the same time, one of the channels pumps an additional proton from the N-side to the P-side for every electron entering the enzyme from cytochrome *c*. The overall reduction of one dioxygen molecule therefore results in eight protons being taken up from the N-side and four protons being released to the P-side. A more detailed explanation of C*c*O's catalytic cycle follows in section 1.4. Both in the inner mitochondrial membrane of eukaryotes and in the inner membrane of bacteria, complexes I, II, and IV associate with one another to form supercomplexes, also called respirasomes. Research into these supercomplexes has shown that they are functional and increase the overall stability of the enzymes while also providing a kinetic advantage.[23] All four respiratory complexes make use of heme cofactors and iron sulfur clusters to accept and transport electrons. Together with other (metal) cofactors like copper ions, they lie at the core of cellular redox chemistry.

Finally, the enzyme ATP synthase[24] harnesses the proton motive force to convert ADP to ATP through a set of mechanical motions. The enzyme itself consists of a movable and an immovable part, termed rotor and stator, respectively. The stator consists of a soluble protein that facilitates ATP production and a transmembrane part that harbors two proton channels, one reaching from the P-side into the center and one reaching from the center to the N-side. Connected to this subunit is the rotor with its 10-14-meric ring structure and an attached central stalk reaching into the soluble part of the stator. A proton entering the first channel of the transmembrane stator part from the P-side protonates and thus neutralizes an aspartic acid side chain of one of the transmembrane rotor subunits. The neutralization of this side chain enables this subunit to leave the protein-protein interface and enter the hydrophobic membrane environment, thus prompting the entire ring structure with stalk to move in a rotational manner within the membrane. The deprotonation of this side chain occurs after nearly a full rotation, when a rotor subunit contacts the transmembrane stator again and the proton leaves through the second channel into the N-side. The central stalk movement induces conformational changes in the hexameric soluble stator protein that catalyze the synthesis of ATP from ADP and phosphate. Three of this protein's six subunits can bind both substrates, which happens in the loose (L) state. [25, 26] Movement of the stalk by 120° causes this subunit to change into the tight (T) form, thus catalyzing the binding of ADP to the phosphate (+1H^+) to produce ATP and water. Another 120° turn induces the open (O) conformation that releases the ATP product and can rebind ADP + P_i. All of the three catalyzing subunits are in a different state of the cycle (O, L, or T), depending on which part of the stalk they face. The number of protons needed to produce one molecule of ATP is thus the number subunits of the transmembrane rotor ring divided by the three catalyzing subunits of the soluble stator protein.

1.2. Pathology of C*c*O dysregulation and dysfunction

Due to its importance as the terminal enzyme of cellular respiration, defects in C*c*O present a variety of severe diseases. One of the most prominent diseases associated with C*c*O dysfunction is Leigh syndrome, which is not the cause of a mutation in a gene of one of the subunits of the oxidase itself, but can be (among others) the result of a mutation in the gene SURF1, an important assembly factor for the oxidase (see section 4.1).[27] The progression of Leigh syndrome leads to increasing damage to the muscular and nervous system and after a few years death, most commonly from respiratory failure.

The correct expression and stoichiometry of the enzyme's individual subunits is critical for its normal function and abnormalities are connected to many diseases of the heart and brain. For example, underexpression of subunit 4 leads to reduced C*c*O content in heart tissue and is associated with dilated cardiomyopathy.[28] The dose-limiting cardiotoxicity of the chemotherapeutic drug doxorubicin is a result of subunit 2 and 5a depletion.[29] The brain cells of AD patients, too, exhibit subunit 2 depletion and subsequent loss of C*c*O content.[30] While the underlying cause of AD is still unclear and there are many competing hypotheses, the impairment of mitochondrial function including of C*c*O is an important part of AD pathology and can serve as an important biomarker of the disease. In humans, the enzyme exists in tissue-specific isoforms with distinct kinetics. Allosteric binding and phosphorylation further regulate the enzyme, and dysregulation is implicated in pathologies like sepsis and ischemia-reperfusion injury (e.g. due to stroke).[31]

As the terminal enzyme of oxidative phosphorylation, cytochrome *c* oxidase has high pharmacological relevance. Drugs inhibiting C*c*O might result in mitochondrial toxicity as an off-target effect.[32] Due to its importance, the enzyme can, however, also serve as an intended drug target. C*c*O is inhibited by and a potential target of the antifungal and antiparasitic drug miltefosine in *Leishmania* parasites, leading to apoptosis-like cell death.[33] Mitigating dysfunction of the enzyme with drugs might also assist in the treatment of the aforementioned neurodegenerative diseases.[34] Depending on the type of cancer, there might either be a dysfunction of C*c*O that increases cancer progression[35] by switching to a glycolytic metabolism (Warburg effect) or an upshift of C*c*O activity[36] to increase ATP production (e.g. inverse Warburg effect). Modulation of C*c*O activity might therefore prove to be beneficial in the treatment of cancer.[37]

1.3. Structure and homology

The current scientific understanding holds that the mitochondrion originated from a symbiotic encounter between a member of the alphaproteobacteria and a host cell that was most probably an archaeon in a process called symbiogenesis.[38] Subsequent gene transfer from the incorporated proto-mitochondrion to the host's nuclear DNA and loss of genes unnecessary for the symbiotic relationship created the relatively compact mitochondrial genome. The human mitogenome consists of only 37 genes on a length of 16569 bp. Among these are 22 genes for transfer RNAs and two genes for ribosomal RNA. The remaining are protein-coding genes for subunits of complexes I, III, and IV of the ETC, ATP synthase, and a very recently discovered protein called humanin.[39] The mammalian mitochondrial genes for C*c*O encode the subunits 1-3, which represent the core subunits of the enzyme. The other 11 protein subunits have to traverse to the inner mitochondrial membrane after their synthesis from genes of the nuclear genome, involving at least 30 assembly factors.[40] Bacterial C*c*O, on the other hand, needs two core subunits (Figure 2) to

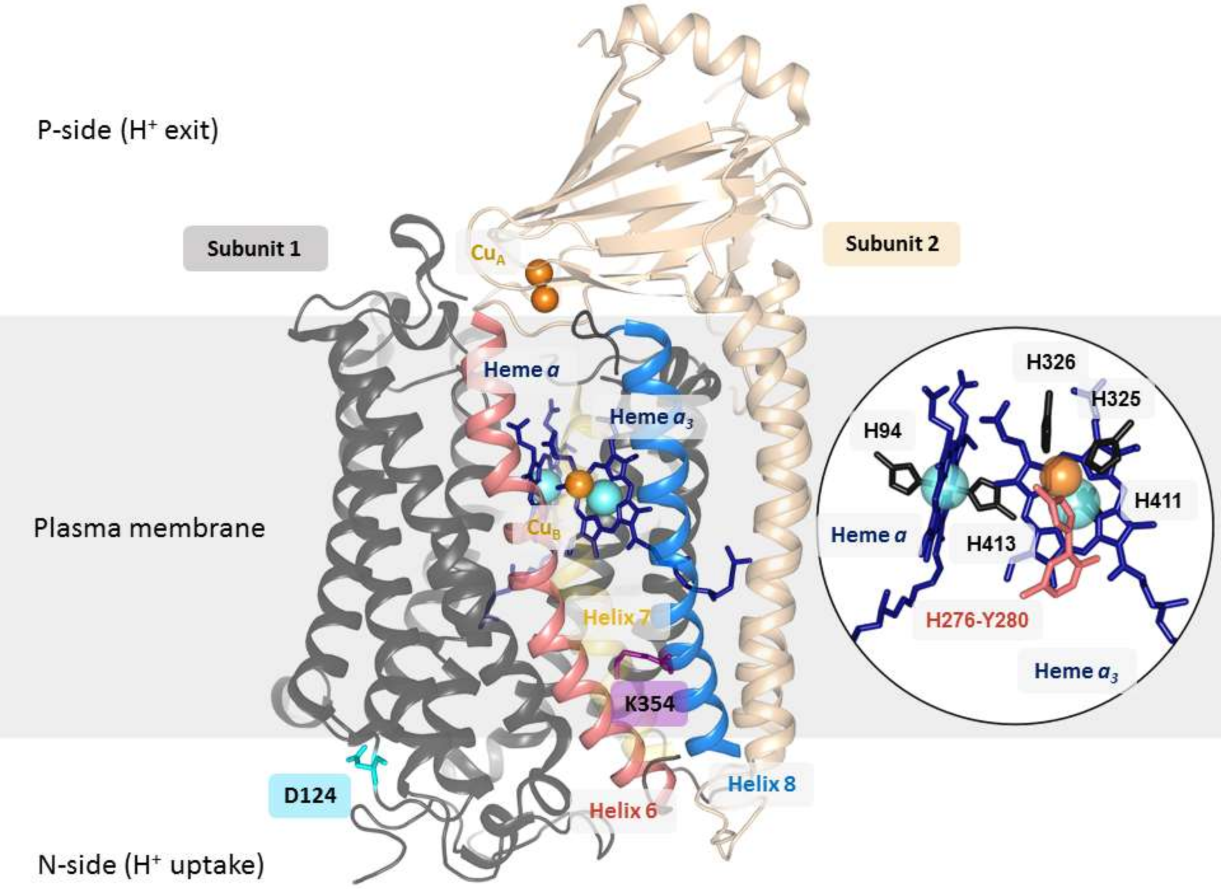


Figure 2: Crystal structure of the two core subunits of *P. denitrificans* cytochrome *c* oxidase.[43] Subunit 1 is shown in gray and subunit 2 in sand color. Transmembrane helices 6, 7 and 8 are highlighted in red, yellow, and blue, respectively. The close up on the right depicts the two hemes (dark blue) with their central irons (cyan), and Cu_B and the histidine side chains coordinating the metal cofactors (H411 is largely obstructed by the iron). Histidine 276 and the catalytically involved tyrosine 280, to which it is covalently bound, are red. PDB code: 3HB3

function[41, 42] and possesses only two additional subunits. The two most studied bacterial C*c*Os come from the organisms *Rhodobacter sphaeroides* and *Paracoccus denitrificans*, the latter of which is the focus of this thesis. Like *R. sphaeroides*, *P. denitrificans* belongs to the class of alphaproteobacteria and—due to its genetic markup containing all the proteins of the ETC as well as other factors—has long been a contender for a close descendent of the proto-mitochondrion.[43] While there exists a large degree of domain-spanning homology regarding structure, function and sequence identity, there are some differences between the bacterial and eukaryotic C*c*O. The main difference lies in the far more complex overall structure and composition of the latter's subunits. Notwithstanding the variations in complexity, major commonalities exist between eukaryotic and prokaryotic ETC proteins like their preference for forming respirasomes as well as many shared catalytic mechanisms.

The crystal structure[44] of the enzyme reveals twelve transmembrane helices (TMHs) in subunit 1 and another two TMHs in subunit 2, which also contains a large soluble domain. Subunit 1 of the enzyme contains both heme cofactors as well as the Cu_B ion of the binuclear center. Several histidine residues coordinate the involved iron and copper ions. One of these histidines is H276 that is covalently bound to a tyrosine residue, Y280, one alphahelical turn further toward the N-side (*P. denitrificans* numbering is used throughout this work). This tyrosine is involved in the enzyme's catalytic mechanism and thought to donate one proton and one electron to the oxygen reduction reaction. The benefit or origin of this bond is not yet understood. One hypothesis holds that the crosslink reduces tyrosine's pK_A and thus increases the oxidase's activity. [45, 46] Oxygen binds to the heme a_3 iron and subsequently splits between this iron and the Cu_B ion. Subunit 2 contains the two Cu_A ions held in place by two cysteine residues.

There are two distinct proton uptake pathways, each named after an essential residue (D124 and K354, respectively) and each of these pathways contributes to proton uptake from the N-side (Figure 2). The K-channel, however, is only active during the reductive phase and only translocates two of the protons needed for the reduction of oxygen to water. This leaves the two remaining protons required for this reaction as well as all protons pumped across the membrane for the D-pathway.[22] The D-channel is well established and starts at D124, which connects to a glutamic acid residue close to the active site (E278) through a series of hydrogen bonded water molecules. Residues around this path are N131, N113, S134, G109, S193, and I104.[44] E278 acts as a valve preventing backflow into channel from the active site.[47] A problematic issue arises due to the necessity of the enzyme to distinguish between pumped and substrate protons, as both types arrive through the D-channel.[22] In order to eliminate the possibility of a proton-leak along the concentration gradient (and thus against the enzyme's pumping direction), there cannot be an open channel from the P-side. For this purpose, the so-called proton-loading site (PLS) serves as a transient storage container for the to-be-pumped protons until electron transfer to the catalytic site causes C*c*O to eject them.[48] Potential candidates for the location of this site are either the heme a_3 propionic acid groups towards the P-site of the membrane[49] or a close by and highly conserved water cluster[50] with variable pK_A.

The K-channel's structure is comparatively more elusive. Notwithstanding experimental difficulties, a few of its features have so far emerged. The channel begins at the N-side and then winds through K354 and T351 towards the catalytic tyrosine Y280.[47] E78 of subunit 2 apparently serves as its start, although experiments[51] seemed to indicate that this was not the case for *P. denitrificans* C*c*O. More recent results have, however, shown that the binding of lipidic ligands to the K-channel can rescue inactivated mutants and in general modulate K-channel activity and thus conflicting results regarding E78 might simply arise from the purification conditions of the enzymes.[52, 53] Due to the selective activity of the K- and D-channels, there has to exist some kind of gating mechanism controlled by the redox-state of the catalytic center.[22] As to the exact nature of this mechanism, there is still plenty disagreement. Section 5.2 of this thesis will address the main competing models.

1.4. Function and catalytic cycle

Through the injection of electrons from four cytochrome *c* molecules per cycle, C*c*O catalyzes the following reaction:

4 cytochrome *c* (Fe^{2+}) + 8 $H^+_{\text{N-side}}$ + O_2 → 2 H_2O + 4 cytochrome *c* (Fe^{3+}) + 4 $H^+_{\text{P-side}}$.[22]

Each electronation and thus redox change of the enzyme's cofactors drives the reactions and adjustments needed for completion of the catalytic cycle. For this to occur, soluble cytochrome *c* has to attach to the P-side of C*c*O and donate an electron from its reduced heme iron to the Cu_A cofactor of C*c*O subunit 2. The interaction between cytochrome *c* and C*c*O has two major contributions[54]: Long-range electrostatic interactions between positively charge lysine residues of cytochrome *c* and the negatively charged carboxylates at the P-side of C*c*O facilitate the correct orientation of the docking cytochrome *c*. Hydrophobic interactions then potentially aid in stabilizing the complex.

While there is still plenty of disagreement on the details of the cycle, some of its individual steps and possible alternative paths, a consensus has formed around the main states and the reduction sites at the catalytic center. The following cycle (Figure 3) is a composite of the models of Michel[55] and Wikström[56] and intended to highlight the most important features of the enzyme's catalytic function. The cycle consists of an oxidative and a reductive phase, each featuring the uptake of two electrons from cytochrome *c*, the addition of two protons to oxygen and the pumping of two additional protons across the membrane. In the fully oxidized ("activated" O_H[57]) state, a water molecule is already bound to Cu_B and a hydroxide at the heme iron. The incoming electron reduces Cu_B. Subsequently, a proton enters the catalytic site through the K-channel and binds to the catalytic tyrosine 280 while the enzyme ejects another proton to the P-side. The resulting state is the electronated (E) state. The next electron reduces the heme iron. A proton coming from the K-channel leads to the formation of the second water and another proton is pumped. The steps from the O_H to the resulting reduced (R) state constitute one-half of the catalytic cycle, referred to as the

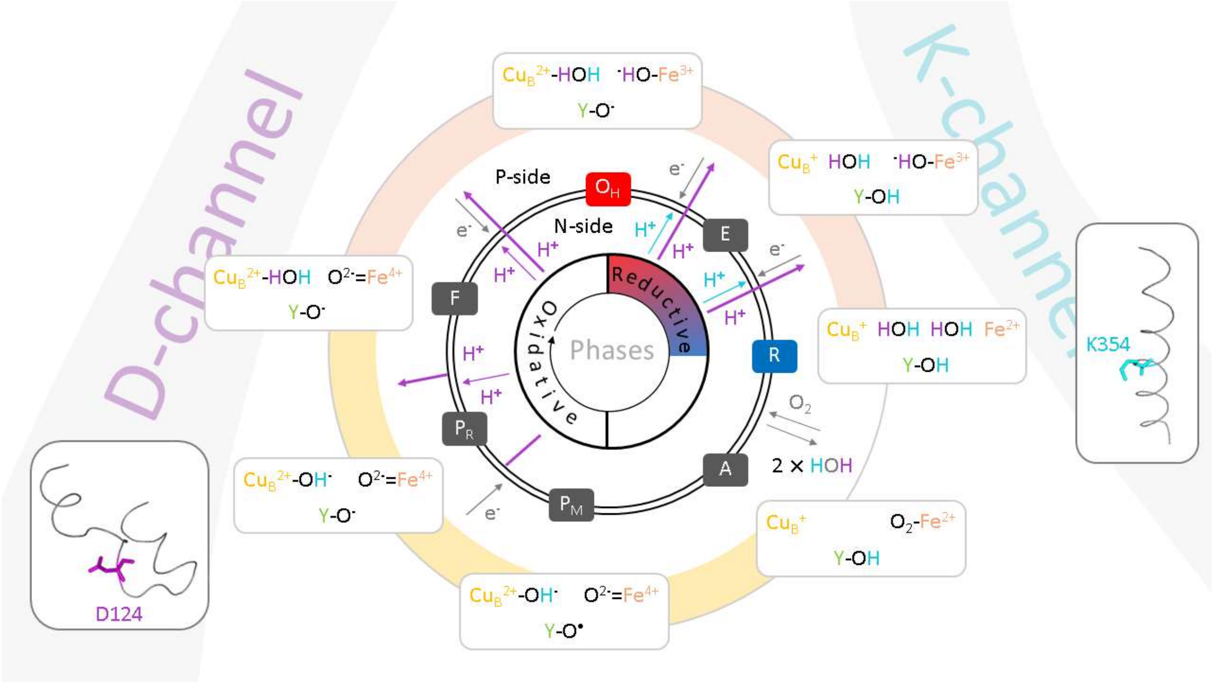


Figure 3: Catalytic cycle of cytochrome *c* oxidase. Electrons arriving via the D- and K-channel are colored pink and cyan, respectively. The double ring represents the double membrane with the N-side and P-side on its respective in- and outside. Arrows crossing the membrane indicate pumped protons. Boxes show the chemical structure of the catalytic site. The outer ring colors indicate the parts of the cycle in which protons react with the oxygen at Cu_B (yellow) or at the heme iron (orange).

reductive part. What follows then, is the oxidative part of the cycle. In the R state, a new dioxygen molecule can enter the catalytic site from the membrane and bind to the heme a_3 iron. The catalytic center subsequently releases both previously created waters. This is only an intermediate state named after its oxygen-adduct (A). The A state decays into the P_M state by breaking the O–O bond with the help of one electron and one proton from Y280. The tyrosine exists as a radical for a short time, until the next electron reduces it, thus creating the P_R state with Y-OH^-. For the creation of the F state, a proton is taken up from the D-channel and reacts with the OH^- at Cu_B to form the first water and another proton is pumped. In the last step, from F → O_H, the fourth electron enters the catalytic site and reduces the heme iron, a proton from the D-channel reacts with the iron-oxygen and, again, a proton is pumped to the P-site.

One of the most pressing issues regarding the mechanics of the cycle lies in the selective activation of the K-channel during the reductive phase in absence of large structural changes to the enzyme[58]. This thesis aims to elucidate this activation and the processes enabling efficient proton uptake. The expression system developed in this thesis further provides the tools to access and identify residues involved in proton release in future studies.

2. Biophysical methods for the investigation of molecular fluorescence

Techniques based on fluorescence rank among the most commonly employed experimental methods in biophysics. Due to its exceptionally high sensitivity as well as its temporal and spatial properties (femtoseconds, femtoliters, femtomoles)[59], fluorescenc provides the ability to detect and analyze even a singular fluorescent molecule, also called fluorophore. Fluorescence enables researchers to study the ultrafast processes occurring on very small scales underlying the mechanisms of life. With an apparent molecular weight of ~126 kDa[60] (i.e. ~ 7 nm in diameter) for the four subunit bacterial enzyme, C*c*O is well suited to be studied using fluorescence. Further, proton and electron transfers in C*c*O have characteristic timescales in the milli–microseconds[61] and protein motions in the pico–nanoseconds. Because the immediate environment of fluorophores influences their fluorescence properties, they can provide information on a diverse set of parameters like charge, temperature, protonation, polarity, viscosity, conformation, complexation and diffusivity, all of which can be investigated with fluorescence spectroscopy.[62-66] Confocal microscopy and total internal reflection fluorescence microscopy (TIRFM) can further transform this method into an imaging technique and thus allow for investigating the distribution of these parameters in cells and biological tissues.

2.1. Absorption and fluorescence

When electromagnetic radiation encounters atoms and molecules, its photons can induce an electronic transition in their electronic orbitals, provided the energy of the photon equals the energy of the transition and the photon's polarization vector and the absorption transition dipole moment vector are aligned sufficiently collinearly. One can calculate the Einstein coefficient of induced absorption B_{12} by applying perturbation theory:

$$B_{12} = \frac{1}{6\,\varepsilon_0\,\hbar^2} \times \left|\underline{M}_{12}\right|^2 \quad . \tag{1}$$

The vector $\underline{M}_{12}$ in this equation is the absorption transition dipole moment:

$$\underline{M}_{12} = \left\langle 1 \middle| \underline{\mu} \middle| 2 \right\rangle \int \Psi_1^* \, \underline{\mu} \, \Psi_2 \; d^3\rho \tag{2}$$

where ρ is a spatial coordinate, Ψ_1 and Ψ_2 are the wave functions of states 1 and 2, and $\underline{\mu}$ is the dipole moment operator. Electronic transitions are also guided by symmetry and spin selection rules, the latter of which are partly overcome through spin-orbit coupling. Intersystem crossing, i.e. transitions between singlet and triplet states, therefore often plays an important role in fluorescence-based experiments and can reduce the overall fluorescence yield.

The determination of concentrations c of absorbing molecules in a homogenous solution can be calculated with the law of Lambert-Beer, giving the absorbance A of a sample depending on the wavelength λ:

$$A(\lambda) = \log\frac{I_0}{I} = \varepsilon(\lambda) \times c \times d \tag{3}$$

where I_0/I is the ratio of incident to the transmitted intensity, $\varepsilon(\lambda)$ is the molar absorption coefficient and d the thickness of the sample. Steady-state absorbance measurements were conducted with a UV-2450-PC UV-Vis spectrometer (Shimadzu, Japan).

2.2. Environmentally sensitive fluorescent probes

The surrounding nanoenvironment of a fluorophore has considerable impact on the molecule's fluorescence. The strength and nature of this effect is dependent on the photophysical properties of the fluorophore and the environment. There are therefore many different applications for fluorophores to probe a wide variety of environmental characteristics. Two of the environmentally sensitive fluorophores used in this work are 5-iodoacetamido-fluorescein (IAF) and 6-bromoacetyl-2-demithylaminonaphtalene (BADAN).

IAF is a derivative of the xanthene dye fluorescein. Due to its thiolreactivity, it can bind to the sulfhydryl moieties of cysteine side chains in proteins. Fluorescein can undergo three protonation-induced transitions that change its photophysical properties (Figure 4 A). Depending on the wavelength, the extinction coefficient ε of the dye increases by three orders of magnitude when the pH is increased from a value of 1 to 10.[67] At the same time, the fluorescence quantum yield increases significantly. Each protonation has a distinct pK_a value that is dependent on the dye's

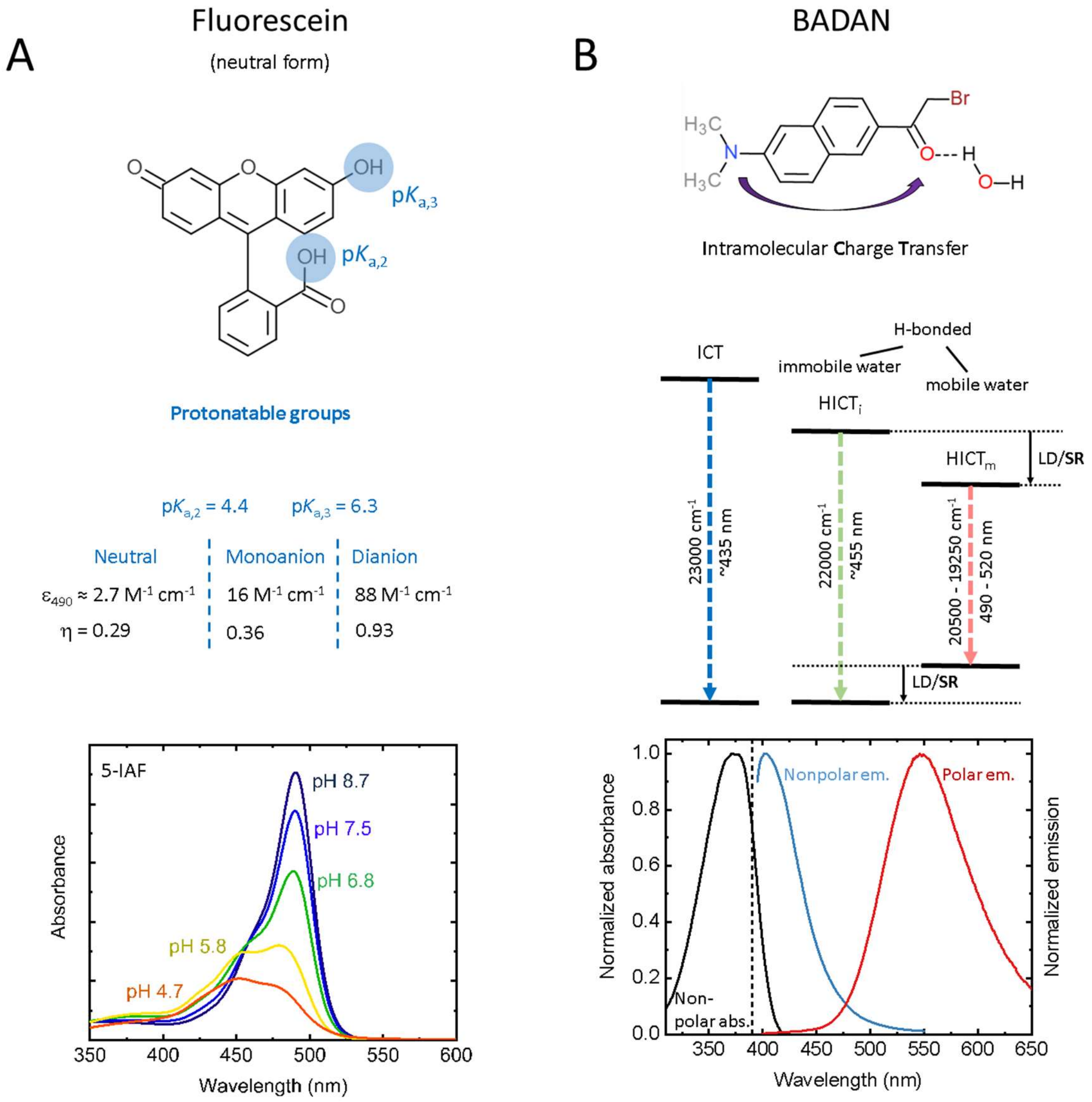


Figure 4: (A) Fluorescein and its thiolreactive derivative IAF change their absorbance and fluorescence properties (values from [66], the cationic form is left out for the purpose of clarity) depending on the protonation state of the dye. (B) BADAN undergoes an intramolecular charge transfer upon excitation as well as hydrogen bonding leading to different energy levels in its emission (energy diagram was adapted from [69]). Due to the dye's solvatochromism, its fluorescence emission depends on the polarity and mobility of the surrounding solvent leading to a higher energy spectrum in a nonpolar solvent like cyclohexane (blue curve) and a red-shifted spectrum in 50%$_{v/v}$ methanol (red curve). The dashed line indicates the excitation wavelength at 390 nm.

nanoenvironment. The pK_a value of a group is the negative decadic logarithm of its acid dissociation constant K_a:

$$K_a = \frac{[A^-][H^+]}{[AH]} \quad , \tag{4}$$

with brackets denoting the concentrations of the chemical species of the dissociated acid (numerator) and undissociated acid (denominator). The pK_a value of a diluted acid and its conjugate base is connected to the pH of the solution through the Henderson-Hasselbalch equation:

$$\mathrm{pH} = \mathrm{p}K_a + \log_{10}\left(\frac{[\mathrm{A}^-]}{[\mathrm{AH}]}\right) \quad . \tag{5}$$

In aqueous solution, the pK_a values of fluorescein are 2.2, 4.4, and 6.3.[67] When IAF is bound to a protein, its pK_a values can be shifted trough interactions with the protein surface, e.g. through hydrogen bonding and due to the local surface polarity and charge and a change in the reaction's associated free energy.[68] The dye can therefore serve to sense small structural changes at the surface of proteins. Additionally, the absorption maximum of the dye depends on the dielectric constant of its nanoenvironment. This feature can therefore be used to sense the polarity of the protein surface.[64]

BADAN is a thiolreactive dye undergoing intramolecular charge transfer (ICT) from the dimethylamine to the carbonyl moiety upon excitation (Figure 4B). This leads to a strong shift in the dipole moment of the dye. Upon excitation, dipolar relaxation of the surrounding solvent lowers the energy of the S_1 energy level and thus leads to a redshift in the emission spectrum.[69] The polarity of the solvent in the nanoenvironment of the dye as well as its local viscosity determine the amplitude and dynamics of this time-dependent fluorescence shift (TDFS). Hydrogen bonding of the carbonyl moiety can further reduce the excited state energy.[70] This hydrogen bonding results in the so-called HICT state. The indices *m* and *i* in Figure 4B indicate whether the solvating molecules in the nanoenvironment of the dye are mobile or immobile. Another important influence on the excited state energy level of BADAN are internal label dynamics (LD), e.g. twist and wobbling motions between the propanoyl moiety and the aromatic ring system.[70] The interplay between these different factors results in a complex spectral behavior of BADAN. Due to the strong influence of solvent dynamics, however, the dye is well suited to sense local changes in hydration and viscosity.

Steady-state fluorescence measurements were conducted with a Fluoromax-3 fluorescence spectrometer (HORIBA Jobin-Yvon, Japan). For information on the time-resolved fluorescence spectroscopy setup, see the next section.

2.3. Picosecond time-resolved fluorescence spectroscopy

In this work, I measured the dynamics of fluorescence decays trough time-correlated single photon counting (TCSPC). This spectroscopic technique is based on the detection of individual photons and allows for the measurement of subpicoseconds to nanoseconds dynamics depending on the pulse duration of the pulsed laser source (LS) employed. Figure 5 displays the optical setup[65, 71], which operates as follows: The measurement of fluorescence decays was performed with a solid-state laser (Millennia VS, Spectra Physics) pumped Titan Sapphire (Ti:Sa, Tsunami 3950, Spectra Physics) laser with a pulse duration of 1-2 ps. Since the output of the Ti:Sa is in the far red to infrared wavelength range of 710-1100 nm, to excite the fluorescein and fluorescein-conjugates used in this work, the frequency is increased by a frequency doubler (Model 3980, Spectra Physics), resulting in a tunable range between 360 to 550 nm, of which 484 nm were chosen to excite fluorescein.

To quasi-eliminate the possibility of overlapping nanoseconds fluorescence decays, a pulse picker integrated into the frequency doubler system reduces the frequency of the laser to 4 MHz. Any remaining low-frequency fundamental is subsequently filtered out by a high-pass filter. The axis of polarization is then rotated through a Fresnel rhomb and the single light path split into two perpendicular paths by a beam splitter. One of the beams is decreased in intensity by a neutral density filter and hits the photodiode (DET210, Thorlabs) giving the STOP (inversed START/STOP mode) signal for the measurement. The other beam is diverted into the encased measurement box where it’s intensity is adjusted for single-photon counting by an adjustable neutral density filter wheel. Before it reaches the sample, a Glan-Thompson polarizer filters out all light that is not polarized in the vertical direction. The sample fluorescence in perpendicular direction then passes an edge filter and is focused on a multichannel plate detector (R3809U, Hamamatsu) giving the START signal after passing an iris diaphragm and the adjustable emission polarizer. For measurements of fluorescein-conjugated C*c*O, the edge filter was exchanged for a 515 nm long pass filter (OG515). Both START and STOP signals are passed through pre-amplifiers before reaching the PCI card containing the TCSPC electronics (SPC-830, Becker&Hickl). There, constant fraction discriminators (CFDs) determine a defined time from the pulses, which have a temporal width. This is facilitated by separating the incoming pulse into two pulses, one of which is amplified and inverted and the other one temporally shifted. After adding both signals, the point of zero crossing determines the time of the measurement. The two CFDs are connected to a time-to-amplitude converter (TAC) that charges a capacitor after the START

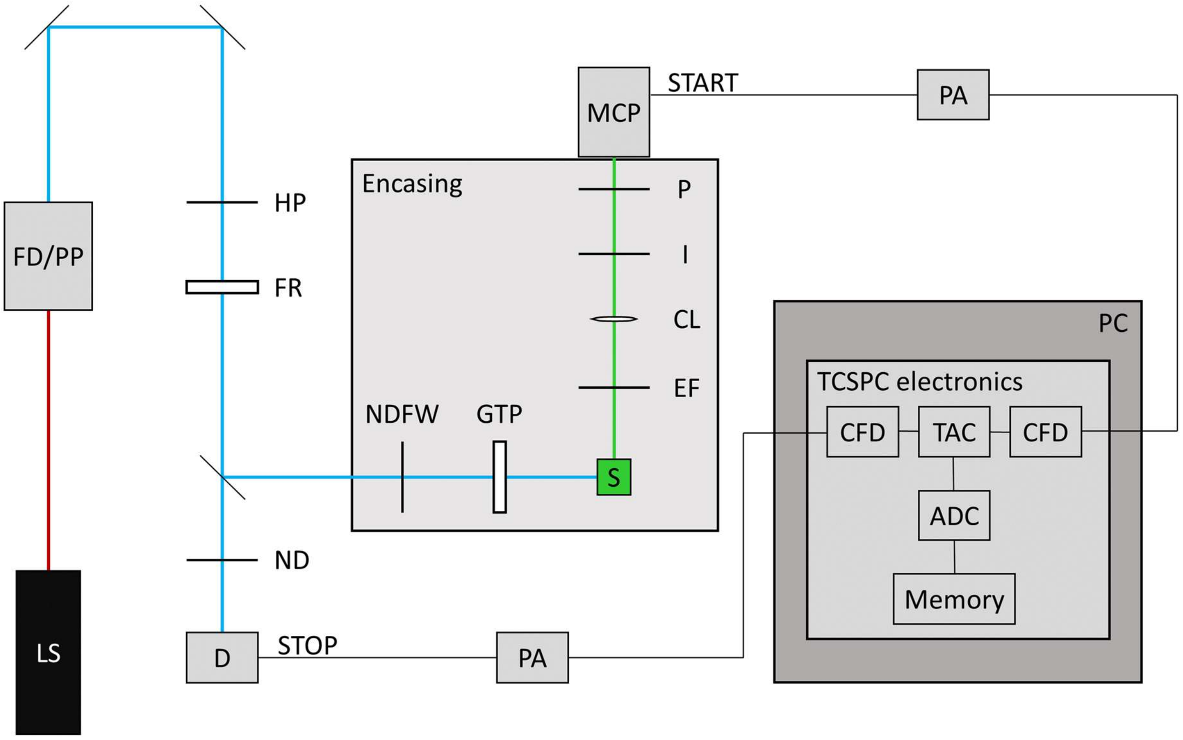


Figure 5: Schematic of the TCSPC setup. The Ti:Sa laser source (LS, black rectangle) produces a 82 MHz polarized beam of long wavelength that doubles in frequency upon passing the frequency doubler (FD) and gets reduced in frequency to a value of 4 MHz by a pulse picker (PP). After passing a high-pass filter (HP) and a Fresenel-rhomb (FR) the beam is split up by a beam splitter. One of the two divergent beam paths reaches the STOP diode (D) after being reduced in intensity by a neutral density (ND) filter. The other optical path leads into the encased measuring box. There, the beam's intensity is reduced by an adjustable neutral density filter wheel (NDFW) and after passing a Glan-Thompson polarizer (GTP) excites the sample. In perpendicular direction the sample's fluorescence passes an edge filter (EF) and is focused on the multi-channel plate (MCP) detector giving the START signal by a collimating lense (CL). Before reaching the detector, the emission light passes an iris filter (I) and its polarized components are filtered by an adjustable polarizer (P). Both the START and STOP signals are amplified by pre-amplifiers (PA) and transferred to a PC containing the TCSPC electronics, where they are converted to decay histograms through passing constant fraction discriminators (CFD), a time-to-amplitude converter (TAC), and finally an analog-to-digital converter (ADC).

signal and stops charging as soon as the STOP signal reaches it. The TCSPC measurements are conducted in reverse mode, i.e. a laser photon stops a measurement initiated by an emitted photon to avoid overcharging the TAC and to improve resolution of the lifetime curve. Finally, an analog-to-digital converter (ADC) converts the voltage output of the TAC into a count in the photon histogram in the device memory.

Fluorescence decay traces were collected in 1024 time channels with a channel width of 19.7 ps. The instrumental response function (IRF) of the time-correlated single photon counting setup was determined with a colloidal silica solution as scattering material (LUDOX, Grace) and had a full width at half-maximum of 30–40 ps. Data analysis was performed with the program GLOBALS Unlimited through deconvolution with the instrumental response function and fitting of the fluorescence decay traces with a sum of exponentials.

2.3.1. Time-resolved fluorescence depolarization spectroscopy

A specialized application of TCSPC is fluorescence depolarization spectroscopy. By applying a polarization filter to the light emitted with intensity I by the sample in parallel and perpendicular direction ($I_{\parallel}$ and $I_{\perp}$, respectively) of the incident excitation light one can calculate the fluorescence anisotropy r(t) as

$$r(t) = I_{\parallel}(t) - I_{\perp}(t) / \left(I_{\parallel}(t) + 2I_{\perp}(t) \right) \quad . \tag{6}$$

The decay of the fluorescence anisotropy is due to the rotational diffusion of the dye (or dye-conjugate) occuring between the absorbance and emission of a photon.[72] The resulting curve can be fitted with the multiexponential expression

$$r(t) = r_0 \ \textstyle\sum_i \beta_i^* \ e^{-t/\varphi_i} \quad , \tag{7}$$

where r_0 is the initial anisotropy and β^*_i and φ_i are the relative amplitudes and decay times for all anisotropy components i.

In case of a fluorophore attached to a protein, the first component belongs to the fluorescent label itself, the second is related to the segment the dye is attached to, and the third component describes the motion of the entire system.[65] The latter component results in a final anisotropy r_∞ if the overall motion of the system occurs on a time scale much longer than the excited state lifetime of the dye. Further knowledge about the steric restriction imposed on the dye can be extracted by applying the wobbling-in-a-cone model[73] approximating the dye's overall position space is by motions within a fixed cone of semiangle θ_M. This angle is connected to the ratio of the final and initial anisotropy:

$$\frac{r_\infty}{r_0} = \left(\frac{1}{2}\cos\theta_\mathrm{M}\,(1+\cos\theta_\mathrm{M})\right)^2 \quad . \tag{8}$$

One can also apply the more complex cone-in-a-cone model[73], in which dye (index 1) and protein (index 2) motion are assumed to be uncoupled and occurring each within their own overlapping cones, each with their own amplitudes A_i and characteristic times φ_i, and all under influence of the motion of the entire protein with characteristic time φ_G:

$$r(t) = r_0\left((1-A_1)e^{-t/\varphi_1} + A_1\right)\left((1-A_2)e^{-t/\varphi_2} + A_2\right)e^{-t/\varphi_G} \quad . \tag{9}$$

By fitting the anisotropy curves with this function, one can retroactively eliminate the tumbling motion of the protein in a mixed micelle in order to compare it to a membrane-embedded protein in the MD simulation data shown in section 5.1. Furthermore, it is possible to determine the semiangles of the inner and outer cone by equating the right side of equation (8) with the amplitudes A_1 and A_2 of equation (9).

2.3.2. Time-dependent fluorescence shift

Further insight into a fluorophore's nanoenvironment can be gained by spectrally resolving its fluorescence decay and studying its time-dependent fluorescence shift (TDFS). In polar solvents, the electric field change from the excitation of a fluorophore leads to a reorientation of nearby solvent molecules (solvent relaxation, SR) in accordance with the new electric field.[74] This polarization decreases the energy of the excited state over the molecules' reorientation time and thus shifts the maximum emission of the spectrum towards higher wavelengths (Figure 6). Since higher solvent viscosities retard the reorientation process, TDFS is dependent on the local viscosity experienced by the probe. The strength of the red-shift further depends on the polarity of the solvent. Especially suited for the detection of TDFS are fluorophores undergoing intramolecular charge transfer (ICT). These dyes usually display strong spectral shifts and are thus refered to as solvatochromic fluorophores.

The spectrally resolved TCSPC measurements of BADAN were conducted similarly to the TCSPC experiments outlined at the beginning of this section, however, employing a white-light laser source

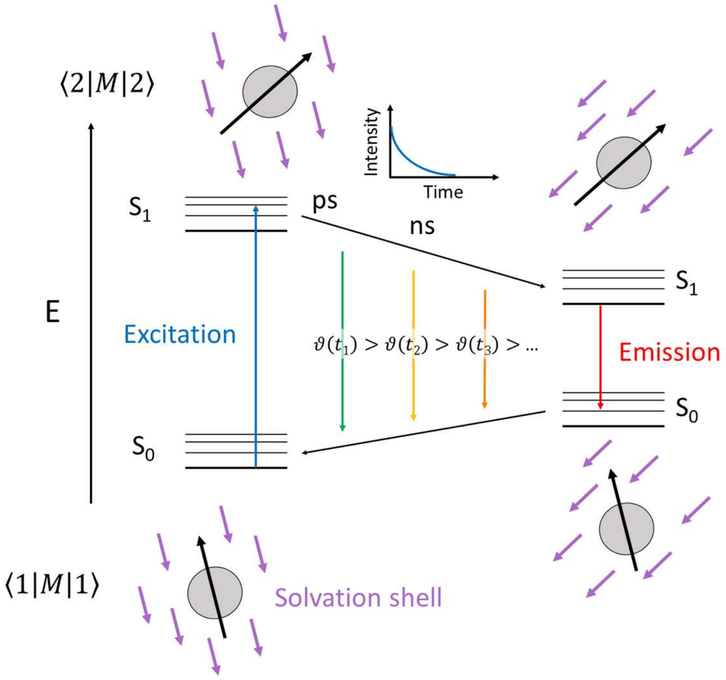


Figure 6: Influence of solvent relaxation on time-dependent fluorescence: The strength and orientation of a solvatochromic fluorophore's (gray sphere) dipole moment (black arrow) change upon excitation, thus leading to a time-dependent rearrangement of its surrounding solvent dipoles (purple arrows). This leads to a red-shifting of the emission spectrum from all emitters during the excited state lifetime of the fluorophores. Figure adapted from [72].

(Supercontinuum, NKT Photonics) at 405 nm excitation. Fluorescence was detected with a spectrally resolved 16-channel MW-FLIM detector (Becker&Hickl) between 422.1 and 612.3 nm. For the analysis of the TDFS, the maximum intensity for each time bin was determined through Gaussian fitting of the peak region observable in the measured spectrum. The resulting peak wavenumbers were then plotted over time to yield the TDFS graph.

3. Molecular biological and biochemical methods

This chapter presents the molecular biological and biochemical methods and protocols used to prepare the fluorophore-conjugated samples in a series of steps starting from DNA modification and culminating in the thiolreactive labeling process.

3.1. Cloning and mutagenesis

For the creation of mutant enzymes, I chose a method based on complementary primers in a thermal cycling (PersonalCycler, Biometra) reaction with Phusion Hot Start Flex polymerase (New England Biolabs). I used the same enzyme to amplify DNA in polymerase chain reactions (PCRs) by employing flanking primers with and without 5’ overhangs for subsequent cloning. The former technique resulted in linear amplification in only a few cycles to minimize undesired mutations and the latter in exponential amplification. The guanine-cytosine content of the template determined the use of either the High Fidelity (HF) or GC buffer. Due to the GC-richness of the *P. denitrificans* genome, it was always amplified in GC buffer. The reactions had a volume of 50 µl and contained 200 ng template DNA (for mutagenesis, 3 ng for PCR reactions), 500 nM of each primer, 200 µM dNTPs, 6 % DMSO, and 1 u polymerase in the appropriate buffer. Table 1 outlines the cycling parameters, which were individually optimized for each reaction.

After mutagenesis cycling, methylation-sensitive DpnI (New England Biolabs) digested the parental DNA in CutSmart buffer at 37°C for 3 h, which was then directly transformed. After PCR amplification, samples were subjected to (depending on fragment size) 1-3%$_{w/v}$ agarose gel electrophoresis at 60V for 1 h (Mini-Sub Cell, Bio-Rad Laboratories, Inc.) in TAE buffer pH 8.5 (40 mM tris base, 20 mM acetic acid, 2 mM EDTA) and the excised bands cleaned up through affinity spin-column chromatography (Macherey-Nagel). I employed two different methods for cloning, either by using restriction enzymes or with a more recent restriction-free technique. Restriction reactions usually contained 5 µg of DNA and 25-50 units of enzymes (New England Biolabs) in 100 µl of double-digest compatible buffer and were incubated for at least 2 h at 37°C. After electrophoresis and cleanup of the linear DNA, I ligated the fragments with either T4 or T7 ligase (New England Biolabs, both) depending on their ends being blunt or sticky.

I performed restriction-free cloning with a commercial kit (In-Fusion HD EcoDry, Takara Bio) in analogy to Gibson’s[75] assembly method. The reaction solution contained the individual DNA

Table 1: Thermal cycling conditions for mutagenesis reactions and PCR (if different, noted after a semicolon)

Cycle Step	Temperature	Time	# of cycles
Initial denaturation	98°C	1–2 min	1
Denaturation Annealing Extension	98°C 60–72°C 72°C	20–30 s 20–40 s 20–40 s/kb	12–18; 30
Final Extension Storage	72°C 4°C	3–30 min; 3–10 min hold	1

fragments (50-200 ng) with 15 bp long complementary ends lining each side and the enzyme-buffer-mix. It was incubated for 15 min at 37° and another 15 min at 50°C. Afterwards, I directly transformed 2.5-5.0 µl of the reaction into *E. coli* HST04 cells.

3.2. Transformation and triparental conjugation

As *P. denitrificans* by itself is not competent, I first transformed all vectors into chemically competent *E.coli* strains: DH5α for restriction cloning and site-directed mutagenesis (NEBα, New England Biolabs), and HST04 for restriction-free cloning (Takara Bio). The transformation procedure is the same for both strains: 1–5 µl of the reaction mixture are added to 50 µl of competent cells and incubated on ice for 30 min. Cells are then immersed in a 42°C water bath for 35 s and put back on ice for 2 minutes. To this I added 950 µl of S.O.C. medium (New England Biolabs and Takara Bio, respectively) and let it shake at 215 rpm and 37°C for 1 h. Afterwards, I plated the cells onto agar plates containing the antibiotic matching the resistance gene encoded in the transformed vector and incubated them overnight at 37°C to yield individual colonies for subsequent sequencing.

For triparental conjugation[76], helper strain *E. coli* MM294 containing the plasmid RP4-7[77] (Amp^R) provided the genes necessary for conjugation *in trans*. Small starter cultures of MM294, the *E coli* colony (Sm^R) containing the C*c*O mutant gene, and the *P. denitrificans* deletion strain (Rif^R, Kan^R, Tet^R) were shaken at 125 rpm overnight in lysogeny broth (LB) containing the appropriate antibiotic at 37°C (for *E. coli*) or 32°C (for *P. denitrificans*). On the next day, 300 µl of the recipient *P. denitrificans* strain were pipetted into an Eppendorf tube together with 100 µl of each *E. coli* culture. Subsequent centrifugation at 7000 g rcf yielded a small pellet that was subsequently dissolved in 200 µl of fresh LB. The liquid was then pipetted onto an agar plate and incubated agar-down at 32°C overnight. The resulting blob was then dissolved in 1 ml of fresh LB and 200 µl 1:20 and 1:40 dilutions plated onto rifampicin-kanamycin-streptomycin triple-antibiotic agar plates and incubated for 48 h at 32°C to eliminate the *E. coli* strains from the culture and allow for sufficient growth of transconjugant *P. denitrificans*. An area on the plate was then marked and the cultures within it scraped off and combined into the finalized *P. denitrificans* culture.

3.3. Microbiological culture

Expression always started with a 2 ml cryogenic culture, which was previously stored at -80°C and added to 200 ml LB supplemented with Kanamycin and Streptomycin. This culture was incubated overnight at 32°C and 165 rpm. The LB was centrifuged (Sorvall LYNX 6000, Thermo Scientific) off at 6000 g rcf for 20 and replaced with twice the amount of kanamycin and streptomycin containing succinate medium. Over the next two days, this culture was successively doubled and each time incubated overnight at 32°C and 165 rpm until the last main culture reached roughly 6 l in volume. After overnight incubation the entire culture was centrifuged at 4°C for 20 min at ≥ 8000 g rcf, the pellets scraped off, weighted and stored until use at -80°C. A typical expression yielded 8 g/l of wet cells.

3.4. Cell disruption and protein purification

For membrane purification, cells were thawed on ice and under continuous stirring in about twice their mass of 20 mM potassium phosphate buffer, 20 mM sodium chloride and supplemented with 200 μg/ml protease inhibitor Pefabloc SC (VWR Germany), a small amount of lyophilized DNase I (Sigma) and 100 μg/ml lysozyme (VWR Germany). The liquid was then loaded into a BeadBeater (BioSpec Products) and filled up to between 1/3-1/2 of the volume with 0.1 mm glass beads. The entire process was done in an 8°C cold room and the beating chamber placed in the manufacturer supplied cooling mantel, which was filled to the top with wet ice. The disruption process consisted of 5 min of total beating in 15 s intervals followed by 45 s of rest to keep temperatures sufficiently low. Afterwards, beads were filtered off by pouring the solution through a fritted glass funnel. To remove any remaining undisrupted cells and mechanical abrasion, the filtrate was subsequently centrifuged at 4°C and ≥ 8000 g rcf for 20 min. Cells were then ultracentrifuged (Sorvall WX Ultra 80, Thermo Scientific) at 4°C and ≥ 200,000 g rcf for 4–5 h. The supernatant was discarded and the dark brown gelatinous membrane weighted and carefully scraped off with a glass rod. The membrane was then supplemented with a small volume of 20 mM potassium phosphate buffer, 20 mM sodium chloride, 200 μg/ml Pefabloc SC and manually homogenized with a glass homogenizer. The solution was then cryogenically frozen with liquid nitrogen (LN_2) and stored at -80°C.

To next step was to solubilize the membrane protein. For this, the membrane was thawed and to this were added 3 % w/v β-dodecyl-maltoside and again 200 µg/ml Pefabloc SC and everything stirred vigorously at 8°C for 30 min. Insolubilized membrane was subsequently removed through ultracentrifugation at 4°C and ≥ 200,000 g rcf for 1 h followed by filtering through a 0.45 µm pipette tip filter. Subsequently, 10 mM imidazole were added. The first purification step was affinity chromatography through a (depending on the volume of the membrane) 1 ml or 5 ml HisTrap FF (GE Healthcare) Ni^{2+}-NTA column. The column was equilibrated in 20 mM potassium phosphate buffer, 20 mM sodium chloride, 20 mM imidazole, 0.05% β-DM (buffer A) and then loaded using a peristaltic pump. Afterwards, the column was washed with buffer A and installed into an ÄKTAprime plus system (GE Healthcare). The ÄKTA ran a linear gradient over 20 min at 1 ml/min from buffer A to buffer B (like buffer A, but containing 250 mM imidazole) and separated the eluate, which was already green in color, into 2 ml fractions. The main fractions were then pooled and washed in 20 mM potassium phosphate buffer, 20 mM sodium chloride, 0.05% β-DM (buffer E) through centrifugation in a 50 kDa concentrator (Amicon, Merck Millipore) to remove excess salt and imidazole. The next step was ion affinity chromatography through a 1 ml Q-Sepharose FF (GE Healthcare) column equilibrated in buffer E. The column was again loaded with a peristaltic pump and washed in buffer E and then the same gradient run as with the HisTrap column, however this time from buffer E to buffer F (same as E, but with 500 mM sodium chloride). The main fractions were again pooled and washed in buffer E. After determining the concentration and thus yield of the purification through UV-Vis spectroscopy, the enzyme was aliquoted, cryogenically frozen with LN_2 and stored at -80°C until use.

3.5. Spectroscopic characterization of mutant C*c*O

A dual-beam UV-Vis spectrometer (UV-2450-PC, Shimadzu) was used to characterize the mutant enzymes. Lambert-Beer's law (Eq. 3) allows for the determination of the concentrations of individual species in a given spectrum. The Soret band of oxidized C*c*O lies at 425 nm and shifts to 442 nm in the reduced state, which can be induced by adding small amounts of sodium dithionite. At the same time, the α-band shifts from 601 nm to 605 nm. The concentration of the enzyme can be determined from the Soret band of the oxidized spectrum with the absorbance coefficient[1] $\varepsilon_{435} = 158\ mM^{-1}\ cm^{-1}$ or from the reduced-minus-oxidized spectrum (henceforth "red-ox spectrum") with $\varepsilon_{606\text{-}630} = 25.7\ mM^{-1}\ cm^{-1}$.[60] For the determination of C*c*O's enzymatic activity, reduced horse heart cytochrome *c* was used at a concentration of ~100 µM in 10 mM Tris-HCl pH 7.4, 40 mM NaCl, and 0.05% β-DM at room temperature. The first order rate of cytochrome *c* oxidation was measured through the change in absorbance at 550 nm. The measurements had an accuracy of ~10 s^{-1}.

3.6. Bioconjugation with fluorescent probes

Labeling of C*c*O with 5-iodoacetamido-fluorescein (IAF, Thermo Scientific) was performed with twice the molar amount of reduction agent dithiothreitol (DTT) and a 100-times excess of IAF in in 60 mM Tris pH 8.0, 50 mM sodium chloride, 0.05% β-DM. The sample was left to react for one hour at room temperature. This was followed by size exclusion chromatography on a 5 ml Sephadex G-25 fine column to remove unreacted dye. For the 2-subunit preparation, 3% lauryldimethylamine oxide (LDAO) were added to the sample and incubated under heavy shaking for 1.5 h at 4°C. Then, a small 0.5 ml Ni^{2+}-NTA column (PureCube, Cube Biotech) was prepared and equilibrated it with 20 mM potassium phosphate buffer pH 8.0, 20 mM sodium chloride, 10 mM imidazole, 1.5 % LDAO. Afterwards, the sample was supplemented with 20 mM imidazole, added it to the column and washed out residual subunit 3 and 4 with two volumes of the equilibration buffer. To remove LDAO, the column and sample were washed thrice with wash buffer (20 mM potassium phosphate buffer pH 8.0, 20 mM sodium chloride, 10 mM imidazole, 0.05% β-DM) with ~15 times the volume of Ni^{2+}-NTA. The sample was eluted with the same buffer, however containing 250 mM imidazole. To remove the imidazole, two rounds of centrifugal concentration were performed with 50 kDa Amicon concentrators in wash buffer minus the imidazole.

Fluorescence labeling with 6-bromoacetyl-2-dimethylaminonaphtalene (BADAN, Thermo Scientific) followed the same procedure with some minor adjustments. Due to the low solubility of the dye, it could only be added to an excess of 5:1 and and the reaction was set up in absence of DTT. The reaction therefore had to incubate for two hours at room temperature. With this protocol, the labeling stoichiometries of C*c*O-ADAN were very similar for all labeled mutants to those of C*c*O-AF.

3.7. Polyacrylamide gel electrophoresis (PAGE)

Denaturing polyacrylamide gel electrophoresis was performed with self-cast discontinuous gels with a 5% acrylamide stacking gel and a 12% acrylamide separation gel in MES SDS running buffer (50 mM MES, 50 mM Tris, 1 mM EDTA, 0.1 % SDS). The sample was diluted in loading buffer (62.5 mM Tris-HCl pH 6.8, 2% SDS, 10% glycerol, 5 % β-mercaptoethanol) and run at 60V for ~90 min at RT (Mini-PROTEAN Tetra Vertical Electrophoresis Cell, Bio-Rad Laboratories, Inc.). For fluorophore-labeled samples, gels were then imaged under UV light and stained overnight in 10% acetic acid, 50% methanol, 0.1% Coomassie Brilliant Blue R 250. Afterwards, gels were destained for 6-8 h in 10% acetic acid, 40% methanol and subsequently washed with destilled water before imaging.

3.8. List of equipment, and media

The following tables list all tools (Table 2) and culture media (Table 3) used in this work. For more information on the time-resolved fluorescence spectroscopy setups, please see section 2.3.

Table 2: Tools and equipment used in this work

Name	**Manufacturer**
Absorbance spectrometer UV-2450-PC	Shimadzu, Japan
Fluorescence spectrometer Fluoromax-3	HORIBA Jobin-Yvon, Japan
Thermocycler Personal Cycler	Biometra, Germany
Cooling centrifuge Sorvall LYNX 6000	Thermo Scientific, United States of America
Ultracentrifuge Sorvall WX Ultra 80	Thermo Scientific, United States of America
Cell disruptor Beadbeater	BioSpec products, United States of America
Chromatography system ÄKTAprime plus	GE Healthcare, United States of America
Autoclav Varioclav 400	Thermo Scientific, United States of America

Table 3: Culture media and trace element solution

Name	**Composition and preparation**
Lysogeny broth (L.B. medium)	$1\%_{w/v}$ tryptone $0.5\%_{w/v}$ yeast extract $1\%_{w/v}$ NaCl for culture plates medium is supplemented with $1.5\%_{w/v}$ agar; sterilized by autoclaving
Succinate medium	50 mM K2HPO4 10 mM NH4Cl 1 mM MgSO4·x 7 H2O 1 mM citric acid 40 mM succinic acid $0.1\%_{v/v}$ trace element solution ad pH 6,2 with KOH; sterilized by autoclaving
Trace element solution	100 mM $CaCl_2$ 10 mM $CoCl_2$ 5 mM $CuCl_2$ 90 mM $FeCl_3$ 5 mM H_3BO_3 50 mM $MnCl_2$ 10 mM Na_2MoO_4 25 mM $ZnCl_2$ in 18.5% HCl
S.O.C. medium	2% vegetable peptone/tryptone 0.5% yeast extract 10 mM NaCl 2.5 mM KCl 10 mM $MgCl_2$ 10 mM $MgSO_4$ 20 mM glucose sterilized by autoclaving

3.9. List of bacterial strains and vectors

The following tables list the organisms (Table 4) and plasmids (Table 5) used in this work. Chapter 4 will go into more detail about the vectors and the individual cloning steps leading to their creation.

Table 4: Bacterial strains used in this work

Species + strain	**Genotype**
E. coli HST04	F- endA1 supE44 thi-1 recA1 relA1 gyrA96 phoA φ80d lacZΔM15 Δ(lacZYA-argF)U169 Δ(mrr-hsdRMS-mcrBC) ΔmcrA λ-
E. coli DH5a	F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80d lacZΔM15 Δ(lacZYA-argF)U169 hsdR17(rK-mK+) λ-
P. denitrificans ATCC13543	wildtype
P. denitrificans AO1	PD1222 derivative, ΔctaDI::Km^R, ΔctaDII::Tet^R, ΔccoN::Gm^R
P. denitrificans OR4	MR31 derivative (itself from PD1222), ΔctaDI::Km^R ΔctaDII::Tet^R Δcta-operon::Gm^R

Table 5: Vectors

Vector	**Important features**	**Source**
pKH163	pBBR1 ori, Sm^R, Cm^R, ctaDII::His$_5$	derivative of pRI2[78]
pAW1	pKH163 derivative, ctaDIIC48SC140S	This work
pAW6	ATCC13543 ctaCBGESurF1 cloned into pAW1	This work
pAW6WT	pAW6 w/ ctaDII::His$_5$ from pKH163	This work
pAW9	pAW6 derivative, ctaCC79SC85S + ctaEC185SC231S	This work
pAW11	pAW6WT derivative, 2nd cta promoter inserted before ctaCBGESurF1	This work
pAW13	pAW11 derivative, 2nd cta terminator inserted after ctaDII::His$_5$, ΔCm^R	This work
pAW16	pAW9 derivative w/ inserted cta terminator-promoter of pAW13	This work

4. Results and Discussion IA: Establishing of an expression system for background-free thiolreactive labeling

Site-specific fluorescence labeling of proteins requires the existence of a base protein that in itself does not get labeled by the fluorophore, which means that the labeling is “background-free”. The establishment of a background-free C*c*O base mutant is the subject of this chapter. Prior research by Kirchberg et al. suggested that subunit 2 of the *P. denitrificans* C*c*O was inaccessible to labeling with the thiolreactive dye 5-IAF.[1] Background-free labeling could thus be established by conservatively mutating the accessible and non-essential cysteines of subunit 1 ($C48^{A}S$ and $C140^{A}S$), since any labeling of subunits 3 and 4 could be removed through the separation of these subunits from the enzyme core, which consists of subunits 1 and 2.[42] This was, however, not feasible, as demonstrated below. The sections in this chapter will describe the progression of steps necessary for achieving site-specific labeling without interfering background fluorescence.

4.1. Recombinant expression of C*c*O subunit 1

Initially, and to perform fluorescence labeling of C*c*O, the vector pAW1 was constructed, which contains subunit 1 of the oxidase and the two cysteine-to-serine mutations, which was subsequently conjugated into a subunit 1-deletion strain (AO1) through triparental conjugation. First labeling experiments of the resulting C*c*O with fluorescein, however, revealed that subunit 2 is labeled with high yield, as one can see in the PAGE image of Figure 7A (left). The cause of this difference in labeling behavior might be the differences between both expression systems, as e.g. the stoichiometry between C*c*O subunits and genes involved assembly was found to have significant impact on the kinetics of C*c*O’s catalase activity.[79]

The presence of unwanted subunit 2 labeling is neither impacted by different concentrations of reduction agent nor can it be avoided by using fluorescein-5-maleimide (FMA, Figure 7A, right). As subunit 3 can be separated from the two core subunits alongside subunit 4, the equally high amount of acetamido-fluorescein bound to it does not impose a barrier to the spectroscopic experiments. Due to subunits 2 and 3 of C*c*O belonging to the same operon, I therefore chose to construct a vector containing the first three subunits of the oxidase to further exchange surface exposed cysteines in the subunits 2 and 3. Such vector is compatible with the host strain OR4, which has deletions for all three genes.

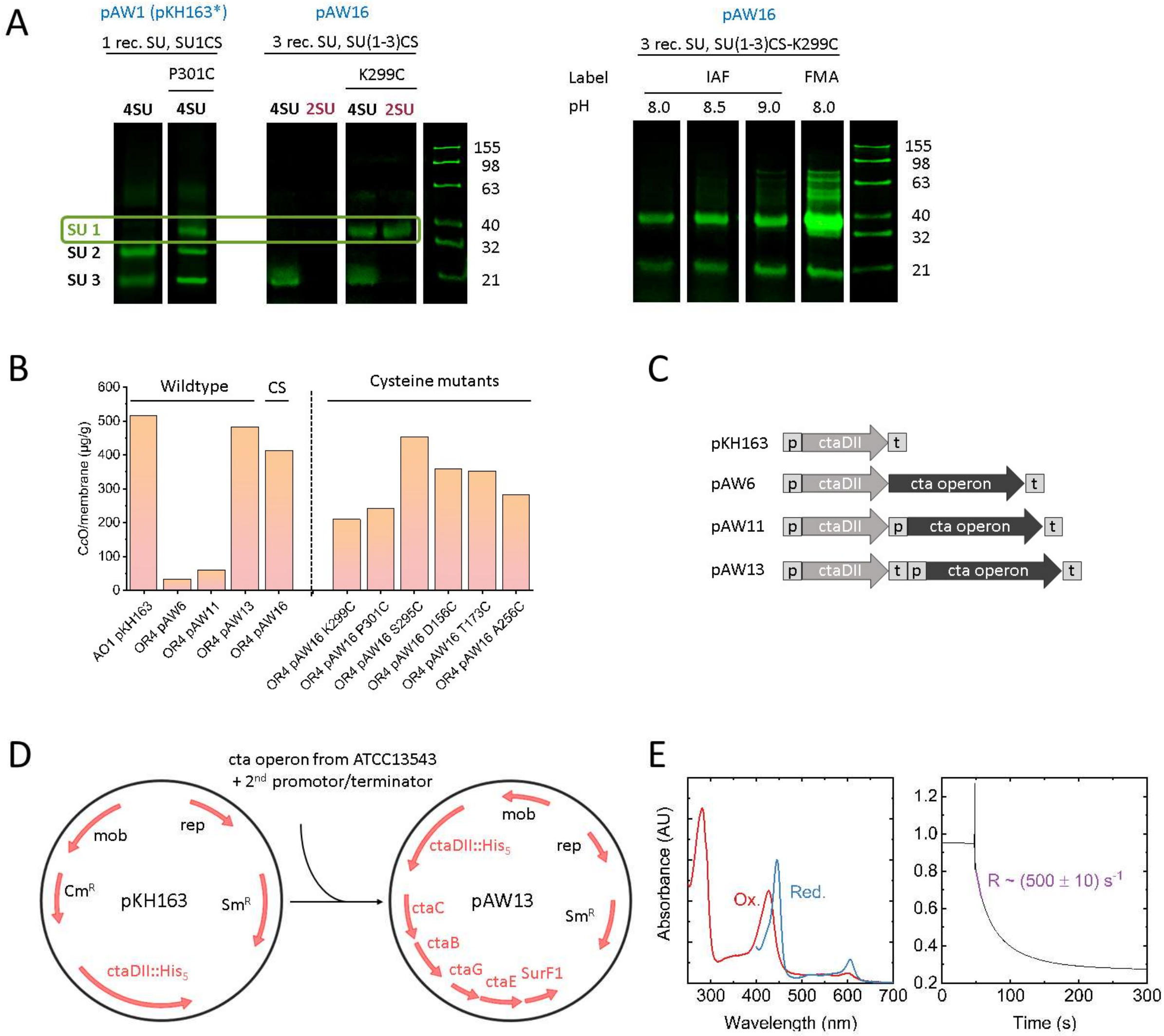


Figure 7: (A) PAGE image of SU1CS and SU(1-3)CS single-cysteine C*c*O constructs incubated with thiolreactive IAF (left) and under different labeling conditions including using FMA instead of IAF (right, 80 pmol C*c*O per lane). (B) Expression yield of C*c*O constructs: Expression levels comparable to AO1 pKH163 with only subunit 1 provided *in trans* are only restored after insertion of both a second promoter and terminator. (C) Structure of the new constructs: p/t denote promoter/terminator elements (light gray), elements belonging to ctaDII are colored in medium gray, those belonging to the cta operon in dark gray. (D) Vector maps of the original vector pKH163 and the finalized product pAW13 after three subsequent cloning steps (the gene conferring chloramphenicol resistance was deleted in the last step to reduce the size of pAW13). (E) Oxidized (red) and reduced (blue) absorbance spectrum and cytochrome *c* reduction kinetics of C*c*O from AO1 pKH163 (only subunit 1 recombinantly expressed).

4.2. Recombinant expression of C*c*O subunits 1-3

Paracoccus denitrificans has two isoforms of subunit 1 of C*c*O encoded in the genes ctaDI and ctaDII. However, ctaDI is not expressed under any known physiological condition even in ΔctaDII mutants, except when the gene is provided *in trans*, possibly due to the copy number of the plasmids.[76] Subunits 2 and 3 as well as proteins involved in C*c*O assembly[80] are encoded in the single ctaCBGESurF1 operon. In this operon, ctaC encodes for subunit 2, ctaB for protoheme-IX-farnesyltransferase, ctaG for an assembly factor implicated in Cu_B insertion, ctaE for subunit 3, and SurF1 for an assembly factor involved in heme *a* insertion.[81] Subunit 4 is not part of this operon as well as other important genes for C*c*O biosynthesis like heme *a* synthase (ctaA), and the Cu_A insertion factors (ScoA and ScoB).[82]

I therefore extracted the ctaCBGESurF1 operon from the genome of *P. denitrificans* strain ATCC13543, a wild-type strain likely from the same original isolate as ATCC17741[83], of which PD1222 is a descendent.[84] All primers used for cloning and mutagenesis can be found in Supplement 1. Sequencing after extraction of the operon revealed that the ctaCBGESurF1 operon is indeed genetically identical to the same operon of PD1222. Interestingly, in formerly used vectors like pKH160 (a fragment containing ctaC from this vector was used by Kirchberg, et al.[1]), the operon appears to have originated in a different strain, most probably S1657 (sequence from [85]). This leads to some differences in the genes ctaB (1 insertion and 1 deletion resulting in a short framshift over 10 bases, 1 additional different base), ctaG (3 deleted nucleotides, 2 additional different bases), and ctaE (1 different base). These mutations amount to six, two, and one changed amino acid, respectively, when compared to the PD1222 genome. A comparison between the two vectors can be found in Supplement 2.

Starting from the vector pKH163/pAW1, which contains ctaDII with a C-terminal hexahistidine-tag (5 histidines added after the final histidine of ctaDII), the ctaCBGESurF1 operon was inserted directly downstream of ctaDII::His_5 before its cta terminator in order to construct a superoperon analogous to the makeup of pKH160. I used restriction-free cloning to produce cloning sites identical to those of pKH160. Unfortunately, expression of this vector pAW6WT in OR4 yields only very small amounts of protein of undesirable purity (Figure 7B). Therefore, the superoperon was separated and both ctaDII::His_5 and the ctaCBGESurF1 operon flanked by a cta promoter and cta operon terminator. This process (Figure 7C/D) involved two sequential cloning steps, the first one adding a second ctaDII promoter in front of the ctaCBGESurF1 operon (pAW11) and the second one refitting ctaDII::His_5 with the cta terminator (pAW13). As can be seen in Figure

7B, adding the second promoter increases the expression yield slightly (and thus improves purity), however only the configuration in pAW13 with two separate promoter/terminator elements for ctaDII::His$_5$ and ctaCBGESurF1 restores the yield to the original level of AO1 pKH163, where only subunit 1 is provided *in trans*. Importantly, C*c*O expressed in OR4 pAW13 shows wild-type activity and spectra (Figure 9/Figure 7E). Thus, a homologous expression system for the mutagenesis of C*c*O subunits 1-3 was established.

4.3. Cysteine mutants for site-specific fluorescence labeling

Site-specific fluorescence labeling with thiolreactive dyes is only possible if the target protein is devoid of any other surface-accessible cysteines. Previous results by Kirchberg, et al.[1] showed that exchange of cysteines 48 and 140 in subunit 1 is sufficient to eliminate background labeling. The other two cysteines in subunit 1, C66^A and C80^A, form a disulfide bridge and are thus inaccessible under physiological conditions. This result was reproduced with vector pAW1 (Figure 7A, first lane). The addition of a single cysteine such as P301C therefore yields background-free labeling of subunit 1 (Figure 7A, second lane). The fluorescence from this band, however, is not significantly stronger than the background fluorescence of subunits 2 and 3. Subunit 4 does not contain any cysteines and is not labeled by fluorescein. Therefore, vector pAW13 was expected to allow for the exchange of all fluorophore-accessible cysteines that cause the background fluorescence.

In subunit 2, C216^B and C220^B bind the Cu$_A$ ligand and are therefore both essential and inaccessible. The other two cysteines, C50^B and C56^B, are close to the protein surface and were therefore exchanged into serines. The same applies to the only two cysteines of subunit, C184^C and C230^C. The resulting vector was pAW16 (full map and sequence as well as the protein sequences for C*c*O subunits 1-3 in the Supplements 3 and 4) and indeed eliminated background fluorescence in subunit 2 (Figure 7A, third lane). However, fluorescence in subunit 3 remains and is therefore not the result of cysteine-labeling, but of a reaction with another amino acid. These unspecific reactions can occur with (mostly terminal) amines and imidazoles and under certain conditions be preferred over binding to sulfhydryls.[86] In those cases, changing the pH of the labeling reaction might reduce the yield of the unspecific reaction. I therefore varied the pH during fluorescein-conjugation between pH 8.0 and 9.0. While these conditions should shift the ratio of specific to unspecific binding due to the favored deprotonation of the protein's sulfhydryl group, there was unspecific binding of the fluorophore under all tested conditions (Figure 7A, right). Labeling with fluorescein-5-maleimide (FMA, Thermo Scientific), a fluorescein derivative with a different linker, at pH 8.0 increased the

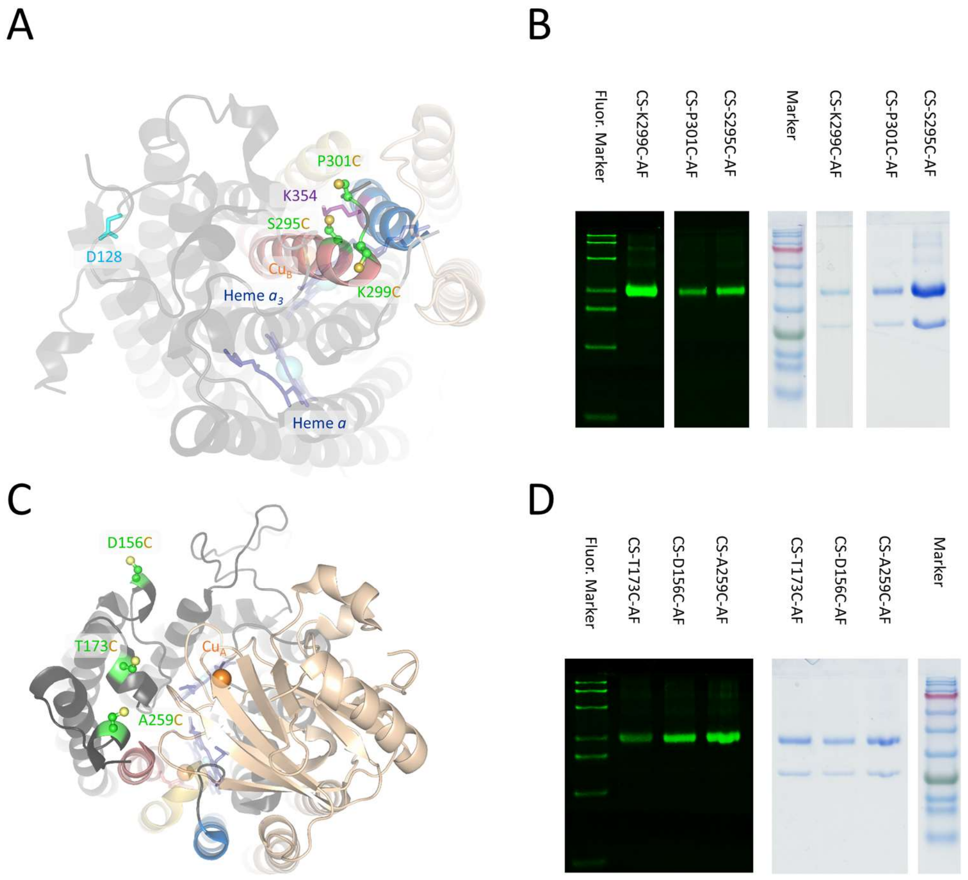


Figure 8: (A) Structural view above the K-channel with cysteine mutant labeling sites S295C, K299C, and P301C in green (sulfur atoms for the thiolreactive labeling reaction in yellow). (B) 12% PAGE gels of the three mutants both under UV illumination (left) and after Coomassie blue staining (right). Fluorescent marker bands (in kDa): 155, 98, 63, 40, 32, 21, 11. Prestained marker bands: 245, 190, 135, 100, 80, 58, 46, 32, 25, 22, 17, 11. Fluorescence gel: 18 pmol K299C-AF, 5 pmol P301C-AF/S295C-AF Coomassie stained gel: 18 pmol K299C-AF, 40 pmol P301C-AF, 150 pmol S295C-AF (C) Structural view of the P-side with labeling mutants D156C, T173C, and A259C. (D) 12% PAGE gels of the P-side mutants. Both gels: 53 pmol T173C-AF, 33 pmol D156C-AF, 49 pmol A259C-AF.

overall labeling stoichiometry including the background labeling stoichiometry and did thus not provide a benefit over using IAF.

For all site-specific labeling experiments, subunits 3 (and 4, which is attached to subunit 3) therefore had to be separated from the enzyme, which is possible by using a strong detergent like LDAO. The separation was always performed after the size-exclusion chromatographic purification of the labeled enzyme, as this further allowed to eliminate any possibly unreacted fluorescein sticking to the enzyme through hydrophobic interaction. The resulting C*c*O was background-free and could be labeled specifically with high yield (Figure 7A, lanes 3-6).

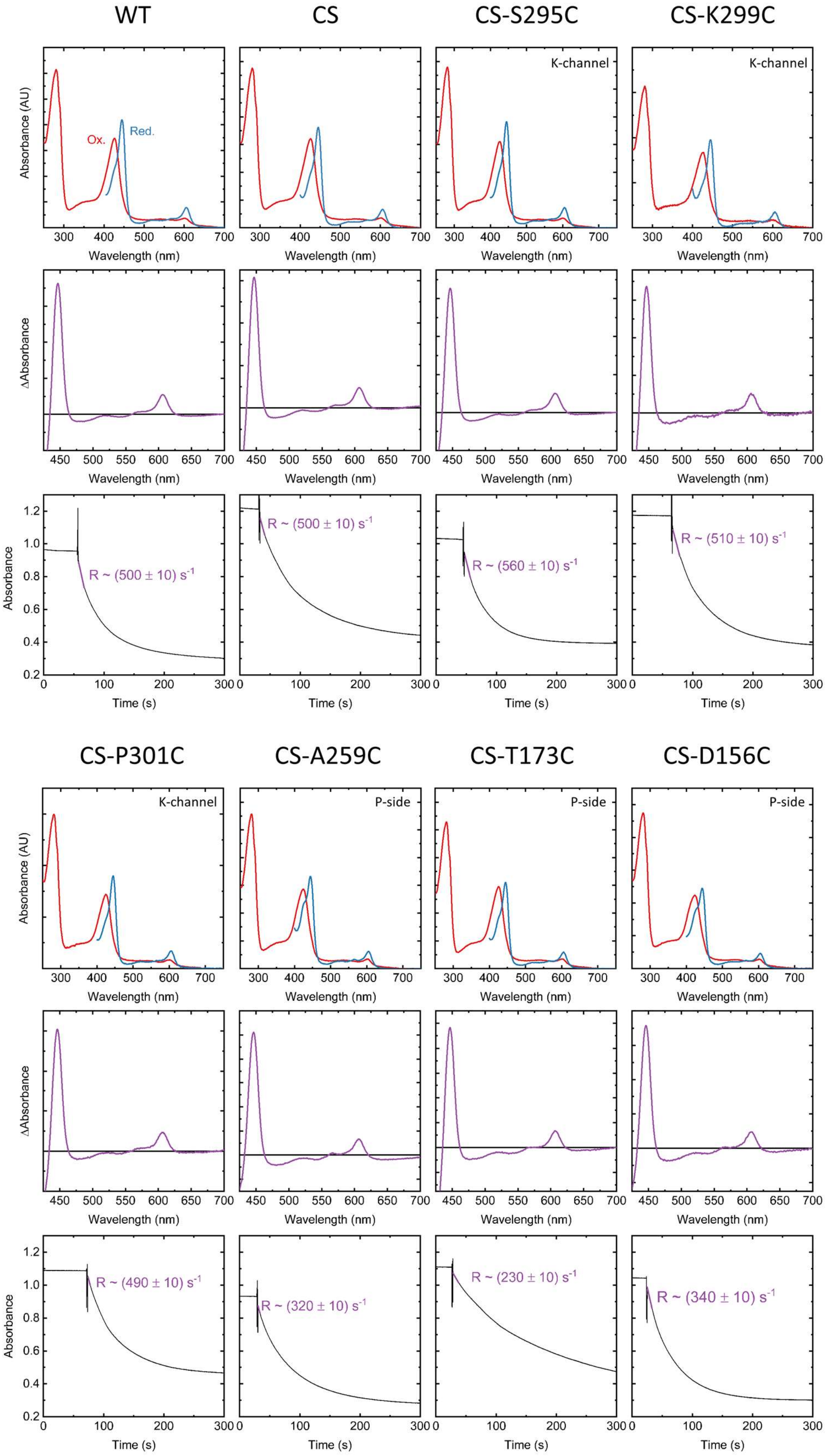


Figure 9: First and fourth row: Reduced and oxidized spectra of WT, CS, and single-cysteine mutant C*c*O from the subunit 1-3 recombinant expression system. Second and fifth row: Reduced-minus-oxidized difference spectra. Third and sixth row: Spectral characterization of cytochrome *c* reduction by the mutants monitored at 550 nm

In order to probe the protein surface above the K-channel, I constructed the N-side mutants S295^{A}C, K299^{A}C, P301^{A}C, which are located at the end of helix 6 and on the loop connecting helix 6 with helix 7 (Figure 8A). Helix 6 harbors the "catalytic" tyrosine 280, which is covalently bound to histidine 276.[44] This tyrosine lends one electron and one proton to the catalytic reaction of the enzyme. These sites are also close to lysine 354 in helix 8, the essential residue of the K-channel and are therefore well suited to monitor changes in and around the K-channel without disturbing its activity. Indeed, all three N-side mutants show wild-type catalytic activity (Figure 9, Table 6).

To perform reference measurements on the side of proton release, for future experiments with covalently attached electron donor molecules, as well as the possibility to attach the enzyme to electrodes, I further created P-side mutants. Results by Kirchberg[87] have indicated that P-side cysteine mutants in subunit 2 are not able to bind IAF. I therefore chose three solvent-exposed loops/helices in subunit 1 as targets, D156^{A}C, T173^{A}C, and A259^{A}C. Mutating residues at the P-side of subunit 1 might interfere with C*c*O assembly or efficient interaction with cytochrome *c*. A potential role in cytochrome *c* binding was proposed for D156^A, which is part of a cluster of negatively charged residues thought to inititate contact between the two proteins.[88] Figure 10 shows the positions of all three residues in bovine C*c*O[89], which are close to the bound cytochrome *c*. Structural changes, or changes in the distribution of surface charges due to the mutations might therefore result in reduced C*c*O-cytochrome *c* electron transfer rates. Indeed, while all three mutants can be expressed with high yield (Figure 7B), they all exhibit reduced levels of cytochrome *c* oxidation activity (Figure 9, Table 6).

Table 6: Single-cysteine C*c*O mutants

Mutant	Soret band maximum (nm)	Turnover rate of cytochrome *c* oxidation (s^{-1})
WT	426 ± 0.5	500 ± 10
CS	426 ± 0.5	500 ± 10
CS-S295C	426 ± 0.5	560 ± 10
CS-K299C	426 ± 0.5	510 ± 10
CS-P301C	425 ± 0.5	490 ± 10
CS-D156C	426 ± 0.5	340 ± 10
CS-T173C	424 ± 0.5	230 ± 10
CS-A259C	425 ± 0.5	320 ± 10

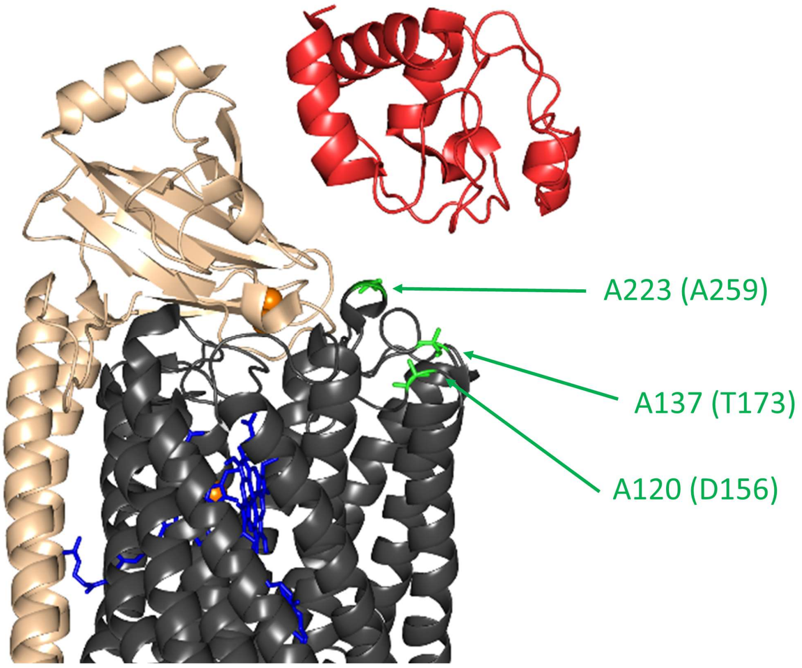


Figure 10: Crystal structure of bovine[88] cytochrome *c* (red) and C*c*O subunits 1 (gray) and 2 (beige) showing the corresponding residues to the P-side mutants of this work (in brackets) on the bovine enzyme. Hemes are shown in stick representation and blue color and coppers as orange spheres. PDB code: 5IY5

5. Results and Discussion IB: Surface protonation, hydration and conformational dynamics

Authorship Statement

All experiments in this chapter were performed by me (Alexander Wolf). For studying the selective activation of the K-channel, a collaboration with the group of Prof. Dr. Ernst-Walter Knapp Knapp (Chemistry Department at the Freie Universität Berlin) was established. Jovan Dragelj of the Chemistry Department performed the MD simulations and electrostatic energy computations. Prof. Dr. Ulrike Alexiev (Physics Department) and Prof. Dr. Ernst-Walter supervised the joint project.

In this chapter, the dye-conjugated mutants established in the previous chapter are characterized and their absorbance and fluorescence parameters used to gather site-specific information about local conditions in the O and R state of C*c*O. Section 5.1. focusses on conformational changes determined with fluorescein-labeled C*c*O. Section 5.2. then presents the results from BADAN-labeled C*c*O showing redox-dependent changes in the local hydration dynamics.

5.1. Redox-dependent conformational changes

The position of fluorescein's absorbance maximum strongly depends on the local electrostatics in the vicinity of the label, more specifically on the local relative permittivity.[64] The higher the wavelength of the absorbance maximum, the less polar is the label's surrounding. This redshift is due to the fact that in the case of fluorescein's n-π* transition the solvent interacts most strongly with the molecule's ground state in the transition from polar to nonpolar medium, where it increases the solvation energy. The steady-state absorbance maximum of the C*c*O-AF conjugates can therefore serve to analyze the protein's surface electrostatics, changes of which might indicate the fluorophore adapting a different conformation in relation to the protein surface, e.g. facing toward the solution or its relation to the hydrophobic/hydrophilic parts of the detergent micelle.[64]

Fluorescein is further suited to sense nanoenvironmental changes due to its pronation state-dependent absorbance and fluorescence characteristics. As the fluorophore interacts with the surface of proteins, e.g. through H-bonds and electrostatic interactions, its pK_a adjusts to the local surface environment. This pK_a change in turn manifests itself in the pH-dependent absorbance spectrum of the dye. One can fit the change in absorbance of fluorescein depending on the solution's pH by applying the Henderson-Hasselbalch equation for one protonatable group:

$$\Delta A_\lambda(\mathrm{pH}) = A_\lambda + B_\lambda \times \frac{10^{-\mathrm{p}K_\mathrm{a}}}{10^{-\mathrm{p}K_\mathrm{a}} + 10^{-\mathrm{pH}}} \quad , \tag{10}$$

where ΔA_λ is the change in absorbance at a wavelength λ, A_λ is the absorbance of the protonated species, and B_λ is the difference in extinction of the deprotonated and protonated species. By normalizing the data between $A_{\lambda,N+1}$ and $B_{\lambda,N}$ one obtains a graph depicting the deprotonation probability of the group.

The N-side mutants $S295^{A}C$, $K299^{A}C$, and $P301^{A}C$ (henceforth without the notation for their location in subunit 1) of C*c*O and their acetamido-fluorescein conjugates are well suited to observe redox-dependent structural changes close to the K-channel entrance. The three residues fulfill important criteria (Figure 8A): (1) they lie close to the essential lysine 354, (2) they are, however, sufficiently far away not to disturb the proton translocation process and are not known to be involved in catalysis, and (3) they are located at the N-terminal tip of transmembrane helix 6 and the helix 6-helix 7 loop. The former harbors the catalytic tyrosine 280, which is covalently bound to histidine 276 on the same helix. Upon reduction of C*c*O, Y280 is thought to dissociate from the hydroxy group of the heme *a*$_3$ farnesyl.[6, 90]

Yet, theoretical analyses of the *P. denitrificans* and *R. sphaeroides* C*c*O crystal structures have suggested that the helix is fairly rigid towards the N-side.[91] The authors, however, found structural changes and an increase in conformational flexibility in the neighboring helices 7, 8, and 9 that also line the K-pathway. As one can see in Figure 8B,D, all three residues of the respective mutants can be labeled with high specificity due to their sites being sufficiently solvent exposed. The stoichiometries under standard labeling conditions (see methods section) are 98% (S295C-AF), 55% (K299C-AF), and 70% (P301C-AF). Under the same conditions, the three P-side mutants (Figure 8C,D) reach labeling stoichiometries of 29% (D156C-AF), 15% (T173C-AF), and 28% (A259C-AF), potentially due to the densely spaced negatively charged carboxylates on CcO's P-side increasing the p*K*$_a$ of the cystein sulfhydryl groups. Since these need to be deprotonated in order for the acetamido-fluorescein binding to occur, an increase in their p*K*$_a$ hinders the labeling process. After labeling, the fluorophore's absorbance results in additional peak at ~500 nm superimposed on the absorbance caused by the heme cofactors of C*c*O (Figure 11A). The unlabeled protein's absorbance spectrum can therefore be subtracted to yield the background-corrected fluorescein spectra and thus the wavelength maximum of the conjugated fluorescein (Table 7). Expectedly, due to its more inward lying position on the enzyme, of the three N-side mutant conjugates S295C-AF senses the most nonpolar nanoenvironment, as apparent from its comparatively red absorbance maximum of (500.7 ± 0.5) nm, which is also larger than or equal to all the P-side mutants when in the O-state (Figure 11B). This value is comparable to fluorescein labeled residues belonging to helices D and E of bacteriorhodopsin while the fluorescein absorbance maxima of K299C-AF and P301C-AF with (497.8 ± 0.5) nmand (499.3 ± 0.5) nm, respectively, resemble those in the DE-loop of bacteriorhodopsin.[64] Interestingly, all P-side spectra

Table 7: Fluorescein peak positions on labeled C*c*O

Mutant	λ_{max} in O state (nm)	λ_{max} in R state (nm)
CS-S295C	500.7 ± 0.5	500.5 ± 0.5
CS-K299C	497.8 ± 0.5	497.5 ± 0.5
CS-P301C	499.3 ± 0.5	498.9 ± 0.5
CS-D156C	499.5 ± 0.5	497.9 ± 0.5
CS-T173C	500.7 ± 0.5	499.8 ± 0.5
CS-A259C	499.7 ± 0.5	497.9 ± 0.5

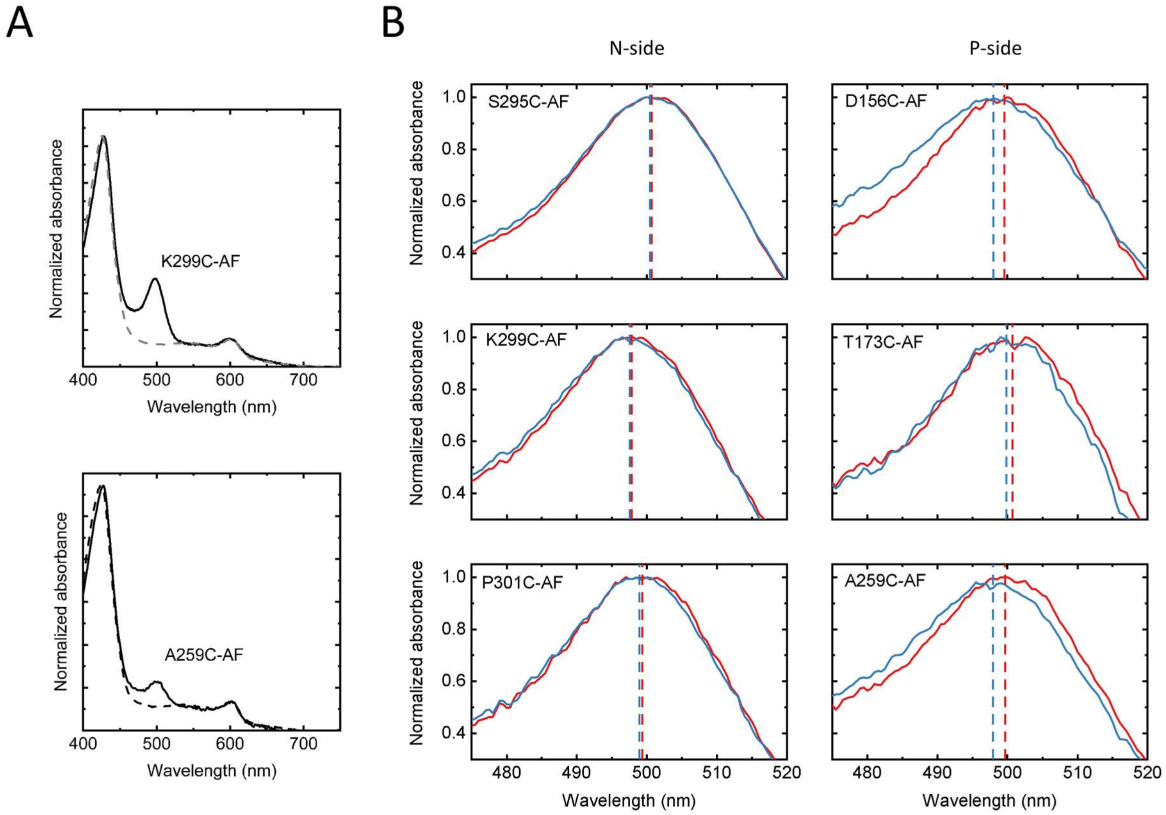


Figure 11: (A) Normalized absorbance spectra of K299C both with the attached fluorescein (2 subunits, full line) and unlabeled (4 subunit enzyme, dashed line). (B) Superimposed oxidized (red) and reduced (blue) spectra of the six IAF-labeled C*c*O mutants with maxima positions in dashed lines. Experimental conditions: 60 mM Tris-HCl pH 9.0, 20 mM sodium chloride, 0.05% β-DM.

are blueshifted in the R-state. The P-side conjugates, D156C-AF, T173C-AF, A259C-AF shift by (-2 ± 1) nm, -(1 ± 1) nm, and (-2 ± 1) nm, respectively. The three N-side mutants, however, exhibit shifts significantly below the measuring accuracy. These results run contrary to those obtained for D-channel and K-channel labeling positions by Kirchberg et al., who saw a global nonpolar shift upon reduction of C*c*O.[1] The reason for these differences is not immediately clear and may be the result of significant differences between the proteins from both expression systems (as became apparent from the dissimilar extent of background labeling observed for both wildtype enzymes). The origin of the small negative shift is not immediately obvious, as one would expect more of a distribution of both batho- and hypsochromic shifts from the oxidized to the reduced state. Possibly, electronation of the enzyme's catalytic center results in a significant change towards a more polar microenvironment at the P-side of the protein, while it does not affect the N-side surface of the K-channel in the same way.

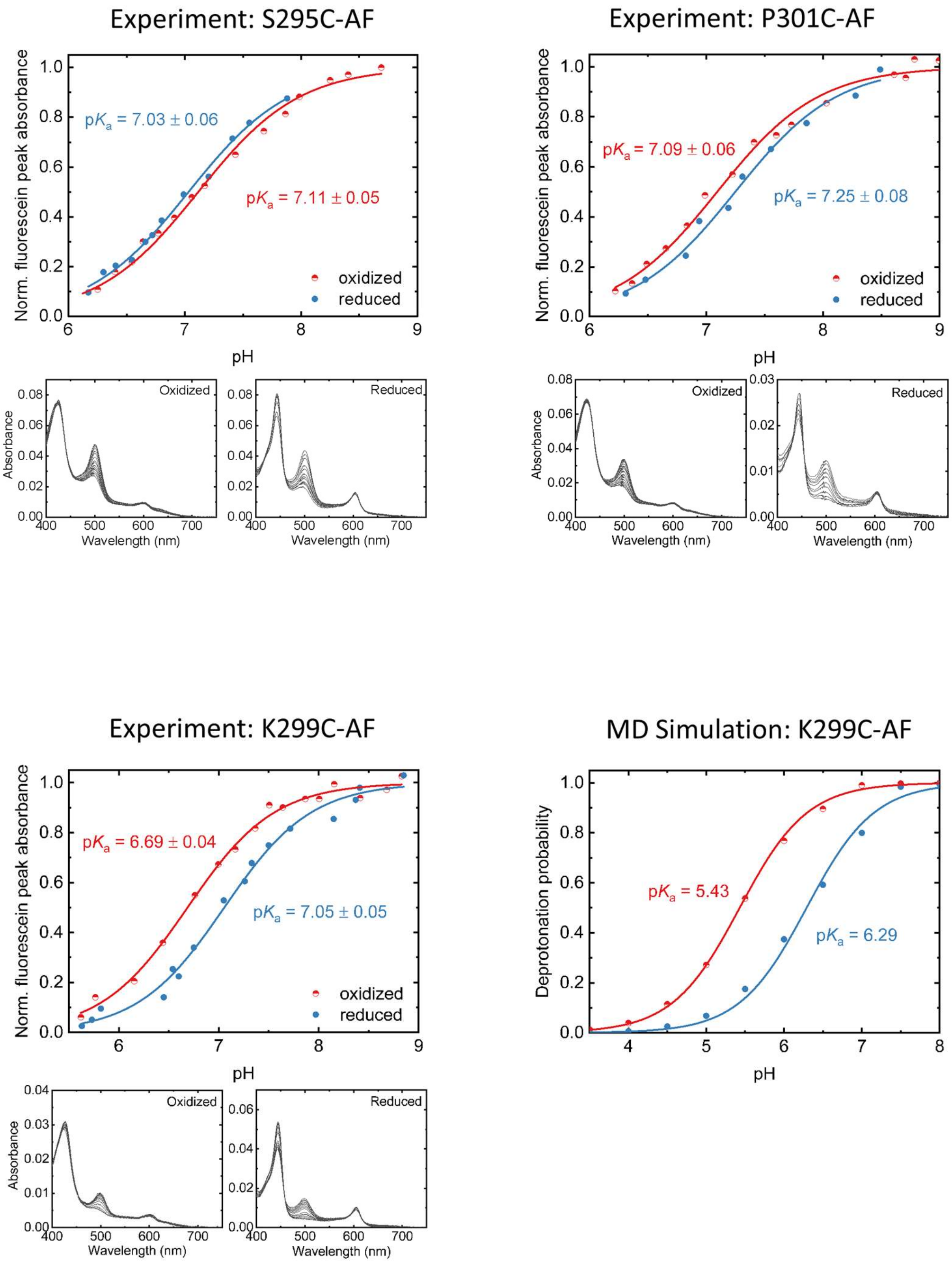


Figure 12: Normalized redox-dependent absorbance titration spectra (2nd and 4th row) and C*c*O-AF intensity data (1st and 3rd row) with Henderson-Hasselbalch fit curves and resulting pK_a values with fit errors for three N-side C*c*O conjugates (top left and right, bottom left). Absorption bands belonging to C*c*O's cofactors do not vary under titration conditions between 485 nm and 505 nm as shown by our group (see e.g. Figure 5C in Kirchberg et al.[1]) and thus do not affect the titration behavior of fluorescein.. Deprotonation probabilities for K299C-AF from MD simulations by J. Dragelj with corresponding pK_a values (bottom right). Experimental conditions: 2 mM potassium phosphate buffer, 100 mM NaCl, 0.05% β-DM.

As apparent in the oxidized peak positions of the P-side mutants, the low labeling stoichiometries are not fully explained by a more hydrophobic N-side surface.

The equilibrium value of fluorescein's de-/protonation transition can serve as an indicator of structural as well as charge changes in the label's nanoenvironment. Due to the low labeling stoichiometry of the P-side mutants, however, pH titrations are not possible, since a lowering in pH results in a further decrease of the label's extinction coefficent. The following experiments will therefore focus on the N-side mutants. Upon titration, the pH-dependent amplitudes of fluorescein's absorbance maximum in the spectra of the three N-side conjugates each display a distinct behavior (Figure 12). In the case of S295C-AF, there is nearly no change in the proton affinity of the dye between the O and R state, as both the oxidized value $\mathrm{p}K_\mathrm{a}(\mathrm{O}) = 7.11 \pm 0.05$ and the reduced $\mathrm{p}K_\mathrm{a}(\mathrm{R}) = 7.03 \pm 0.06$ lie within the margin of error after fitting with equation (10). P301C-AF exhibits a small upshift of $\Delta\mathrm{p}K_\mathrm{a} = +0.16 \pm 0.14$ from $\mathrm{p}K_\mathrm{a}(\mathrm{O}) = 7.09 \pm 0.06$ to $\mathrm{p}K_\mathrm{a}(\mathrm{R}) = 7.25 \pm 0.08$. The largest upshift of $\Delta\mathrm{p}K_\mathrm{a} = +0.36 \pm 0.09$ is observed for K299C-AF, which increases from $\mathrm{p}K_\mathrm{a}(\mathrm{O}) = 6.69 \pm 0.04$ to $\mathrm{p}K_\mathrm{a}(\mathrm{R}) = 7.05 \pm 0.05$. By comparing the six $\mathrm{p}K_\mathrm{a}$ values, it becomes clear that the biggest difference seems to be the low value for oxidized K299C-AF.

Molecular Dynamics (MD) simulations are a well-established method for investigating structural changes in proteins. Together with electrostatic energy computations, one can gain insight into both structural changes in the protein and the $\mathrm{p}K_\mathrm{a}$ value of individual groups, which was demonstrated in C*c*O for the essential K354.[4] Here, the $\mathrm{p}K_\mathrm{a}$ values calculated from MD simulations are $\mathrm{p}K_\mathrm{a}(\mathrm{O}) = 5.43$ and $\mathrm{p}K_\mathrm{a}(\mathrm{R}) = 6.29$ and agree very well with the ones obtained in the experiment (Figure 12, bottom). While there is a quantitative difference, the resulting redox-dependent shift of $\Delta\mathrm{p}K_\mathrm{a} = +0.86$ has the same sign and does not exceed one pH unit. Due to the nature of the MD simulations, where the charge of C*c*O's cofactors can be directly manipulated, the cause of the $\mathrm{p}K_\mathrm{a}$ change was suggested to be a conformational change starting at the enzyme's cofactors and propagating through the enzyme to the surface, where fluorescein is bound. This conformational change correlates with a change in hydrogen bonds (H-bonds) as even small conformational changes can rearrange the structure of a protein's H-bond network[92] and thus change the $\mathrm{p}K_\mathrm{a}$ of the bound probe.

To establish the existence of the conformational changes at the surface one can measure the time-resolved fluorescence depolarization decay of the sample with picosecond time-resolved fluorescence spectroscopy (see chapter 2.3.1. for the experimental details). Measuring the time-resolved fluorescence anisotropy of K299C-AF with TCSPC indeed reveals an increase in conformational flexibility of the label and its attachment site upon enzyme reduction (Figure 13A) manifesting itself in the stronger decay of the system's anisotropy. By fitting the curves with equation (9) and re-plotting them without the tumbling motion of the system one can compare the

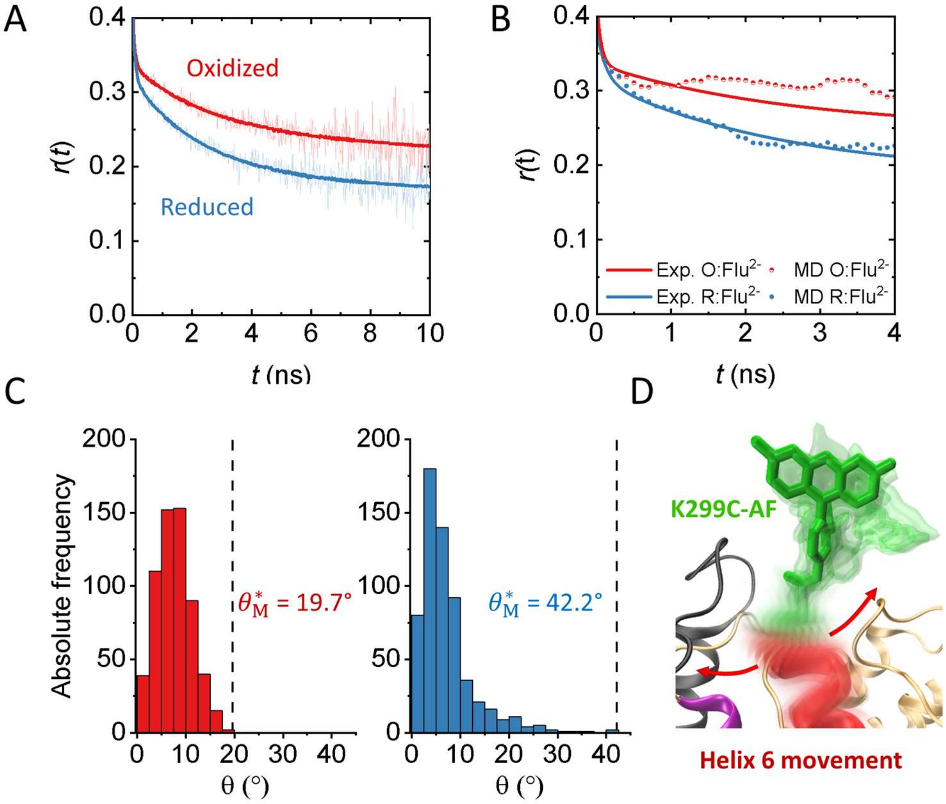


Figure 13: (A) Time-resolved fluorescence anisotropy of oxidized (red) and reduced (blue) K299C-AF with corresponding fit curves. (B) Constructed experimental anisotropies from (A) without whole-protein rotation and corresponding anisotropies generated from MD simulations for C*c*O-bound dianionic fluorescein. (C) Distribution of angular frequencies from the MD simulations and maximal semicone angles θ_M (dotted lines) reached within the simulation time. (D) Structural view of the wobbling motion of K299C-AF. All MD simulation data and the C*c*O-AF conjugated structure were provided by J. Dragelj. Experimental conditions: 20 mM potassium phosphate buffer pH 8.0, 20 mM NaCl, 0.05% β-DM.

experimental data (Figure 13B, lines) to one calculated from the simulation (dotted curves). One can immediately observe the good qualitative agreement between both methods within the first 4 ns. For longer times, the simulated anisotropies start to strongly fluctuate due to the ever-decreasing amount of available data points. For the reduced enzyme, both curves practically align, while the simulation slightly overestimates the anisotropy of oxidized C*c*O, potentially due to the limited amount of available data. Drawing a vector between the cysteine-299 sulfur and the central distal oxygen in fluorescein's triple ring system and tracking the deviation of that vector from its initial position over the course of the simulation produces two angular frequency distributions for both enzymatic states (Figure 13C). The distribution for oxidized K299C-AF is much narrower than that of the reduced enzyme, which has a broad slope towards larger angles. The maximal angle θ^*_M reached during the simulation is 19.7° for oxidized and 42.2° for reduced C*c*O. We can compare these angles to the angles θ_M calculated after equation (8) from the experimental anisotropy decays,

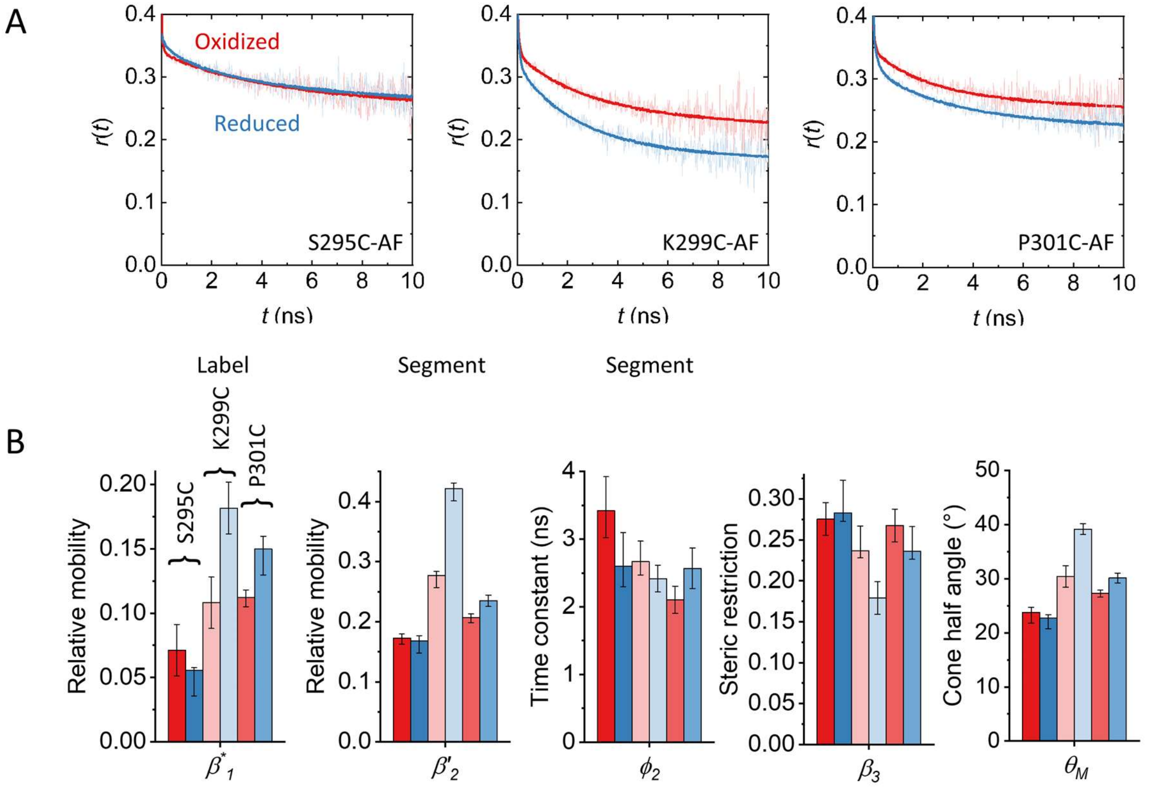


Figure 14: (A) Time-resolved fluorescence anisotropiesof oxidized (red) and reduced (blue)S295C-AF, K299C-AF, and P301C-AF with corresponding fit curves. (B) Experimental parametersdetermined from the anisotropy fits for all three fluorescein-conjugated mutants. Segmental (helix 6) relative mobility is calculated as $\beta_2' = \beta_2/(\beta_2+\beta_3)$. Experimental conditions: 20 mM potassium phosphate buffer pH 8.0, 20 mM NaCl, 0.05% β-DM. Experimental errors were determined through exhaustive error analysis[71] (see Table 8).

which increase from from 30^{+2}_{-2} ° to 39^{+1}_{-1} ° upon reduction of the enzyme (Figure 14B). Again, both the experimental and MD derived values are nearly identical for reduced C*c*O, while the angle of O-state C*c*O generated from the MD simulation is slightly lower than its experimental counterpart. In conclusion, both in experiment and simulation, the fluorescent dye-helix-6-conjugate experiences lower steric restriction from its protein surrounding when C*c*O's cofactors are reduced. To determine the component causing this expansion in the conjugate's conformational space one can use the cone-in-a-cone model. From this it becomes apparent that the increase is not due to the interaction of the dye itself, since the semiangle of the inner cone, which directly belongs to the fluorophore, does not shift strongly between redox states ((21.6 ± 0.3)° in the oxidized, (20.5 ± 0.2)° in the reduced state). What does change, however, is the semiangle of the outer cone attributable to the movement of the protein segment the dye is attached to. This angle significantly increases from (25.4 ± 0.1)° to (33.3 ± 0.1)° upon chemical reduction and thus signifies an increase in flexibility of the helix 6 tip and the beginning of the helix 6-7 loop (Figure 13D, red helix and arrows).

One can further characterize the observable conformational changes by examining the experimental anisotropy parameters of K299C-AF and of the nearby labeling sites of S295C-AF and P301C-AF (Figure 14A,B, Table 8). Expectedly, S295C-AF, which is situated deeper within the enzyme on helix 6, has the lowest conformational flexibility of all three systems. Most of the parameters describing the relative mobility of the system only slightly decrease upon enzyme reduction. This is true for the motions of the label itself (β^*_1) and also for the helix segment it is attached to ($\beta_2' = \beta_2/(\beta_2+\beta_3)$). At the same time, however, the correlation time φ_2 of the segments' motion stays remains within measurement uncertainty ($3.4^{+0.5}_{-0.4}$ ns in O, $2.6^{+0.5}_{-0.3}$ ns in R) indicating that it is mostly the steric restriction (absolute amplitude $\beta_3 = r_0 \times \beta^*_3$) imposed by neighboring helices that limits the conformational space of the S295 helix segment. This is not the case anymore one helix turn outwards at labeling position 299. Both the relative mobility of the label itself and of the helix segment (in this case its tip) increase dramatically by ~70% and ~50%, respectively, upon C*c*O reduction. At the same time, the segmental dynamics remain within measurement uncertainty ($2.7^{+0.3}_{-0.3}$ ns in O, $2.4^{+0.2}_{-0.2}$ in R) and the overall steric restriction decreases by more than ~20%. On the helix 6-7 loop, at position 301, only the label's mobility increases from O→R by ~30% while the loop segments' mobility decreases slightly and dynamics remain within measurement uncertainty.

Taken together, the time-resolved anisotropy results at different protein positions point to a pronounced reduction-induced uptake in conformational flexibility at the tip of helix 6, potentially connected to a partial unwinding of its N-side tip, and partially extending into the helix 6-7 loop. This increase is potentially linked to higher flexibility and heightened movement of the nearby helices 7 and 8, which was suggested by MD simulations.[7, 91] Especially helix 8, which harbors the essential K-channel residue K354, is thought to undergo a conformational change between the O

Table 8: Anisotropy fit parameters. φ_3 was always fixed at 135 ns between measurements. Confidence levels were determined with exhaustive error analysis using the program Global Unlimited V2.2 (Laboratory for Fluorescence Dynamics, University of Illinois, USA). The probabilities p of the confidence levels of the standard deviation are applied, here with p = 68.27% (1σ).

Sample	State	r_0	β_1^*	β_2'	β_3^*	φ_1 (ns)	φ_2 (ns)	Θ_M (°)	χ^2_{red}
S295C-AF	Ox.	$0.36^{+0.03}_{-0.003}$	$0.07^{+0.02}_{-0.02}$	$0.17^{+0.007}_{-0.01}$	$0.77^{+0.02}_{-0.02}$	$0.08^{+0.03}_{-0.02}$	$3.4^{+0.5}_{-0.4}$	24^{+1}_{-2}	1.18
	Red.	$0.36^{+0.007}_{-0.03}$	$0.06^{+0.002}_{-0.02}$	$0.17^{+0.009}_{-0.02}$	$0.79^{+0.01}_{-0.03}$	$0.25^{+0.06}_{-0.06}$	$2.6^{+0.5}_{-0.3}$	$23^{+0.6}_{-2}$	1.21
K299C-AF	Ox.	$0.37^{+0.02}_{-0.009}$	$0.11^{+0.02}_{-0.02}$	$0.28^{+0.007}_{-0.02}$	$0.64^{+0.03}_{-0.03}$	$0.09^{+0.03}_{-0.02}$	$2.7^{+0.3}_{-0.3}$	30^{+2}_{-2}	1.19
	Red.	$0.38^{+0.03}_{-0.006}$	$0.18^{+0.02}_{-0.02}$	$0.42^{+0.01}_{-0.02}$	$0.47^{+0.02}_{-0.02}$	$0.10^{+0.02}_{-0.02}$	$2.4^{+0.2}_{-0.2}$	39^{+1}_{-1}	1.04
P301C-AF	Ox.	$0.38^{+0.02}_{-0.02}$	$0.112^{+0.006}_{-0.007}$	$0.207^{+0.006}_{-0.009}$	$0.70^{+0.02}_{-0.02}$	$0.08^{+0.02}_{-0.02}$	$2.1^{+0.2}_{-0.2}$	$27.3^{+0.6}_{-0.7}$	1.03
	Red.	$0.36^{+0.04}_{-0.001}$	$0.15^{+0.01}_{-0.02}$	$0.24^{+0.009}_{-0.01}$	$0.65^{+0.02}_{-0.02}$	$0.16^{+0.02}_{-0.006}$	$2.6^{+0.3}_{-0.3}$	$30.1^{+0.9}_{-0.9}$	1.09

and R states of C*c*O.[4, 5, 7, 90] All of these changes affect the conformational space of helix 6 lining the K-channel. An important aspect of conformational flexibility is how it impacts the hydration of nearby sections on the surface and inside an enzyme, which is of special relevance to the K-channel and for which results will be presented in the next section.

5.2. Hydration dynamics at the enzyme surface

In its function as a proton pump C*c*O is dependent on a unidirectional flow of hydrogen ions through its proton pathways. These pathways need to connect both sides of the protein in a way that prevents the backflow of protons along the chemical gradient. Hydrogen bonds between titratable side chains and occupant water molecules provide the infrastructure for passive proton transport. This configuration can clearly be observed in the D-channel of C*c*O from the N-side surface to the interior E278.[44, 93] Additionally, water wires may form in hydrophobic cavities inside of proteins due to the preferred interaction of each water molecule with its two neighbors. Water wires facilitate a net transport of protons significantly faster than in bulk water via a Grotthuss-like mechanism.[94, 95] This mechanism was proposed for proton transport from E278 in C*c*O to either the BNC or the PLS under control of the latter's changing electric field depending on the redox state of the enzyme.[96] Importantly, the diffusivity of these protons depends on the geometry of the water network. Thus, as both the orientation of water molecules in water wires and the hydration of cavities and channels inside the protein are naturally linked to its conformation, the proton conductivity of these pathways is not necessarily static and can be gauged through state-dependent conformational changes. An example for the importance of selective hydration for protein function can be found in the light-driven proton pump bacteriorhodopsin. In the M_2 state of bacteriorhodopsin's photocycle, a water molecule enters a hydrophobic cavity via random walk and thus facilitates intra-protein proton transfer. [97]

While the degree of retardation in the dynamics of water molecules in the first hydration shell of proteins is the matter of an ongoing controversy[98], distinct functional roles were found for this hydration shell, e.g. in the transient interaction of cytochrome *c* with C*c*O[89]. Covalent attachment of solvatochromic probes to the the surface of proteins is well suited to study the characteristics of the hydration shell. One common application to this technique is the investigation of changes in the solvent microenvironment of proteins upon ligand binding.[99] The scope of this technique extends farther than just proteins, as the conjugation of solvatochromic dyes with DNA was used to study the binding of several DNA binding proteins[100] and the formation of lipoplexes[101]. The

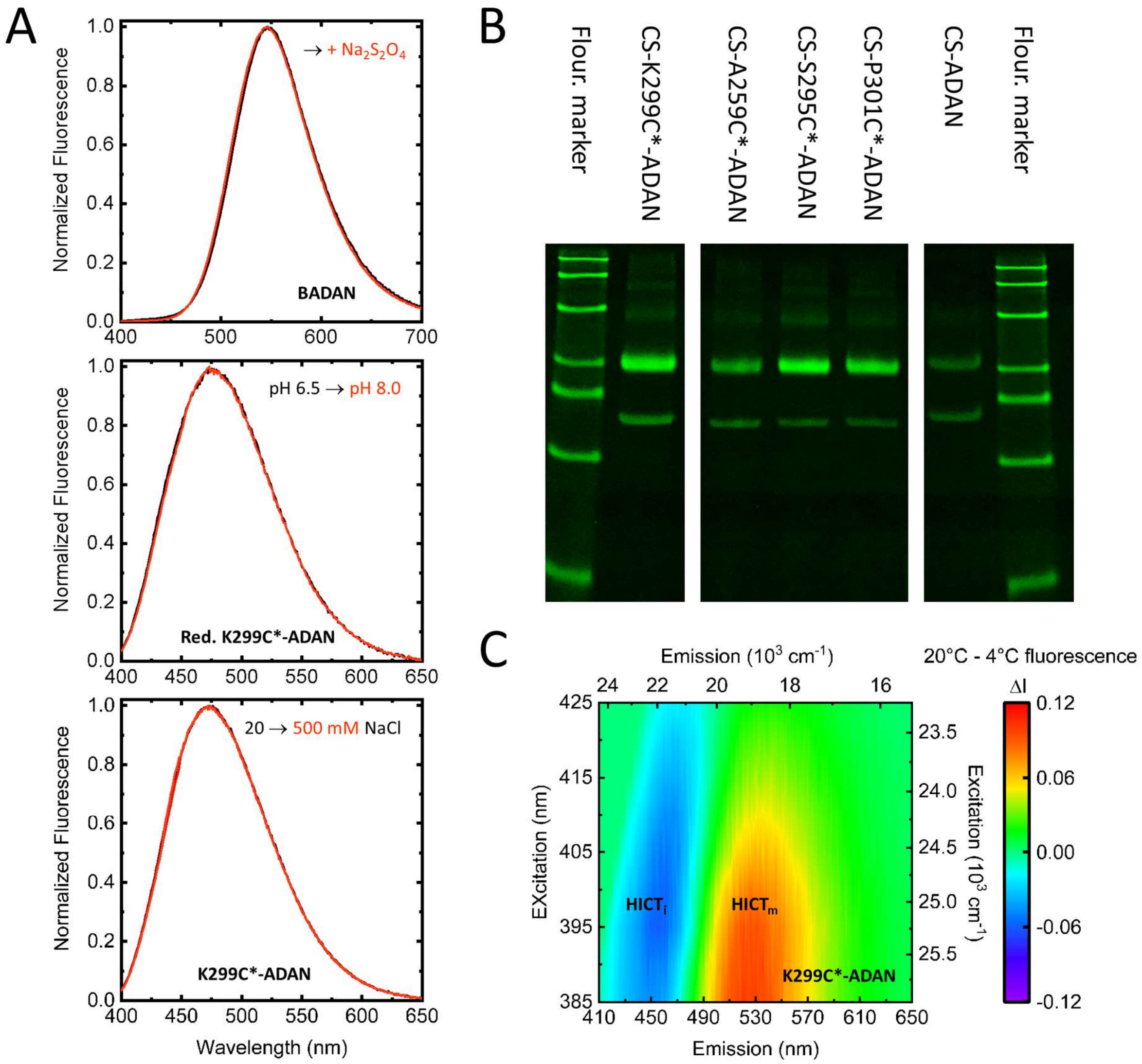


Figure 15: (A) Normalized emission spectra of BADAN in 50%$_{v/v}$ methanol, 10 mM potassium phosphate buffer pH 8.0, 10 mM NaCl with and without sodium dithionite added (top) and C*c*O-K299C*-ADAN at different pH values (center) and NaCl concentrations (bottom). (B) PAGE image under UV illumination of BADAN-labeled two-subunit C*c*O mutants (30 pmoles of enzyme per lane). Fluorescent marker bands (in kDa): 155, 98, 63, 40, 32, 21, 11. (C) Excitation-emission temperature difference scans of oxidized C*c*O-K299C*-ADAN. The star denotes spectra without CS-ADAN background subtraction. Experimental conditions for K299C*-ADAN (if not stated otherwise): 20 mM potassium phosphate buffer pH 8.0, 20 mM NaCl, 0.05% β-DM.

binding of ligands generally results in a hypsochromic shift due to the displacement of (bulk-like) water in the microenvironment of the label.

Conjugation of C*c*O with BADAN was successfully performed for the three N-side mutants S285C, K299C, and P301C as well as for the P-side mutant A259C. Due to the low extinction coefficient of BADAN[99] (ε_{387} = 21 mM^{-1} cm^{-1}) and its position within C*c*O's chromophore absorbance (Soret band maximum ε_{425} = 158 mM^{-1} cm^{-1}), determination of the labeling stoichiometry can only performed in an approximate manner. For CS-K299C*-ADAN, CS-

S295C*-ADAN, and CS-P301C*-ADAN stoichiometries lie between 50-60%. CS-A259C*-ADAN is labeled to 35-40%. The star denotes that there is also unspecific binding to other positions in both subunit 1 and 2 of the protein, as one can observe in the PAGE bands of Figure 15B. The tendency of this dye to bind to off-site residues was also seen for other single-cysteine mutant proteins.[69] By labelling the accessible-cysteine-free CS mutant it was possible to determine the fluorescence from unspecific labeling, which only reached a third of the fluorescence emitted from the N-side BADAN-labeled mutants and about half of the overall fluorescence of A259C*-ADAN. The CS-ADAN fluorescence was subsequently always subtracted yielding the spectra henceforth denoted without the star.

To investigate potential redox-induced changes in C*c*O-ADAN fluorescence, it was first necessary to rule out unspecific spectral variations due to slight differences in the experimental conditions. Sodium dithionite, the reduction agent used to prepare R(-like)-state C*c*O, does not influence the shape of the emission spectrum of free BADAN in $50\%_{v/v}$ methanol (Figure 15A, top). There is, however, a homogenous increase in fluorescence. When investigating the free dye, it is necessary not to perform the experiments in purely aqueous solution, as the dye is only poorly soluble in H_2O and the resulting aggregates have a more complex and blue-shifted emission than the free dye. Notwithstanding the absence of a spectral shift, it should be noted that dithionite addition limits the spectral range of excitation due to its increasingly strong absorbance below 390 nm. Further, it was necessary to measure the fluorescence of C*c*O-ADAN in low (20 mM) and high (500 mM) salt conditions, since the addition of sodium dithionite also increases the ionic strength of the solution. There was no observable spectral change between both conditions (Figure 15A, bottom). Expectedly, the shape of C*c*O-ADAN emission spectra is also independent of pH (shown for K299C*-ADAN in Figure 15A, center) for all four labeled mutants in the range where the enzyme is stable in detergent micelles (pH 6.5-8.8). The emission of BADAN is further dependent on the temperature of the system, as lower temperatures decrease the dynamics of the solvent. This dynamics can also be observed for C*c*O-ADAN (Figure 15C). In the temperature difference excitation-emission scan generated by measuring the sample's emission at different excitation wavelengths at temperatures of 4°C and 20°C and subtracting the former fluorescence from the latter, a decrease in the emission around 22000 cm^{-1} can be observed in favor of a broad increase in fluorescence centered around 19000 cm^{-1}. As this change can be directly attributed to the viscosity of the solvent, it can serve as an appropriate comparison to spectral differences observed in state-dependent CcO-ADAN measurements. To rule out solvent temperature effects, the temperature of all redox-dependent fluorescence measurements was stabilized at 20°C.

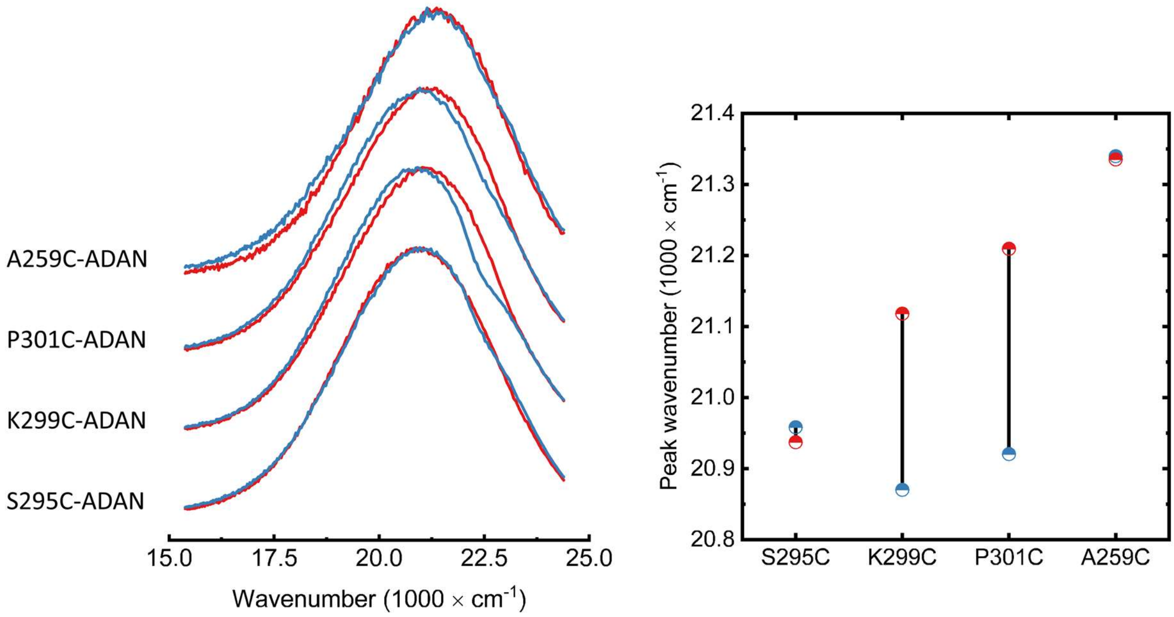


Figure 16: Left: Normalized oxidized (red) and reduced (blue) C*c*O-ADAN emission spectra at 390 nm excitation. Right: Peak wavenumbers of the spectra. Experimental conditions: 20 mM potassium phosphate buffer, 20 mM NaCl, 0.05% β-DM.

All four probed C*c*O-ADAN samples exhibit different fluorescence emission maxima in the oxidized state of the enzyme, thus hinting at differing solvent microenvironments (Figure 16, red curves). All maxima lie roughly centered between the emission of BADAN in nonpolar and polar solvents (Figure 4B). The most red-shifted (compared to nonpolar emission) maximum can be found on S295C-ADAN with 20940 cm^{-1}. The other maxima are 21120 cm^{-1} for K299C-ADAN, 21210 cm^{-1} for P301C-ADAN, and the least shifted maximum with 21340 cm^{-1} for A259C-ADAN. Notably, emission maxima do appear not to correlate with solvent exposure, as the most red-shifted emission can be observed for S295C-ADAN, which should be the least solvent exposed labeling position according to the crystal structure[44]. Especially A259C-ADAN on the P-side of the enzyme should be highly solvent exposed. These results, however, correlate with the low labeling stoichiometry of A259C-ADAN and the three P-side fluorescein-conjugated mutants D156C-AF, T173C-AF, and A259C-AF. Upon C*c*O reduction (Figure 16, blue curves), emission maxima of both S295C-ADAN with (+20 ± 30) cm^{-1} and A259C-ADAN with (+0 ± 30) cm^{-1} do not undergo a significant shift. The emission spectra of K299C-ADAN and P301C-ADAN, however, experience significant red-shifts of –(250 ± 30) cm^{-1} and (-290 ± 30) cm^{-1}, respectively. In contrast to the shifts observed for free BADAN in solvents of different polarities, these shifts do not encompass the entire emission spectrum, but originate below 23000 cm^{-1} and produce a noticeable blue shoulder.

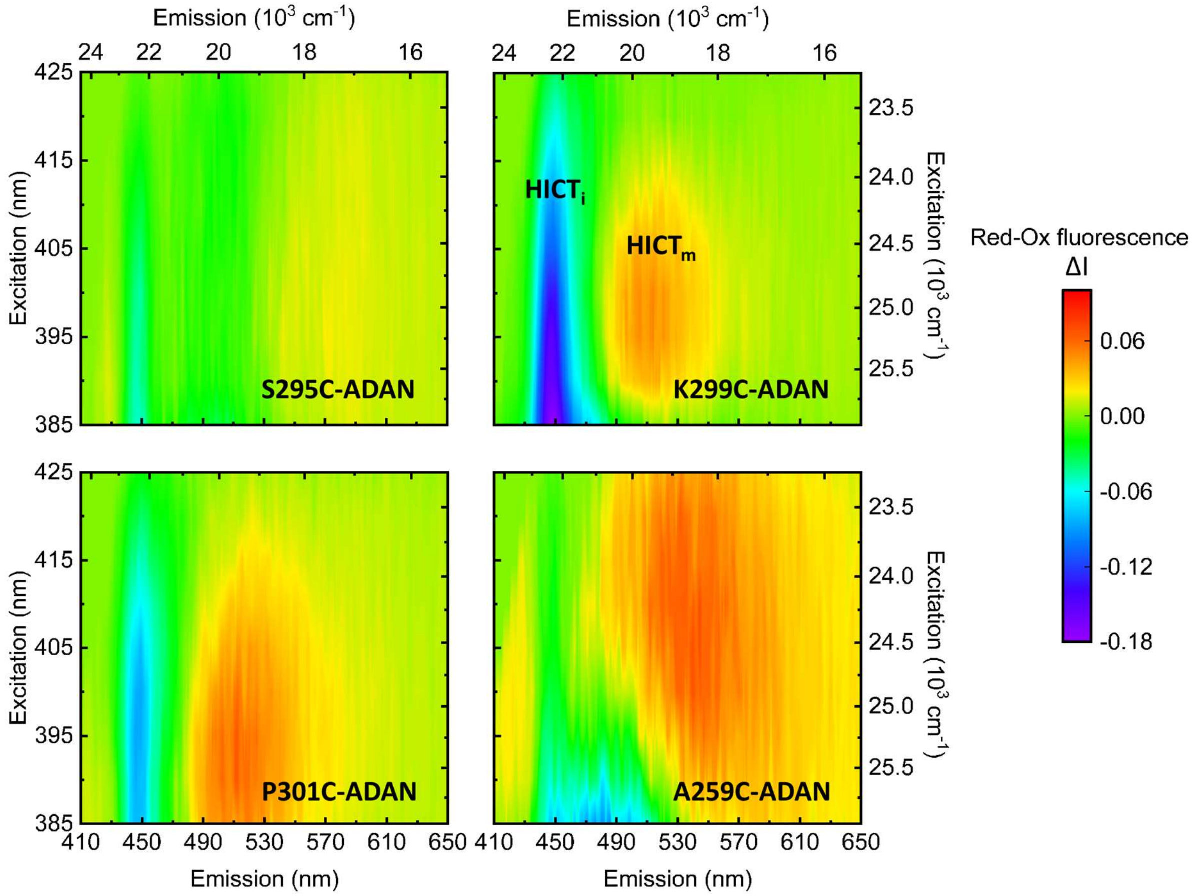


Figure 17: Red-Ox difference excitation-emission scans of the four BADAN-labeled mutants. Experimental conditions: 20 mM potassium phosphate buffer, 20 mM NaCl, 0.05% β-DM.

To further investigate the nature of these shifts, I performed red-ox difference excitation-emission scans, i.e. a subtraction of the oxidized from the reduced normalized excitation-emission spectra of C*c*O-ADAN (Figure 17). In the case of S295C-ADAN, there is no significant change within the probed spectral range, except for a slight decrease in a narrow band at ~22300 cm^{-1} emission. The difference scan of K299C-ADAN, on the other hand, sees the emergence of a broad peak centered around 25000 cm^{-1} excitation and 19500 cm^{-1} emission with a width on both axes of ~2000 cm^{-1}. At the same time, a sharp decrease in fluorescence can be observed at 22300 cm^{-1} stretching over nearly the entire probed excitation range. Comparing these values to the approximate wavenumbers for BADAN's energetic transitions (Figure 4B) it emerges that the fluorescence increase can be assigned to the HICT$_m$ species while the strong decrease belongs to the HICT$_i$ transition. The same effect is observable for P301C-ADAN with, however, a less pronounced decrease in the HICT$_i$ band and slightly more blue-shifted HICT$_m$ peak. Both the P301C-ADAN and K299C-ADAN red-ox difference scans are very similar to the temperature difference scan (Figure 15C), hinting at an increase in solvent mobility in the vicinity of both labels upon enzyme reduction. The red-ox difference scan at position 259 on the P-side of the enzyme displays a different behavior. There is only a miniscule decrease in fluorescence in the HICT$_i$ band and a very broad and strongly red-shifted increase centered around 18000 cm^{-1} emission and 24000-24500 cm^{-1} excitation. This hints at a different mechanism as the cause for the spectral change, potentially connected to a larger conformational rearrangement.

An oft-used tool to study the components of a complex fluorescence spectrum is its decomposition with a linear combination of log-normal functions[102]:

$$
\begin{aligned}
I_j(\vartheta) &= I_{m,j} \times \exp\left(-\left(\frac{\ln 2}{\ln^2 \rho_j}\right) \times \ln^2\left(\frac{a_j - \vartheta}{a_j - \vartheta_{m,j}}\right)\right), \qquad \vartheta < a_j \\
I_j(\vartheta) &= 0, \qquad \vartheta \geq a_j
\end{aligned}
\tag{11}
$$

$I_{m,j}$ is the peak intensity of a fluorescence species j, $\vartheta_{m,j}$ is its maximum intensity wavenumber, a_j is a function limiting point, and $\rho_j = (\vartheta_{m,j} - \vartheta_{-,j})/(\vartheta_{+,j} - \vartheta_{m,j})$ is the band asymmetry parameter with the wavenumber positions of the left and right half maximal amplitudes $\vartheta_{-,j}$, and $\vartheta_{+,j}$. This method has also been used to investigate the fluorescent species of the lipophilic BADAN analog LAURDAN.[103] I chose the three transition wavenumbers for the ICT, HICT$_i$, and HICT$_m$ transitions, 23000 cm^{-1}, 22000 cm^{-1}, and 20000 cm^{-1}, respectively, as initial values for the intensity maxima. In contrast to similar, albeit unmodified Gaussian fits of protein-conjugated BADAN, it was not sufficient to keep these values for the ICT and HICT$_i$ species, $\vartheta_{m,1}$ and $\vartheta_{m,2}$, fixed (they were, however, fixed between the measurements of each mutant in the oxidized and reduced enzymatic state), which leads to a blue-shift of the ICT species maximum and a red-shift for that of the HICT$_i$ species. Notably, however, these values, $\vartheta_{m,1}$ ~ 23300 cm^{-1} and $\vartheta_{m,2}$ ~ 21800 cm^{-1}, were highly consistent for all four mutants and similar to their initial values

Table 9: Fit parameters obtained from the spectral decomposition of oxidized (O) and reduced (R) C*c*O-ADAN samples according to equation (11). Corresponding fits can be found in Figure 18A and Suppement 5.

Sample	$I_{m,1}$	ϱ_1	a_1 (10^3 cm^{-1})	$\nu_{m,1}$ (10^3 cm^{-1})	$I_{m,2}$	ϱ_2	a_2 (10^3 cm^{-1})	$\nu_{m,2}$ (10^3 cm^{-1})	$I_{m,3}$	ϱ_3	a_3 (10^3 cm^{-1})	$\nu_{m,3}$ (10^3 cm^{-1})
S295C:O	0.237	0.988	121.2	23.25	0.656	0.933	38.56	21.75	0.718	0.787	26.12	19.95
S295C:R	0.252	0.986	98.65	23.25	0.604	0.997	477.8	21.75	0.733	0.781	26.25	20.04
K299C:O	0.165	0.933	43.10	23.28	0.692	0.995	270.8	21.82	0.693	0.779	25.95	19.96
K299C:R	0.306	0.965	49.59	23.28	0.461	0.997	377.7	21.82	0.867	0.763	25.92	20.25
P301C:O	0.203	0.984	93.60	23.30	0.696	0.928	37.86	21.84	0.662	0.782	26.31	20.01
P301C:R	0.290	0.994	173.5	23.30	0.556	0.937	37.93	21.84	0.792	0.774	26.25	20.14
A259C:O	0.232	0.967	59.53	23.31	0.736	0.920	36.45	21.79	0.554	0.783	26.00	20.05
A259C:R	0.310	0.994	219.9	23.31	0.646	0.937	39.52	21.79	0.606	0.765	26.16	20.13

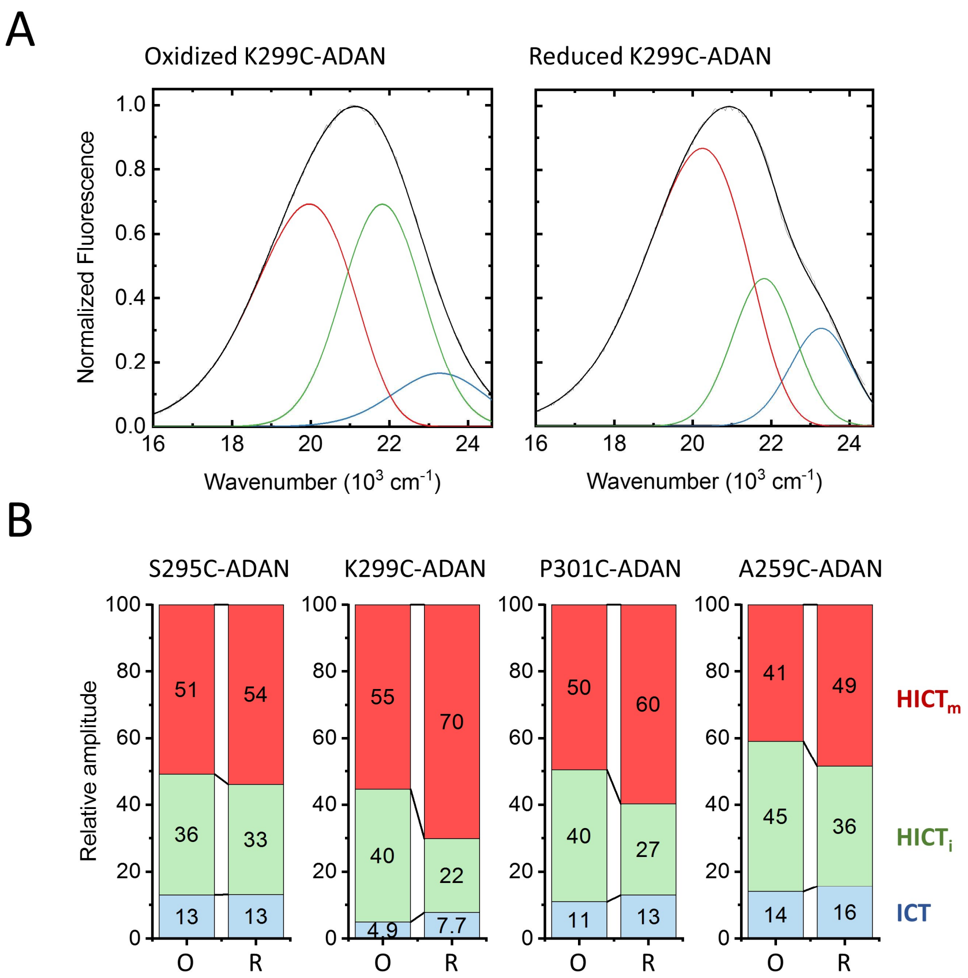


Figure 18: (A) Spectral decomposition of K299C-ADAN according to equation (11) with three Log-normal components colored in red, green and blue in accordance with the ICT, HICT$_i$, and HICT$_m$ components of BADAN. (B) Relative amplitudes generated from spectral decomposition of the C*c*O-ADAN mutants. Experimental conditions: 20 mM potassium phosphate buffer, 20 mM NaCl, 0.05% β-DM.

(Table 9) and thus remained assigned to their respective species. Figure 18A shows the spectral decomposition for K299C-ADAN in the oxidized and reduced state of the enzyme with the three components color-coded in blue (ICT), green (HICT$_i$), and red (HICT$_m$), respectively. Through visual comparison, it becomes apparent that the amplitudes of the HICT$_m$ species increases in the reduced state while the amplitude of the HICTi curve decreases significantly.

Performing this analysis on all four mutants and integrating over the respective species yields their relative amplitudes (Figure 18B). All four labeled positions appear to be in contact with the solvent and the label H-bonded to 84-95%. Importantly, this does not mean that 5-16% of labels are not in contact with the solvent. Rather, the ICT-HICT$_i$ contributions represent a time-averaged picture of hydrogen-bonding to the negatively charged carbonyl of BADAN upon intramolecular charge transfer after excitation.[104] For S295C-ADAN, there is nearly no difference between O and R-state C*c*O. This, however, is different for labeling positions 299 and 301, where there is a reduction-induced decrease in the $HICT_i$ species' frequency by 45% and 33% of their values in the oxidized state, respectively. Both decreases go mostly in favor of the amplitude of the $HICT_m$ species. Altogether, spectral decomposition shows that the bathochromic shift of P301C-ADAN and K299C-ADAN is due to a more mobile solvent microenvironment around the fluorophores, reaching 70% in the reduced state of the latter mutant. While there are some minor differences in the amplitudes of the ICT state, none of the fluorophores appear to be buried within the protein and the observed variations in their fluorescence maxima are mostly due to more or less dynamic local hydration shells. .

Final indications as to the nature of the solvatochromic shifts can be gathered from the time-dependent fluorescence shift (TDFS) of the fluorophores. As mentioned before, the steady-state spectra do not only include heterogeneities of the dynamic protein-fluorophore system, e.g. conformations of individual fluorophores, but also represent a temporal averaging of processes occurring on different timescales. Among these are hydration dynamics, internal label dynamics, and fluorophore-protein interactions. As these processes have characteristic timescales, measuring the time-dependent spectra with TCSPC provides a means to distinguish potential contributions to the overall Stokes shift. Importantly, due to the exponential nature of fluorescence decays, the slower the dynamics of a spectrally shifting process the smaller its contribution will be. Figure 19A shows the TDFS of oxidized and reduced K299C*-ADAN. Due to the emission profile of the white-light laser source, the first retrievable data point is at 312 ps. At this point the fluorescence maximum, as determined from a Gaussian fit of the fluorescence counts from the discrete detector wavelength bands, is 21230 cm^{-1} for oxidized and 21040 cm^{-1} for reduced K299C*-ADAN. From there on out, both curves run mostly parallel to each other and can be fit with a biexponential function yielding ν_∞ = 19900 cm-1 (O) and ν_∞ = 19950 cm^{-1} (R) as extrapolated final values for the Stokes shift.

To determine the bathochromic shift in the lower picosecond and sub-picosecond range, one needs to determine the initial wavenumber ν_0. This can be done through the following procedure published by Jurkiewicz et al.[105] In short, the lineshape functions of absorbance $g(\nu)$ and emission $f(\nu)$ of free BADAN are determined in a nonpolar solvent (in this case cyclohexane, Figure 19D). Then, one fits the absorbance spectrum A_p of C*c*O-ADAN in the polar solvent the TDFS is measured in (20 mM potassium phosphate buffer, 20 mM NaCl, 0.05% β-DM) with the convolution $(g * p)(\nu)$, where $p(\delta,\mu,\sigma)$ is a Gaussian introducing the spectral shift stemming from the solvent:

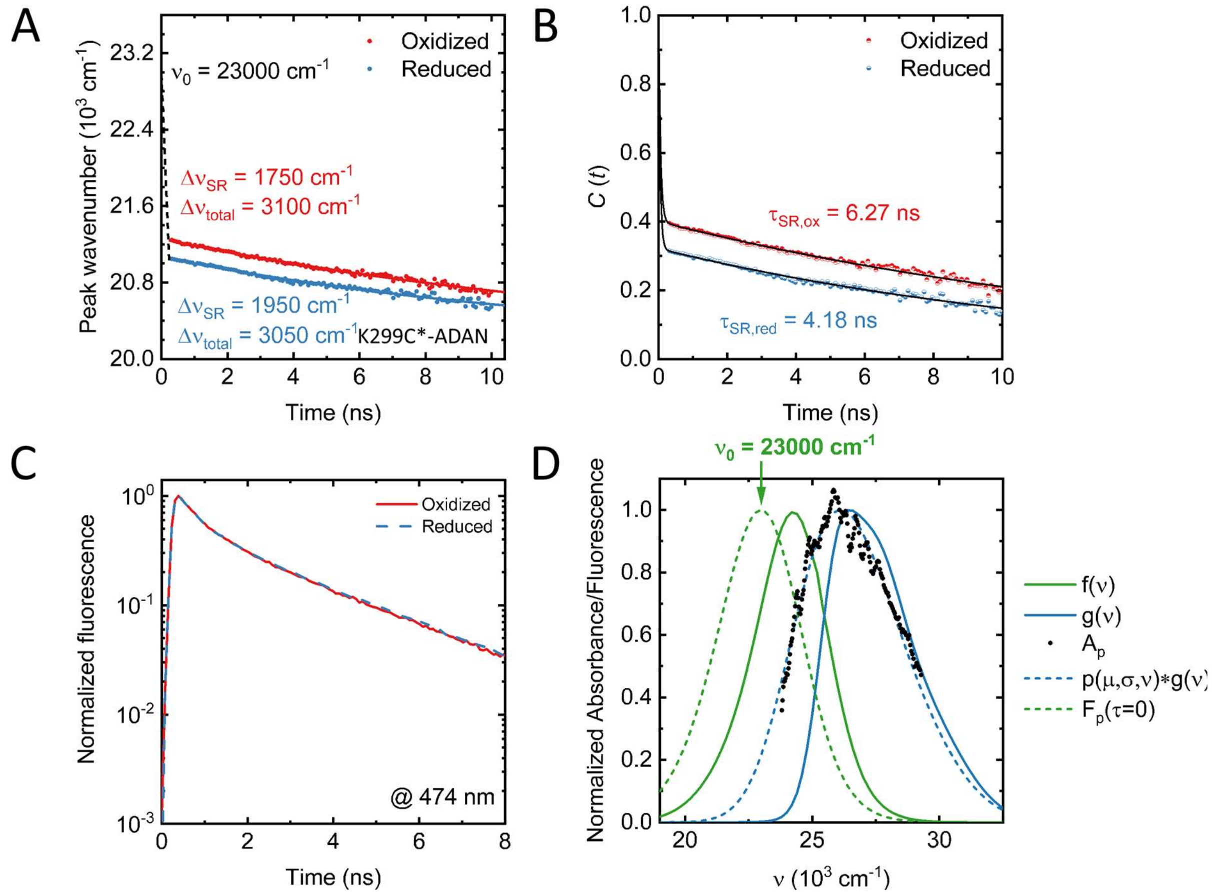


Figure 19: (A) TDFS of oxidized (red) and reduced (blue) K299C*-ADAN. (B) Correlation generated from the TDFS of (A) with respective correlation times. (C) Fluorescence lifetimes of K299C*-ADAN. (D) Lineshape functions of nonpolar absorbance *g* and fluorescence *f* of BADAN (blue and green full lines, respectively), polar absorbance A_p of K299C*-ADAN and fit with convolution $p*g$, and time-zero spectrum F_p (dashed green line) with emission maximum ν_0.

$$p(\delta, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(\delta-\mu)/(2\sigma^2)} \quad , \tag{12}$$

$$A_{\mathrm{p}}(\nu) \propto \nu \int g(\nu - \delta)\, p(\delta)\, \mathrm{d}\delta \quad . \tag{13}$$

Knowing the fit parameters one can generate the time-zero spectrum *F*p:

$$F_{\mathrm{p}}(\nu, \tau = 0, \nu_{\mathrm{exc}}) \propto \nu^3\, \nu_{\mathrm{exc}} \int g(\nu_{\mathrm{exc}} - \delta)\, p(\delta)\, f(\nu - \delta)\, \mathrm{d}\delta \quad , \tag{14}$$

with ν_{exc} being the excitation wavelength in the TDFS experiment. The emission maximum from the reconstructed *F*p is the time-zero emission wavenumber ν_0 with an accuracy of about 200 cm^{-1}.[105] In the case of oxidized K299C*-ADAN ν_0 was found to be ~23000 cm^{-1}.

Unfortunately, it was not possible to apply the same procedure to the reduced enzyme-fluorophore conjugate due to the dithionite absorbance overlapping with the absorbance A_P. However, since the absorbance of BADAN should be less sensitive to hydration than the dye's emission (due to ground state dipole moment being 2.9 times smaller than that of the excited state[106]), it can reasonable be assumed that the time-zero emission of reduced K299C*-ADAN is roughly equal to that of the oxidized enzyme. With ν_0 and ν_∞ known, one can calculate the correlation function $C(t)$ corresponding to the TDFS (Figure 19B) and the characteristic solvent relaxation times τ_SR as[105]:

$$C(t) = \frac{\nu(t) - \nu_\infty}{\nu_0 - \nu_\infty} \quad , \tag{15}$$

$$\tau_\mathrm{SR} = \int_0^\infty C(t)\,\mathrm{d}t \quad . \tag{16}$$

Upon reduction of the enzyme, solvent relaxation dynamics increase by 33%, as τ_SR decreases from 6.27 ns in the oxidized to 4.18 ns in the reduced state of *Cc*O.

Despite observed differences over several orders of magnitude between bulk water and water molecules on the corrugated surface of proteins, both the rotational as well as the translational diffusion of the solvent water molecules are very fast.[107] This means that the ultrafast (in this case everything below 312 ps) component of BADANs TDFS in aqueous solution is dominated by solvent relaxation and the longer components by processes such as dye-protein dynamics[104] such as the movement of close by polar amino acid residues, when the fluorophore is attached to a protein. It should be mentioned that twist-motions and conversion between BADAN's planar and twisted ICT states (PICT and TICT, respectively) are potentially also involved in the label's time-dependent fluorescence emission[104], although there is some debate about their contribution[108]. The timescale of this conversion makes it largely indistinguishable from solvent relaxation processes[70]. However, due to the small size of the label, the large solvent accessibility at the labeling site, and the nearly identical (albeit shifted) dynamics of the TDFS between oxidized and reduced K299C*-ADAN indicating similar label-protein matrix interactions[109], there is no indication of conformation-dependent restrictions to this conversion between O and R-state *Cc*O. Its contribution to the observed differences in the TDFS of K299C*-ADAN should therefore be negligible.

Interestingly, between 60% (O) to 70% (R) of the overall stokes shift occurs within the first ~300 ps (Figure 19B). A similar behavior was observed for solvatochromic nucleoside analogs

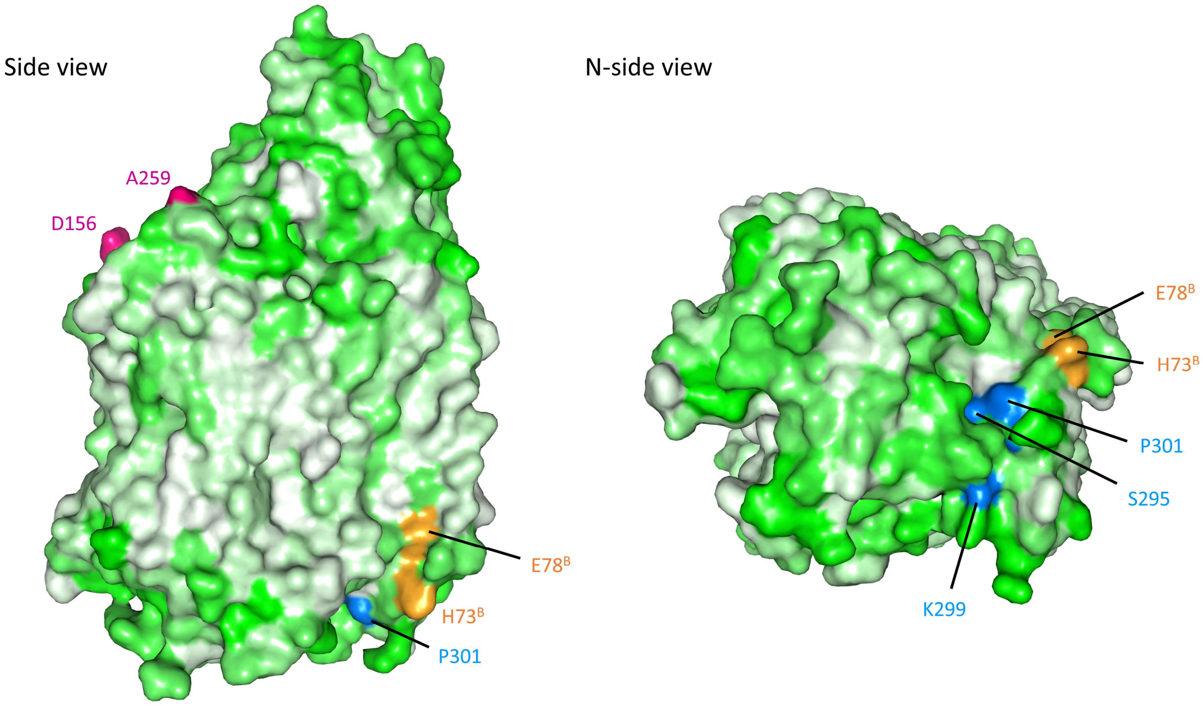


Figure 20: Surface of 2-subunit *Cc*O (PDB code 3HB3) with mutation sites as viewed from the membrane (left) and from the N-side (right). Residue hydrophilicity is mapped from hydrophobic (white) to hydrophilic (green) according to the Eisenberg scale[111].

incorporated into DNA when the sample was dissolved in buffer solution without the addition of lipids.[101] Lipid binding to the DNA retarded the dynamics significantly through the interaction of the dye with the hydrated polar head groups of the lipids. The positioning of the label on K299C*-ADAN indicates that nearly all solvent-dye interaction should come from the hydration shell of the protein and not of the membrane. The high amount of fast dynamics in the TDFS does not hint at a high microviscosity in the label's vicinity and thus a very mobile solvent environment. The micropolarity, however, as can be estimated from the position of the emission maximum (Figure 16) and the low overall Stokes shift (Figure 19B) is limited. The relative permittivity ε in the microenvironment of the label for all mutants lies between 23 and 24 as estimated from the position $\nu_{m,3}$ of the lowest wavenumber emission species (Table 9) according to the calibration curve published by Fischermeier et al. (Supplementary Information)[69]. The surface of proteins itself generally exhibits a permittivity of 20-30 and the permittivity of hydration water in protein cavities is usually much lower than the bulk value of $\varepsilon = 80$.[110] Since the three N-side mutants are located inside a corrugation, albeit in a very hydrophilic microenvironment, as can be seen in Figure 20 indicating the Eisenberg scale[111] of *Cc*O's surface, their estimated values seem sensible. A low permittivity for BADAN-labeled position 259 on the P-side, however, is not expected.

Finally, the influence of the redox-induced conformational change observed for K299C-AF in section 5.1 in the differences in both magnitude and dynamics of the Stokes shift of K299C*-ADAN has to be considered. Aromatic quenching by excited state energy transfer from nearby tryptophan residues might vary for both conformations. One way to rule out this effect is by comparing the fluorescence decays of both O and R-state K299C*-ADAN (Figure 19C). Both decays are approximately identical and display two fluorescence lifetimes, one short with ~600 ps and one longer lifetime that is ~3 ns. Aromatic quenching would result in a decrease of both fluorescence lifetimes[112] and here none of them are affected.

Taken together, labeling of C*c*O with the solvatochromic fluorophore BADAN on three different locations on the enzyme's N-side and one position on its P-side revealed distinct solvent environments in each case. Upon reduction of C*c*O, BADAN conjugated to positions 299 and 301 at the N-side of the enzyme exhibited bathochromic shifts of 250 cm^{-1} and 290 cm^{-1}, respectively, while no significant shift could be observed for residues 295 further into the protein and 259 at the P-side. Spectral decomposition showed that the observed differences were the result of differences in the ratio of immobile and mobile hydration water as all labeled positions are sufficiently hydrated and do not appear to be partially buried within the enzyme. Furthermore, excitation-emission fluorometric scans revealed that the 2D red-ox difference spectra of K299C-ADAN and P301C-ADAN show spectral signatures comparable to a temperature increase to the protein-dye system. Finally, analysis of the time-dependent spectral behavior of K299C*-ADAN showed that between 60% (O) and 70% (R) of the fluorophore's Stokes shift occurs within the first 300 ps and thus within the time regime of most hydration shell dynamics[113], ruling out large-scale effects e.g. from interactions of the label with the protein surface[109] that probably account for most TDFS in the nanoseconds time domain.

The mechanism of selective hydration, i.e. the wetting and/or dewetting of parts of a protein, is an active constituent not just in the functionality of the aforementioned proton pump bacteriorhodopsin, but has also been widely discussed as a key feature in several of the gated steps in C*c*O's catalytic cycle—at different sites of the protein. C*c*O's D-channel possesses a hydrogen-bonded network from N131, close to the N-side, to the conserved proton-donating residue E278 nearby the catalytic center.[44] However, in order for an H-bond network to facilitate proton transport, there has to be some maintenance of directionality. MD simulations suggest that the protonation E278 drives dehydration in this cavity by causing the so-called asparagine gate between N131 and N113 to close through the formation of an H-bond.[114] Drying of the D-channel cavity would therefore stop proton transport as soon a proton reaches E278 at the end of the D-channel and thus prohibit proton backflow. Interestingly, the authors also observed increased water

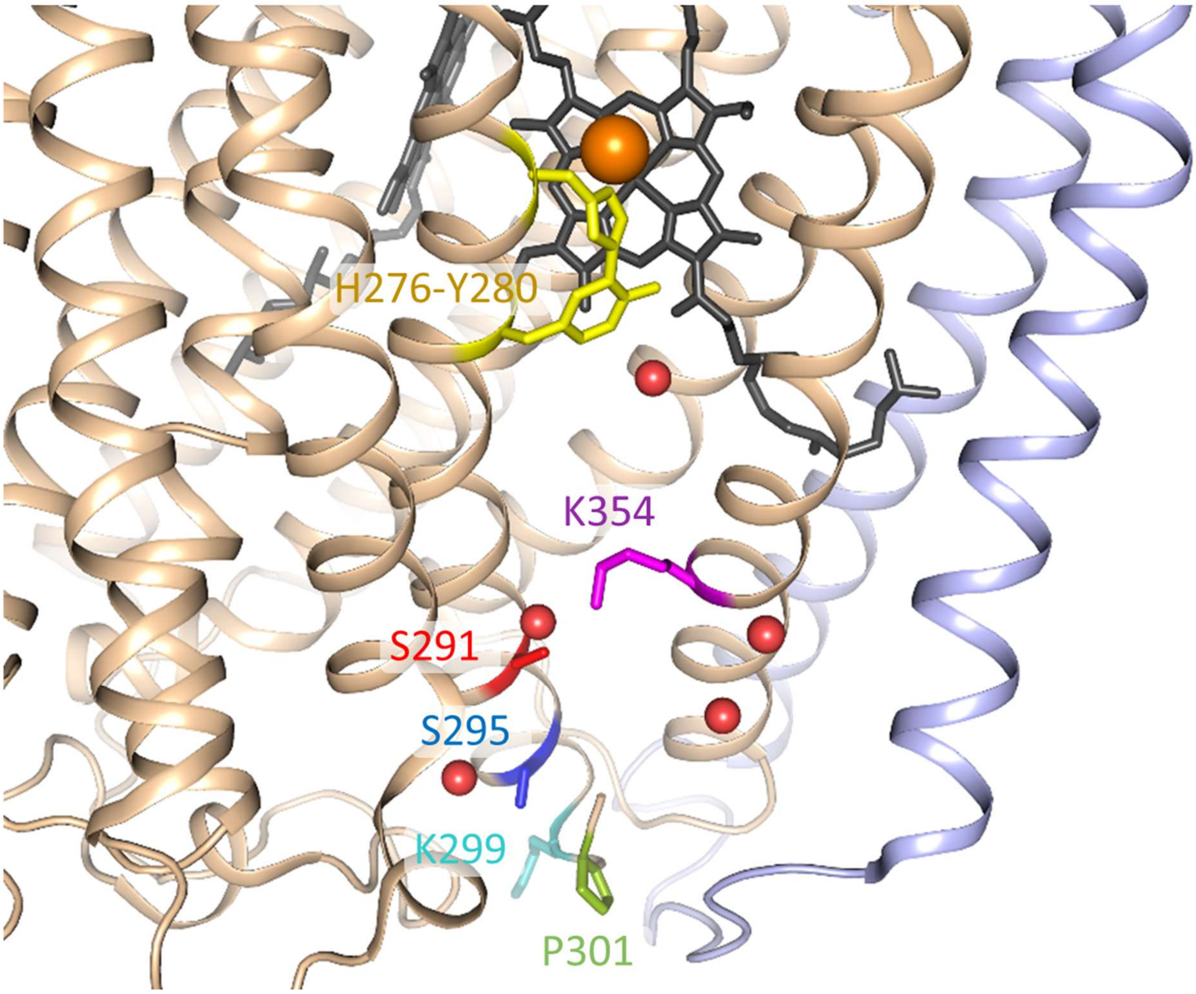


Figure 21: **S**ubunits 1 (beige) and 2 (blue) of *P. denitrificans* C*c*O (PDB code 3HB3) with N-side mutation sites and important residues. Orange spheres are waters within 10 Å of K254. The water molecule trapped between S2291 and K354 as well as the one close to Y280 directly belong to the K-channel. Helix 7 was removed to improve visibility of the channel.

diffusion in the more restrained D-channel cavity while E278 was protonated, leading to shorter hydrogen bond lifetimes.[114] In another MD-derived model[115], the protonation state of the heme a_3 propionic acid causes a positive change in hydration of the hydrophobic cavity near E278. This wetting of the cavity in turn lowers the pK_a of E278 and thus its deprotonation necessary for enzymatic turnover. Yet another theoretical study[116] confirms the latter mechanism but further adds (albeit to a lesser degree) the redox state of heme *a* and the protonation state of E278 itself to the driving forces of transient hydration in the central cavity. One of the most intriguing aspects of the above models is the absence of large-scale conformational changes. Rather, small rearrangements, mostly involving H-bonds, propagate through the enzyme through side chains and cavity waters to cause the change responsible for channel gating somewhere upstream of the proton-conducting network. Another important aspect of C*c*O hydration in both the D-channel and the central hydrophobic cavity is its strong coupling to proton transfer.[117]

In contrast to the well-hydrated D-channel, there are only two crystallographic waters observable in the interior part of the K-channel (see Figure 21, the two uppermost water molecules directly belong to the K-channel).[44, 91] However, it is clear that limiting the analysis of *Cc*O's proton pathways to the static hydration levels observed in the crystal structure underserves the investigation of the proton conducting mechanism as the dynamics of selective hydration have to be included.[116] The most general question regarding the K-channel is how protons are transported through it in absence of a continuous H-bond network and, more specifically, how the channel is temporarily activated during the enzyme's reductive phase. Recent results from MD simulations suggest that transient protonation of K354 leads to a conformational widening of the K-channel and thus to an increase in its hydration.[4] These results agree with computational and crystallographic observations[7, 90, 91] of conformational changes in the K-channel, especially of heme a_3, the catalytic Y280, and helix 8 (for a more detailed discussion of conformational changes in *Cc*O, please browse back to section 5.1). Recently, MD simulations further quantified the extent of selective hydration in the K-channel. In the work by Supekar et al.[5], alongside the widening of the K-channel, the authors observe an increase from four to seven water molecules in its upper section, which extends from K354 to Y280, upon reduction of the enzyme. At the same time, hydration decreases in the nonpolar cavity between E278 and the BNC. The authors conclude that this selective hydration mechanism switches between activity of the D- and K-channel by modulating the connectivity to the BNC necessary for proton transfer depending on the charge state of the BNC.

As indicated above, there are plenty of theoretical contributions to the mechanistic aspects of selective hydration in *Cc*O. Crucially missing, however, are experimental results that could corroborate these findings. Data from Fourier transform infrared spectroscopy has revealed a reorganization of weakly bound water in bovine, yeast and *P. denitrificans* *Cc*O.[8, 9] Yet, to date these changes could not be structurally located. This lack in spatial resolution can be overcome through site-specific labeling with solvatochromic fluorophores. While this method has been limited to sites close to the enzyme surface, the results presented in this section point to an increase in hydration dynamics at the N-side surface of the K-channel. This increase is felt strongest at positions 299 and 301 closer to the surface, while position 295, which is close to a structurally resolved water molecule, experiences a far smaller effect. It can reasonably be assumed that the observed changes at the surface are connected to the conformational changes associated with the widening of the K-channel discussed in section 5.1 as well as to its increase in hydration and that they are implicated in the channel's function. A reduction-induced increase in water accessibility and diffusion due to a lower rigidity at the channel's N-side surface as compared to the O state might be considered as beneficial to proton uptake from this side. This is in line with previous flash spectroscopy experiments, which have shown that fast proton uptake into the K-channel is initiated through electron injection at the P-side.[118] An important consideration is the location of the sites probed in this work, which are located on the N-side segment of helix 6 and thus close to the essential K354 (Figure 21) and the helix 6-7 loop close to the subunit 1-2 interface. Analysis of the entirety of the acquired fluorescence data from these positions suggests that the effect is local in the sense that it is limited to the surface "above" the K-channel, however comprises significant area along the surface. One would therefore arrive at a funnel-like mechanism with an adjustable funnel size. As the catalytic center is reduced from the fully oxidized state, sidechain conformational changes traverse to the N-side surface and increase the local conformational flexibility between helices 6 and 7. This leads to water filling the "funnel" and facilitating hydration of the inner channel[5] leading to conditions favorable for proton conductivity. This does not preclude that the channel's proton entrance is E78 on subunit 2 as the local wetting mere facilitates proton conductivity and is not necessarily connected to translational movement of hydronium ions, however, might partially explain the resiliency[51] of K-channel conductivity to mutation of the residue observed for the *P. denitrificans* enzyme.

Further investigation into the hydration dynamics changes at the surface of the K-channel will require more mutants for site-specific labeling, e.g. at the N-side tips of helices 8 and 9 and their connecting loop as well as on subunit 2. These sites will aid in mapping the extent of the changes in surface water diffusivity as well as provide data on redox induced conformational changes of nearby helices. Improvements can also be made with regard to the experimental parameters. A laser source within the wavelength range between 360-410 nm with a shorter pulse duration would enable analysis of the < 300 ps TDFS signal. Further temporal resolution might be gained by capturing the labels' fluorescence with a Streak camera, thus extending the temporal resolution into the subpicosecond time domain. Employing novel solvatochromic dyes[119] with a different chemical linker group, e.g. haloacetamide, which performed well in attaching fluorescein to 2-subunit C*c*O, might also overcome some of the limitations imposed by usage of BADAN, such as the dye's tendency towards unspecific labeling. Finally, it might be possible to overcome the restriction to sites at the protein surface by incorporating solvatochromic amino acid analogues[120, 121] at sites within the enzyme's hydrophilic channels through genetic code expansion[122].

6. Conclusions

The dynamics of cytochrome *c* oxidase's K-channel are among the most pressing questions regarding the function of this essential respiratory enzyme. Several studies have attributed the selective activation of the channel during the reductive phase of CcO's catalytic cycle to the protonation and conformation of the channel's essential residue K354[4] and of the active site Y280[6]. Further, the mechanism of selective hydration between the D-channel and the otherwise relatively dry K-channel comprising only two crystallographically determined water molecules was introduced to explain the lowering of the energetic barrier for proton transport.[5] In this work, a new expression system was engineered to produce C*c*O mutants suitable for site-specific labeling of several residues of the K-channel entry and the proton release site. The WT 3-subunit expression system had an expression yield similar to that of WT with only one subunit expressed recombinantly. From this system, a background cysteine-free mutant was constructed yielding a background-free 2-subunit enzyme with WT-like expression and catalytic activity. The subsequently created single-cysteine mutants could all be expressed with high yield and had WT catalytic activity, with the exception of the P-side mutant, which showed only about 50% activity. Conjugation with the protonation-sensitive fluorophore fluorescein was successful for all cysteine mutants and yielded labeling stoichiometries between 15% and 98%. Steady-state absorbance spectroscopy of the dye-protein conjugates revealed pK_a changes of fluorescein between the oxidized and chemically reduced state of the enzyme in positions 301 and more strongly in position 299 above the K-channel. The latter residue's pK_a shift was further analyzed in a collaboration with Prof. Ernst Walter-Knapp and Jovan Dragelj of the FU Berlin Chemistry Department, who performed MD simulations and electrostatics computations of the K299C-AF enzyme. The simulations reproduced the pK_a downshift of K299C-AF upon reduction of the catalytic center and revealed that the mechanism for this are conformational changes traversing from the catalytic center to the N-side surface depending on the redox chemistry of the enzyme's cofactors. The analysis of time-dependent fluorescence anisotropy measurements showed increased conformational flexibility of the protein segment at position 299, potentially related to a partial unwinding of the helical structure of the helix tip, and overall less steric restriction imposed on the fluorophores in positions 299 and 301. MD simulations of K299C-AF reproduced the differences in anisotropy decay, showing that the redox state of the catalytic center not only influences the pK_a of the fluorescent probe at CcO's surface, but also regulates surface-near steric restriction and conformational flexibility. The same sites were further analyzed by conjugating C*c*O with the hydration-sensitive solvatochromic dye BADAN. The labeling showed highly solvent accessible sites for the three N-side mutants and one P-side mutant (A259C-ADAN), that was probed as a reference. Reduction of the enzyme from the O-state led to pronounced redshifts for positions 299 and 301, while the effect was negligible for positions 295, further inside the enzyme, and 259 on the P-side. Spectral decomposition revealed the observed redshift to be the result of a decrease in low-diffusivity water in favor of more bulk-like highly diffusive water, comparable to a temperature increase of the dye-protein system. Time-resolved fluorescence measurements of K299C-ADAN confirmed that the majority of the difference in dynamics between the oxidized and reduced enzyme state occurs within the first 300 ps after fluorescence excitation and is thus not due to slow protein-dye interactions that might be caused by the conformational changes between the states. Measurements of the system's fluorescence lifetime further ruled out differences in aromatic quenching as source for the redshift. The accumulation of the fluorescence data points to a reduction-induced conformational change at the enzymatic surface that increases the local flexibility and thus allows for more bulk-like water diffusion. The analysis shows that even in absence of large-scale conformational changes, the redox chemistry deep inside the enzyme can control the conditions at the proton-uptake surface to facilitate the proton transport necessary for catalysis and adds experimental results to simulations showing increases in hydration[4, 5] in the inner K-channel and increased conformational flexibility of helices 7 and 8[7, 91].

Partial results from this and the preceeding chapter are currently under revision for publishing in the FEBS Journal under the title "Electronation dependent structural change at the proton exit side of cytochrome *c* oxidase as revealed by site-directed fluorescence labeling." Prior to the work outlined in this part of the thesis, I further contributed to the following publication using another C*c*O expression system:

Wolf, A., Schneider, C., Kim, T.Y., Kirchberg, K., Volz, P., Alexiev, U., A simulation-guided fluorescence correlation spectroscopy tool to investigate to protonation dynamics of cytrochrome c oxidase. *Phys. Chem. Chem. Phys.* **2016**, 18, 12877-12885.

PART II: DANAE—a non-tracking algorithm for the analysis of diffusive modes in fluorescence microscopy video data

7. Diffusion in biophysical microscopy studies

Investigating the modes of diffusion molecules exhibit in biological matter is key to understanding biological interactions. In this part the of thesis I present a new method to analyze diffusion in single molecule video microscopy experiments that does not require prior assumptions about the diffusive properties of a sample. At first, I underline the significance diffusion plays in biology (section 8.1) and present already existing biophysical methods to analyze it (section 8.2). The following chapter 9 presents the computational tools employed to perform both the diffusional analysis itself and to test the novel DANAE algorithm, which is presented in chapter 10. Within this chapter, section 10.1 lays out the DANAE algorithm, while section 10.2 derives the equations to fit the diffusive spectra generated by DANAE including boundary effects, which become highly significant depending on the outline of the experiment. Section 10.3 tests DANAE against varying experimental conditions, such as high particles densities, fast diffusion, photo-state transitions, and others. Finally, section 10.4 introduces a two-color variation of the DANAE algorithm and highlights its implications for studying biomolecular interactions.

7.1. Diffusion at the interface of biology and physics

Every interaction of individual biomolecules is the result of an encounter initiated through the diffusion of at least one of the interaction partners. This involves every aspect of cellular life, from the synthesis of essential biopolymers such as DNA, RNA, and peptides to the movement of ions and molecules within a cell and across its membrane. Regularly, as it is the case in bacterial respiration[123], the rate of enzymatic reactions is directly modulated through the rate of these encounters and therefore by the diffusivity of the partaking molecules. Thus, understanding e.g. receptor-ligand interactions[66, 124] on a cellular level necessarily requires examining the properties of the diffusion undergone by those molecules. This, however, is no easy task, as the cellular environment is complex and largely inhomogeneous. This leads to a variety of diffusive behaviors ranging from passive transport and facilitated diffusion to active transport processes and due to the discontinuous nature of cells often involves confinements and hop diffusion among subcellular compartments. The parameters of the observed diffusion contain the underlying interactions of the target molecules. Understanding these interactions thus necessitates the application of a suitable physical model of the occurring dynamical processes.

The simplest method to study the diffusion of particles is to record their trajectories from (usually) fluorescence microscopy video data. Fitting the airy disk of a particle's fluorescence allows for the determination of its position with nanometer accuracy (hence the alternative name nanoscopy), which is well below the diffraction limit. The recording and interconnecting of a particle's position over time is known as single particle tracking (SPT) and has first been used to determine Avogadro's constant from the movement of mercury beads.[125] One of the first biological applications of this technique was the tracking of kinesin moving along microtubules.[126] Single virus tracking has become a prominent tool for the study of viral infection.[127-129] There are, however, limitations to SPT. In order to reconstruct the trajectories a molecule has taken, there should be no ambiguity regarding the identity of individual particles. This imposes upper limits on either the concentration of particles within the observed image frame, the maximum velocity in distance/frame of observable molecules, or the minimal temporal resolution of the measurement apparatus. Reducing the concentration of fluorescent molecules is not always feasible or desired as it reduces the amount of data available for statistical evaluation of the results, especially considering ageing effects (more in section 10.3) of the biological system. Too much movement of particles within frames can be compensated for with an increase in temporal resolution. There exists, however, a natural tradeoff between temporal and spatial resolution as the detector needs to detect a sufficient amount of photons within the exposure time of the camera to determine the position of a particle (Figure 22). Recently, algorithms were developed that include more data than the position when determining the identity of particles within frames, a method known as feature point tracking.[11] To overcome the general limitation of observing individual trajectories several statistical methods have been devised to analyze the diffusion of particles in video microscopy data.

7.2. Statistical methods for diffusional analysis

In a wide-field fluorescence microscopy tracking experiment, the position of a small fluorescent molecule can be located with a precision well below the wavelength of the emitted light, even

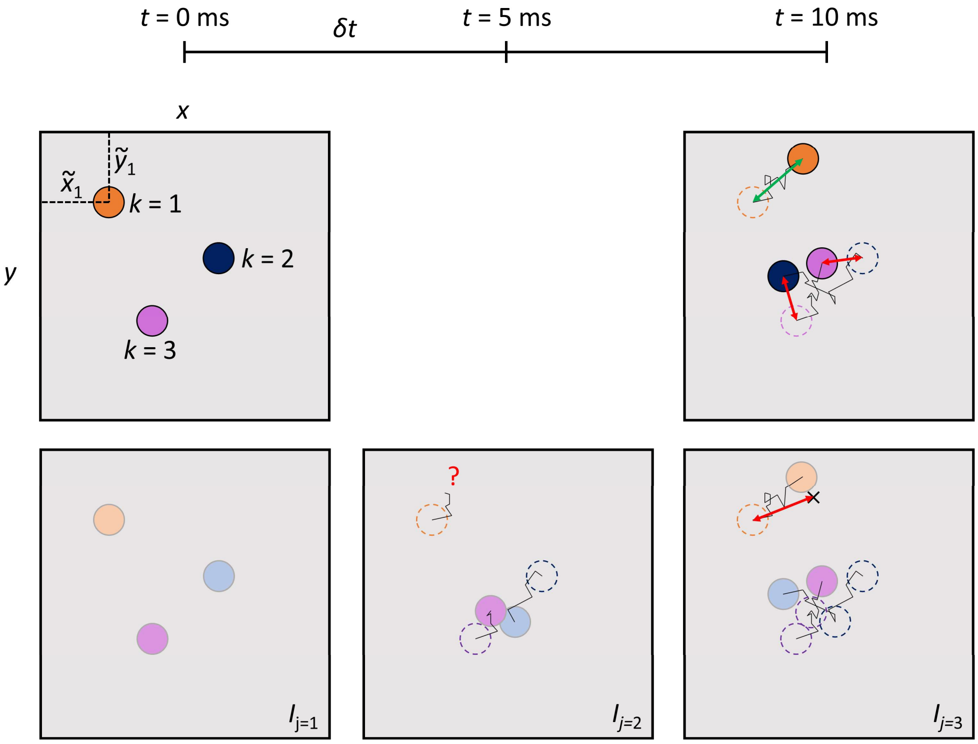


Figure 22: Video nanoscopy of moving fluorescent particles. Top row: If the trajectories of particles *k* intersect between two consecutive microscopy images I_j, the probability to correctly identify particles, in this case blue and purple, decreases sharply. This leads to the calculation of wrong distances (red arrows). Bottom row: Reducing the frame difference time δt by half (in this case from 10 ms down to 5 ms) deceases the probability of wrong assignments at the cost, however, of intensity, which in turn increases the amount of "holes" in trajectories due to particles not exceeding the intensity threshold necessary for their fitting and causes a loss in spatial accuracy in the detection of particle positions as seen for the orange particle.

though two molecules at such separation cannot be resolved. One can obtain the locus of the molecule by calculating the centroid or fitting a 2D Gaussian to the intensity profile of the diffraction limited Airy disk (Figure 23), thus yielding the localization coordinates $\{\tilde{x}, \tilde{y}\}$. Determining the positions of particles in a microscopy image is possible as long as the Rayleigh criterion is satisfied, which states that two light sources can be resolved as long as their distance is at least such that the first minimum of the Airy disk of source 1 lies outside of the central maximum of source 2 (Figure 23, right). The limiting spatial accuracy of such a measurement is determined by the signal to noise ratio of the image, the mechanical stability of the apparatus and the motion of the mobile particle during the acquisition time of the image.

Let us assume a number of molecules m in the image I. Each molecule is located in the two-dimensional I by its coordinates $\underline{\tilde{r}} = \{\tilde{x}, \tilde{y}\}$. A movie therefore is an image stack I_j with j running from 1 to N as the maximum number of frames. The temporal separation δt between two

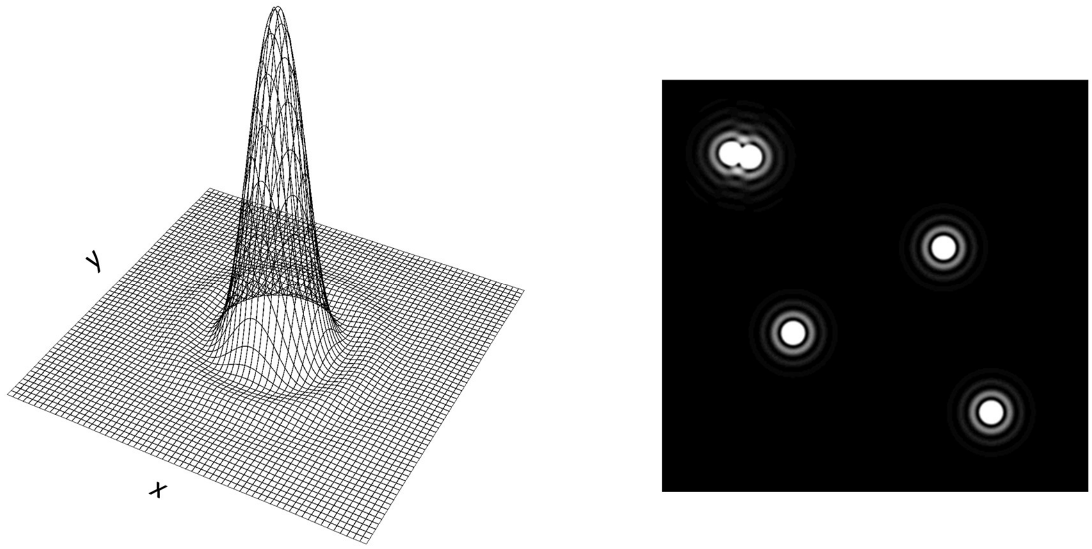


Figure 23:. Airy disk intensity profile of a single point emitter (left) and of multiple emitters in the plane (right). The two emitters in the top left corner of the right image are just resolved, as they fulfill the Rayleigh criterion. The intensity in the right image was nonlinearly amplified to better visualize the diffraction pattern of the airy disk.

consecutive images is determined by the sum of the image exposure time and camera electronic delay time. A good measure for the diffusional behavior of the sample is the mean square displacement $\langle r^2(\Delta t)\rangle$. It depends on the lag time $\Delta t = q\ \delta t$, which is defined as the time between two frames I_j and I_{j+q}.

With the coordinates $\underline{\tilde{r}_k}$ one can calculate the distance a particle k has moved between I_j and I_{j+q}

$$(r_k^j)_q = \left|\underline{\tilde{r}}_{k,j} - \underline{\tilde{r}}_{k,j+q}\right| \tag{17}$$

where $|\cdot|$ denotes the Euclidian norm. For this calculation, it is necessary to identify identical particles between adjoining frames, a method called single particle tracking (SPT).

The expectation value $\langle r^2(\Delta t)\rangle$ of the distances is the mean squared displacement (MSD) of the diffusional process. Depending on the system at hand (expectation of ergodicity, the existence of subspecies) one might chose the time and/or ensemble average of $r^2(\Delta t)$. The outcome of such analysis is very sensitive to the averaging process. In the case of ensemble averaging, it might disguise the existence of different diffusive modes or fabricate them from subsets of trajectories that actually lie within the statistical spread of the main diffusive mode. The same is true for the time average if the system is not in equilibrium, which might be induced by chemical reactions. The MSD for Brownian motion is

$$\langle r^2(\Delta t)\rangle = 4\,\mathrm{D}\,\Delta t \quad . \tag{18}$$

In the case of anomalous diffusion, the dependency of the MSD on Δt will not be linear but with an exponent either smaller than 1 (subdiffusion) or larger than 1 (superdiffusion). These anomalous diffusion processes are often found within cells and on cellular membranes.[130-135]

It is also possible to avoid ensemble averaging and construct a histogram from all distances r_k[136, 137]. The frequency distribution of these "steps" is termed step length distribution (SLD). Known as a jump step/distance analysis (JSD/JDA) it allows for resolving different diffusive species by fitting the frequency distribution with an appropriate probability model function. For a population *i* of particles undergoing two dimensional Brownian motion the model function for the SLD per radial element d*r* in the infinite plane is[138]

$$A_i(r, D_i, \Delta t)\;\mathrm{d}r \;=\; \frac{1}{4\,D_i\,\Delta t}\,\exp\left(-\frac{r^2}{4\,D_i\,\Delta t}\right)\,2\,r\,\mathrm{d}r \quad . \tag{19}$$

It is a measure of the probability to find a particle on a radial segment d*r* in the distance $|\underline{r}|$ of its starting position after a lag time Δt. While a JDA circumvents many of the problems induced by ensemble averaging, it cannot prevent artefacts stemming from single particle tracking, i.e. the generation of false distances from mismatched particles. A possible solution to this problem is an algorithm that does not determine particle identities and operates without relying on the corresponding assumptions, which will be presented in chapter 10.

8. Computational methods

I wrote the DANAE program in the Python programing language (Version 3.2). For easy execution of large-scale data analyses, as was done for the validation of the algorithm (section 10.2), the program is fully functional from a console through the input of system arguments. Alternatively, a simple UI assists in entering the correct parameters before execution of the DANAE algorithm. The program requires input microscopy positional data to have the format: frame number <TAB> y position <TAB> x position </n>. Obviously, x and y positions may be switched. The program further requires the user to enter the longer of the two sides of the filed-of-view, the desired binsize of the generated DANAE distribution, and the frame lags up to which distributions shall be calculated. For each input file, the program generates two output files, one with background and one with the background subtracted. Output files will have the structure [[bins],[zero lag distr.], [distr for lag 1], [...], ...] and all values will have the units of the input files. The entire program is attached in Supplement 6 of the thesis.

For the validation of the DANAE algorithm, I generated synthetic microscopic data sets from Monte-Carlo simulations (MCS) with a self-written script in Python3. The simulations contained m_{bath} individual particles in a quadratic two dimensional box with side length a_{bath}.The particles were set to move following a free diffusion random walk model with each step as the result of a pseudo random number generated from a Gaussian probability density function with $\mu = 0$ and $\sigma^2 = 2 \times D \times \delta t$. Periodic boundary conditions were set at the edges of the box representing the bath. Centered within the bath, a quadratic field of view (fov) with side length a_{fov} was defined and positional data only extracted within the fov, thus reducing the number of extracted particles to a value approximately $\langle m_{fov} \rangle = \mathrm{m}_{\text{bath}} \times (a_{\text{fov}}/a_{\text{bath}})^2$. The accuracy of the exported positions was 16 digits and did not contain any information about the identity of a particle. If not stated otherwise, all simulations were performed with a pixel size of 74 × 74 nm^2, $a_{\text{fov}} = 512 \times 512$ px, and a frame interval of $\delta t = 5$ ms for 4000 frames in total.

For the SPT analyses of the data, I used a self-written Python3 script. The algorithm finds nearest neighbors between consecutive frames by assigning an index to each particle and then calculates all inter-particle distances. Distances are then sorted from the smallest to the longest distance up to a cutoff distance. Nearest neighbors are then assigned in a bijective manner. The algorithm then sorts the indices of consecutive nearest neighbors in lists representing traces (e.g. {*Particle 43* in frame 15, *Particle 8* in frame 16, *Particle 112* in frame 17 …}). These traces then serve as the basis of calculating jump distance histograms for a given t_{lag} (frame lag q).

Fitting and averaging of the SPT and DANAE data was performed by Pierre Volz-Rakebrand with Wolfram Mathematica (Version 10).

9. Results and Discussion II: Diffusion Analysis of NAnoscopic Ensembles (DANAE)

Authorship Statement

This research was performed collaboratively. The DANAE algorithm was programmed and mostly executed by me (Alexander Wolf). In collaboration with Pierre Volz-Rakebrand (Physics Department, AG. Prof. Alexiev) I developed the mathematical background, which involves the background subtraction method. I further programmed and executed the Monte Carlo simulations to test the algorithm. Automated analysis of the simulations was mostly performed by Pierre Volz-Rakebrand. Prof. Ulrike Alexiev (Physics Department) envisioned and supervised the project. Initial work on the project was performed by Thomas Schlieter (AG Prof. Alexiev). To further test the accuracy of DANAE in the blind, 3 simulations were performed by Prof. Helmut Grubmüller (Max Planck Institute for Biophysical Chemistry in Göttingen, Germany).

9.1. Statistics and algorithm

The fundamental idea behind DANAE is the indiscriminate calculation of distances between particle pairs as it was proposed by Semrau and Schmidt[139] with their Particle Image Correlation Spectroscopy (PICS) algorithm, which is based on a cumulative distribution and an approximation to subtract the background. Unlike single particle tracking, where only a subset of distances e.g. those given as nearest neighbors are determined, which can lead to false assignments of particle identity and therefore distorted diffusive parameters, both DANAE and the PICS analysis calculate all distances between all possible particle pairs without cut off or any establishment of a particle identity register (Figure 24). Therefore, the DANAE algorithm calculates all distances $|\underline{r}| = r$ between all particle pairs with indices k, l for all frames I_j with frame lag q:

$$(r_{k,l}^{j})_q = \left|\tilde{\underline{r}}_{k,j} - \tilde{\underline{r}}_{l,j+q}\right| \quad . \tag{20}$$

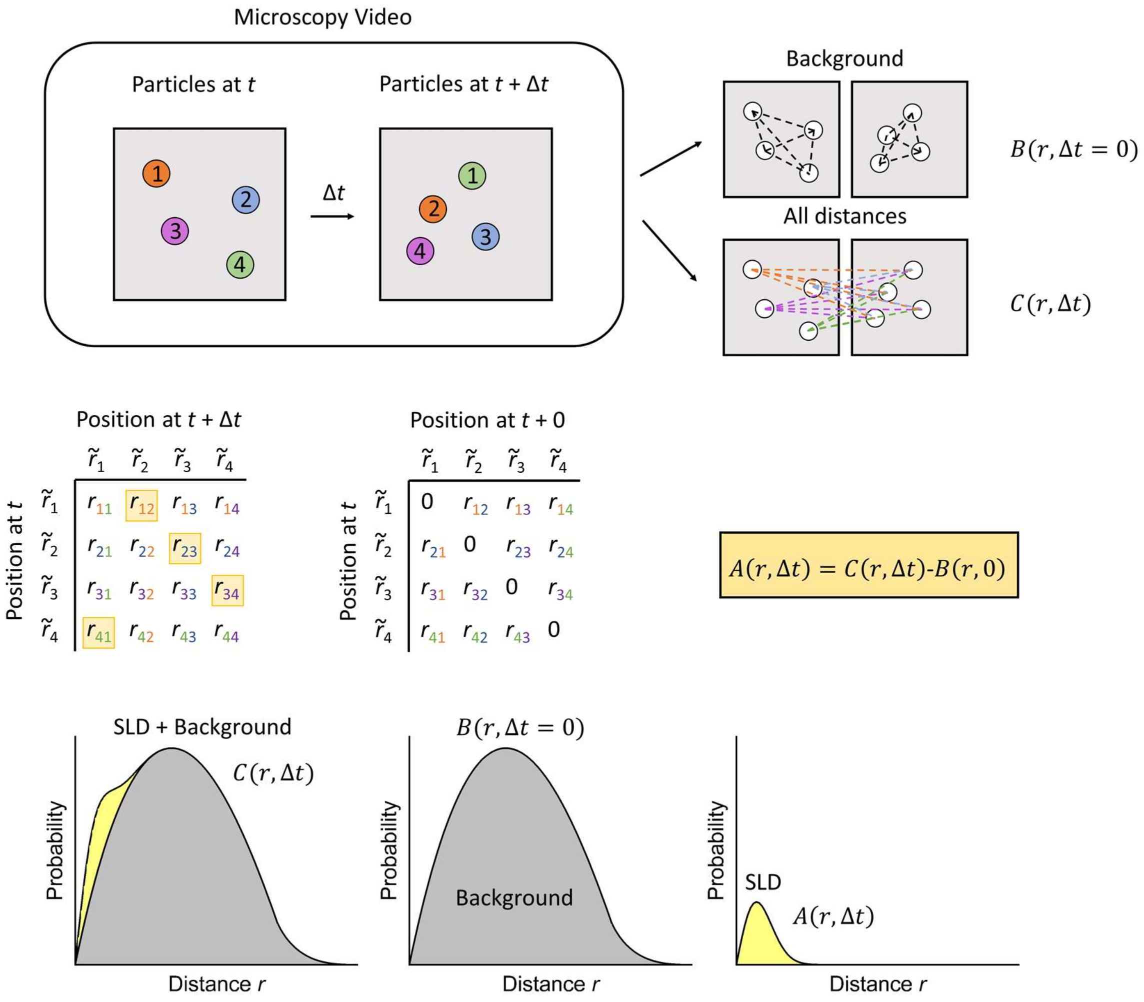


Figure 24: DANAE algorithm's generation of the SLD from inter-particle distances. Particle identities cannot be established with high confidence in dense video microscopy localization data of molecules with high diffusivity. This can lead to mismatching in SPT experiments when trajectories intersect inbetween frames (top left). DANAE therefore ignores the degree of proximity of positions between frames (top right) and calculates all distances between all particles without trying to keep the same index assigned to the same particle. The background is generated from particle distances within the same frames and subtracted to yield the SLD (bottom row).

From the calculation of all distances $|\underline{r}|$ a histogram is generated which corresponds to an absolute frequency distribution $C_q(r) \triangleq C(r,\Delta t)$ for every frame lag *q*. Since all possible distances are used, it naturally follows that $C(r,\Delta t)$ consists of the distribution of distances belonging to identical particles i.e. the step length distribution $A(r,\Delta t)$ and a background distribution of all other distances $B(r,\Delta t)$:

$$C(r,\Delta t) = A(r,\Delta t) + B(r,\Delta t) \quad . \tag{21}$$

The next task then lies in determining the background distribution so that it can be subtracted from the absolute frequency distribution $C(r, \Delta t)$ to yield the SLD $A(r, \Delta t)$. The background histogram will be determined by the local distribution of the particles in the movie. In the case of uniformly distributed particles in a quadratic field of view with side length a, it is possible to fit the absolute distribution function with the sum of a SLD and the square line picking distribution (SLPD, for more information see the next subsection):

$$B(r, \Delta t) = \frac{2r}{a^2} \begin{cases} \pi - 4\frac{r}{a} + \frac{r^2}{a^2} & , \quad 0 < r < a \\ 4\sin^{-1}\left(\frac{a}{r}\right) + 4\sqrt{\frac{r^2}{a^2} - 1} - \frac{r^2}{a^2} - \pi - 2 & , \quad a < r < \sqrt{2}a \end{cases} \tag{22}$$

.

This, however, only holds for truly uniformly distributed particles, which in reality is likely not the case. Also, it has to be considered that the additional degrees of freedom from the SLPD will naturally decrease the reliability of the fit.

Figure 25 visualizes the dependence of the background frequency distribution on the local distribution of particles in the field of view on three exemplary cases: a circular distribution, a quadratic distribution (from the boundaries of the FOV) and the same quadratic distribution with circular "hole" in the middle. Patches without fluorescence (e.g. where there exists an impenetrable barrier like cellular organelles for fluorophores) lead to distributions that miss a quantity of distances that is directly linked to the size and geometric shape of the patches. The shape of the background can thereby serve as a measure of the geometric shape of the observed particle ensemble.

To avoid any approximations in fitting the background the DANAE algorithm directly acquires the necessary parameters about both the shape and size from the positional data itself. The former is recovered from the zero lag distribution $(\Delta t = 0)$ and the latter by a counting based normalization procedure. Since all distances of "identical" particles vanish in the zero lag distribution (Figure 24) it contains the frequency distribution of all particle distances that are not part of the SLD and in the language of SPT could be attributed to un-identical particles. Any structures of particles or restrains on particle mobility in the movie will therefore add corresponding distances to the zero lag distribution that can then be subtracted without introducing artifacts.

The second step for subtracting the background is the determination of its correct amplitude. The normalization is then carried out by multiplying the zero lag distribution with a factor *f* to adjust its amplitude to the amplitude of the background for each Δt. Since for a given frame pair the number of particles in the SLD can only be the minimal number of particles between the two frames, the algorithm simply counts how many particles are supposed to take part in the SLD:

$$M_{\Delta t} = \sum_{j=1}^{j_{max}-q} \min\left(m_j, m_{j+q}\right) \quad . \tag{23}$$

The correctional factor for the zero lag distribution therefore becomes

$$f(\Delta t) = \left(\sum_{\text{bins } i} C_i(r, \Delta t) - M_{\Delta t}\right) / \sum_{\text{bins } i} C_i(r, 0) \quad . \tag{24}$$

The algorithm can therefore calculate the SLD for a given lag time Δt:

$$A(r, \Delta t) = C(r, \Delta t) - f(\Delta t)\, C(r, 0) \quad . \tag{25}$$

The relation between the three distributions becomes obvious by visualizing the matrix of pairwise distances between positions $\underline{\tilde{r}_i}$ and $\underline{\tilde{r}_j}$ (Figure 24). For the zero lag distribution the diagonal is zero and all other elements belong to distances of particles that are not identical. The absolute distribution $C(r, \Delta t > 0)$ contains all pairwise distances for a respective Δt, including those of corresponding particles and those between un-identical particles, whereas the SLD only contains distances of corresponding particles. Naturally, as long as there are particles that are localized in

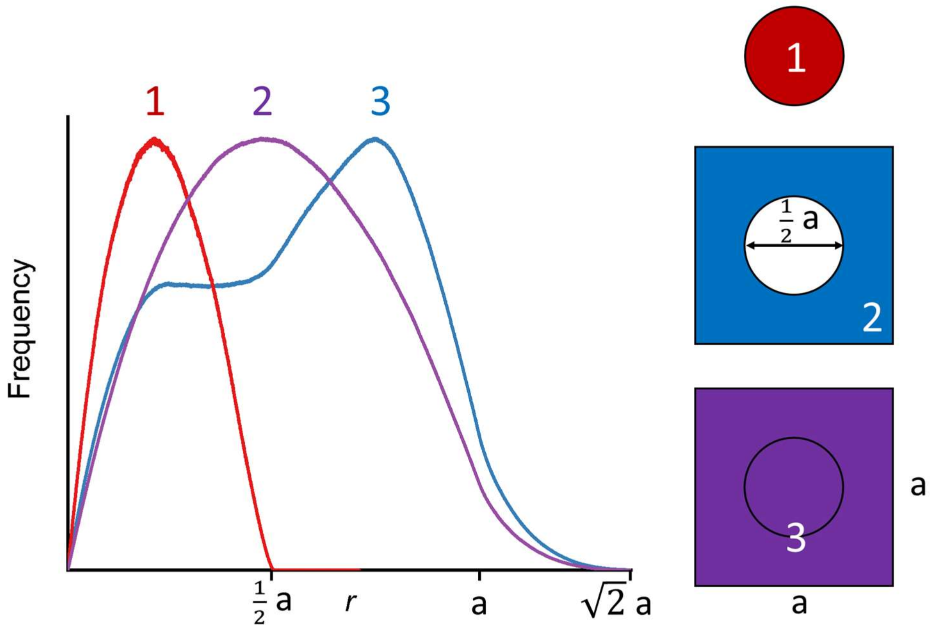


Figure 25: The shape of the background in DANAE directly corresponds to the distribution of particles in the microscopic data. Colored areas are accessible to particles. Observation of white areas is blocked.and thus no distances can be calculated to or from the corresponding area.

the movie for more than one frame, their step length will always be part of the absolute distribution $C(r, \Delta t)$.

The fundamental advantage of DANAE lies in its pure ensemble nature. The indices of particles are reset for every frame and indexing is performed simply through enumeration without any assumption of proximity in the positions of particles from frame to frame. Without the error-prone assignment of particle identities, diffusional artefacts arising from the naturally limited temporal resolution of fluorescence microscopy camera setups vanish in the DANAE approach. For this reason, it is exceptionally well suited to discriminate populations with different mobility (e.g. confined, free, polymerized etc.) in a movie and assign them to different states of molecular interaction.

9.2. Boundary effects in time-resolved single molecule microscopy

Up to this point, particle diffusion occurred under idealized conditions, i.e. a static number of particles and isotropic diffusivity. Experimenters can face four general types of deviations from the ideal situation: (1) the diffusion of particles in- and out of the field of view (FOV), (2) particle mismatching due to interweaving trajectories, (3) changes in photo-state due to statistical transitions between fluorescing and non-fluorescing states by individual particles (blinking) or the photochemical destruction of fluorophores by excitation light (bleaching), and (4) ageing effects during the measurement changing the experimental parameters while they are being measured. This section will focus on the first effect and establish the basis for analyzing single molecule microscopic data under the effect of boundary transitions. Effects (2)-(4) and their influence on both SPT and DANAE will be discussed in the following section.

The amount of particles diffusing in and out of the FOV is often neglected as an influence on apparent diffusivities determined with any SPT based method. This process, however, can have considerable consequences depending on the size of the field of view in relation to the diffusive velocity. One can easily picture a very small FOV with a mean of only one particle in the FOV per frame due to high rate of entering and leaving, thus causing a completely false determination of the SLD. The number of particles leaving the FOV per frame follows a binomial distribution, $M_{out} \sim \mathcal{B}(m, p_{out})$ with m being the number of particles inside the FOV and p_{out} the generalized probability one particle has to leave the FOV. Since the field of view does not present an actual barrier to particles, both the interior of the FOV as well as its surrounding are in equilibrium so that transitions into and out of the FOV are equally probable ($p_{in} = p_{out} \equiv p_{\Delta m}$). For an appropriate amount of particles inside the FOV, their number distribution can be approximated by a normal distribution with mean $\mu_m = \langle m \rangle$ and a standard deviation σ_m. One can derive the probability $p_{\Delta m}$ of a particle entering/leaving the FOV through the Green's function G, which solves the Fokker-Planck equation for a freely diffusing particle in one dimension

$$\frac{\partial G(\tilde{x}_2,\ \Delta t \mid \tilde{x}_1)}{\partial t} = \mathrm{D}\,\frac{\partial^2}{\partial x^2}\, G(\tilde{x}_2,\ \Delta t \mid \tilde{x}_1) \tag{26}$$

with the initial condition $G(\tilde{x}_2,\ 0 \mid \tilde{x}_1) = \delta(\tilde{x}_2 - \tilde{x}_1)$. The Green's function G discribes a particle moving from position $\tilde{x}_1$ to position $\tilde{x}_2$ within a certain lag time Δt and is given by

$$G(\tilde{x}_2,\ \Delta t \mid \tilde{x}_1) = \frac{1}{\sqrt{4\pi \mathrm{D}\Delta t}} \exp\left(-\frac{(\tilde{x}_2 - \tilde{x}_1)^2}{4\mathrm{D}\Delta t}\right)\ . \tag{27}$$

In order to obtain the probability p_a of a particle staying within the field of view, i.e. between 0 and *a*, one has to integrate over the dimensions of the FOV. Since in a homogenous sample the initial distribution is $G(\tilde{x}_1,\ 0) = 1/a$, the staying probability becomes

$$p_a(\Delta t) \ = \int_0^a \int_0^a G(\tilde{x}_2,\ 0)\, G(\tilde{x}_2,\ \Delta t \mid \tilde{x}_1)\, \mathrm{d}\tilde{x}_1\, \mathrm{d}\tilde{x}_2 = \frac{\exp(-\zeta^2) - 1}{\sqrt{\pi}\zeta} + \operatorname{erf}(\zeta) \tag{28}$$

with $\zeta(\mathrm{D}, \Delta t, a) = \ \sqrt{a^2/4\mathrm{D}\Delta t} \ \equiv \ \zeta$.

To account for these transitions it becomes necessary to adjust the expression for both the MSD and SLD alike in order to include field of view induced boundary effects. The MSD in general is the expectation value $\langle x^2\rangle(\Delta t)$ of the squared distances particles move in a given dimension. Using the Green's function, it can be calculated as

$$\begin{aligned}\langle x^2\rangle(\Delta t) \ = \ \langle(\tilde{x}_1 - \ \tilde{x}_2)^2\rangle \ &= \frac{1}{a} \int_0^a \int_0^a (\tilde{x}_1 - \ \tilde{x}_2)^2\, G(\tilde{x}_2,\ \Delta t \mid \tilde{x}_1)\ \mathrm{d}\tilde{x}_1\, \mathrm{d}\tilde{x}_2 \\ &= 2\ \mathrm{D}\ \Delta t \left(\frac{2\ \exp(-\zeta^2) - 1}{\sqrt{\pi}\zeta} + \ \operatorname{erf}(\zeta)\right) \ .\end{aligned} \tag{29}$$

For $a \rightarrow \ \infty$ we get $\langle x^2\rangle(\Delta t) \ = \ 2\ \mathrm{D}\ \Delta t$, which is the standard expression for diffusion in one dimension. This approach however only applies to particles that start inside the FOV and will approach zero for higher Δt since there are no particles left inside the FOV and influx is not considered. We can adjust this by dividing this result by p_a to account for a constant mean particle density inside the FOV and get

$$\langle x^2\rangle_{FOV}(\Delta t) = \frac{\langle x^2\rangle(\Delta t)}{p_a(\Delta t)} = 2\mathrm{D}\Delta t \left(1 \ + \ \frac{\exp(-\zeta^2) - 1}{\exp(-\zeta^2) - 1 \ + \ \sqrt{\pi}\zeta\ \operatorname{erf}(\zeta)}\right) \tag{30}$$

for the one-dimensional and

$$\langle r^2\rangle_{FOV}(\Delta t) = 4\mathrm{D}\Delta t\left(1 + \frac{\exp(-\zeta^2) - 1}{\exp(-\zeta^2) - 1 + \sqrt{\pi}\zeta\,\operatorname{erf}(\zeta)}\right) \tag{31}$$

for the two-dimensional case. Figure 26D shows the mean of 100 one-dimensional simulations with D = 100 µm^2/s and without an imposed field of view which where fitted with free MSD and $\langle x^2\rangle_{FOV}(\Delta t)$ respectively. A fit without regard for the confines of the FOV will always underestimate the diffusivity D.

The isotropic SLD was introduced as the probability distribution of distances in the histogram of correlated step lengths. In experimental terms, the probability to generate a distance r can be expressed in polar coordinates as

$$R(r,\beta_i)\,\mathrm{d}r = P(\beta_i)\,\Psi_{FOV}(r)\,r\,\mathrm{d}r\ , \tag{32}$$

where $P(\beta_i)$ is the probability function of distances r depending on some diffusive parameters β_i and $\Psi_{FOV}(r)$ the angular frequency function due to the influence of the field of view (i.e. the mean angle of a circular cut out inside a quadratic FOV). This becomes obvious in Figure 26A: The greater the distance r and the closer a particle is to the boundaries of the FOV, the smaller the angle $\varphi_{FOV}(r)$ becomes enabling the measurement of such particle inside the FOV.

For a quadratic FOV the $\Psi_{FOV}(r)$ is defined as[140]

$$\Psi_{FOV}(r) = 2\begin{cases} \pi - \left(4\frac{r}{a} - \frac{r^2}{a^2}\right) & ,\quad 0 < r < a \\ 4\sin^{-1}\left(\frac{a}{r}\right) + 4\sqrt{\frac{r^2}{a^2} - 1} - \frac{r^2}{a^2} - \pi - 2 & ,\quad a < r < \sqrt{2}a \end{cases} \tag{33}$$

For randomly generated distances inside the FOV, we have a uniform distribution $P = 1/a^2$ and $R(r)$ becomes the SLPD of equation (22). For the SLD one uses the Greens function of diffusion, i.e. $P(\beta_1,\beta_2) = G(r,\ \Delta t)$ and gets

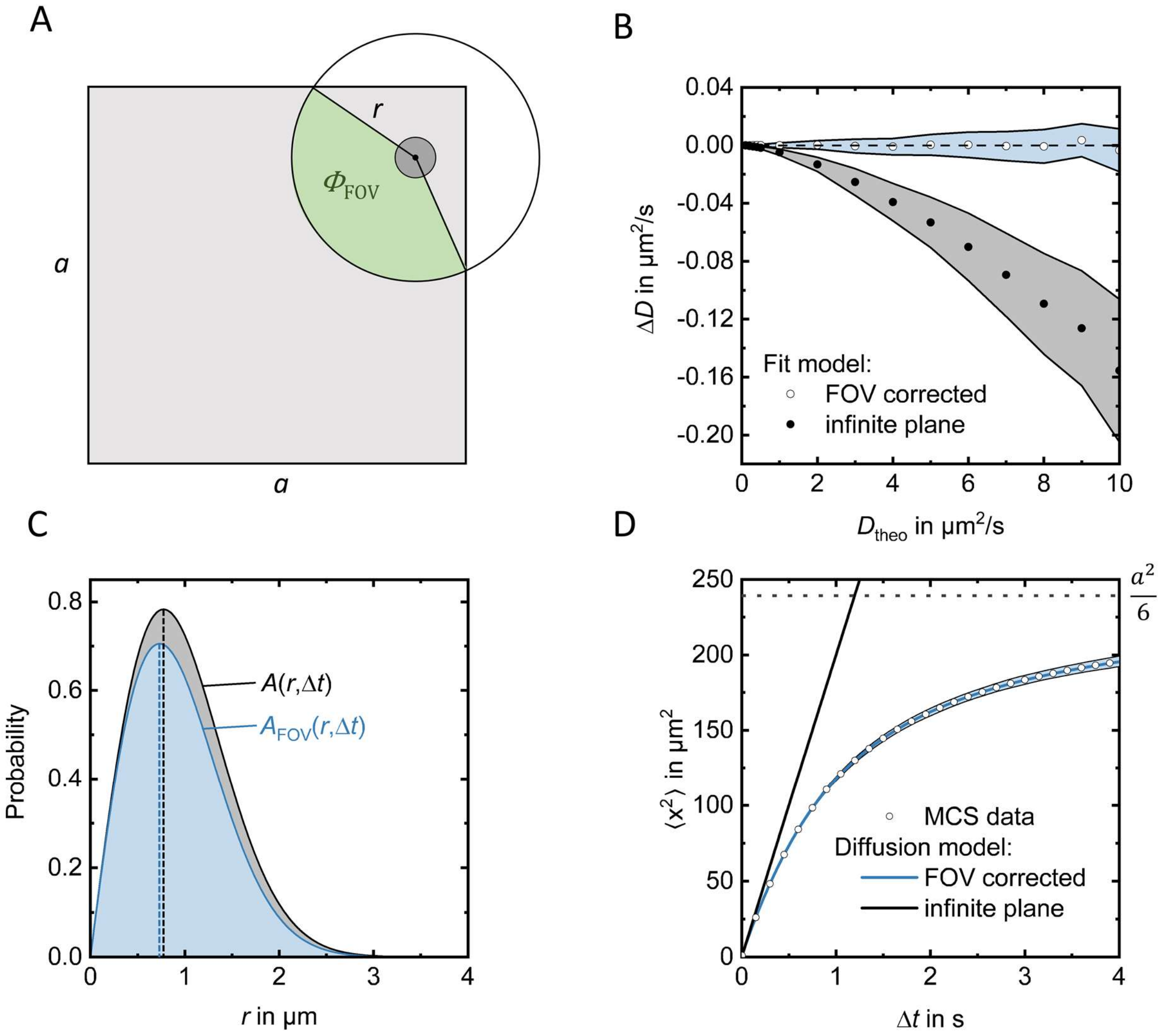


Figure 26: Boundary effects on time-resolved single molecule microscopy data. (A) A particle close to the edges of the FOV moving a distance *r* between two temporally separated frames can only be measured within the angle Φ_{FOV}. (B) Difference in diffusivites derived by fitting MCS localization data with either an SLD based on the assumption of an infinite plane (filled circles, standard deviation as gray area) or and SLD accounting for boundary transitions (non-filled circles, standard deviation as blue area). (C) SLD functions with (blue) and without (black) FOV correction for a diffusivity of 10 µm²/s and a lag time of 30 ms. (D) 1D mean squared displacement of MCS dervived localization data directly taken from the simulation before particle anonymization (non-filled circles) and diffusion model functions including FOV effects (blue line, $\langle r^2 \rangle_{FOV}$) and without them (black line, $\langle r^2 \rangle = 4\,\mathrm{D}\,\Delta t$).

$$\mathrm{A}_{FOV}(r,\Delta t) = \frac{1}{4\pi \mathrm{D}\Delta t} \exp\left(-\frac{r^2}{4\mathrm{D}\Delta t}\right) 2r$$
$$\times \begin{cases} \left(\pi - \left(4\,\frac{r}{a} - \frac{r^2}{a^2}\right)\right) \mathrm{d}r & , \quad 0 < r < a \\ \left(4\,\sin^{-1}\left(\frac{a}{r}\right) + 4\sqrt{\frac{r^2}{a^2} - 1} - \frac{r^2}{a^2} - \pi - 2\right) \mathrm{d}r & , \quad a < r < \sqrt{2}a \end{cases} \tag{34}$$

as the expression for the step length distribution of particles measured inside a quadratic FOV. Depending on the values of D, a, and Δ*t* this function differs both in shape and amplitude from the uncorrected SLD of equation (19) as seen in Figure 26C (the maximum of the distribution is shown as a dashed line to highlight the change in shape). The higher the product D × Δ*t* (or the smaller the FOV), the stronger becomes the impact of FOV effects leading to an underestimation of the experimental value of D as the boundary crossings become more likely. For practical purposes only the first part of equation (34) is of importance, since the SLD amplitude falls to zero well before r approaches the side length a for most reasonable values of D and a. Numerical integration shows that the normalized MSD of equation (31) is indeed the center of mass of the SLD of equation (34) according to

$$\langle r^2 \rangle_{FOV}(\Delta t) = \frac{\int_0^{\sqrt{2}a} \mathrm{A}_{FOV}(r,\Delta t)\, r^2\, \mathrm{d}r}{\int_0^{\sqrt{2}a} \mathrm{A}_{FOV}(r,\Delta t)\, \mathrm{d}r} \quad . \tag{35}$$

Importantly, both expressions for the MSD and SLD are equally applicable to SPT analyses and can be applied without increasing the complexity of the problem.

9.3. Testing the accuracy of DANAE with Monte Carlo simulations

Single particle tracking relies on the correct assignment of identical particles between different frames. This requirement, however, becomes increasingly difficult to meet for particles that are densely distributed and/or rapidly diffusing. Mismatching probabilities can then grow to a level that makes it impossible for SPT to produce correct diffusivities, as demonstrated later in this section. In contrast, DANAE is unaffected by mismatching since it does not match particles at all. Artefactual distances that arise from a distance calculation between two non-identical particles fall into the background distribution and are subtracted at the end of the analysis. The result of this difference between SPT using the nearest neighbor criterion in a JDA and DANAE is shown in Figure 27A in an example with a moderate input diffusivity of 1 $\mu m^2/s$ and increasing frame difference times δt between 5 and 300 ms at a constant mean particle density of 500 particles per FOV of 1436 μm^2. Even with a frame difference time of 30 ms, which is common in nanoscopy, the SPT derived distribution produces fits with a diffusivity that is underestimated to about 92% of its true value. At 50 ms, the diffusivity has dropped to 87%. For increasing frame difference times, the diffusivity is underestimated more and more and the fit model increasingly deviates from the experimental SLDs. At the same time, DANAE leads to the determination of the correct diffusivities in the entire tested range from analyzing the same localization data.

In SPT, when the algorithm has no other means of distinguishing particles than their relative distance, it will choose the distances that are smallest, even if they belong to different particles. When the algorithm operates in a Boolean manner, i.e. it does not apply fuzzy logic[141], which could slightly mitigate the severeness of the effect, the error from one wrong assignment propagates, as this wrongly assigned particle is now missing from its own trace and so forth. As a result, SPD produces a distribution that is shifted to smaller distances, which DANAE would assort into the background distribution, and broadens the long distance slope (as shown in Figure 27B for an exemplary distribution), eventually leading to wrong fits and an apparent smaller value for the diffusivity. For higher values of $\delta t,$ this change in the SPT derived frequency distribution becomes progressively more pronounced while the DANAE derived SLD stays on the actual distance frequency from the MCS (Figure 27A). It should also be noted that due to the mismatching induced shape of the SPT generated distribution a fit with the SLD equation is practically nonsensical for high lag times. As a consequence an experimenter might be tempted to interpret the shape of the SPT distance distribution as being the result of additional particle populations i with different D_i (e.g. one really fast and a slower one) for this assumption of course significantly improves the fit of the distribution, while in fact no such additional populations exist. Figure 27C shows this on the example of the SPT derived distribution of Figure 27B as two subpopulations (SP) appear to produce a good fit, leading to a fundamentally wrong conclusion about the data.

An extensive analysis with DANAE employing MCS generated data sets confirms that the mean diffusivities determined with the DANAE algorithm lie well within the intrinsic spread of the simulations themselves. Neither high input diffusivities D_{theo} (Figure 28A) nor high mean particle

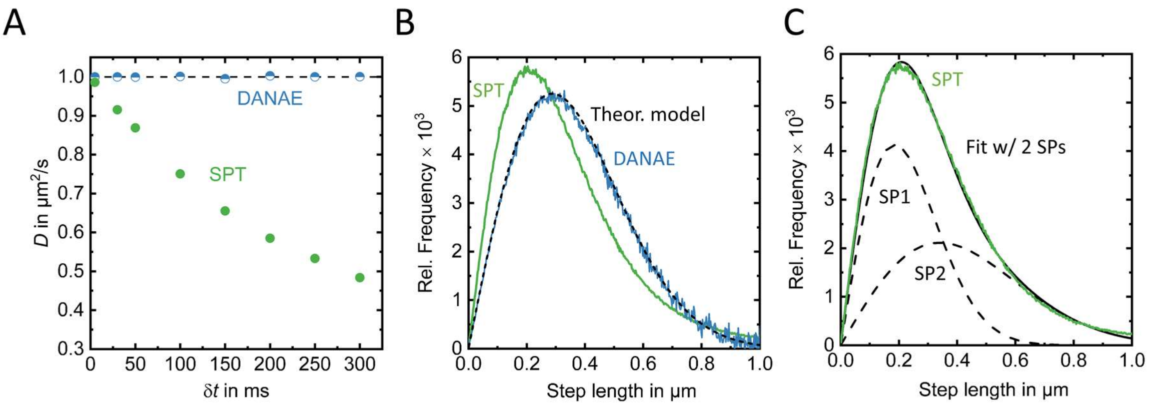


Figure 27: (A) Difference in mean diffusivites D of 50 simulations for each data point from the input value of 1 µm²/s (dashed line) determined by fitting the SLDs generated by DANAE (blue) and SPT (green) for increasing frame difference times δ*t*. (B) SPT (green) and DANAE (blue) derived SLD of a simulation with 800 particles in a FOV of 655.4 µm² with D = 1.4 µm²/s. The dashed black line shows the distribution when values for the parameters of the simulation are put into equation (34). (C) Fit of the same SPT generated distribution (green) with two subpopulations (black line). Individual subpopulations shown as dashed lines.

numbers (Figure 28B) lead to wrong results in DANAE. On the contrary, the results are even improved by higher particle numbers due to better statistics. For this matter, DANAE can be applied where a SPT based analysis is not feasible, thus making it possible to gain insights into processes formerly unresolvable.

To further test the viability of DANAE under extreme conditions, Prof. Helmut Grubmüller from the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany generated three positional files from Monte Carlo simulations and sent to me for analysis using DANAE without knowing the input diffusivities of the simulations. In these simulations, the root mean squared displacement $\sqrt{r^2}$ was set to be significantly larger than the mean inter-particle distance *d* (Figure 28C). The conditions for these simulations were: a_{bath} = 400 µm, a_{FOV} = 200 µm, m_{bath} = 10000, δ*t* = 100 ms. The total number of exported frames was 1000. Under these conditions, there can be no expectation of any SPT-based algorithm (without further knowledge than particle localizations) being able to produce correct results. Yet, in each of the tested simulations the relative error from an analysis with DANAE was smaller than 5%. A closer look at the resulting distributions shows a high degree of noise. However, step length distributions are clearly visible (Figure 28D-F), which can be fitted with a single population.

Up to this point, particles where simulated under idealized conditions in that they were regarded as observable as long as they resided within the FOV. In an experimental setting, however, an experimenter will nearly always need to deal with the statistical nature of fluorescence from single

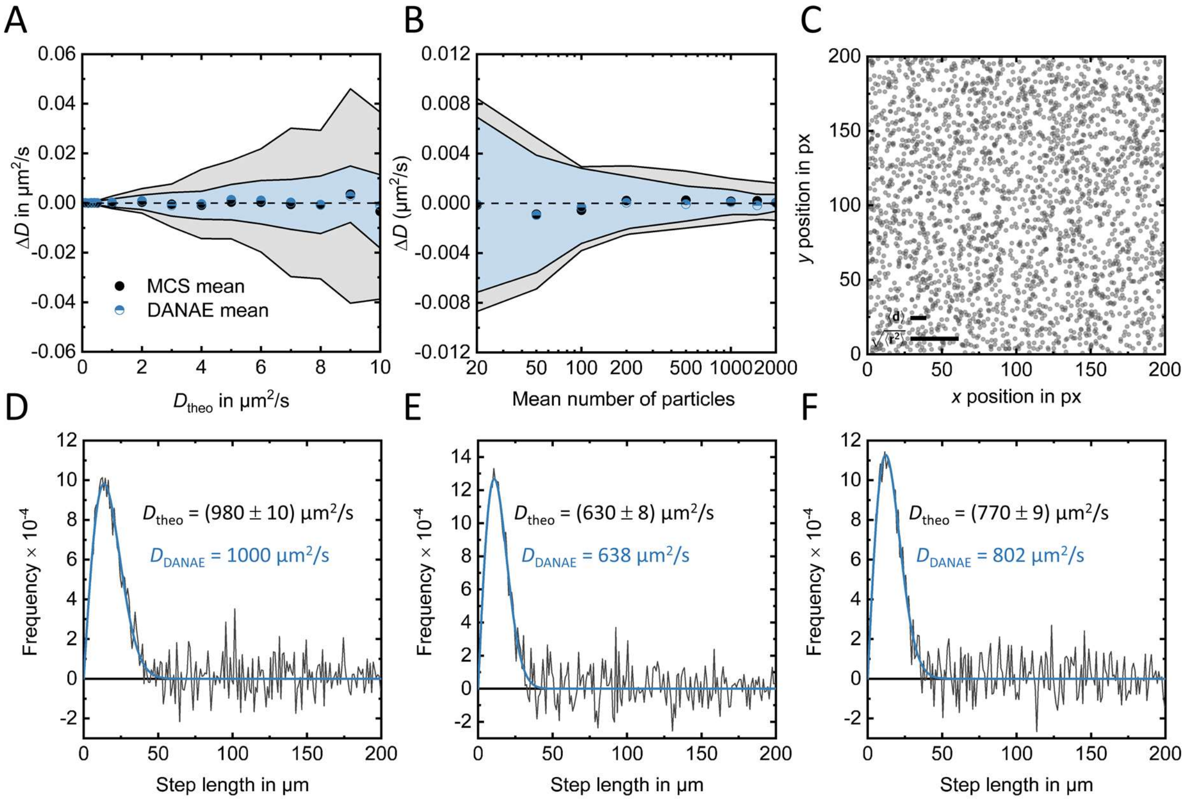


Figure 28: (A) Accurate determination of particle diffusion at increasing diffusivity. Difference from mean values from 50 simulations per data point as determined with DANAE (blue) and directly calculated from the simulations (gray) with respective standard deviations. The particle density was 500 particles per quadratic FOV of 1436 µm^2 and the frame difference time 5 ms (B) Difference from mean values of diffusivity from 50 simulations per particle number as determined with DANAE (blue) and directly calculated from the simulations (gray) with respective standard deviations. The simulation's input diffusivity was 1 µm^2/s and the FOV 1436 µm^2 with a frame difference time of 5 ms. (C) One frame with particle positions from the simulation in (D) with an input diffusivity of D = 980 µm^2/s. (D-F) SLDs of three simulations generated with DANAE (gray) with respective fits (blue). All three simulations had high diffusivities and a high particle density of 2500 particles per FOV of 40,000 µm^2 at frame difference times of 100 ms.

particles in the form of photochemical transitions between fluorescent and non-fluorescent states of particles. While these processes are often times put to excellent use, as in stochastic optical reconstruction microscopy (STORM)[142] and related nanoscopy techniques, they can introduce artifacts into the mobility analysis and are often only avoidable to a degree (e.g. by minimizing excitation or working in deoxidized solutions). Transitions into the dark state can also be permanent, which is referred to as photobleaching. Artifacts induced by these effects can be categorized in mismatching of particles pairs after dis-/appearing events and the apparent depletion of the SLD amplitude for higher lag times Δt (this of course also holds for transitions in and out of the field of view). Both effects do not appear in DANAE since there is neither particle matching nor trace construction where "holes" in the timeline could reduce the frequency of particle connections between time points before and after the gap.

To analyze the impact of photo-state transitions, I performed additional MCSs. In the case of blinking, before exporting the particle positions from the MCS, a population of particles $\{m_{\mathrm{SC}}\}$ was subjected to temporary deletions (simulated blinking). Two characteristic blinking times, τ_{on} and τ_{off}, were set for each simulation and individual on- and off-times of particles were generated from exponential probability density functions. 50 simulations were performed with $\langle m_{\mathrm{FOV}} \rangle = 2000$ and D = 1 µm^2/s. Bleaching simulations were conducted similar to blinking, but without particles returning to the on-state. Two different bleaching populations $\{m_{\mathrm{destr}}\}$ were selected representing 50% and 95% bleached (i.e. destroyed) particles out of 2000 particles. The characteristic off-times were 50% and 75%, respectively, of the overall simulation time of 4000 frames. As was done with blinking, 50 simulations were performed for each scenario.

Figure 29A shows the apparent amplitude depletion in SPT as compared to DANAE, where intermittent non-fluorescence does not remove particle distances from the SLDs. In SPT, whenever traces are interrupted, distances are lost for lag times greater than the time a particle was continuously fluorescing. Any attempt at interconnecting separate trajectories necessarily increases the likelihood of mismatches. Since DANAE does not establish trajectories, its SLD matches the distribution directly generated from the conducted simulation. Furthermore, the disappearance or returning into the light state of a particle can lead to mismatching, if it suddenly satisfies the condition given to SPT for connecting particles, e.g. having the closest distance to the position of a particle in the previous frame. Figure 29B shows the consequence of this blinking induced mismatching on an exemplary simulated measurement with 50 % of 2000 particles undergoing blinking with characteristic times $\tau_{\mathrm{on}} = 50$ ms and $\tau_{\mathrm{off}} = 25$ ms. The mean diffusivity determined with SPT deviates most from the input diffusivity of 1 µm^2/s as compared to DANAE and individual diffusivities vary most between different lag times. Figure 29C shows particle numbers for 1 of 50 simulations with stable photostate, i.e. with fluctuations only arising from FOV boundary transitions, and under the influences of blinking and bleaching. In the most severe case of bleaching, particle numbers quickly decrease to only a small fraction (100 particles) of their initial value of 2000 particles. Yet, even undergoing this amount of bleaching, there is no systematic deviation in the mean diffusivity determined by DANAE after evaluation of 50 simulations per scenario (Figure 29D). Only the standard deviation increases slightly due to drastically lower particle numbers. The same is true for blinking. As can be seen in Figure 29D, all mean diffusivities from particles with changes in photo-state lie within the standard deviation resulting from the

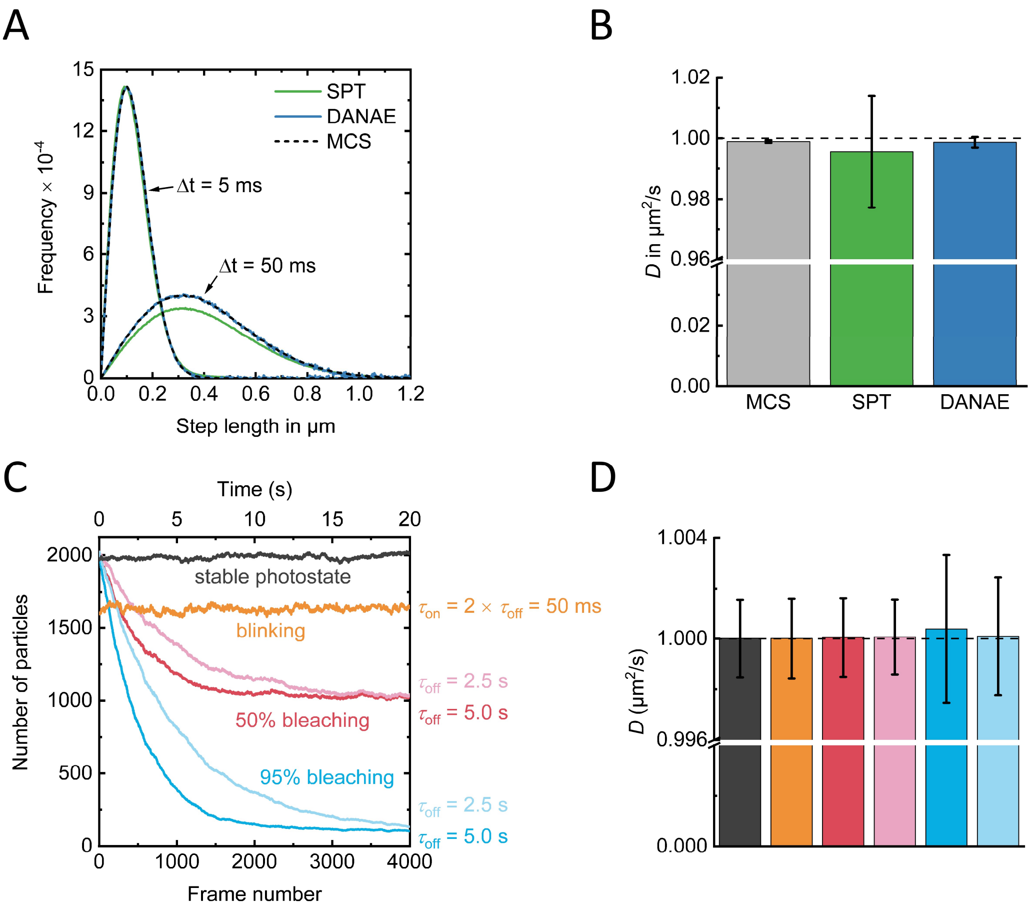


Figure 29: (A) Amplitude loss in SLDs from simulated movies with 50% of 2000 particles blinking with τ_{on} = 50 ms and τ_{on} = 25 ms analyszed with SPT (green line) and DANAE (blue line) for two lag times, 5 ms and 50 ms. The SLDs directly gained from the simulation are shown dashed lines. (B) Determined diffusivity and standard deviation from fitting 10 lag times of (A), each 5 ms apart. (C) Exemplary particle numbers in 1of 50 simulations with differing degrees of blinking and bleaching. (D) Analysis of 50 simulations each under the conditions of (C) with standard deviations and the simulation input diffusivity 1 µm²/s.

statistical nature of the simulations themselves. It can therefore be concluded that the DANAE algorithm is robust against intermittent and permanent non-detectability of particles during the observation time, be it from transitions at the borders of the FOV or transitions in the photo-state of the individual molecules. These processes can analogously be described in terms of partial occlusion. The intrinsic reason why DANAE is not affected is that it does not establish trajectories of particles with assigned identities. Partial occlusion will therefore not result in an increased rate of misidentification due to either a lack of available "suspects" (disappearance) or an overabundance thereof (reappearance).

Another important aspect of measuring diffusion in biological systems is the possibility of non-stationarity not in the fluorescence of the molecules used to detect motion but rather of the target

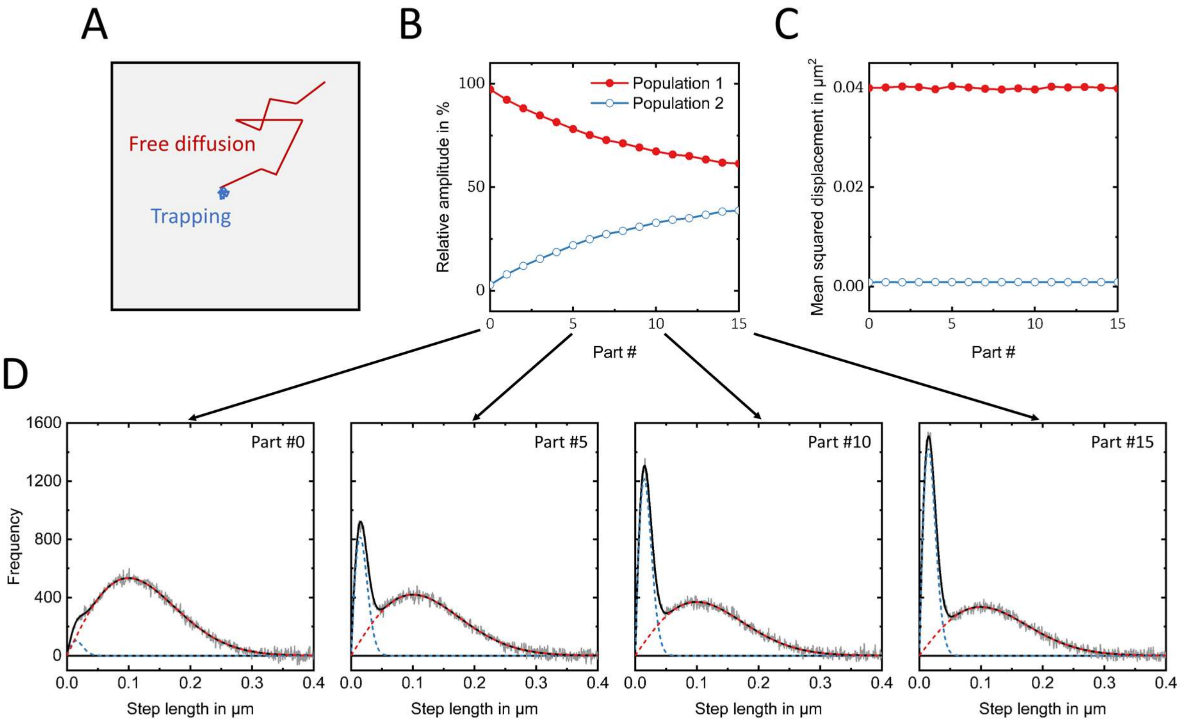


Figure 30: Ageing effects on DANAE. (A) Time-dependent diffusion behavior of particles. A simulated movie begins with 100% of mobile particles of constant diffusivity, of which an increasing subpopulation gets trapped over time. Relative amplitudes (B) and diffusivities (C) resulting from a fit with two subpopulations after partitioning the movie into 15 consecutive sequences and analysis with DANAE. (D) SLDs and fits of four of the histograms from which the parameters in (B) and (C) were extracted.

molecules' diffusivities themselves. This might be the result of structural changes (e.g. dimerization, loss of subunits), the trapping of molecules (receptor binding, adhesion), or differences in the environment (viscosity, membrane charge, chemical composition, temperature). These effects can be summed up under the terminology "ageing". In SPT-based JDA, changes in diffusive parameters over time can lead to smearing between the involved subpopulations' SLDs. Ageing does not affect DANAE in the same way, as subpopulations, which are made up of individual distances, are always resolved on their own. Again, MCSs can provide an illustration of this behavior. The simulation of these trapping-based non-stationary effects was essentially performed the same as for bleaching, however, 50% of particles were not destroyed but forced to move around the trapping position in a small Gaussian shaped probability volume that mimics the localization accuracy in microscopy experiments. As for the simulations of blinking and bleaching a_{bath} was 4048 × 4048 px.

Figure 30A illustrates a particle diffusing freely (red trajectory) until it gets trapped (blue trajectory) and can only move around its trapping site. This results in two subpopulations, one for each type of movement. By splitting the movie into equal parts and applying DANAE, it becomes possible to analyze the amplitudes and mean squared displacements of these subpopulations during the duration of the movie, while particles undergo the transition from Brownian motion to restricted motion. Both parameters are displayed in Figures 30B and 30C. Clearly, the relative amplitude of the freely diffusing subpopulation decreases over time in favor of the amplitude of the trapped subpopulation. At the same time, both MSDs stay the same over the entire movie, meaning that there is no mixture between both diffusive modes in the histograms. This can be seen in Figure 30D, which shows the histograms from four parts of the movie, alongside the fits of both subpopulations. Both subpopulations are well resolved and provide an excellent fit (linear combination shown in black) to the overall distribution (gray), regardless of the transitional stage during the movie.

Depending on the experiment, analyzing ageing effects might be beneficial on its own. It can provide important biological inferences on molecular interaction, e.g. regarding the desensitization of receptors, or the stability of the interacting molecules. More generally, however, taking into consideration the biological stability during the measurement time will yield results that are overall more reliable.

9.4. Two-color DANAE (2cDANAE)

An intuitive modification of DANAE is the transition from calculating the time-separated distances within one data file to calculating distances between two distinct files that can be understood as two color channels in fluorescence microscopy (2cDANAE). Analogous to the Particle Image Cross Correlation Spectroscopy (PICCS) variation of their algorithm by Semrau et. al.[143], this is especially useful when studying the interactions of particles in movies, e.g. the binding of fluorescently labeled proteins in cells. While co-localization[144] is a useful tool for analyzing individual images, co-diffusion is complicated by the addition of the temporal dimension. The basic problem behind this analysis lies in determining what is co-diffusion and what is only a random crossover between the paths of individual molecules, which is in itself highly reminiscent of the problem solved by one-color DANAE. There, only those distances appeared in the SLD that are part of the particles' motion and one could thus ascribe as "self-correlated", while all other distances were subtracted with the background.

To test the concept of 2cDANAE I performed two MCSs with a particle density of 500 particles per quadratic FOV of 1436 μm^2, a frame difference time of 5 ms and a diffusion coefficient of

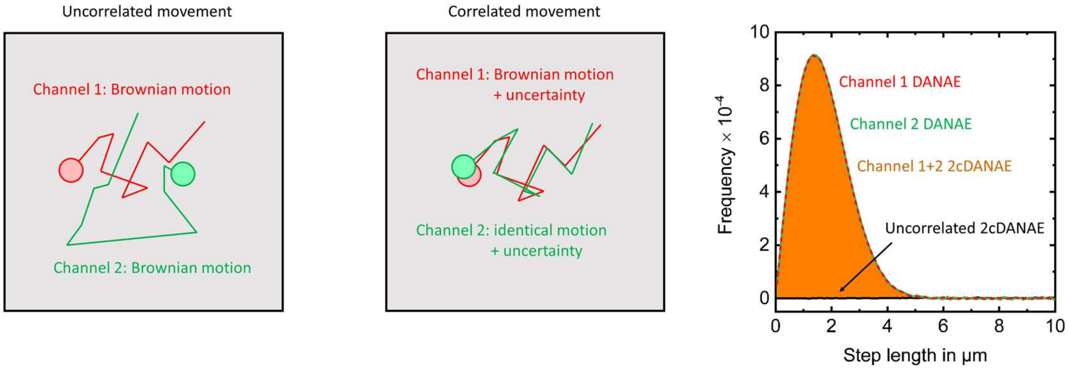


Figure 31: 2cDANAE. In a two-color video microscopy measurement the two extrema of particle behavior are either no interaction and thus uncorrelated movement (left) or fully correlated motion (middle). Uncorrelated diffusion results in no signal after executing 2cDANAE (right, black line). Correlated motion (i.e. co-diffusion, orange area) results in an SLD that is nearly identical in shape and amplitude to the elements of the one-color DANAE SLDs (red dashed line, green dashed line) that are part of the co-diffusional process (in this case 100% of particles co-diffuse).

1 $\mu m^2/s$. Since both simulations had a different random seed, the movements of the particles in each movie should not be correlated to those of particles in the other one (Figure 31A, left). After 2cDANAE, there is indeed no SLD in the step length histogram (Figure 31, right, black line). The easiest way to simulate co-diffusion would be to perform 2cDANAE on the same movie in both channels, which is identical to performing regular DANAE on that movie. Since a real microscopic measurement is limited to the localization accuracy necessarily leading to different values even when measuring the same molecules in two channels, two different normally distributed random uncertainties with $\sigma = 0.0148$ µm were added to the same MCS (Figure 31, middle). 2cDANAE of these two MCSs reveals that the resulting SLD (orange area) is nearly identical in both shape and amplitude to the SLDs generated by performing regular DANAE on the two correlated simulations (red and green dashed lines). Measurement uncertainties therefore do not preclude a successful 2cDANAE analysis.

In a biological context, this means that one can perform 2cDANAE between two channels of fluorescently labeled molecules as well as DANAE on the individual channels and thus precisely determine the amount of interacting particles from the number of distances in the 2cDANAE SLD in relation the number of distances in the one-color SLDs. Since diffusion in living matter is usually more complex than Brownian motion with a single diffusivity, determining the parameters of co-diffusion is of immense value to aid the assignment of diffusive modes in the individual DANAE histograms. Additionally, the decay of amplitudes in the 2cDANAE SLDs for longer lag times Δt contains the kinetics of the underlying interaction between the particles in both channels.

10. Conclusions

The diffusive modes molecules display in biological samples contain valuable information about their underlying biomolecular interactions. Correct analysis of their diffusive properties is therefore key in the identification and quantification of the biological processes observable in single-molecule microscopy. Yet, analysis methods that rely on interconnecting microscopy localization data with limited information can produce highly erroneous jump distance spectra and thus lead to incorrect models. DANAE overcomes this fundamental problem by approaching localization data in a purely statistical way that does not make assumptions about the identity of particles in a movie, which by necessity is affected by its limited means to keep up this identification as the microscopic movie progresses and particles diffuse away. By modifying the equations for the free diffusion SLD and MSD to include boundary effects, DANAE can determine accurate diffusivities in samples with extremely high particle density and high diffusivities, even when the root mean squared displacement is a multiple of the average distance between particles. The DANAE algorithm was shown to be robust against typical experimental influences like intersystem crossing and photodestruction as well as ageing effects potentially exhibited by biological samples. In samples with several subpopulations undergoing different types of diffusion, DANAE produced SLDs that are a linear combination of each diffusive mode. Finally, two-color DANAE was introduced as an analysis method for microscopy experiments with two channels. The algorithm produces SLDs only in case of co-diffusion and can thus be used to directly quantify the modes of interaction between the two fluorescent target molecule species as well as aid the analysis of complex single-channel step length histograms.

Initial application of the algorithm was used to perform analysis in the following publication:

Volz, P.,Boreham, A., Wolf, A., Kim, T.Y., Balke, J., Frombach, J., Hadam, S., Afraz, Z., Rancan, F., Blume-Peytavi, U., Vogt, A., Alexiev, U., Application of Single Molecule Fluorescence Microscopy to Characterize the Penetration of a Large Amphiphilic Molecule in the Stratum Conerum of Human Skin. *Int. J. Mol. Sci.* **2015**, 16(4), 6960-6977.

References

1. Kirchberg, K.; Michel, H.; Alexiev, U., Exploring the entrance of proton pathways in cytochrome c oxidase from Paracoccus denitrificans: Surface charge, buffer capacity and redox-dependent polarity changes at the internal surface. *Biochim. Biophys. Acta, Bioenerg.* **2013,** *1827* (3), 276-284.

2. Arnold, S., Cytochrome c oxidase and its role in neurodegeneration and neuroprotection. *Adv. Exp. Med Biol* **2012,** *748*, 305-39.

3. Rak, M.; Benit, P.; Chretien, D.; Bouchereau, J.; Schiff, M.; El-Khoury, R.; Tzagoloff, A.; Rustin, P., Mitochondrial cytochrome c oxidase deficiency. *Clin. Sci.* **2016,** *130* (6), 393-407.

4. Woelke, A. L.; Galstyan, G.; Knapp, E. W., Lysine 362 in cytochrome c oxidase regulates opening of the K-channel via changes in pK_a and conformation. *Biochim. Biophys. Acta, Bioenerg.* **2014,** *1837* (12), 1998-2003.

5. Supekar, S.; Kaila, V. R. I., Dewetting transitions coupled to K-channel activation in cytochrome c oxidase. *Chem. Sci.* **2018,** *9* (32), 6703-6710.

6. Sharma, V.; Wikström, M., The role of the K-channel and the active-site tyrosine in the catalytic mechanism of cytochrome c oxidase. *Biochim. Biophys. Acta, Bioenerg.* **2016,** *1857* (8), 1111-1115.

7. Qin, L.; Liu, J.; Mills, D. A.; Proshlyakov, D. A.; Hiser, C.; Ferguson-Miller, S., Redox-Dependent Conformational Changes in Cytochrome a Oxidase Suggest a Gating Mechanism for Proton Uptake. *Biochemistry* **2009,** *48* (23), 5121-5130.

8. Marechal, A.; Rich, P. R., Water molecule reorganization in cytochrome c oxidase revealed by FTIR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (21), 8634-8638.

9. Marechal, A.; Hartley, A. M.; Warelow, T. P.; Meunier, B.; Rich, P. R., Comparison of redox and ligand binding behaviour of yeast and bovine cytochrome c oxidases using FTIR spectroscopy. *Biochim. Biophys. Acta, Bioenerg.* **2018,** *1859* (9), 705-711.

10. Shen, H.; Tauzin, L. J.; Baiyasi, R.; Wang, W. X.; Moringo, N.; Shuang, B.; Landes, C. F., Single Particle Tracking: From Theory to Biophysical Applications. *Chem. Rev.* **2017,** *117* (11), 7331-7376.

11. Sbalzarini, I. F.; Koumoutsakos, P., Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.* **2005,** *151* (2), 182-195.

12. Och, L. M.; Shields-Zhou, G. A., The Neoproterozoic oxygenation event: Environmental perturbations and biogeochemical cycling. *Earth-Sci. Rev.* **2012,** *110* (1-4), 26-57.

13. Stryer, L.; Berg, J. M.; Tymoczko, J. L., *Biochemie*. 7th ed.; 2014, 529.

14. Mitchell, P., Coupling of Phosphorylation to Electron and Hydrogen Transfer by a Chemi-Osmotic Type of Mechanism. *Nature* **1961,** *191* (478), 144-&.

15. Hirst, J., Mitochondrial complex I. *Annu. Rev. Biochem.* **2013,** *82*, 551-75.

16. Esposti, M. D., Genome Analysis of Structure-Function Relationships in Respiratory Complex I, an Ancient Bioenergetic Enzyme. *Genome Biol. Evol.* **2016,** *8* (1), 126-147.

17. Murphy, M. P., How mitochondria produce reactive oxygen species. *Biochem. J.* **2009,** *417*, 1-13.

18. Poprac, P.; Jomova, K.; Simunkova, M.; Kollar, V.; Rhodes, C. J.; Valko, M., Targeting Free Radicals in Oxidative Stress-Related Human Diseases. *Trends Pharmacol. Sci.* **2017,** *38* (7), 592-607.

19. Cecchini, G., Function and structure of complex II of the respiratory chain. *Annu. Rev. Biochem.* **2003,** *72*, 77-109.

20. Berry, E. A.; Guergova-Kuras, M.; Huang, L. S.; Crofts, A. R., Structure and function of cytochrome bc complexes. *Annu. Rev. Biochem.* **2000,** *69*, 1005-75.

21. Stryer, L.; Berg, J. M.; Tymoczko, J. L., *Biochemie.* 7th ed.; 2014, 547.

22. Brzezinski, P.; Gennis, R. B., Cytochrome c oxidase: exciting progress and remaining mysteries. *J. Bioenerg. Biomembr.* **2008,** *40* (5), 521-531.

23. Genova, M. L.; Lenaz, G., Functional role of mitochondrial respiratory supercomplexes. *Biochim. Biophys. Acta* **2014,** *1837* (4), 427-43.

24. Junge, W.; Nelson, N., ATP synthase. *Annu. Rev. Biochem.* **2015,** *84*, 631-57.

25. Cross, R. L., The Mechanism and Regulation of Atp Synthesis by F1-Atpases. *Annu. Rev. Biochem.* **1981,** *50*, 681-714.

26. Boyer, P. D., The Binding Change Mechanism for Atp Synthase - Some Probabilities and Possibilities. *Biochim. Biophys. Acta* **1993,** *1140* (3), 215-250.

27. Wedatilake, Y.; Brown, R. M.; McFarland, R.; Yaplito-Lee, J.; Morris, A. A.; Champion, M.; Jardine, P. E.; Clarke, A.; Thorburn, D. R.; Taylor, R. W.; Land, J. M.; Forrest, K.; Dobbie, A.; Simmons, L.; Aasheim, E. T.; Ketteridge, D.; Hanrahan, D.; Chakrapani, A.; Brown, G. K.; Rahman, S., SURF1 deficiency: a multi-centre natural history study. *Orphanet J. Rare. Dis.* **2013,** *8*, 96.

28. Vogt, S.; Ruppert, V.; Pankuweit, S.; Paletta, J. P. J.; Rhiel, A.; Weber, P.; Irqsusi, M.; Cybulski, P.; Ramzan, R., Myocardial insufficiency is related to reduced subunit 4 content of cytochrome c oxidase. *J. Cardiothorac. Surg.* **2018,** *13* (1), 95.

29. Chandran, K.; Aggarwal, D.; Migrino, R. Q.; Joseph, J.; McAllister, D.; Konorev, E. A.; Antholine, W. E.; Zielonka, J.; Srinivasan, S.; Avadhani, N. G.; Kalyanaraman, B., Doxorubicin inactivates myocardial cytochrome c oxidase in rats: cardioprotection by Mito-Q. *Biophys. J.* **2009,** *96* (4), 1388-98.

30. Chaturvedi, R. K.; Flint Beal, M., Mitochondrial diseases of the brain. *Free. Radic. Biol. Med.* **2013,** *63*, 1-29.

31. Huttemann, M.; Helling, S.; Sanderson, T. H.; Sinkler, C.; Samavati, L.; Mahapatra, G.; Varughese, A.; Lu, G.; Liu, J.; Ramzan, R.; Vogt, S.; Grossman, L. I.; Doan, J. W.; Marcus, K.; Lee, I., Regulation of mitochondrial respiration and apoptosis through cell signaling: cytochrome c oxidase and cytochrome c in ischemia/reperfusion injury and inflammation. *Biochim. Biophys. Acta* **2012,** *1817* (4), 598-609.

32. Hargreaves, I. P.; Al Shahrani, M.; Wainwright, L.; Heales, S. J., Drug-Induced Mitochondrial Toxicity. *Drug. Saf.* **2016,** *39* (7), 661-74.

33. Dorlo, T. P.; Balasegaram, M.; Beijnen, J. H.; de Vries, P. J., Miltefosine: a review of its pharmacology and therapeutic efficacy in the treatment of leishmaniasis. *J. Antimicrob. Chemother.* **2012,** *67* (11), 2576-97.

34. Arnold, S., Cytochrome c oxidase and its role in neurodegeneration and neuroprotection. *Adv. Exp. Med. Biol.* **2012,** *748*, 305-39.

35. Srinivasan, S.; Guha, M.; Dong, D. W.; Whelan, K. A.; Ruthel, G.; Uchikado, Y.; Natsugoe, S.; Nakagawa, H.; Avadhani, N. G., Disruption of cytochrome c oxidase function induces the Warburg effect and metabolic reprogramming. *Oncogene* **2016,** *35* (12), 1585-95.

36. Whitaker-Menezes, D.; Martinez-Outschoorn, U. E.; Flomenberg, N.; Birbe, R. C.; Witkiewicz, A. K.; Howell, A.; Pavlides, S.; Tsirigos, A.; Ertel, A.; Pestell, R. G.; Broda, P.; Minetti, C.; Lisanti, M. P.; Sotgia, F., Hyperactivation of oxidative mitochondrial metabolism in epithelial cancer cells in situ: visualizing the therapeutic effects of metformin in tumor tissue. *Cell Cycle* **2011,** *10* (23), 4047-64.

37. Oliva, C. R.; Markert, T.; Ross, L. J.; White, E. L.; Rasmussen, L.; Zhang, W.; Everts, M.; Moellering, D. R.; Bailey, S. M.; Suto, M. J.; Griguer, C. E., Identification of Small Molecule Inhibitors of Human Cytochrome c Oxidase That Target Chemoresistant Glioma Cells. *J. Biol. Chem.* **2016,** *291* (46), 24188-24199.

38. Gray, M. W., Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* **2012,** *4* (9), a011403.

39. Hashimoto, Y.; Niikura, T.; Tajima, H.; Yasukawa, T.; Sudo, H.; Ito, Y.; Kita, Y.; Kawasumi, M.; Kouyama, K.; Doyu, M.; Sobue, G.; Koide, T.; Tsuji, S.; Lang, J.; Kurokawa, K.; Nishimoto, I., A rescue factor abolishing neuronal cell death by a wide spectrum of familial Alzheimer's disease genes and A beta *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98* (22), 12854-12854.

40. Timon-Gomez, A.; Nyvltova, E.; Abriata, L. A.; Vila, A. J.; Hosler, J.; Barrientos, A., Mitochondrial cytochrome c oxidase biogenesis: Recent developments. *Semin. Cell Dev. Biol.* **2018,** *76*, 163-178.

41. Solioz, M.; Carafoli, E.; Ludwig, B., The Cytochrome-C Oxidase of Paracoccus-Denitrificans Pumps Protons in a Reconstituted System. *J. Biol. Chem.* **1982,** *257* (4), 1579-1582.

42. Nicoletti, F.; Witt, H.; Ludwig, B.; Brunori, M.; Malatesta, F., Paracoccus denitrificans cytochrome c oxidase: a kinetic study on the two- and four-subunit complexes. *Biochim. Biophys. Acta, Bioenerg.* **1998,** *1365* (3), 393-403.

43. John, P.; Whatley, F. R., Paracoccus denitrificans and evolutionary origin of mitochondrion. *Nature* **1975,** *254* (5500), 495-498.

44. Koepke, J.; Olkhova, E.; Angerer, H.; Muller, H.; Peng, G.; Michel, H., High resolution crystal structure of Paracoccus denitrificans cytochrome c oxidase: new insights into the active site and the proton transfer pathways. *Biochim. Biophys. Acta* **2009,** *1787* (6), 635-45.

45. Liu, X. H.; Yu, Y.; Hu, C.; Zhang, W.; Lu, Y.; Wang, J. Y., Significant Increase of Oxidase Activity through the Genetic Incorporation of a Tyrosine-Histidine Cross-Link in a Myoglobin Model of Heme-Copper Oxidase. *Angew. Chem., Int. Ed.* **2012,** *51* (18), 4312-4316.

46. Yu, Y.; Lv, X.; Li, J.; Zhou, Q.; Cui, C.; Hosseinzadeh, P.; Mukherjee, A.; Nilges, M. J.; Wang, J.; Lu, Y., Defining the role of tyrosine and rational tuning of oxidase activity by genetic incorporation of unnatural tyrosine analogs. *J. Am. Chem. Soc.* **2015,** *137* (14), 4594-7.

47. Kaila, V. R. I.; Verkhovsky, M.; Hummer, G.; Wikstrom, M., Prevention of leak in the proton pump of cytochrome c oxidase. *Biochimica Et Biophysica Acta-Bioenergetics* **2008,** *1777* (7-8), 890-892.

48. Belevich, I.; Verkhovsky, M. I.; Wikstrom, M., Proton-coupled electron transfer drives the proton pump of cytochrome c oxidase. *Nature* **2006,** *440* (7085), 829-32.

49. Dragelj, J. Computing the $\mathrm{pK_A}$ Values of Functional Residues in Proteins: The Case of Proton Transfer in Cytochrome cOxidase. Freie Universität Berlin, Berlin, Germany, 2018.

50. Supekar, S.; Gamiz-Hernandez, A. P.; Kaila, V. R. I., A Protonated Water Cluster as a Transient Proton-Loading Site in Cytochrome c Oxidase. *Angew. Chem., Int. Ed.* **2016,** *55* (39), 11940-11944.

51. Richter, O. M.; Durr, K. L.; Kannt, A.; Ludwig, B.; Scandurra, F. M.; Giuffre, A.; Sarti, P.; Hellwig, P., Probing the access of protons to the K pathway in the Paracoccus denitrificans cytochrome c oxidase. *FEBS J.* **2005,** *272* (2), 404-12.

52. Ferguson-Miller, S.; Hiser, C.; Liu, J., Gating and regulation of the cytochrome c oxidase proton pump. *Biochim. Biophys. Acta* **2012,** *1817* (4), 489-94.

53. Hiser, C.; Buhrow, L.; Liu, J.; Kuhn, L.; Ferguson-Miller, S., A conserved amphipathic ligand binding region influences k-path-dependent activity of cytochrome C oxidase. *Biochemistry* **2013,** *52* (8), 1385-96.

54. Chertkova, R. V.; Brazhe, N. A.; Bryantseva, T. V.; Nekrasov, A. N.; Dolgikh, D. A.; Yusipovich, A. I.; Sosnovtseva, O.; Maksimov, G. V.; Rubin, A. B.; Kirpichnikov, M. P., New insight into the mechanism of mitochondrial cytochrome c function. *PLoS One* **2017,** *12* (5).

55. Michel, H., Cytochrome c oxidase: Catalytic cycle and mechanisms of proton pumping-A discussion. *Biochemistry* **1999,** *38* (46), 15129-15140.

56. Wikstrom, M.; Sharma, V., Proton pumping by cytochrome c oxidase - A 40 year anniversary. *Biochim. Biophys. Acta, Bioenerg.* **2018,** *1859* (9), 692-698.

57. Vilhjalmsdottir, J.; Gennis, R. B.; Brzezinski, P., The electron distribution in the "activated" state of cytochrome c oxidase. *Sci. Rep.* **2018,** *8*.

58. Harrenga, A.; Michel, H., The cytochrome c oxidase from Paracoccus denitrificans does not change the metal center ligation upon reduction. *J. Biol. Chem.* **1999,** *274* (47), 33296-9.

59. Valeur, B.; Berberan-Santos, M. N., *Molecular Fluorescence.* 2nd ed.; 2013, 23.

60. Angerer, H. Functional and Structural Studies on the Atmungsferment Cytochrome c Oxidase from Paracoccus denitrificans. Johann Wolfgang Goethe-Universität Frankfurt am Main, 2008.

61. Bjorck, M. L.; Brzezinski, P., Control of transmembrane charge transfer in cytochrome c oxidase by the membrane potential. *Nat. Commun.* **2018,** *9*.

62. Shvadchak, V. V.; Kucherak, O.; Afitska, K.; Dziuba, D.; Yushchenko, D. A., Environmentally sensitive probes for monitoring protein-membrane interactions at nanomolar concentrations. *Biochim. Biophys. Acta, Biomembr.* **2017,** *1859* (5), 852-859.

63. Klymchenko, A. S., Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Acc. Chem. Res.* **2017,** *50* (2), 366-375.

64. Alexiev, U.; Marti, T.; Heyn, M. P.; Khorana, H. G.; Scherrer, P., Covalently Bound Ph-Indicator Dyes at Selected Extracellular or Cytoplasmic Sites in Bacteriorhodopsin .2. Rotational Orientation of Helix-D and Helix-E and Kinetic Correlation between M-Formation and Proton Release in Bacteriorhodopsin Micelles. *Biochemistry* **1994,** *33* (46), 13693-13699.

65. Alexiev, U.; Rimke, I.; Pohlmann, T., Elucidation of the nature of the conformational changes of the EF-interhelical loop in bacteriorhodopsin and of the helix VIII on the cytoplasmic surface of bovine rhodopsin: A time-resolved fluorescence depolarization study. *J. Mol. Biol.* **2003,** *328* (3), 705-719.

66. Kim, T. Y.; Schlieter, T.; Haase, S.; Alexiev, U., Activation and molecular recognition of the GPCR rhodopsin - Insights from time-resolved fluorescence depolarisation and single molecule experiments. *Eur. J. Cell Biol.* **2012,** *91* (4), 300-310.

67. Klonis, N.; Sawyer, W. H., Spectral properties of the prototropic forms of fluorescein in aqueous solution. *J. Fluoresc.* **1996,** *6* (3), 147-57.

68. Moller, M.; Alexiev, U., Surface Charge Changes upon Formation of the Signaling State in Visual Rhodopsin. *Photochem. Photobiol.* **2009,** *85* (2), 501-508.

69. Fischermeier, E.; Pospisil, P.; Sayed, A.; Hof, M.; Solioz, M.; Fahmy, K., Dipolar Relaxation Dynamics at the Active Site of an ATPase Regulated by Membrane Lateral Pressure. *Angew. Chem., Int. Ed.* **2017,** *56* (5), 1269-1272.

70. Koehorst, R. B. M.; Spruijt, R. B.; Hemminga, M. A., Site-directed fluorescence labeling of a membrane protein with BADAN: Probing protein topology and local environment. *Biophys. J.* **2008,** *94* (10), 3945-3955.

71. Kim, T. Y. Zeitaufgelöste fluoreszenzspektroskopische Untersuchungen an der cytoplasmatischen Oberfläche von Rhodopsin Freie Universität Berlin, Berlin, Germany, 2012.

72. Alexiev, U.; Farrens, D. L., Fluorescence spectroscopy of rhodopsins: insights and approaches. *Biochim. Biophys. Acta* **2014,** *1837* (5), 694-709.

73. Schröder, G. F.; Alexiev, U.; Grubmuller, H., Simulation of fluorescence anisotropy experiments: Probing protein dynamics. *Biophys. J.* **2005,** *89* (6), 3757-3770.

74. Pospíšil, p. Fluorescence Spectroscopy: Advanced methods and their defined applications in protein science. J. Heyrovský Institute of Physical Chemistry of the CAS, v. v. i., Prague, Czech Republic, 2017.

75. Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009,** *6* (5), 343-U41.

76. de Gier, J. W. L.; Lubben, M.; Reijnders, W. N. M.; Tipker, C. A.; Slotboom, D. J.; Vanspanning, R. J. M.; Stouthamer, A. H.; Vanderoost, J., The Terminal Oxidases of Paracoccus-Denitrificans. *Mol. Microbiol.* **1994,** *13* (2), 183-196.

77. Gerhus, E.; Steinrucke, P.; Ludwig, B., Paracoccus-Denitrificans Cytochrome-C1 Gene Replacement Mutants. *J. Bacteriol.* **1990,** *172* (5), 2392-2400.

78. Lucioli, S.; Hoffmeier, K.; Carrozzo, R.; Tessa, A.; Ludwig, B.; Santorelli, F. M., Introducing a novel human mtDNA mutation into the Paracoccus denitrificans COX I gene explains functional deficits in a patient. *Neurogenetics* **2006,** *7* (1), 51-57.

79. Hilbers, F.; von der Hocht, I.; Ludwig, B.; Michel, H., True wild type and recombinant wild type cytochrome c oxidase from Paracoccus denitrificans show a 20-fold difference in their catalase activity. *Biochim. Biophys. Acta, Bioenerg.* **2013,** *1827* (3), 319-327.

80. Schimo, S.; Wittig, I.; Pos, K. M.; Ludwig, B., Cytochrome c Oxidase Biogenesis and Metallochaperone Interactions: Steps in the Assembly Pathway of a Bacterial Complex. *PLoS One* **2017,** *12* (1).

81. Bundschuh, F. A.; Hoffmeier, K.; Ludwig, B., Two variants of the assembly factor Surf1 target specific terminal oxidases in Paracoccus denitrificans. *Biochim. Biophys. Acta* **2008,** *1777* (10), 1336-43.

82. Dash, B. P.; Alles, M.; Bundschuh, F. A.; Richter, O. M. H.; Ludwig, B., Protein chaperones mediating copper insertion into the Cu-A site of the aa(3)-type cytochrome c oxidase of Paracoccus denitrificans. *Biochim. Biophys. Acta, Bioenerg.* **2015,** *1847* (2), 202-211.

83. de Vries, G. E.; Harms, N.; Hoogendijk, J.; Stouthamer, A. H., Isolation and Characterization of Paracoccus-Denitrificans Mutants with Increased Conjugation Frequencies and Pleiotropic Loss of a (Ngatcn) DNA-Modifying Property. *Arch. Microbiol.* **1989,** *152* (1), 52-57.

84. Goodhew, C. F.; Pettigrew, G. W.; Devreese, B.; VanBeeumen, J.; VanSpanning, R. J. M.; Baker, S. C.; Saunders, N.; Ferguson, S. J.; Thompson, I. P., The cytochromes c-550 of Paracoccus denitrificans and Thiosphaera pantotropha: A need for re-evaluation of the history of Paracoccus cultures. *FEMS Microbiol. Lett.* **1996,** *137* (1), 95-101.

85. Raitio, M.; Jalli, T.; Saraste, M., Isolation and Analysis of the Genes for Cytochrome-C-Oxidase in Paracoccus-Denitrificans. *EMBO J.* **1987,** *6* (9), 2825-2833.

86. Boja, E. S.; Fales, H. M., Overalkylation of a protein digest with iodoacetamide. *Anal. Chem.* **2001,** *73* (15), 3576-3582.

87. Kirchberg, K. Untersuchung der Protonenbewegung während des O→E Schrittes im katalytischen Zyklus der Cytochrom-*c*-Oxidase von *Paracoccus denitrificans*. Johann Wolfgang Goethe-Universität, Frankfurt am Main, Germany, 2006.

88. Witt, H.; Malatesta, F.; Nicoletti, F.; Brunori, M.; Ludwig, B., Tryptophan 121 of subunit II is the electron entry site to cytochrome-c oxidase in Paracoccus denitrificans - Involvement of a hydrophobic patch in the docking reaction. *J. Biol. Chem.* **1998,** *273* (9), 5132-5136.

89. Shimada, S.; Shinzawa-Itoh, K.; Baba, J.; Aoe, S.; Shimada, A.; Yamashita, E.; Kang, J. Y.; Tateno, M.; Yoshikawa, S.; Tsukihara, T., Complex structure of cytochrome c-cytochrome c oxidase reveals a novel protein-protein interaction mode. *EMBO J.* **2017,** *36* (3), 291-300.

90. Liu, J.; Hiser, C.; Ferguson-Miller, S., Role of conformational change and K-path ligands in controlling cytochrome c oxidase activity. *Biochem. Soc. Trans.* **2017,** *45*, 1345-1345.

91. Buhrow, L.; Ferguson-Miller, S.; Kuhn, L. A., From Static Structure to Living Protein: Computational Analysis of Cytochrome c Oxidase Main-chain Flexibility. *Biophys. J.* **2012,** *102* (9), 2158-2166.

92. Oliveira, A. S. F.; Campos, S. R. R.; Baptista, A. M.; Soares, C. M., Coupling between protonation and conformation in cytochrome c oxidase: Insights from constant-pH MD simulations. *Biochim. Biophys. Acta, Bioenerg.* **2016,** *1857* (6), 759-771.

93. Rich, P. R.; Marechal, A., Functions of the hydrophilic channels in protonmotive cytochrome c oxidase. *J. R. Soc., Interface* **2013,** *10* (86).

94. Rasaiah, J. C.; Garde, S.; Hummer, G., Water in nonpolar confinement: from nanotubes to proteins and beyond. *Annu. Rev. Phys. Chem.* **2008,** *59*, 713-40.

95. Agmon, N., The Grotthuss Mechanism. *Chem. Phys. Lett.* **1995,** *244* (5-6), 456-462.

96. Wikstrom, M.; Verkhovsky, M. I.; Hummer, G., Water-gated mechanism of proton translocation by cytochrome c oxidase. *Biochim. Biophys. Acta, Bioenerg.* **2003,** *1604* (2), 61-65.

97. Sass, H. J.; Buldt, G.; Gessenich, R.; Hehn, D.; Neff, D.; Schlesinger, R.; Berendzen, J.; Ormos, P., Structural alterations for proton translocation in the M state of wild-type bacteriorhodopsin. *Nature* **2000,** *406* (6796), 649-653.

98. Mondal, S.; Mukherjee, S.; Bagchi, B., Protein Hydration Dynamics: Much Ado about Nothing? *J. Phys. Chem. Lett.* **2017,** *8* (19), 4878-4882.

99. Owenius, R.; Osterlund, M.; Lindgren, M.; Svensson, M.; Olsen, O. H.; Persson, E.; Freskgard, P. O.; Carlsson, U., Properties of spin and fluorescent labels at a receptor-ligand interface. *Biophys. J.* **1999,** *77* (4), 2237-2250.

100. Riedl, J.; Pohl, R.; Ernsting, N. P.; Orsag, P.; Fojta, M.; Hocek, M., Labelling of nucleosides and oligonucleotides by solvatochromic 4-aminophthalimide fluorophore for studying DNA-protein interactions. *Chem. Sci.* **2012,** *3* (9), 2797-2806.

101. Dziuba, D.; Pospisil, P.; Matyasovsky, J.; Brynda, J.; Nachtigallova, D.; Rulisek, L.; Pohl, R.; Hof, M.; Hocek, M., Solvatochromic fluorene-linked nucleoside and DNA as color-changing fluorescent probes for sensing interactions. *Chem. Sci.* **2016,** *7* (9), 5775-5785.

102. Burstein, E. A.; Emelyanenko, V. I., Log-normal description of fluorescence spectra of organic fluorophores. *Photochem. Photobiol.* **1996,** *64* (2), 316-320.

103. Vincent, M.; de Foresta, B.; Gallay, J., Nanosecond dynamics of a mimicked membrane-water interface observed by time-resolved stokes shift of LAURDAN. *Biophys. J.* **2005,** *88* (6), 4337-4350.

104. Koehorst, R. B. M.; Laptenok, S.; van Oort, B.; van Hoek, A.; Spruijt, R. B.; van Stokkum, I. H. M.; van Amerongen, H.; Hemminga, M. A., Profiling of dynamics in protein-lipid-water systems: a time-resolved fluorescence study of a model membrane protein with the label BADAN at specific membrane depths. *Eur. Biophys. J.* **2010,** *39* (4), 647-656.

105. Jurkiewicz, P.; Sykora, J.; Olzynska, A.; Humplickova, J.; Hof, M., Solvent relaxation in phospholipid bilayers: Principles and recent applications. *J. Fluoresc.* **2005,** *15* (6), 883-894.

106. Kawski, A.; Kuklinski, B.; Bojarski, P., Ground and excited state dipole moments of BADAN and ACRYLODAN determined from solvatochromic shifts of absorption and fluorescence spectra. *Z. Naturforsch. A* **2001,** *56* (5), 407-411.

107. Laage, D.; Elsaesser, T.; Hynes, J. T., Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017,** *117* (16), 10694-10725.

108. Samanta, A.; Fessenden, R. W., Excited state dipole moment of PRODAN as determined from transient dielectric loss measurements. *J. Phys. Chem. A* **2000,** *104* (39), 8972-8975.

109. Amaro, M.; Brezovsky, J.; Kovacova, S.; Maier, L.; Chaoupkova, R.; Sykora, J.; Paruch, K.; Damborsky, J.; Hof, M., Are Time-Dependent Fluorescence Shifts at the Tunnel Mouth of Haloalkane Dehalogenase Enzymes Dependent on the Choice of the Chromophore? *J. Phys. Chem. B* **2013,** *117* (26), 7898-7906.

110. Li, L.; Li, C.; Zhang, Z.; Alexov, E., On the 'Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013,** *9* (4), 2126-2136.

111. Eisenberg, D., Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem.* **1984,** *53*, 595-623.

112. Pospíšil, P.; Luxem, K. E.; Ener, M.; Sýkora, J.; Kocábová, J.; Gray, H. B.; Vlček, A.; Hof, M., Fluorescence Quenching of (Dimethylamino)naphthalene Dyes Badan and Prodan by Tryptophan in Cytochromes P450 and Micelles. *J. Phys. Chem. B* **2014,** *118* (34), 10085-10091.

113. Qiu, W. H.; Zhang, L. Y.; Kao, Y. T.; Lu, W. Y.; Li, T. P.; Kim, J.; Sollenberger, G. M.; Wang, L. J.; Zhong, D. P., Ultrafast hydration dynamics in melittin folding and aggregation: Helix formation and tetramer self-assembly. *J. Phys. Chem. B* **2005,** *109* (35), 16901-16910.

114. Ghane, T.; Gorriz, R. F.; Wrzalek, S.; Volkenandt, S.; Dalatieh, F.; Reidelbach, M.; Imhof, P., Hydrogen-Bonded Network and Water Dynamics in the D-channel of Cytochrome c Oxidase. *J. Membr. Biol.* **2018,** *251* (3), 299-314.

115. Goyal, P.; Lu, J. X.; Yang, S.; Gunner, M. R.; Cui, Q., Changing hydration level in an internal cavity modulates the proton affinity of a key glutamate in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (47), 18886-18891.

116. Son, C. Y.; Yethiraj, A.; Cui, Q., Cavity hydration dynamics in cytochrome c oxidase and functional implications. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114* (42), E8830-E8836.

117. Liang, R. B.; Swanson, J. M. J.; Peng, Y. X.; Wikstrom, M.; Voth, G. A., Multiscale simulations reveal key features of the proton-pumping mechanism in cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113* (27), 7420-7425.

118. Kirchberg, K.; Michel, H.; Alexiev, U., Net proton uptake is preceded by multiple proton transfer steps upon electron injection into cytochrome c oxidase. *J. Biol. Chem.* **2012,** *287* (11), 8187-93.

119. Pal, K.; Koner, A. L., Rationally Designed Solvatochromic Fluorescent Indoline Derivatives for Probing Mitochondrial Environment. *Chemistry* **2017,** *23* (36), 8610-8614.

120. Strizhak, A. V.; Postupalenko, V. Y.; Shvadchak, V. V.; Morellet, N.; Guittet, E.; Pivovarenko, V. G.; Klymchenko, A. S.; Mely, Y., Two-Color Fluorescent L-Amino Acid Mimic of Tryptophan for Probing Peptide-Nucleic Acid Complexes. *Bioconjugate Chem.* **2012,** *23* (12), 2434-2443.

121. Venkatraman, P.; Nguyen, T. T.; Sainlos, M.; Bilsel, O.; Chitta, S.; Imperiali, B.; Stern, L. J., Fluorogenic probes for monitoring peptide binding to class II MHC proteins in living cells. *Nat. Chem. Biol.* **2007,** *3* (4), 222-228.

122. Budisa, N., *Engineering the Genetic Code: Expanding the Amino Acid Repertoire for the Design of Novel Proteins*. 2006; p 1-296.

123. Budin, I.; de Rond, T.; Chen, Y.; Chan, L. J. G.; Petzold, C. J.; Keasling, J. D., Viscous control of cellular respiration by membrane lipid composition. *Science* **2018,** *362* (6419), 1186-1189.

124. Kim, T. Y.; Uji-i, H.; Moller, M.; Muls, B.; Hofkens, J.; Alexiev, U., Monitoring the Interaction of a Single G-Protein Key Binding Site with Rhodopsin Disk Membranes upon Light Activation. *Biochemistry* **2009,** *48* (18), 3801-3803.

125. Nordlund, I., Eine neue Bestimmung der Avogadroschen Konstante aus der Brownschen Bewegung kleiner, in Wasser suspendierter Quecksilberkugelchen. *Z. Phys. Chem.* **1914,** *87*, 40-62.

126. Gelles, J.; Schnapp, B. J.; Sheetz, M. P., Tracking Kinesin-Driven Movements with Nanometre-Scale Precision. *Nature* **1988,** *331* (6155), 450-453.

127. Brandenburg, B.; Zhuang, X. W., Virus trafficking - learning from single-virus tracking. *Nat. Rev. Microbiol.* **2007,** *5* (3), 197-208.

128. Seisenberger, G.; Ried, M. U.; Endress, T.; Buning, H.; Hallek, M.; Brauchle, C., Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. *Science* **2001,** *294* (5548), 1929-1932.

129. Wang, Z. G.; Liu, S. L.; Zhang, Z. L.; Tian, Z. Q.; Tang, H. W.; Pang, D. W., Exploring Sialic Acid Receptors-Related Infection Behavior of Avian Influenza Virus in Human Bronchial Epithelial Cells by Single-Particle Tracking. *Small* **2014,** *10* (13), 2712-2720.

130. Speil, J.; Baumgart, E.; Siebrasse, J. P.; Veith, R.; Vinkemeier, U.; Kubitscheck, U., Activated STAT1 Transcription Factors Conduct Distinct Saltatory Movements in the Cell Nucleus. *Biophys. J.* **2011,** *101* (11), 2592-2600.

131. Kusumi, A.; Sako, Y.; Yamamoto, M., Confined Lateral Diffusion of Membrane-Receptors as Studied by Single-Particle Tracking (Nanovid Microscopy) - Effects of Calcium-Induced Differentiation in Cultured Epithelial-Cells. *Biophys. J.* **1993,** *65* (5), 2021-2040.

132. Ritchie, K.; Shan, X. Y.; Kondo, J.; Iwasawa, K.; Fujiwara, T.; Kusumi, A., Detection of non-Brownian diffusion in the cell membrane in single molecule tracking. *Biophys. J.* **2005,** *88* (3), 2266-2277.

133. Bates, I. R.; Hebert, B.; Luo, Y. S.; Liao, J.; Bachir, A. I.; Kolin, D. L.; Wiseman, P. W.; Hanrahan, J. W., Membrane lateral diffusion and capture of CFTR within transient confinement zones. *Biophys. J.* **2006,** *91* (3), 1046-1058.

134. Smith, P. R.; Morrison, I. E. G.; Wilson, K. M.; Fernandez, N.; Cherry, R. J., Anomalous diffusion of major histocompatibility complex class I molecules on HeLa cells determined by single particle tracking. *Biophys. J.* **1999,** *76* (6), 3331-3344.

135. Simson, R.; Sheets, E. D.; Jacobson, K., Detection of Temporary Lateral Confinement of Membrane-Proteins Using Single-Particle Tracking Analysis. *Biophys. J.* **1995,** *69* (3), 989-993.

136. Anderson, C. M.; Georgiou, G. N.; Morrison, I. E. G.; Stevenson, G. V. W.; Cherry, R. J., Tracking of Cell-Surface Receptors by Fluorescence Digital Imaging Microscopy Using a Charge-Coupled Device Camera - Low-Density-Lipoprotein and Influenza-Virus Receptor Mobility at 4-Degrees-C. *J. Cell. Sci.* **1992,** *101* (2), 415-425.

137. Kues, T.; Dickmanns, A.; Luhrmann, R.; Peters, R.; Kubitscheck, U., High intranuclear mobility and dynamic clustering of the splicing factor U1 snRNP observed by single particle tracking. *P. Natl. Acad. Sci. U.S.A.* **2001,** *98* (21), 12021-12026.

138. Crank, J., *The mathematics of diffusion*. 2d ed.; Clarendon Press: Oxford, Eng, 1975; p viii, 414 p.

139. Semrau, S.; Schmidt, T., Particle image correlation spectroscopy (PICS): Retrieving nanometer-scale correlations from high-density single-molecule position data. *Biophys. J.* **2007,** *92* (2), 613-621.

140. Garwood, F., The Variance of the Overlap of Geometrical Figures with Reference to a Bombing Problem. *Biometrika* **1947,** *34* (1-2), 1-17.

141. Jiang, S.; Zhou, X. B.; Kirchhausen, T.; Wong, S. T. C., Tracking molecular particles in live cells using fuzzy rule-based system. *Cytometry, Part A* **2007,** *71a* (8), 576-584.

142. Rust, M. J.; Bates, M.; Zhuang, X. W., Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **2006,** *3* (10), 793-795.

143. Semrau, S.; Holtzer, L.; Gonzalez-Gaitan, M.; Schmidt, T., Quantification of Biological Interactions with Particle Image Cross-Correlation Spectroscopy (PICCS). *Biophys. J.* **2011,** *100* (7), 1810-1818.

144. Dunn, K. W.; Kamocka, M. M.; McDonald, J. H., A practical guide to evaluating colocalization in biological microscopy. *Am. J. Physiol.-Cell Physiol.* **2011,** *300* (4), C723-C742.

Publications

The following articles were published during the course of this thesis:

Volz, P.,Boreham, A., Wolf, A., Kim, T.Y., Balke, J., Frombach, J., Hadam, S., Afraz, Z., Rancan, F., Blume-Peytavi, U., Vogt, A., Alexiev, U., Application of Single Molecule Fluorescence Microscopy to Characterize the Penetration of a Large Amphiphilic Molecule in the Stratum Conerum of Human Skin. *Int. J. Mol. Sci.* **2015**, 16(4), 6960-6977.

Wolf, A., Schneider, C., Kim, T.Y., Kirchberg, K., Volz, P., Alexiev, U., A simulation-guided fluorescence correlation spectroscopy tool to investigate to protonation dynamics of cytrochrome c oxidase. *Phys. Chem. Chem. Phys.* **2016**, 18, 12877-12885.

Döge, N., Hadam, S.,Volz, P., Wolf, A., Schönborn, K.H., Blume-Peytavi, U., Alexiev, U., Vogt, A., Identification of polystyrene nanoparticle penetration across intact skin barrier as rare event at sites of focal particle aggregations. *J. Biophotonics* **2017**, 110, 39-46.

Balke, J., Volz, P., Neumann, F., Brodwolf, R., Wolf, A., Pischon, H., Radbruch, M., Mundhenk, L., Gruber, A.D., Ma, N., Alexiev, U., Visualizing Oxidative Cellular Stress Induced by Nanoparticles in the Subcytotoxic Range using Fluorescence Lifetime Imaging. *Small* **2018**, 14 (23), 1800310.

The following articles are soon to be published or in revision:

Wolf, A., T., Balke, Wonneberg, J., Alexiev, U., Electronation dependent structural change at the proton exit side of cytochrome *c* oxidase as revealed by site-directed fluorescence labeling. *FEBSJ. (In revision)*

Wolf, A., Dragelj, J., Wonneberg, J., Stellmacher, J., Woelke, A.L., Hodoscek, M., Knapp, E.W., Alexiev, U., Structurally induced protonation and water dynamics changes accompany K-channel activation in cytochrome *c* oxidase. *Manuscript in preparation*

Wolf, A., Volz-Rakebrand, P., Schlieter, T., Balke, J., Alexiev, U., DANAE/UPICS: A tracking-free diffusivity analysis for single molecule imaging. *Manuscript in preparation*

Supplement 1: List of primers

Table SI1: Primers used in this work (Residue umbers refer to unmodified gene product, not the crystal structure)

Primer name	**Sequence**
ctaDII_C48S_f	5‘-GACGGTGAAGGATACCGAGATCAG-3‘
ctaDII_C48S_r	5‘-CTGATCTCGGTATCCTTCACCGTC-3‘
ctaDII_C140S_f	5‘-CAGGGCCACGCCCGAGACATACATC-3‘
ctaDII_C140S_r	5‘-GATGTATGTCTCGGGCGTGGCCCTG-3‘
ctaC_C(79/85)S_f	5‘-GAAACGGACGATCGAGATCAGCAGCAGCAGCGAGACGAAGATCGTC-3‘
ctaC_C(79/85)S_r	5‘-GACGATCTTCGTCTCGCTGCTGCTGCTGATCTCGATCGTCCGTTTC -3‘
ctaE_C185S_f	5‘-CTTGCAAGCCGGTGAACGAGACGCCCAGGATCAC-3‘
ctaE_C185S_f	5‘-GTGATCCTGGGCGTCTCGTTCACCGGCTTGCAAG-3‘
ctaE_C231S_f	5‘-GCAGGCGGATCAGCGAGACGAACAGGAAGATC-3‘
ctaE_C231S_f	5‘-GATCTTCCTGTTCGTCTCGCTGATCCGCCTGC-3‘
ctaDII_A259C_f	5‘-CGCCGCCGCCGCACGGATCGAAGAAC-3‘
ctaDII_A259C_r	5‘-GTTCTTCGATCCGTGCGGCGGCGGCG-3‘
ctaDII_T173C_f	5‘-GAATAGCCCGCCTCGCAGGTCGAGAGC -3‘
ctaDII_T173C_r	5‘-GCTCTCGACCTGCGAGGCGGGCTATTC-3‘
ctaDII_D156C_f	5‘-GAACCCATCTGGCAGTTGCCGCCCGGCG-3‘
ctaDII_D156C_r	5‘-CGCCGGGCGGCAACTGCCAGATGGGTTC-3‘
ctaDII_S295C_f	5‘-CTTGGCGAAGGTGCAGATGACGTGGCTG-3‘
ctaDII_S295C_r	5‘-CAGCCACGTCATCTGCACCTTCGCCAAG-3‘
ctaDII_K299C_f	5‘-CGAAGATCGGCTTGCAGGCGAAGGTCGAG-3‘
ctaDII_K299C_r	5‘-CTCGACCTTCGCCTGCAAGCCGATCTTCG -3‘
ctaDII_P301C_f	5‘-CAGGTAGCCGAAGATGCACTTCTTGGCGAAGG-3‘
ctaDII_P301C_r	5‘-CCTTCGCCAAGAAGTGCATCTTCGGCTACCTG-3‘
pAW/5371_revN2_2	5‘-GCGGGCTCCGGTCCTGGACTTCGTGGCATGATGCGGCCCCGGAG-3‘
pAW/9766_fwdN2_2	5‘-CGGCAGGCCCCTGTCCTGATAGGGCATCGTTGGATCAATGC-3‘
C2/9734_revN2_2	5‘-GACAGGGGCCTGCCGGCGGCATG-3‘
C2/5371_fwdN2_2	5‘-AGGACCGGAGCCCGCCCCCGTG-3‘
opp_pAW_fwd1	5‘-AGCCCTAGGCGGGGCCTGATAGGGCATCGTTGGATCAATGCGGTG-3‘
opp_pAW_fwd2	5‘-TGCCGTGCGGGACGACTGATAGGGCATCGTTGGATCAATGCGGTG-3‘
opp_c2_fwd	5‘-TATCATGCCGCCGGCAGGCCCCTGTC-3‘
opp_c2_rev1	5‘-GCCCCGCCTAGGGCTTTTTGACAAG-3‘
opp_c2_rev2	5‘-TCGTCCCGCACGGCACGCAGGC-3‘
2t_pAW11_rev	5‘-ATCCAACGATGCCCTATCAG-3‘
2t_pAW11_fwd	5‘-AGCTCCTGAAAATCTCGATAACTC-3‘
2t_pKH163_rev	5‘-AGATTTTCAGGAGCTCCAAGCGCGCAATTAACCCTC-3‘
2t_pKH163_fwd	5‘-AGGGCATCGTTGGATTCGAGCCGAAAGCATGCG-3‘

Supplement 2: Comparison of cta operons of vectors pAW13 and pKH160

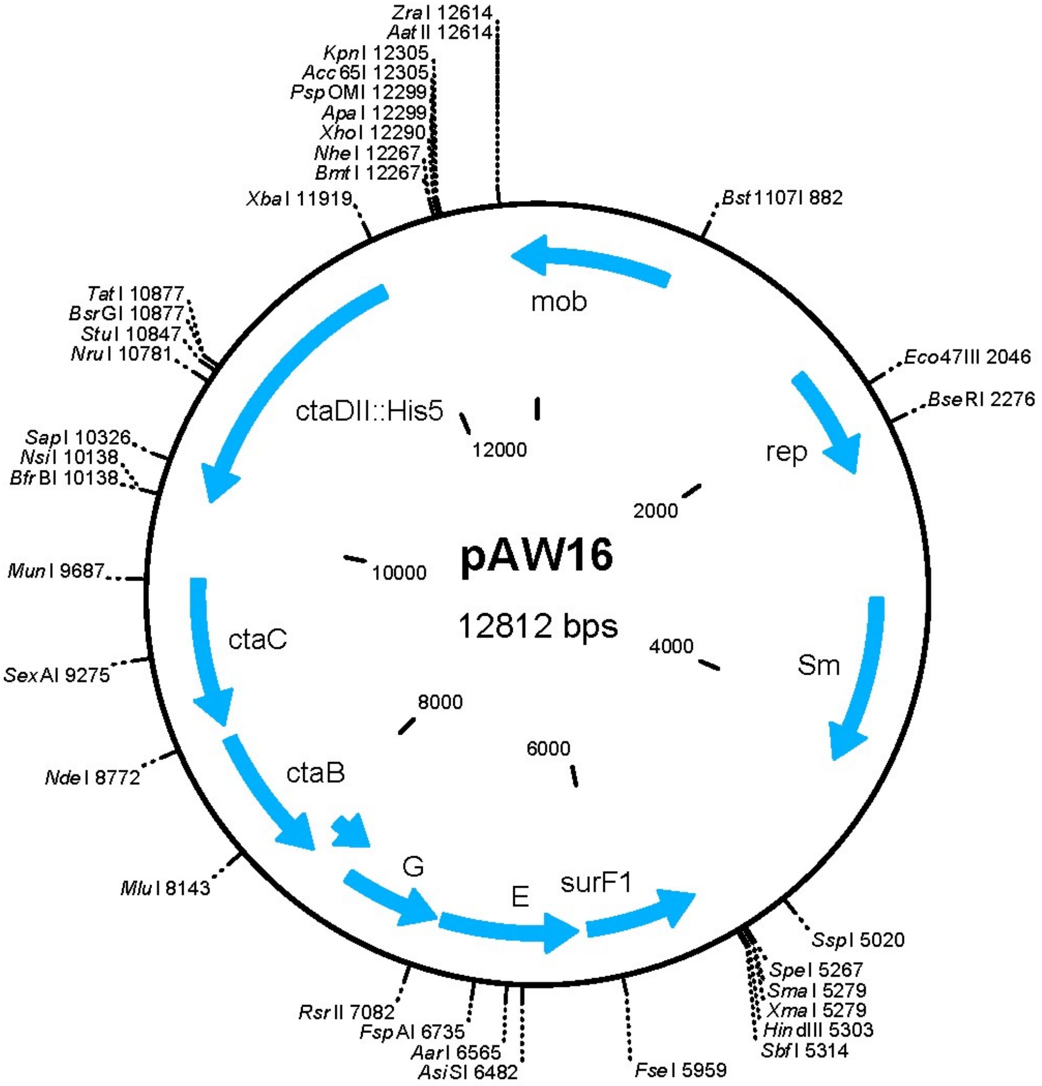
```
                                                                              SurF1
                                                         <--------------------------------
pAW13                   5400 tgtcgaaacgtaacgcatcgcgtcccccctagaattgccgctgcctgatacgccagatcag
pKH160                  5400 tgtcgaaacgtaacgcatcgcgtcccccctagaattgccgctgcctgatacgccagatcag

pAW13                   5460 cgccactgtcatccctgcccaaaccgcagcgatcatgaaccattgcgcggcatagctcag
pKH160                  5460 cgccactgtcatccctgcccaaaccgcagcgatcatgaaccattgcgcggcatagctcag

pAW13                   5520 gtggttgttcggaatgccctcgaccgcgaccgggatcggccgcacccccctgcgcatcgcc
pKH160                  5520 gtggttgttcggaatgccctcgaccgcgaccgggatcggccgcacccccctgcgcatcgcc

pAW13                   5580 ccgcacctcggccgccacgaccagcaccggctcggtccccagctgcgcggccatggcggg
pKH160                  5580 ccgcacctcggccgccacgaccagcaccggctcggtccccagctgcgcggccatggcggg

pAW13                   5640 cacgtcgcgggcgaaccagacattctcggtcaggttgggctcgggcgtggcgctgcccctt
pKH160                  5640 cacgtcgcgggcgaaccagacattctcggtcaggttgggctcgggcgtggcgctgcccctt

pAW13                   5700 ttcgtcgggccagtgcaggttgcccgccacctccagccgcaccggcgggcgcggggcacg
pKH160                  5700 ttcgtcgggccagtgcaggttgcccgccacctccagccgcaccggcgggcgcggggcacg

pAW13                   5760 cttgtggtcctgatcgacgaagccccgatccagcaggatcctgcgcccgtcatcggtgac
pKH160                  5760 cttgtggtcctgatcgacgaagccccgatccagcaggatcctgcgcccgtcatcggtgac

pAW13                   5820 gaaacccgagacgacctgatagccgccgcccgcctcgcgcgtgcccgacagcacgtcgat
pKH160                  5820 gaaacccgagacgacctgatagccgccgcccgcctcgcgcgtgcccgacagcacgtcgat

pAW13                   5880 ctcctgacccgtggtctggcccgaaaccagcaccggcatgtatttcatcgacggattcac
pKH160                  5880 ctcctgacccgtggtctggcccgaaaccagcaccggcatgtatttcatcgacggattcac

pAW13                   5940 cgccgccggcaggggcacgggccggccctcgatcctttgctgaatctgggcgatcagccc
pKH160                  5940 cgccgccggcaggggcacgggccggccctcgatcctttgctgaatctgggcgatcagccc

pAW13                   6000 ttctttccagtccagccgctgaagctgccacatgcccagcgaaatcaggatggcgcagcc
pKH160                  6000 ttctttccagtccagccgctgaagctgccacatgcccagcgaaatcaggatggcgcagcc

pAW13                   6060 caccacaccgacgatcagcgggaacaggtaacggcgcatgggtgggcctccaaagcaaac
pKH160                  6060 caccacaccgacgatcagcgggaacaggtaacggcgcatgggtgggcctccaaagcaaac
                                                                                     ctaE
                                                                          <---------------
pAW13                   6120 ggcgcgggaacatctgccccgcgccgtggaaatctctgtcgggatcagcggccccagata
pKH160                  6120 ggcgcgggaacatctgccccgcgccgtggaaatctctgtcgggatcagcggccccagata

pAW13                   6180 tagatcacgacaaagaggaacagccagaccacatcgacgaaatgccaataccaggcggcg
pKH160                  6180 tagatcacgacaaagaggaacagccagaccacatcgacgaaatgccaataccaggcggcg

pAW13                   6240 gcctcgaagccgacatgctgcttctgggtcatctggcccttgagcaggcggatcaggcag
pKH160                  6240 gcctcgaagccgacatgctgcttctgggtcatctggcccttgagcaggcggatcaggcag

pAW13                   6300 acgaacaggaagatcgtgccgatgatgacatgggcaccgtggaacccggtggccatgtag
pKH160                  6300 acgaacaggaagatcgtgccgatgatgacatgggcaccgtggaacccggtggccatgtag

pAW13                   6360 aaggcgccggcatagaccgtatcggccaggccaaacgccgcgtggctgtattcataggct
pKH160                  6360 aaggcgccggcatagaccgtatcggccaggccaaacgccgcgtggctgtattcataggct

pAW13                   6420 tgcaagccggtgaagcagacgcccaggatcaccgcgacgatcagcccgttgatcgtggtc
pKH160                  6420 tgcaagccggtgaagcagacgcccaggatcaccgcgacgatcagcccgttgatcgtggtc

pAW13                   6480 ttgcgatcgccttcatgcacgaaggcatggtgcgcccaggtcacggcaacgcccgaaagc
pKH160                  6480 ttgcgatcgccttcaagcacgaaggcatggtgcgcccaggtcacggcaacgcccgaaagc

pAW13                   6540 agcaggatcagcgtgttgatcagcggcaggtgccaggggtcgaaggtaacgatgccctcg
pKH160                  6540 agcaggatcagcgtgttgatcagcggcaggtgccaggggtcgaaggtaacgatgccctcg

pAW13                   6600 ggcggccagacgccgtccttgatcgggctgtccggtcccatcggatacagggcgttcttg
pKH160                  6600 ggcggccagacgccgtccttgatcgggctgtccggtcccatcggatacagggcgttcttg
```

```
pAW13            6660 atgaaggcccagaaccaggcgacgaagaacatcacctcggacatgatgaacaggatgaag
pKH160           6660 atgaaggcccagaaccaggcgacgaagaacatcacctcggacatgatgaacaggatgaag

pAW13            6720 ccgtattgcaggccgatgcgcaccaccggggtatgttcgcccgtctcgccctcgttcacg
pKH160           6720 ccgtattgcaggccgatgcgcaccaccggggtatgttcgcccgtctcgccctcgttcacg

pAW13            6780 acatcggcccaccagccgaacatgacgtagagcacgcccaccagaccgatcaggaacatc
pKH160           6780 acatcggcccaccagccgaacatgacgtagagcacgcccaccagaccgatcaggaacatc

pAW13            6840 cacggcccctcgacaggaagcccgaagaaggtgatgcccttcatccaggcgaccgcgccg
pKH160           6840 cacggcccctcgacaggaagcccgaagaaggtgatgcccttcatccaggcgaccgcgccg

pAW13            6900 gtcagcatcacgaaggcgccgatcgcgccgaagaacggccagatcgatggcggcaggatc
pKH160           6900 gtcagcatcacgaaggcgccgatcgcgccgaagaacggccagatcgatggcggcaggatc

pAW13            6960 tgatagtcgtggttctttacatgggccatggctgcatttccggctgttgctggttcttgt
pKH160           6960 tgatagtcgtggttctttacatgggccatggctgcatttccggctgttgctggttcttgt
                                                          ctaG
pAW13            7020 tgctcagtttacggtcggttccgttttcgcgtccagcgccgcctgtttgggcgcgggcgg
pKH160           7020 tgctcagtttacggtcggttccgttttcgcgtccagcgccgcctgtttgggcgcgggcgg

pAW13            7080 atcggtccgatggaaggtataggacagggtgatgtcgcggatgcgaccggcgtcgcggtc
pKH160           7080 atcggtccgatggaaggtataggacagggtgatgtcgcggat---cacggcgtcgcggtc

pAW13            7140 gttcaccagatcggcatcgacgaagaagctgaccggcatctcgacccgctcccccggctg
pKH160           7137 gttcaccagatcggcatcgacgaagaagctgaccggcatctcgacccgctcccccggctg

pAW13            7200 cagggtctgctcggtaaagcagaagcattcaatcttgttgaagaaatagccggccgcatc
pKH160           7197 cagggtctgctcggtaaagcagaagcattcaatcttgttgaagaaatagccggccgcatc

pAW13            7260 gggcgccacgttatagctcgccgtgccggtcaccggctcgtcggtgttgttgatcgcctc
pKH160           7257 gggcgccacgttatagctcgccgtgccggtcaccggctcgtcggtgttgttgatcgcctc

pAW13            7320 gtagaaggcgatggcgttctcgccgatcttcagctccatctcgcgctgcatcgggcggaa
pKH160           7317 gtagaaggcgatggcgttctcgccgatcttcagctccatctcgcgctgcatcgggcggaa

pAW13            7380 ggtccagccgagattgctgtccgcattggcgtcgaagcgcacccggatcttttcgtccag
pKH160           7377 ggtccagccgagattgctgtccgcattggcgtcgaagcgcacccggatcttttcgtccag

pAW13            7440 caccgtgtccgacgcggcctcggccacgttggtggtgccggcaaagccggttaccttgca
pKH160           7437 caccgtgtccgacgcggcctcggccacgttggtggtgccggcaaagccggttaccttgca

pAW13            7500 gaaccaggaatagaacggcaccgccgcccagctgagcgcacccatcagcaccacgacgcc
pKH160           7497 gaaccaggaatagaacggcaccgccgcccagctgagcgcacccatcagcaccacgacgcc

pAW13            7560 ggccagcatggccacggtgcgggtgttcgaccttggcttcccgccgctcattggttcgcc
pKH160           7557 ggccagcatggccacggtgcgggtgttcgaccttggcttcccgccgctcattggttcgcc

pAW13            7620 tcctgcgtgacgccctgcgtggcggcaggggtgcccggcgcggccacgggcgcggcattc
pKH160           7617 tcctgcgtgacgccctgcgtggcggcaggggtgcccggcgcggccacgggcgcggcattc

pAW13            7680 ggcgccggcggatcgggatcgggcggcagcatcgaggcgcgcggccggtggtcatagccc
pKH160           7677 ggcgccggcggatcgggatcgggcggcagcatcgaggcgcgcggccggtggtcatagccc

pAW13            7740 tgcatcatgtcgccctgggtgatcttgaccaccgagagcccgaagaccagcgcgacgaag
pKH160           7737 tgcatcatgtcgccctgggtgatcttgaccaccgagagcccgaagaccagcgcgacgaag

pAW13            7800 gccagcagcacgaccagcaggccgatgttccggctgcggcggcgcttgtgcaactcatgc
pKH160           7797 gccagcagcacgaccagcaggccgatgttccggctgcggcggcgcttgtgcaactcatgc
                                                                      ctaB
pAW13            7860 tcgacctggggcagcatcaccatcctccgacccagtgctggaccagcagcgccaggaaat
pKH160           7857 tcgacctggggcagcatcaccatcctccgacccagtgctggaccagcagcgccaggaaat
```

```
pAW13          7920 gcaggaaggtgtaatagagcgacaggcggaaataacgcttctcgacccggtagccatcgg
pKH160         7917 gcaggaaggtgtaatagagcgacaggcggaaataacgcttctcgacccggtagccatcgg

pAW13          7980 cctgggcctggtcctcgctgcggcgcaggatctgccagccgccggcgatgaacagcgcat
pKH160         7977 cctgggcctggtcctcgctgcggcgcaggatctgccagccgccggcgatgaacagcgcat

pAW13          8040 tcagcaccaccgagaccgccaggtagagcggcccgccgaccgaggtgaagcccagccaga
pKH160         8037 tcagcaccaccgagaccgccaggtagagcggcccgccgaccgaggtgaagcccagccaga

pAW13          8100 gcgcaaagggcgcgagcaccagcgtataggcgaagatatggcaacgcgtgaccttgcggc
pKH160         8097 gcgcaaagggcgcgagcaccagcgtataggcgaagatatggcaacgcgtgaccttgcggc

pAW13          8160 catgggtcacggtcagcatcggcacgccggccttggaatagtcgtccttcatgaacagcg
pKH160         8157 catgggtcacggtcagcatcggcacgccggccttggaatagtcgtccttcatgaacagcg

pAW13          8220 ccaaggcccagaaatgaggcggcgtccagaagaagatcagcgcgaacatcaacagcgact
pKH160         8217 ccaaggcccagaaatgaggcggcgtccagaagaagatcagcgcgaacatcaacagcgact

pAW13          8280 cgatgccgatgccgccggtggcgcag-gcccagccgatcatgggcgggaaggccccggcc
pKH160         8277 cgatgccgatgccgccggtgg-gcagcgcccagccgatcatgggcgggaaggccccggcc

pAW13          8339 gcgccgccgatgacgatgttctgcggcgtcgagcgtttcagccagatcgtatagaccacg
pKH160         8336 gcgccgccgatgacgatgttctgcggcgtcgagcgtttcagccagatcgtatagaccacg

pAW13          8399 gcatagaagaagatggtgaaggccaggaacccggccgcgaaccagttcgccgccagc-cc
pKH160         8396 gcatagaagaagatggtgaaggccaggaacccggccgcgaaccagtt-gccgccagcgcc

pAW13          8458 cagcatcatcaccgacaggcccgacagcgcgatgccgaccgccagcgcctcttgcgaggt
pKH160         8455 cagcatcatcaccgacaggcccgacagcgcgatgccgaccgccagcggctcttgcgaggt

pAW13          8518 gacgcggcccgagggcacgggacggcccgcggtgcgacgcatcaccgcatcgatgtcggc
pKH160         8515 gacgcggcccgagggcacgggacggcccgcggtgcgacgcatcaccgcatcgatgtcggc

pAW13          8578 gtcataccacatgttgagggcgcccgaggccccgcccccagcgcgatgaacagcaccgc
pKH160         8575 gtcataccacatgttgagggcgcccgaggccccgcccccagcgcgatgaacagcaccgc

pAW13          8638 gcagaaggcgacgaaggggttgaccggctgcggcgcgatccacagccccacgaaggccgt
pKH160         8635 gcagaaggcgacgaaggggttgaccggctgcggcgcgatccacagccccacgaaggccgt

pAW13          8698 gaacaccaccagcgacatgacgcgcggcttcagaagcgcgacatagtccccgaagcccgc
pKH160         8695 gaacaccaccagcgacatgacgcgcggcttcagaagcgcgacatagtccccgaagcccgc
                                                                       ctaC
                                                          ←----------------
pAW13          8758 ctcggcaggcccctcatatgcgttgatatcggccacggcccggccttactctgccgaggc
pKH160         8755 ctcggcaggcccctcatatgcgttgatatcggcca--ggccggccttactctgccgaggc

pAW13          8818 cagcttgacggggctggcgggcaggtagtccgaggcgtcggcggcgaattcctctttcgc
pKH160         8813 cagcttgacggggctggcgggcaggtagtccgaggcgtcggcggcgaattcctctttcgc

pAW13          8878 gccggcaagccaggcttcgtatttctcctggctgaccgccttgacgacgatgggcatata
pKH160         8873 gccggcaagccaggcttcgtatttctcctggctgaccgccttgacgacgatgggcatata

pAW13          8938 ggcgtggttgatgccgcaaagctcggagcactggccgaaatagacgccctcctggtcgac
pKH160         8933 ggcgtggttgatgccgcaaagctcggagcactggccgaaatagacgccctcctggtcgac

pAW13          8998 cgagaaccacagctgcgcgatgcggcccggcacggcgtcctgcttgacggcgaaggcggg
pKH160         8993 cgagaaccacagctgcgcgatgcggcccggcacggcgtcctgcttgacggcgaaggcggg

pAW13          9058 gatggtccaggcgtggatcacgtcggtcgcggtcacctggaccaggaccttcttgccgac
pKH160         9053 gatggtccaggcgtggatcacgtcggtcgcggtcacctggaccaggaccttcttgccgac

pAW13          9118 cggaaccacgaccgggttgtcggtcgccagcagatactcgtcctcggaatagccggcatc
pKH160         9113 cggaaccacgaccgggttgtcggtcgccagcagatactcgtcctcggaatagccggcatc

pAW13          9178 ggccagggcctctttctccagcatcagcgcgtcgaaggcgacgccgtcattggggtattc
pKH160         9173 ggccagggcctctttctccagcatcagcgcgtcgaaggcgacgccgtcattggggtattc
```

```
pAW13           9238 ataggaccagtaccactggtggccgatggccttgatcaccaggtcgggatcgttcggcat
pKH160          9233 ataggaccagtaccactggtggccgatggccttgatcaccaggtcgggatcgttcggcat

pAW13           9298 ctcctggctgcggaacaggatcggcagcgagaaggccccgatcgccaccaggatcagcac
pKH160          9293 ctcctggctgcggaacaggatcggcagcgagaaggccccgatcgccaccaggatcagcac

pAW13           9358 cgggaccagggtccagatcacctcgatcggggtgttgtgggtaaagcgcgccggcaccgg
pKH160          9353 cgggaccagggtccagatcacctcgatcggggtgttgtgggtaaagcgcgccggcaccgg

pAW13           9418 attcgcccggcggttgaaacggacgatgcagatcagcagcagcaggcagacgaagatcgt
pKH160          9413 attcgcccggcggttgaaacggacgatgcagatcagcagcagcaggcagacgaagatcgt

pAW13           9478 cacggccgtgatgatgtagagcacgaaatgatccagccattgctggtcatgcgccagcgg
pKH160          9473 cacggccgtgatgatgtagagcacgaaatgatccagccattgctggtcatgcgccagcgg

pAW13           9538 gctcgacgccggttggaagttcatcccgccgttcaccggcttgccgatcaccggcagatc
pKH160          9533 gctcgacgccggttggaagttcatcccgccgttcaccggcttgccgatcaccggcagatc

pAW13           9598 gcccagcacgtcctgcgccaaggccggcacggccgtcatcgtcgcaactcccaggctcat
pKH160          9593 gcccagcacgtcctgcgccaaggccggcacggccgtcatcgtcgcaactcccaggctcat

pAW13           9658 caccgccgccaccccgcggcgcttggtcgcaattgccatcatcccatcccgttgcctatc
pKH160          9653 caccgccgccaccccgcggcgcttggtcgcaattgccatcatcccatcccgttgcctatc
```

Supplement 3: Vector map and full DNA sequence of pAW16



Supplement 2 Figure: Vector map of the pAW16 expression vector with relevant restriction sites and genes. Map exported from Clone Manager 7 (Scientific & Educational Software, United States of America).

```
1      ctcgggccgt ctcttgggct tgatcggcct tcttgcgcat ctcacgcgct cctgcggcgg
61     cctgtagggc aggctcatac ccctgccgaa ccgcttttgt cagccggtcg gccacggctt
121    ccggcgtctc aacgcgcttt gagattccca gcttttcggc caatccctgc ggtgcatagg
181    cgcgtggctc gaccgcttgc gggctgatgg tgacgtggcc cactggtggc cgctccaggg
241    cctcgtagaa cgcctgaatg cgcgtgtgac gtgccttgct gccctcgatg ccccgttgca
301    gccctagatc ggccacagcg gccgcaaacg tggtctggtc gcgggtcatc tgcgctttgt
361    tgccgatgaa ctccttggcc gacagcctgc cgtcctgcgt cagcggcacc acgaacgcgg
421    tcatgtgcgg gctggtttcg tcacggtgga tgctggccgt cacgatgcga tccgccccgt
481    acttgtccgc cagccacttg tgcgccttct cgaagaacgc cgcctgctgt tcttggctgg
541    ccgacttcca ccattccggg ctggccgtca tgacgtactc gaccgccaac acagcgtcct
601    tgcgccgctt ctctggcagc aactcgcgca gtcggcccat cgcttcatcg gtgctgctgg
661    ccgcccagtg ctcgttctct ggcgtcctgc tggcgtcagc gttgggcgtc tcgcgctcgc
721    ggtaggcgtg cttgagactg gccgccacgt tgcccatttt cgccagcttc ttgcatcgca
781    tgatcgcgta tgccgccatg cctgcccctc cctttggtg tccaaccggc tcgacggggg
```

```
841   cagcgcaagg cggtgcctcc ggcgggccac tcaatgcttg agtatactca ctagactttg
901   cttcgcaaag tcgtgaccgc ctacggcggc tgcggcgccc tacgggcttg ctctccgggc
961   ttcgccctgc gcggtcgctg cgctcccttg ccagcccgtg gatatgtgga cgatggccgc
1021  gagcggccac cggctggctc gcttcgctcg gcccgtggac aaccctgctg gacaagctga
1081  tggacaggct gcgcctgccc acgagcttga ccacagggat tgcccaccgg ctacccagcc
1141  ttcgaccaca tacccaccgg ctccaactgc gcggcctgcg gccttgcccc atcaattttt
1201  ttaattttct ctggggaaaa gcctccggcc tgcggcctgc gcgcttcgct tgccggttgg
1261  acaccaagtg gaaggcgggt caaggctcgc gcagcgaccg cgcagcggct tggccttgac
1321  gcgcctggaa cgacccaagc ctatgcgagt gggggcagtc gaaggcgaag cccgcccgcc
1381  tgcccccga gcctcacggc ggcgagtgcg ggggttccaa gggggcagcg ccaccttggg
1441  caaggccgaa ggccgcgcag tcgatcaaca agccccggag gggccacttt ttgccggagg
1501  gggagccgcg ccgaaggcgt gggggaaccc cgcaggggtg cccttctttg ggcaccaaag
1561  aactagatat agggcgaaat gcgaaagact taaaaatcaa caacttaaaa aaggggggta
1621  cgcaacagct cattgcggca ccccccgcaa tagctcattg cgtaggttaa agaaaatctg
1681  taattgactg ccacttttac gcaacgcata attgttgtcg cgctgccgaa aagttgcagc
1741  tgattgcgca tggtgccgca accgtgcggc accctaccgc atggagataa gcatggccac
1801  gcagtccaga gaaatcggca ttcaagccaa gaacaagccc ggtcactggg tgcaaacgga
1861  acgcaaagcg catgaggcgt gggccgggct tattgcgagg aaacccacgg cggcaatgct
1921  gctgcatcac ctcgtggcgc agatgggcca ccagaacgcc gtggtggtca gccagaagac
1981  actttccaag ctcatcggac gttctttgcg gacggtccaa tacgcagtca aggacttggt
2041  ggccgagcgc tggatctccg tcgtgaagct caacggcccc ggcaccgtgt cggcctacgt
2101  ggtcaatgac cgcgtggcgt ggggccagcc ccgcgaccag ttgcgcctgt cggtgttcag
2161  tgccgccgtg gtggttgatc acgacgacca ggacgaatcg ctgttggggc atggcgacct
2221  gcgccgcatc ccgaccctgt atccgggcga gcagcaacta ccgaccggcc ccggcgagga
2281  gccgcccagc cagcccggca ttccgggcat ggaaccagac ctgccagcct tgaccgaaac
2341  ggaggaatgg gaacggcgcg ggcagcagcg cctgccgatg cccgatgagc cgtgttttct
2401  ggacgatggc gagccgttgg agccgccgac acgggtcacg ctgccgcgcc ggtagcactt
2461  gggttgcgca gcaacccgta agtgcgctgt tccagactat cggctgtagc cgcctcgccg
2521  ccctatacct tgtctgcctc cccgcgttgc gtcgcggtgc atggagccgg gccacctcga
2581  cctgaatgga agccagcttt atgcttgtaa accgttttgt gaaaaaattt ttaaaataaa
2641  aaaggggacc tctagggtcc ccaattaatt agtaatataa tctattaaag gtcattcaaa
2701  aggtcatcca ccggatcaat tcccctgctc gcgcaggctg ggtgccaagc tctcgggtaa
2761  catcaaggcc cgatccttgg agcccttgcc ctcccgcacg atgatcgtgc cgtgatcgaa
2821  atccagatcc ttgacccgca gttgcaaacc ctcactgatc cgcatgcccg ttccatacag
2881  aagctgggcg aacaaacgat gctcgccttc cagaaaaccg aggatgcgaa ccacttcatc
2941  cggggtcagc accaccggca agcgccgcga cggccgaggt cttccgatct cctgaagcca
3001  gggcagatcc gtgcacagca ccttgccgta gaagaacagc aaggccgcca atgcctgacg
3061  atgcgtggag accgaaacct tgcgctcgtt cgccagccag gacagaaatg cctcgacttc
3121  gctgctgccc aaggttgccg ggtgacgcac accgtggaaa cggatgaagg cacgaaccca
3181  gtggacataa gcctgttcgg ttcgtaagct gtaatgcaag tagcgtatgc gctcacgcaa
3241  ctggtccaga accttgaccg aacgcagcgg tggtaacggc gcagtggcgg ttttcatggc
3301  ttgttatgac tgtttttttg gggtacagtc tatgcctcgg gcatccaagc agcaagcgcg
3361  ttacgccgtg ggtcgatgtt tgatgttatg gagcagcaac gatgttacgc agcagggcag
3421  tcgccctaaa acaaagttaa acatcatgag ggaagcggtg atcgccgaag tatcgactca
3481  actatcagag gtagttggcg tcatcgagcg ccatctcgaa ccgacgttgc tggccgtaca
3541  tttgtacggc tccgcagtgg atggcggcct gaagccacac agtgatattg atttgctggt
3601  tacggtgacc gtaaggcttg atgaaacaac gcggcgagct ttgatcaacg acctttgga
3661  aacttcggct tcccctggag agagcgagat tctccgcgct gtagaagtca ccattgttgt
3721  gcacgacgac atcattccgt ggcgttatcc agctaagcgc gaactgcaat ttggagaatg
3781  gcagcgcaat gacattcttg caggtatctt cgagccagcc acgatcgaca ttgatctggc
3841  tatcttgctg acaaaagcaa gagaacatag cgttgccttg gtaggtccag cggcggagga
3901  actctttgat ccggttcctg aacaggatct atttgaggcg ctaaatgaaa ccttaacgct
3961  atggaactcg ccgcccgact gggctggcga tgagcgaaat gtagtgctta cgttgtcccg
4021  catttggtac agcgcagtaa ccggcaaaat cgcgccgaag gatgtcgctg ccgactgggc
4081  aatggagcgc ctgccggccc agtatcagcc cgtcatactt gaagctagac aggcttatct
4141  tggacaagaa gaagatcgct tggcctcgcg cgcagatcag ttggaagaat ttgtccacta
4201  cgtgaaaggc gagatcacca aggtagtcgg caaataatgt ctaacaattc gttcaagccg
4261  acgccgcttc gcggcgcggc ttaactcaag cgttagatgc actaagcaca taattgctca
4321  cagccaaact atcaggtcaa gtctgctttt attattttta agcgtgcata ataagcccta
4381  cacaaattgg gagatatatc atgaaaggct ggctttttct tgttatcgca atagttggcg
4441  aagtaatcgc aacatccgca ttaaaatcta gcgagggctt tactaagctg atccggtgga
```

4501 tgacctttg aatgaccttt aatagattat attactaatt aattggggac cctagaggtc
4561 cccttttta ttttaaaaat tttttcacaa aacggtttac aagcataaag ctggcggcac
4621 ctcgctaacg gattcaccgt ttttatcagg ctctgggagg cagaataaat gatcatatcg
4681 tcaattatta cctccacggg gagagcctga gcaaactggc ctcaggcatt tgagaagcac
4741 acggtcacac tgcttccggt agtcaataaa ccggtaaacc agcaatagac ataagcggct
4801 atttaacgac cctgccctga accgacgacc gggtcgaatt tgctttcgaa tttctgccat
4861 tcatccgctt attatcactt attcaggcgt agcaccaggc gtttaagggc accaataact
4921 gccttaaaaa aattacgccc cgccctgcca ctcatcgcag tcggcctatt ggttaaaaaa
4981 tgagctgatt taacaaaaat ttaacgcgaa ttttaacaaa atattaacgc ttacaatttc
5041 cattcgccat tcaggctgcg caactgttgg gaagggcgat cggtgcgggc ctcttcgcta
5101 ttacgccagc tggcgaaagg gggatgtgct gcaaggcgat taagttgggt aacgccaggg
5161 ttttcccagt cacgacgttg taaaacgacg gccagtgagc gcgcgtaata cgactcacta
5221 tagggcgaat tggagctcca ccgcggtggc ggccgctcta gctagaacta gtggatcccc
5281 cgggctgcag gaattcgata tcaagcttgc atgcctgcag gtcgagccga aagcatgcgg
5341 ccccggaatt ctccggggcc gcatcatgcc acgaagtcca ggaccggagc ccgcccccgt
5401 gtcgaaacgt aacgcatcgc gtccccctag aattgccgct gcctgatacg ccagatcagc
5461 gccactgtca tccctgccca aaccgcagcg atcatgaacc attgcgcggc atagctcagg
5521 tggttgttcg gaatgccctc gaccgcgacc gggatcggcc gcacccctg cgcatcgccc
5581 cgcacctcgg ccgccacgac cagcaccggc tcggtcccca gctgcgcggc catggcgggc
5641 acgtcgcggg cgaaccagac attctcggtc aggttgggct cgggcgtggc gctgcccttt
5701 tcgtcgggcc agtgcaggtt gcccgccacc tccagccgca ccggcgggcg cggggcacgc
5761 ttgtggtcct gatcgacgaa gccccgatcc agcaggatcc tgcgcccgtc atcggtgacg
5821 aaacccgaga cgacctgata gccgccgccc gcctcgcgcg tgcccgacag cacgtcgatc
5881 tcctgacccg tggtctggcc cgaaaccagc accggcatgt atttcatcga cggattcacc
5941 gccgccggca ggggcacggg ccggccctcg atcctttgct gaatctgggc gatcagccct
6001 tctttccagt ccagccgctg aagctgccac atgcccagcg aaatcaggat ggcgcagccc
6061 accacaccga cgatcagcgg gaacaggtaa cggcgcatgg gtgggcctcc aaagcaaacg
6121 gcgcgggaac atctgccccg cgccgtggaa atctctgtcg ggatcagcgg ccccagatat
6181 agatcacgac aaagaggaac agccagacca catcgacgaa atgccaatac caggcggcgg
6241 cctcgaagcc gacatgctgc ttctgggtca tctggccctt gagcaggcgg atcagcgaga
6301 cgaacaggaa gatcgtgccg atgatgacat gggcaccgtg gaacccggtg gccatgtaga
6361 aggcgccggc atagaccgta tcggccaggc caaacgccgc gtggctgtat tcataggctt
6421 gcaagccggt gaacgagacg cccaggatca ccgcgacgat cagcccgttg atcgtggtct
6481 tgcgatcgcc ttcatgcacg aaggcatggt gcgcccaggt cacggcaacg cccgaaagca
6541 gcaggatcag cgtgttgatc agcggcaggt gccaggggtc gaaggtaacg atgccctcgg
6601 gcggccagac gccgtccttg atcgggctgt ccggtcccat cggatacagg gcgttcttga
6661 tgaaggccca gaaccaggcg acgaagaaca tcacctcgga catgatgaac aggatgaagc
6721 cgtattgcag gccgatgcgc accaccgggg tatgttcgcc cgtctcgccc tcgttcacga
6781 catcggccca ccagccgaac atgacgtaga gcacgcccac cagaccgatc aggaacatcc
6841 acggcccctc gacaggaagc ccgaagaagg tgatgccctt catccaggcg accgcgccgg
6901 tcagcatcac gaaggcgccg atcgcgccga agaacggcca gatcgatggc ggcaggatct
6961 gatagtcgtg gttctttaca tgggccatgg ctgcatttcc ggctgttgct ggttcttgtt
7021 gctcagttta cggtcggttc cgttttcgcg tccagcgccg cctgtttggg cgcgggcgga
7081 tcggtccgat ggaaggtata ggacagggtg atgtcgcgga tgcgaccggc gtcgcggtcg
7141 ttcaccagat cggcatcgac gaagaagctg accggcatct cgacccgctc cccggctgc
7201 agggtctgct cggtaaagca gaagcattca atcttgttga agaaatagcc ggccgcatcg
7261 ggcgccacgt tatagctcgc cgtgccggtc accggctcgt cggtgttgtt gatcgcctcg
7321 tagaaggcga tggcgttctc gccgatcttc agctccatct cgcgctgcat cgggcggaag
7381 gtccagccga gattgctgtc cgcattggcg tcgaagcgca cccggatctt ttcgtccagc
7441 accgtgtccg acgcggcctc ggccacgttg gtggtgccgg caaagccggt taccttgcag
7501 aaccaggaat agaacggcac cgccgcccag ctgagcgcac ccatcagcac cacgacgccg
7561 gccagcatgg ccacggtgcg ggtgttcgac cttggcttcc cgccgctcat tggttcgcct
7621 cctgcgtgac gccctgcgtg gcggcagggg tgcccggcgc ggccacgggc gcggcattcg
7681 gcgccggcgg atcgggatcg ggcggcagca tcgaggcgcg cggccggtgg tcatagccct
7741 gcatcatgtc gccctgggtg atcttgacca ccgagagccc gaagaccagc gcgacgaagg
7801 ccagcagcac gaccagcagg ccgatgttcc ggctgcggcg gcgcttgtgc aactcatgct
7861 cgacctgggg cagcatcacc atcctccgac ccagtgctgg accagcagcg ccaggaaatg
7921 caggaaggtg taatagagcg acaggcggaa ataacgcttc tcgacccggt agccatcggc
7981 ctgggcctgg tcctcgctgc ggcgcaggat ctgccagccg ccggcgatga acagcgcatt
8041 cagcaccacc gagaccgcca ggtagagcgg cccgccgacc gaggtgaagc ccagccagag
8101 cgcaaagggc gcgagcacca gcgtataggc gaagatatgg caacgcgtga ccttgcggcc

```
8161  atgggtcacg gtcagcatcg gcacgccggc cttggaatag tcgtccttca tgaacagcgc
8221  caaggcccag aaatgaggcg gcgtccagaa gaagatcagc gcgaacatca acagcgactc
8281  gatgccgatg ccgccggtgg cgcaggccca gccgatcatg ggcgggaagg ccccggccgc
8341  gccgccgatg acgatgttct gcggcgtcga gcgtttcagc cagatcgtat agaccacggc
8401  atagaagaag atggtgaagg ccaggaaccc ggccgcgaac cagttcgccg ccagccccag
8461  catcatcacc gacaggcccg acagcgcgat gccgaccgcc agcgcctctt gcgaggtgac
8521  gcggcccgag ggcacgggac ggcccgcggt gcgacgcatc accgcatcga tgtcggcgtc
8581  ataccacatg ttgagggcgc ccgaggcccc gccccccagc gcgatgaaca gcaccgcgca
8641  gaaggcgacg aaggggttga ccggctgcgg cgcgatccac agccccacga aggccgtgaa
8701  caccaccagc gacatgacgc gcggcttcag aagcgcgaca tagtccccga agcccgcctc
8761  ggcaggcccc tcatatgcgt tgatatcggc cacggcccgg ccttactctg ccgaggccag
8821  cttgacgggg ctggcgggca ggtagtccga ggcgtcggcg gcgaattcct ctttcgcgcc
8881  ggcaagccag gcttcgtatt tctcctggct gaccgccttg acgacgatgg gcatataggc
8941  gtggttgatg ccgcaaagct cggagcactg gccgaaatag acgccctcct ggtcgaccga
9001  gaaccacagc tgcgcgatgc ggcccggcac ggcgtcctgc ttgacggcga aggcggggat
9061  ggtccaggcg tggatcacgt cggtcgcggt cacctggacc aggaccttct tgccgaccgg
9121  aaccacgacc gggttgtcgg tcgccagcag atactcgtcc tcggaatagc cggcatcggc
9181  cagggcctct ttctccagca tcagcgcgtc gaaggcgacg ccgtcattgg ggtattcata
9241  ggaccagtac cactggtggc cgatggcctt gatcaccagg tcgggatcgt tcggcatctc
9301  ctggctgcgg aacaggatcg gcagcgagaa ggccccgatc gccaccagga tcagcaccgg
9361  gaccagggtc cagatcacct cgatcggggt gttgtgggta aagcgcgccg gcaccggatt
9421  cgcccggcgg ttgaaacgga cgatcgagat cagcagcagc agcgagacga agatcgtcac
9481  ggccgtgatg atgtagagca cgaaatgatc cagccattgc tggtcatgcg ccagcgggct
9541  cgacgccggt tggaagttca tcccgccgtt caccggcttg ccgatcaccg gcagatcgcc
9601  cagcacgtcc tgcgccaagg ccggcacggc cgtcatcgtc gcaactccca ggctcatcac
9661  cgccgccacc ccgcggcgct tggtcgcaat tgccatcatc ccatcccgtt gcctatcatg
9721  ccgccggcag gcccctgtcc tgaggccgcc ggcattcggt cctgcaccag aaccatatct
9781  taccccggcg ggcaagccat acctatcgac ttcttgtcaa aaaagcccta ggcggggctt
9841  catgacaagg accggttcca tgacagacgc ttcgttttca cccttccgga cgcatctcga
9901  tcaggaccgc gcgctgcgca tcctgcgcga cgccctggcc ggcgcggatg acggcgaact
9961  tttcctggaa cgctcccgtt ccgaggcgct ggtcttcgac gacggccggc tgcgcaccgc
10021 aagctatgat gccgagcagg gcttcggcct gcgtgccgtg cgggacgact gatagggcat
10081 cgttggattc gagccgaaag catgcggccc cggaattctc cggggccgca tcatgccatg
10141 catagggcat cgttggatca atgcggtgat cagtgatggt gatggtgatg cgcgtgtgcg
10201 cgatcccagt cctcgcgctt gggcagggtc tcgaaggtat gctcgggcgg cggcgagggc
10261 agggtccatt ccagcgtgtc ggcatgctcg ttccagtagt tcggcacgtt cacgcgcttg
10321 ccggcgaaga gcgtgtagaa cacgatgccg atgaagaaca ggaaggacgc gaaggagata
10381 taggcgccga tcgacgagat gttgttccaa taggcgaact cgaccggata gtcgatatag
10441 cgccgcggca tgccctggcg gcccaggaag tgctgcggga agaagatcag gttcgagccg
10501 atgaacatca tccagaaatg cagctggccc gcccattccg ggtattgccg gcccgacatc
10561 ttgccgatcc agtaatagac ccggcgaag atgccgaaca ccgcgcccag cgacatcacg
10621 tagtggaagt gggccacgac gtaataggtg tcgtgataga cccggtccag cggcgcctgg
10681 ctcagcacca cgccggtcac gccgccgacg gtgaacagga acaggaagcc gaaggcccag
10741 agcatcggcg tcttgaactc gatgctgccg ccccacatgg tcgcgatcca cgagaagacc
10801 ttgatgccgg tgggcaccgc gatggtcatg gtcgccagca tgaaataggc ctgctgggtc
10861 agcgacatgc cggccgtgta catgtggtgc gcccagacga cgaagcccag gatgccgatc
10921 gccgccatgg ccagcaccat cggcaggtag ccgaagatcg gcttcttggc gaaggtcgag
10981 atgacgtggc tgatgatgcc gaagcccggc aggatgatga tatagacctc gggatggccg
11041 aagaaccaca ggatgtgctg gtaaagcacc gggtcgccgc cgccggccgg atcgaagaac
11101 tgcgtgccga agttgcggtc catcagcagc atggtgatcg cgcccgccag aaccggcagc
11161 gacagcagga tcagccaggc ggtgatgaag accgaccagg caaacagcgg caccttgaac
11221 agcgtcatgc ccggtgcgcg catgttgagg aaggtggtga tgatgttgat cgcgcccagg
11281 atcgacgagg cacccgagac gtggacggca aagatcgcca ggtccatgga atagcccgcc
11341 tcggtggtcg agagcggcgg gtagagcacc cagccgacgc ccgaacccat ctggtcgttg
11401 ccgcccggcg ccagcagcga ggcgacgccc agggccacgc ccgagacata catccagtag
11461 gagaggttgt tcagccgcgg gaaggccatg tccggggcgc cgatatgcag cggcatgaaa
11521 tagttgccga aaccgccgaa cagtgccggg atcacgacga agaacatcat gagcacgccg
11581 tggtaggtga tcatgacgtt ccacaggtgt ccgttcgggg tgcattccgc cgaggcgtcg
11641 gcgatgagac gcgcgccttc caggcacatg tattgcacgc ccggatgctg cagttccatc
11701 cgcatataga cggtgaagga taccgagatc aggccgacga tgccggccgt gaacaggtaa
11761 aggataccga tatccttgtg gtttgttgac atgaaccagc gggtgaagaa cccgcgggtg
```

```
11821 tcatgatggt caccgtggcc gtgaacggct gcgtctgcca tgcgtggact ccctagacat
11881 gcggctggcc ggcgcaaagg ccgacggact cgcctgtttc tagaactaga ggatccctga
11941 ggccgccggc attcggtcct gcaccagaac catatcttac cccggcgggc aagccatacc
12001 tatcgacttc ttgtcaaaaa agccctaggc ggggcttcat gacaaggacc ggttccatga
12061 cagacgcttc gttttcaccc ttccggacgc atctcgatca ggaccgcgcg ctgcgcatcc
12121 tgcgcgacgc cctggccggc gcggatgacg gcgaactttt cctggaacgc tcccgttccg
12181 aggcgctggt cttcgacgac ggccggctgc gcaccgcaag ctatgatgcc gagcagggct
12241 tcggcctgcg tgccgtgcgg gacgaagcta gcttatcgat accgtcgacc tcgagggggg
12301 gcccggtacc cagcttttgt tccctttagt gagggttaat tgcgcgcttg gagctcctga
12361 aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt
12421 ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc
12481 ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacaggtat
12541 ttattcgaag acgaaagggc ctcgtgatac gcctattttt ataggttaat gtcatgataa
12601 taatggtttc ttagacgtca ggtggcactt ttcggggaaa tgtgcgcgcc cgcgttcctg
12661 ctggcgctgg gcctgtttct ggcgctggac ttcccgctgt tccgtcagca gcttttcgcc
12721 cacggccttg atgatcgcgg cggccttggc ctgcatatcc cgattcaacg gccccagggc
12781 gtccagaacg ggcttcaggc gctcccgaag gt
```

Supplement 4: Background cysteine-free C*c*O primary sequences

Precursory amino acid sequences from pAW16 with exchanged cysteines highlighted in orange:

C*c*O subunit 1 (ctaDII):

```
  1   madaavhghg dhhdtrgfft rwfmstnhkd igilylftag ivglisvsft vymrmelqhp
 61   gvqymclega rliadasaec tpnghlwnvm ityhgvlmmf fvvipalfgg fgnyfmplhi
121   gapdmafprl nnlsywmyvs gvalgvasll apggndqmgs gvgwvlyppl stteagysmd
181   laifavhvsg assilgaini ittflnmrap gmtlfkvplf awsvfitawl illslpvlag
241   aitmllmdrn fgtqffdpag ggdpvlyqhi lwffghpevy iiilpgfgii shvistfakk
301   pifgylpmvl amaaigilgf vvwahhmyta gmsltqqayf mlatmtiavp tgikvfswia
361   tmwggsiefk tpmlwafgfl flftvggvtg vvlsqapldr vyhdtyyvva hfhyvmslga
421   vfgifagvyy wigkmsgrqy pewagqlhfw mmfigsnlif fpqhflgrqg mprryidypv
481   efaywnniss igayisfasf lffigivfyt lfagkrvnvp nywnehadtl ewtlpspppe
541   htfetlpkre dwdrahahhh hhh
```

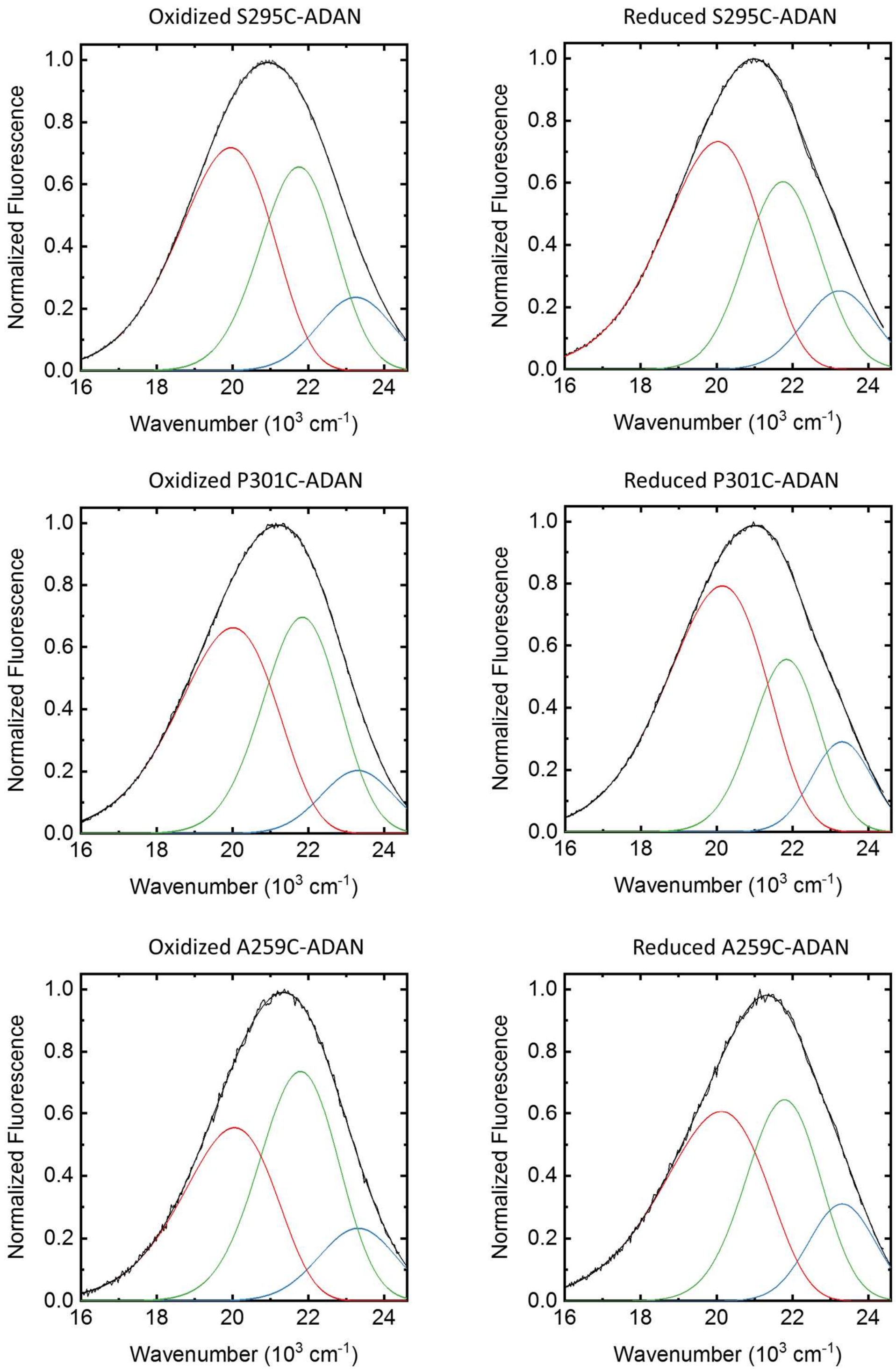
C*c*O subunit 2 precursor (ctaC):

```
  1   mmaiatkrrg vaavmslgva tmtavpalaq dvlgdlpvig kpvnggmnfq passplahdq
 61   qwldhfvlyi itavtifvsl lllisivrfn rranpvparf thntpieviw tlvpvlilva
121   igafslpilf rsqempndpd lvikaighqw ywsyeypndg vafdalmlek ealadagyse
181   deyllatdnp vvvpvgkkvl vqvtatdvih awtipafavk qdavpgriaq lwfsvdqegv
241   yfgqcselcg inhaympivv kavsqekyea wlagakeefa adasdylpas pvklasae
```

C*c*O subunit 3 (ctaE):

```
  1   mahvknhdyq ilppsiwpff gaigafvmlt gavawmkgit ffglpvegpw mfliglvgvl
 61   yvmfgwwadv vnegetgeht pvvriglqyg filfimsevm ffvawfwafi knalypmgpd
121   spikdgvwpp egivtfdpwh lplintlill lsgvavtwah hafvhegdrk ttinglivav
181   ilgvsftglq ayeyshaafg ladtvyagaf ymatgfhgah viigtiflfv slirllkgqm
241   tqkqhvgfea aawywhfvdv vwlflfvviy iwgr
```

Supplement 5: Spectral decomposition of C*c*O-ADAN



Supplement 5 Figure: Spectral decomposition of BADAN-labeled C*c*O samples fitted with equation (11) having three components.

Supplement 6: The DANAE program

File 1: danae_utils.py

```
# coding: utf-8
# DANAE utils (2015)

"""
Small module with basic utilities required for running DANAE.
"""

import numpy as np
import logging
from logging import handlers

LOG_FORMATTER = logging.Formatter('%(asctime)s %(levelname)s
%(funcName)s(%(lineno)d) %(message)s')
LOG_HANDLER = handlers.RotatingFileHandler("danae.log",
mode='a', maxBytes=5 * 1024 * 1024,
                                           backupCount=2,
encoding=None, delay=0)
LOG_HANDLER.setFormatter(LOG_FORMATTER)
LOG_HANDLER.setLevel(logging.INFO)
LOGGER = logging.getLogger(__name__)
LOGGER.setLevel(logging.INFO)
LOGGER.addHandler(LOG_HANDLER)


def import_picking_data(path_to_file):
    """
    Imports localization data from a file
    :param path_to_file: Path to the file to be imported
    :return: A numpy array containing the imported
localization data
    """
    data = np.loadtxt(path_to_file)
    i, j = 0, 0
    counts = np.empty((int(data[data.shape[0] - 1, 0]) +
1)).astype(int)
    while i < data[data.shape[0] - 1, 0] + 1:
        counts[i] = np.extract(data[:, 0] == i, data).shape[0]
        i += 1
    results = np.empty((int(data[data.shape[0] - 1, 0] + 1),
int(np.amax(counts)), data.shape[1] - 1))
    results.fill(np.nan)
    counter = 0
    while j < data[int(data.shape[0]) - 1, 0] + 1:
        results[j, :counts[j]] = data[counter:counter +
counts[j], 1:data.shape[1]]
```

```
        counter += counts[j]
        j += 1
    LOGGER.info("Successfully imported file %s.",
path_to_file)
    return results


def generate_histograms(distances, time_lags, bin_size,
max_distance):
    """
    Generates DANAE step length distributions
    :param distances: A matrix of distances with the
dimensions Time x max(N_particles) x max(N_particles)
    :param time_lags: Array of time lags to calculate
histograms for
    :param bin_size: Bin size for the generated histograms
    :param max_distance: Maximum distance for the histograms
    :return:
    """
    time_lag_index = 0
    max_len = np.ceil(max_distance).astype(int)
    bins = (np.linspace(0, max_len, int(max_len / bin_size +
1)) + bin_size / 2)[:int(max_len / bin_size)]
    histograms = np.empty([len(time_lags) + 1, bins.shape[0]])
    histograms.fill(np.nan)
    histograms[0] = bins
    scale_factor = np.zeros([len(time_lags)])

    for time_lag in time_lags:
        bin_counts = np.zeros(len(bins))
        temp_data = np.empty((distances.shape[1]) ** 2)

        for frames in range(distances.shape[0] - time_lag):
            temp_data.fill(np.nan)

            first =
np.sum(np.logical_not(np.isnan(distances[frames, :, 0])))
            second =
np.sum(np.logical_not(np.isnan(distances[frames + time_lag, :,
0])))
            both = np.append(first, second)
            scale_factor[time_lag] += np.amin(both)

            all_distances =
calculate_inter_particle_distances(distances[frames],
distances[frames + time_lag],

time_lag == 0)
            temp_data[:len(all_distances)] =
np.sqrt(all_distances)
```

```
            histogram_data = np.floor_divide(temp_data,
bin_size)
            histogram_data =
histogram_data[np.logical_not(np.isnan(histogram_data))].astyp
e(int)
            try:
                bin_counts += np.bincount(histogram_data,
minlength=len(bins))
            except Exception as N:
                LOGGER.error('There are distances greater than
your FOV!')
                raise N
        histograms[time_lag_index + 1, :] = bin_counts
        LOGGER.info("Histogram generated for time lag %s.",
time_lag_index)
        time_lag_index += 1
    return histograms, scale_factor


def normalize_histograms(in_data):
    """
    Calculates a normalization factor corresponding to the
amount of particles in the localization data and multiplies
    it with the input distributions
    :param in_data: DANAE histograms
    :return: Normalized DANAE histograms
    """
    data = np.copy(in_data)
    for norm in range(data[0].shape[0] - 2):
        norm_factor = (np.sum(data[0][2 + norm]) -
data[1][norm + 1]) / np.sum(data[0][1])
        data[0][2 + norm] = data[0][2 + norm] - data[0][1] *
norm_factor
    return data[0]


def calculate_inter_particle_distances(frame_one, frame_two,
is_zero_lag_distribution):
    """
    Calculates a distance matrix between two frames
    :param frame_one: Localization data of the first frame
    :param frame_two: Localization data of the second frame
    :param is_zero_lag_distribution: Whether both frames are
identical
    :return: An inter-particle distance matrix
    """
    positions = np.empty([frame_one.shape[1] * 2,
frame_one.shape[0], frame_one.shape[0]])
    for i in range(frame_one.shape[1]):
        positions[2 * i:2 * i + 2] = np.meshgrid(frame_two[:,
i], frame_one[:, i])[0:2]
```

```
    positions[1::2] *= -1
    distances = (positions[::2] + positions[1::2]) ** 2
    distances = np.sum(distances, axis=0)
    if is_zero_lag_distribution:
        for j in range(distances.shape[0]):
            distances[j, j] = np.nan
    return distances.ravel()
```

File 2: danae_gui.py

```
# coding: utf-8
# DANAE gui (2015)

"""
Simple graphical interface for running DANAE.
"""

import glob2
import os
import math
import numpy as np
from tkinter import *
from danae_utils import import_picking_data,
normalize_histograms, generate_histograms

__author__ = "A. Wolf"
__version__ = "1.0"

TIME_LAGS = "time_lags"
PATHS = "paths"
BIN_SIZE = "bin_size"
MAX_DIST = "max_dist"


def configure_start_button():
    if all(user_input.go_flags_.values()):
        start_button.config(state="normal")
    else:
        start_button.config(state="disabled")


def process_time_lags_entry(event):
    try:
        user_input.time_lags = lag_entry.get()
        user_input.update_go_flags(TIME_LAGS, True)
        if user_input.go_flags_.get(TIME_LAGS):
            lag_entry.configure(bg='lightgreen')
            message_display.configure(text="OK")
```

```
        else:
            lag_entry.configure(bg='pink')
            start_button.config(state="disabled")
    except:
        message_display.configure(text="ERROR: Positive 
integer required!")
        lag_entry.configure(bg='pink')
        start_button.config(state="disabled")
        user_input.update_go_flags(TIME_LAGS, False)
    configure_start_button()


def process_path_entry(event):
    try:
        user_input.paths = path_entry.get()
        user_input.update_go_flags(PATHS, 
len(user_input.paths) > 0)
        if user_input.go_flags_.get(PATHS):
            path_entry.configure(bg='lightgreen')
            message_display.configure(text="OK")
        else:
            path_entry.configure(bg='pink')
            start_button.config(state="disabled")
        message_display.configure(text="There are %i Files for 
DANAE." % (len(user_input.paths)))
    except:
        message_display.configure(text="ERROR: Must enter 
path!")
        path_entry.configure(bg='pink')
        start_button.config(state="disabled")
    user_input.update_go_flags(PATHS, len(user_input.paths) > 
0)
    configure_start_button()


def process_bin_size_entry(event):
    try:
        user_input.bin_size = bin_entry.get()
        bin_entry.configure(bg='lightgreen')
        message_display.configure(text="OK")
        user_input.update_go_flags(BIN_SIZE, True)
    except:
        message_display.configure(text="ERROR: Improper bin 
size!")
        bin_entry.configure(bg='pink')
        start_button.config(state="disabled")
        user_input.update_go_flags(BIN_SIZE, False)
    configure_start_button()


def process_fov_size_entry(event):
```

```
    try:
        user_input.max_dist = fov_entry.get()
        fov_entry.configure(bg='lightgreen')
        message_display.configure(text="OK")
        user_input.update_go_flags(MAX_DIST, True)
    except:
        message_display.configure(text="ERROR: Improper FOV
Side!")
        fov_entry.configure(bg='pink')
        start_button.config(state="disabled")
        user_input.update_go_flags(MAX_DIST, False)
    configure_start_button()


def run_danae_for_files():
    for path in user_input.paths:
        run_danae(path, user_input)
    message_display.configure(text="Done!")


def run_danae(path, entries):
    data = import_picking_data(path)
    raw_results = generate_histograms(data, entries.time_lags,
entries.bin_size, entries.max_dist)
    np.savetxt('%s_DANAE_native_data.txt' % (path[:-4]),
raw_results[0], fmt='%.10g', delimiter='  ', newline='\n')
    normalized_results = normalize_histograms(raw_results)
    np.savetxt('%s_DANAE_data.txt' % (path[:-4]),
normalized_results, fmt='%.10g', delimiter='  ', newline='\n')


class Input:

    def __init__(self):
        self._time_lags_ = np.array([])
        self._paths_ = []
        self._bin_size_ = .0
        self._max_dist_ = .0
        self.go_flags_ = {"time_lags": False, "paths": False,
"bin_size": False, "max_dist": False}

    @property
    def time_lags(self):
        return self._time_lags_

    @property
    def paths(self):
        return self._paths_

    @property
    def bin_size(self):
```

```
        return self._bin_size_

    @property
    def max_dist(self):
        return self._max_dist_

    @time_lags.setter
    def time_lags(self, entry):
        entry = int(entry)
        self._time_lags_ = np.linspace(0, entry, entry +
1).astype(int)

    @paths.setter
    def paths(self, entry):
        entry = str(entry)
        self._paths_ = sorted(glob2.glob(entry),
key=os.path.getmtime)

    @max_dist.setter
    def max_dist(self, entry):
        entry = float(entry)
        self._max_dist_ = math.ceil(np.sqrt(2) * entry)

    @bin_size.setter
    def bin_size(self, entry):
        entry = float(entry)
        self._bin_size_ = entry

    def update_go_flags(self, flag, value):
        if flag in self.go_flags_.keys():
            self.go_flags_.update({flag: value})
        else:
            raise KeyError("Flag unknown.")


user_input = Input()

w = Tk()
w.title("DANAE")

Label(w, text="Path to File/s").grid(row=0)
path_entry = Entry(w)
path_entry.bind("<Return>", process_path_entry)
path_entry.grid(row=0, column=1)

Label(w, text="Max Lag Index").grid(row=2)
lag_entry = Entry(w)
lag_entry.bind("<Return>", process_time_lags_entry)
lag_entry.grid(row=2, column=1)

Label(w, text="Binsize").grid(row=3)
```

```
bin_entry = Entry(w)
bin_entry.bind("<Return>", process_bin_size_entry)
bin_entry.grid(row=3, column=1)

Label(w, text="Largest FOV Side").grid(row=4)
fov_entry = Entry(w)
fov_entry.bind("<Return>", process_fov_size_entry)
fov_entry.grid(row=4, column=1)

message_display = Label(w)
message_display.grid(row=5, column=1)

start_button = Button(w, state="disabled", text='RUN DANAE',
bg='lightgray', fg='black', command=run_danae_for_files)
start_button.grid(row=5, column=0)

if __name__ == "__main__":
    w.mainloop()
```

File 3: environment.yml

```
name: danae_env
channels:
  - defaults
dependencies:
  - blas=1.0=mkl
  - ca-certificates=2020.7.22
  - certifi=2020.6.20
  - glob2=0.7
  - icc_rt=2019.0.0
  - intel-openmp=2020.2
  - mkl=2020.2
  - mkl-service=2.3.0
  - mkl_fft=1.1.0
  - mkl_random=1.1.1
  - numpy=1.18.1
  - numpy-base=1.18.1
  - openssl=1.1.1g
  - pip=20.2.2
  - python=3.8.5
  - setuptools=49.6.0
  - six=1.15.0
  - sqlite=3.33.0
  - vc=14.1
  - vs2015_runtime=14.16.27012
  - wheel=0.35.1
  - wincertstore=0.2
  - zlib=1.2.11
```

Acknowledgements

First and foremost, I would like to thank my supervisor Prof. Dr. Ulrike Alexiev for giving me the opportunity to conduct the research that cumulated in this thesis in her workgroup. I am especially grateful for the opportunity and encouragement I have received to apply a wide range of exciting techniques.

I would also like to thank Prof. Dr. Karsten Heyne for agreeing to by my second supervisor and for valuable discussions during the time of my thesis.

Special thanks go to all the people I had the opportunity to collaborate on projects with or who have otherwise aided this research: Prof. Dr. Ernst-Walter Knapp and Jovan Dragelj from the Chemistry Department for the great work on the C*c*O project, Prof. Dr. Hartmut Michel from the Max-Planck Institute for Biophysics for providing me with the opportunity to visit his lab in the beginning of my work on C*c*O and Dr. Hao Xie for his support and all the highly valuable discussions, as well as Prof. Dr. Helmut Grubmüller from the Max-Planck Institute for Biophysical Chemistry for creating an exciting challenge for the DANAE algorithm and the bottle of wine he gifted our group as a winning prize.

I would also like to thank my wonderful coworkers: Pierre Volz-Rakebrand with whom I had the pleasure to collaborate on the DANAE project and with whom I had countless helpful discussions, Juliane Wonneberg for being an irreplaceable assistant to the group, Johannes Stellmacher for the valuable assistance on the TCSPC setup, and Jens Balke for agreeing to share many teaching responsibilities with me. My time in the group was a great experience and I cannot credit these people enough for it.

Finally, I would like to try to express the gratitude I have for my family and friends, who have always supported and believed in me.

Selbstständigkeitserklärung

Hiermit bestätige ich, Alexander Wolf, dass ich die vorliegende Dissertation mit dem Titel: „Conformational, hydration, and protonation dynamics of cytochrome *c* oxidase and a tracking-free analysis method for diffusion in fluorescence microscopy“ selbstständig und ohne unerlaubte Hilfe angefertigt habe. Ich versichere, dass ich ausschließlich die angegebenen Quellen und Hilfen in Anspruch genommen habe.

Berlin, den 30.08.2019