Technische Universität München
Institut für Organische Chemie und Biochemie

Max-Planck-Institut für Biochemie
Abteilung Strukturforschung
Biologische NMR-Arbeitsgruppe

Structural Investigations on Green Fluorescent Protein Variants and the Adenylyl Cyclase-Associated Protein

Dorota Ksiazek

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. W. Hiller
Prüfer der Dissertation:
1. apl. Prof. Dr. L. Moroder
2. Univ.-Prof. Dr. Dr. A. Bacher

Die Dissertation wurde am 30.01.2003 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 20.03.2003 angenommen.

Publications

Parts of this thesis have been or will be published in due course:

Markus H. J. Seifert, **Dorota Ksiazek**, M. Kamran Azim, Pawel Smialowski, Nedilijko Budisa and Tad A. Holak
Slow Exchange in the Chromophore of a Green Fluorescent Protein Variant
J. Am. Chem. Soc. 2002, *124*, 7932-7942

Markus H. J. Seifert, Julia Georgescu, **Dorota Ksiazek**, Pawel Smialowski, Till Rehm, Boris Steipe and Tad A. Holak
Backbone Dynamics of Green Fluorescent Protein and the effect of Histidine 148 Substitution
In press (Biochemistry)

Dorota Ksiazek, Hans Brandstetter, Lars Israel, Gleb P. Bourenkov, Galina Katchalova, Hans D. Bartunik, Michael Schleicher and Tad A. Holak
The Crystal Structure of the *N*-terminal Domain of the Adenylyl Cyclase-Associated Protein (CAP) from *Dictyostelium discoideum*
submitted (Nature Structural Biology)

Other publications:

Lewinski K., Chruszcz M., **Ksiazek D.**, Laidler P.
Crystallization and preliminary crystallographic analysis of a new crystal form of arylsulfatase A isolated from human placenta
Acta Crystallogr D Biol Crystallogr. 2000, *56,* 650-652

Litynska A, Przybylo M, **Ksiazek D**. Laidler P.
Differnces of alpha3beta1 integrin glycans from different human bladder cell lines
Acta Biochim. Pol. 2000; *47*, 427-34.

Laidler P., Gil D., Pituch-Noworolska A., Ciolczyk D., **Ksiazek D.**, Przybylo M., Litynska A.

Expression of beta1-integrins and N-cadherin in bladder cancer and melanoma cell lines

Acta Biochim. Pol. 2000; *47,* 1159-1170

Contents

1. Introduction

The work of this thesis has been carried out from October 1999 to December 2002 at the Department of Structural Research of the Max Planck Institute for Biochemistry. The scope of this thesis is to give a structural and dynamic characterization of fluorescent proteins and to determine the structure of the N-terminal domain of the adenylyl cyclase-associated protein (CAP-N) from *Dictyostelium discoideum*. The ^{15}N and ^{1}H-^{15}N nuclear magnetic resonance (NMR) studies done on the green fluorescent protein (GFPuv) and its mutant His148Gly show a substantial conformational flexibility and a strong impact on fluorescence properties of GFPs and suggests the presence of two conformations in slow exchange on the NMR time scale in this mutant. The structure of the CAP-N determined in the present thesis is the first for an N-terminal domain of any CAP and can be useful for defining specific functions for these domains.

The family of fluorescent proteins is one of the most widely studied and exploited protein families in biochemistry and cell biology which represents basic tools for monitoring gene expression, protein localization, movement and interaction in living cells (Chalfie et al., 1994; Tsien, 1998; Garcia-Parajo et al., 2000). This kind of monitoring is minimally perturbing the cell under investigation. Fluorescent proteins provide also a system rich in photophysical and photochemical phenomena of which an understanding is crucial for the development of new and optimized variants of the GFP (Lossau et al., 1996). Its amazing ability to generate a highly visible, efficiently emitting internal fluorophore is intrinsically fascinating and tremendously valuable (Tsien, 1998; Ward, 1981). Up to now the family of fluorescent proteins comprises about 30 cloned and spectroscopically characterized proteins (Labas et al., 2002; Ormoe et al., 1996; Yang et al., 1996). High-resolution crystal structures of GFPs offer opportunities to understand and manipulate the relation between protein structure and spectroscopic function. But there is still a continuing effort to develop by mutagenesis and engineering new GFP variants with better properties and to open up new ways to monitor protein-protein interactions (Voityuket et al., 1998).

Cyclase associated proteins (CAPs) are multifunctional proteins with several structural domains, that are present in a wide range of organisms. Two domains are highly conserved; one of them helps to activate the catalytic activity of the adenylyl cyclase in the cyclase-bound state through interaction with Ras, which binds to the cyclase in a different region (Hubberstey & Mottilo, 2002). The second conserved domain of CAP can bind monomeric actin and thus CAP has also a cytoskeletal function. CAP is involved in the Ras/cAMP-

dependent signal transduction and most likely serves as an adaptor protein translocating the adenylyl cyclase complex to the actin cytoskeleton. The CAP of *Dictyostelium discoideum* is involved in the microfilament reorganization at anterior and posterior plasma membrane regions. But the full CAP function still presents a mystery. CAP interaction with actin maybe controlled through phospholipid binding in a similar fashion as profilin (Gottwald et al., 1996). Phospholipid interactions may regulate the interaction between the amino- and carboxyl-terminal domains. The specific residues within CAP that interact with actin must be defined, and the possibility that actin binding may involve interaction between the amino and carboxyl termini cannot be excluded (Wesp et al., 1997).

This intoduction is followed by *Chapter 2,* which provides a short introduction to nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography – the two most powerful methods of structural studies. The used materials and methods are described in *Chapter 3*. *Chapter 4* of this thesis describes the work carried out on GFPs – dynamic studies and the crystal structure determination of DsRed. *Chapter 5* deals with the structure determination of the N-terminal domain of CAP and specific functions of this domain and the whole protein.

2. Methods for Structural Studies

In this chapter the two most powerful techniques for structural studies are presented: X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. X-ray crystallography is the main method for elucidation of the three-dimensional macromolecular structures at the atomic level. NMR, which has the disadvantage of being more time consuming and restricted to smaller molecular weight proteins (up to 30 kDa), has, however, many advantages in comparison to X-ray crystallography, as it can be useful for dynamic studies and can provide many other useful information about the protein in solution (Sali, 1998).

2.1 X-ray Crystallography

2.1.1 General Background

The central role of protein crystallography in structural analysis is illustrated by the increasingly high number of structures determined by X-ray diffraction techniques deposited in the Brookhaven Protein Data Bank (PDB). Until December 2002 a total of 19464 protein structures have been deposited, 16448 of them have been determined with the help of X-ray diffraction techniques and 3021 by NMR (Berman et al., 2000). Although crystallography, when compared with NMR gives a more static description of the macromolecular structures, there are no limits in the size of the molecule to be analyzed. This makes X-ray crystallography the method of choice for studying large macromolecular complexes at the atomic level. Otherwise, the structures determined by these two techniques are not much different. Differences arise when exposed protein regions are hindered by contacts in the crystalline lattice. In recent years, however, the advances in radiation detection and computing power have made it possible to study enzyme catalysis and associated conformational changes in the crystalline state. The so-called time-resolved crystallography overcomes this major disadvantage of X-ray protein crystallography (Moffat, 2001).

The main problem in X-ray crystal structure analysis is to find not only the amplitudes of all the diffracted X-rays (usually known as reflections), but also their phases. Knowledge of both, amplitudes and phases, allows the reconstitution of the electron density of the crystal. The amplitudes can be deduced from the intensities of the diffracted X-rays but the phases cannot be directly measured. This is known as the “phase problem” (see below). To determine

proteins three-dimensional structure, one has to first obtain good diffracting crystals of the protein in what is mainly a trial-and-error process.

2.1.2 What is a Protein Crystal?

Crystals are regular, three-dimensional arrays of atoms, ions, molecules or molecular assemblies. Ideally, a crystal can be described as an infinite array in which the building blocks (the symmetric units) are arranged according to well-defined symmetries (forming one of 230 space groups) into unit cells that are repeated in three dimensions by translation. Proteins and nucleic acids do not crystallize in space groups with inversion symmetries because they are composed of enantiomers (L-amino acids and D-sugars, respectively), thus reducing the number of possible space groups to 65.

Why are protein crystals needed for X-ray three dimensional structure determination? In practice, the reflection pattern of a single molecule cannot be observed, but only that of many in an ordered crystalline array. The maximum achievable resolution of any microscopic technique is limited by the applied wavelength. The radiation needed to analyze atomic distances lies within the spectral range of X-rays that are used in crystal studies because their wavelength (1.542 Å for copper Kα radiation) is comparable to the planar separation of atoms in a crystal lattice.

2.1.3 Crystal Growth

Growth of high quality single crystals is the basis of X-ray structure determination. It is also sometimes the primary difficulty in the determination of a macromolecular structure. Proteins and nucleic acids are structurally dynamic systems, often micro-heterogenous, whose properties are influenced by environmental conditions such as pH, temperature, ionic strength and a number of other factors. Protein purity and homogeneity is essential to the growth of single protein crystals. Crystallization of macromolecules is a multiparametric process involving three main steps: nucleation, growth and cessation of growth. It is indispensable for crystallization to bring the protein to a supersaturated state (Fig. 2.1), which will force the macromolecules into the solid state - the crystal.

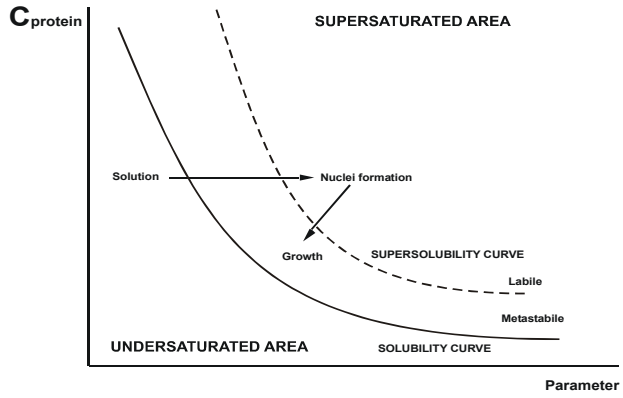

Fig. 2.1 *Solubility diagram. It represents the different steps occurring during crystal formation. The supersolubilyty curve separates the labile region, where nucleation occurs, from the metastable region, where crystals grow.*

The Debye-Hückel theory describes the solubility of a protein (C) as a function of the ionic strength (I) in the precipitant solution:

$$\ln C = \ln C_0 + \{AZ^2 I^{1/2} / (1+aB)\, I^{1/2}\} - K_S I \qquad \text{(eq 2.1)}$$

''Salting in'' *''Salting out''*

Herein, C_0 is the solubility of the protein in water, I is the ionic strength ($I = \frac{1}{2} \Sigma_i c_i z_i^2$ with concentration c and charge z of the ion), Z is the total charge of the protein, a is the sum of the radii of the protein and the salt ion, K_s is an empirical salting-out constant and the constants A and B depend on temperature and dielectricity. For high salt concentrations, the salting-out term will be dominant, and it can be derived that ions with a high charge density will have stronger influence on the solubility, as described in the Hofmeister series (Hofmeister, 1888).

K_s for anions: citrate>tartrate>sulfate>acetate>chloride>nitrate
K_s for cations: $Li^+ > K^+ > NH_4^+ > Ca^{2+} > Sr^{2+} > Ba^{2+} > Al^{3+}$

In addition to salts, commonly used precipitants are polyethylene glycols or organic solvents like ethanol, isopropanol or methylpentane diol. At low ionic strength (low ionic concentration), the solubility of protein is higher if the amount of electrolytes is increased – termed "salting in". At high ionic strength the ions start to compete with each other for water molecules, resulting in a decrease in solubility. This process is known as "salting out". The crystallographer can shift the equilibrium to supersaturation by increasing or reducing the ionic strength of the protein solution. The Hofmeister series indicates that at a high ion concentration small ions with a high charge are generally most effective. For proteins, much larger and with complicated surface charge distributions, this theory is not sufficient to

explain the phenomenon of crystal formation. For instance, the difference between the free energies (ΔG) of the solid and soluble states will indicate the favorable trend. In general the electrostatic interactions in crystals are much more favorable than the ‘‘interactions’’ in amorphous precipitates. As shown in Fig 2.1, crystal growth can be divided into two steps. First, a spontaneous nucleus formation occurs in the supersaturation area followed by formation of small aggregates. After the critical amount of aggregated molecules (10-200) is surpassed, the crystal growth becomes an energetically privileged process. Crystal growth always needs a lower degree of supersaturation than nucleus formation. Crystals should grow slowly to achieve the maximum possible internal order. It is clear that crystal morphology is not a direct synonym of crystal quality. ‘‘Good looking’’ crystals can have disordered crystal packing limiting their diffraction properties. Contrary to this, some ‘‘awful’’ crystals can give a positive surprise when they are measured. Different methods for growing crystals, as batch crystallization, dialysis, liquid-liquid diffusion and vapor diffusion, have been established and developed in the direction of using less amounts of material. The general role is that the sample must be pure and homogenous, it should be cleaned from small undesirable molecules and should contain a population with the same protein conformation. Microheterogenity of the sample can occur due to different reasons: variation in primary structure (genetic mutations), secondary structure (unfolding), tertiary structure (conformers), quaternary structure (oligomerization), partial oxidation of some groups, fragmentation by proteolysis or molecular dynamics of flexible parts. For data collection only best quality crystals should be used in order to obtain structural data. When different crystal forms are obtained, the best diffracting crystals with the highest symmetry should be chosen for further work.

2.1.4 Crystalline Lattice

The unit cell is defined as the minimal structural part that repeats in all three dimensions to build up the crystal. Within the unit cell, a crystal can contain further symmetry elements, dividing it into several asymmetric units, which form the most basic structural element, which is related to all other identical objects in the unit cell by symmetry elements. The geometry of the unit cell together with the possible symmetry operations defines the space group of the crystal. As discussed before, only 65 space groups in seven crystal systems (triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic) are therefore feasible for chiral molecules such as proteins. The combination of the four crystal lattices, primitive (P), body centred (I, from the German Innenzentrierte), face centred (F) or centred

in the (010) planes (C), with the seven crystal systems allows a total of 14 Bravais lattices. Identification of the correct space group is essential for correct indexing of diffraction patterns and is therefore the first step of understanding a crystal structure.

2.1.5 X-ray Diffraction by Crystals

The diffraction or scattering of X-rays is based on the interference phenomenon. In crystallography, X-rays are used to ‘‘visualize’’ atoms in a macromolecular structure, since the radiation has to be in the same range as the object of interest. In molecules interatomic distances are 0.15 nm = 1.5 Å. In the electromagnetic spectrum, this wavelength corresponds to the X-ray region. Bragg showed that diffraction from single crystals could be mathematically treated as a reflection from sets of equivalent parallel planes of atoms in the crystals (Fig. 2.2)

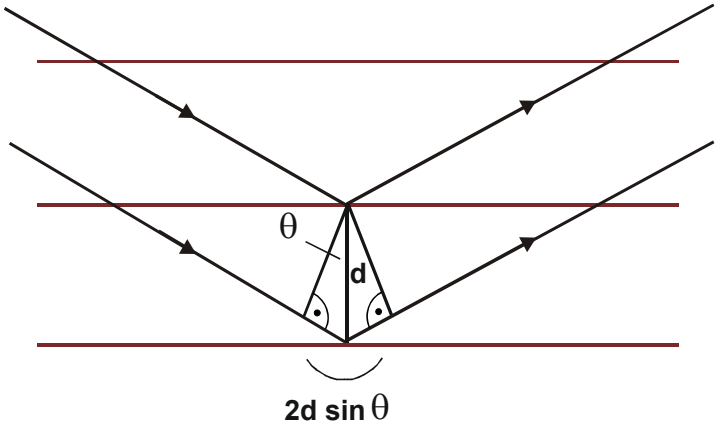

Fig. 2.2 *Scheme that explains Bragg's law. Two waves that are reflected by two adjacent lattice planes with distance d have a difference in path length that is equal to 2dsinθ, as it can easily be derived from the scheme. A prerequisite for constructive interference is, that this difference in path is an integer multiple of the wavelength used:* $2d \sin\theta = n\lambda$.

Constructive interference between the scattered X-ray from successive planes will take place if the path difference between these rays is equivalent to an integral number of wavelengths. A set of equivalent, parallel planes at an interplanar distance *d* produces a diffracted beam when X-rays of wavelength λ incide on the planes at an angle θ and are reflected at the same angle only when θ meets the condition (Bragg's law (Bragg and Bragg, 1913)):

$$2d\sin\theta = n\lambda \qquad \text{(eq. 2.2)}$$

The diffraction principle, which follows Bragg's law, can be interpreted by a geometrical construction proposed by Ewald (Ewald, 1921).

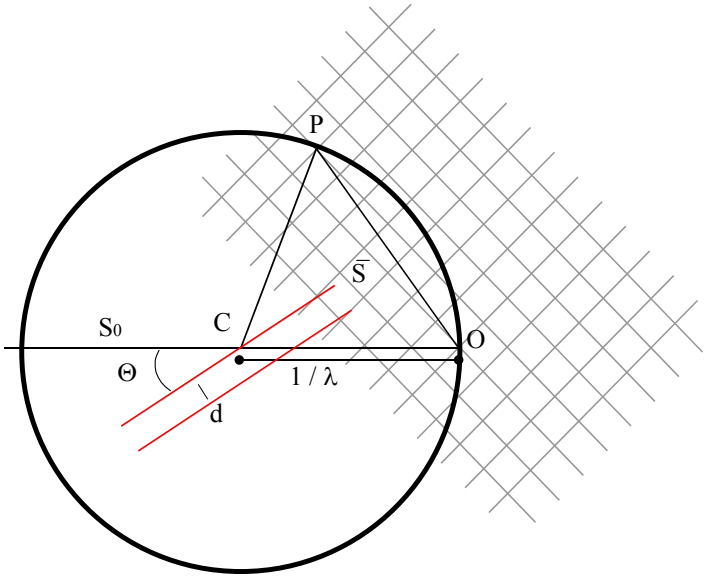

Fig. 2.3 *Ewald's construction is a convenient way to construct the direction of the scattered beam. In reciprocal space, the crystal (C) is placed in the center of a sphere (here, in two dimensions, a circle) with radius of 1/λ, called the Ewald sphere. The origin of the reciprocal lattice, i.e. reflection (0,0,0), is placed in (O). The reciprocal lattice will rotate as the crystal does and only those reciprocal lattice points that intersect with the Ewald sphere will be in diffraction condition and will be recorded on an image plate detector in real space.*

The Ewald sphere (Fig. 2.3) is a sphere of radius 1/ λ with the crystal in its center. The point where the incident beam s_0 enters the sphere and the origin O of the reciprocal lattice is on opposite sides of the center. Bragg's law is fulfilled for every reciprocal lattice point that lies on the Ewald sphere. A rotation of the crystal rotates the reciprocal lattice in the same way allowing different reciprocal lattice points to intersect with the sphere. The reciprocal lattice, is a theoretical lattice that is useful for constructing the directions of diffraction because it rotates exactly as the crystal does. In the reciprocal lattice the unit cell axes are the inverse of the unit cell axes in the crystal: a*=1/a, b*=1/b, c*=1/c. A set of planes (denoted by the Miller indices *hkl*) produces a reflection *hkl* in the direction CP (s) when the reciprocal point P_{hkl} contacts the sphere. The collected data will reveal the space group and unit cell dimensions, and – based on this - an intensity measurement I(h,k,l) for every reflection (h,k,l).

2.1.6 The Phase Problem

The aim of a crystallographic experiment is to calculate the distribution of electrons in the asymmetric unit of the crystal in order to place an atomic model of the crystallized

molecule therein. While the amplitude of the beam can be derived from the measured intensity (eq. 2.3), information about the phase angle is lost.

$$I(hkl) = |F(hkl)|^2 \qquad \text{(eq. 2.3)}$$

No phase information can be directly extracted from the intensities of the diffracted X-rays and without correct phase angles; the calculation of an interpretable electron density is impossible (eq. 2.4).

$$F(hkl) = \underbrace{|F(hkl)|}_{\text{Amplitude}} \underbrace{\exp i\alpha(hkl)}_{\text{Phase angle}} \qquad \text{(eq. 2.4)}$$

This is the so-called phase-problem. To overcome this problem five approaches are applicable:

- Molecular Replacement (MR)
- Single Isomorphous Replacement (SIR)
- Multiple Isomorphous Replacement (MIR)
- Multiple-Wavelength Anomalous Dispersion (MAD) and
- Direct Methods

The method of Molecular Replacement (MR) depends on the availabity of a sufficiently homologous model structure. That is, if the structure of a protein (search molecule) is known that is homologous to the crystallized protein (target molecule) with unknown structure, the former can be used as a model to calculate a starting set of phases that can be interactively refined (Hoppe, 1957; Huber, 1965). For that, the search molecule must be oriented and positioned in the unit cell of the target molecule in a way that maximum overlap of the models calculated diffraction pattern with the observed diffraction pattern is achieved. Thus, the related structure (phasing model) is used to obtain phase information after orienting the model in the unit cell of the new protein. Homologous proteins, which are derived by divergent evolution from an ancestral protein, often have a large proportion of their tertiary structure arranged in similar ways. Hemoglobins and myoglobins form a homologous series, as do chymotrypsin, trypsin and elastase. These proteins, despite the similarity in tertiary structures, crystallize in quite different space-groups. In these various situations of identical structures or parts of the structures in different crystallographic environments, similarities

between their diffraction patterns would be expected. Thus the relative orientation and position of these molecules can be derived from the known structure. The problem is to place the crystalline arrangement of the known protein onto the crystal of the protein for which the structure is not yet known.

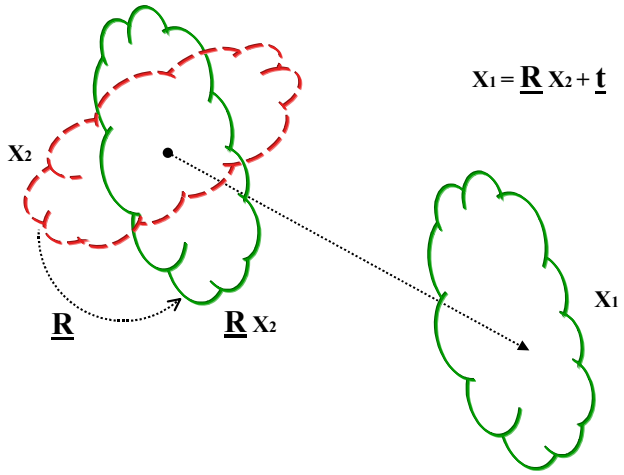

Fig. 2.4 *Scheme illustrates MR method. Allignment of the molecule in the target unit cell with the search model requires its proper orientation and precise positioning. This involves two steps: rotation and translation. (X_1 – target molecule, X_2 – search molecule with known structure; $\underline{R}$, $\underline{t}$ – rotation and translation vectors)*

In the rotation step, the spatial orientation of the known and unknown molecule with respect to each other is determined, while in the next step the translation needed to superimpose the now correctly oriented molecule onto the other is calculated. The basic principle of the MR method can be understood with help of the Patterson function of a protein crystal structure. The Patterson map is a vector map: vectors between atoms in the real structure show up as vectors from the origin to maxima in the Patterson map. If the pairs of atoms belong to the same molecule, then the corresponding vectors are relatively short and their end-points are found not too far from the origin in the Patterson map; they are called self-Patterson vectors. If there were no intermolecular vectors (cross-Patterson vectors), this inner region of the Patterson map would be equal for the same molecule in different crystal structures, apart from a rotation difference. For homologous molecules it is not exactly equal but very similar. Therefore, the self-Patterson vectors can supply us with the rotation relationship between the known and the unknown molecular structures. The solutions of the rotation and translation functions are not always found in a straightforward way. In some cases, it can be necessary to modify the model, for instance, by ignoring the side chains and deletions/insertions in the model, and/or by systematically varying the resolution range of the X-ray data used in the search. With the rapid increase in the number of successful protein structure determinations,

molecular replacement has become an extremely used technique for protein phase angle determination.

Without any previous knowledge of the structure Multiple Isomorphous Replacement (MIR) is still the most commonly used method. In this method the new X-ray scatterer of a heavy atom is introduced into the macromolecule under study without disrupting its structure or packing in the crysal. It is essential that native and derivative structures are isomorphous. The way to introduce such an atom is either soaking the crystal or co-crystallization along with the protein. Soaking with the help of the large solvent channels present in protein crystals is usually the method of choice. Addition of one or more heavy atoms to a macromolecule introduces differences in the diffraction pattern of the derivative crystal in relation to that of the native:

$$F_{PH} = F_P + F_H \qquad \text{(eq. 2.5)}$$

where F_P is the structure factor of the protein (with amplitude F_P and phase α_P). F_H and F_{PH} are the structure factors of the heavy atom and the corresponding derivative (protein + heavy atom, respectively). The position of the heavy atom can be determined from a difference Patterson map between the derivative and the native protein crystals (Fig. 2.5).

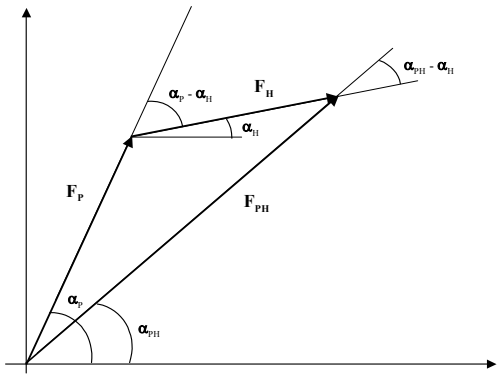

Fig. 2.5 *The phase problem. A vector diagram illustrating the native protein (F_P) and heavy atom (F_H) that contributes to the structure factor for the heavy atom derivative of the protein (F_{PH}). The structure factor amplitudes and phases defined in the text are indicated.*

After adequate refinement of this heavy atom position(s) the protein phases can be calculated:

$$\alpha_P = \alpha_H + \cos^{-1}((F^2_{PH} - F^2_P - F^2_H)/(2F_P F_{H)}) = \alpha_H \pm \alpha' \qquad \text{(eq. 2.6)}$$

As shown in equation 2.6, there are always two possible values for α_P, which cannot be distinguished with a single isomorphous derivative. The phase ambiguity for acentric reflection can be illustrated geometrically in a Harker diagram (Harker, 1956) (Fig. 2.6).

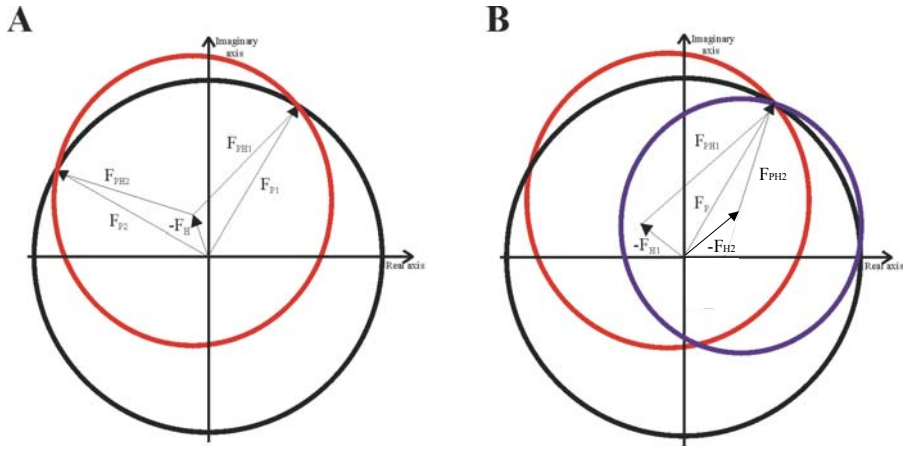

Fig. 2.6 *The Harker diagram for protein phase determination.* ***A.*** *Case where only one heavy atom is present. A circle of radius* $|F_P|$ *representing the native protein is drawn (black). A second circle (in red) with radius* $|F_{PH}|$, *representing the heavy atom derivative, is drawn. Its center is displaced by the vector* $-F_H$. *The intersections of the two circles correspond to two equally probable protein phase angles. Both triangle fulfil the condition* $F_{PH} = F_P + F_H$
B. *Case where two different heavy atoms are present. In this situation it is easier to elucidate which is the correct solution because the drawn circles ideally intersect at the same position.*

The ambiguity of the phase is theoretically solved when a second, unrelated heavy atom derivative is obtained. In practice the experimental data are affected by a number of errors arising from the crystal growth procedure, the soaking (or co-crystallisation) process, non-isomorphism, and errors derived from the collection of X-ray diffraction data. To obtain an interpretable electron density map usually more than two isomorphous heavy derivatives are necessary. Additionally, the phasing ambiguity can be solved by incorporating anomalous scattering data from different wavelength. By density modification methods (e.g. solvent flattening) or by averaging the calculated electron density when noncrystalographic symmetry is present, phases can be improved.

The Multiwavelength Anomalous Diffraction (MAD) method is insensitive to systematic errors due to non-isomorphism but the differences in intensities in this case will be less than for the Isomorphous Replacement (IR) method. MAD depends on precisely tuneable synchrotron radiation that has only been available for the last years, but is becoming more and more standard (Hendricson et al., 1988). For heavy atoms, the energy of an absorbed X-ray photon promotes the transition from a ground state orbital to an excited level, and the corresponding electronic acceleration leads to anomalous scattering. The characteristic energy for such a transition is known as an absorption edge. Edge energies depend on atomic orbital levels, which are element specific. K edges originate from 1s orbitals, L edges are based in 2p orbitals, and M edges derive from 3d orbitals. Owing to the multiplicity of energy states within 2p and 3d orbital levels, there are three L edges and five M edges. When the X-ray wavelength is near the heavy atom absorbing edge, a fraction of the radiation is absorbed by the heavy atom and reemitted with altered phase. Information about the phase of the scattered X-rays can be derived from the resonance effects or anomalous scattering. Anomalous scattering is an atomic property and thus enters the equations of X-ray diffraction in the expression for the atomic scattering factor (f), which is the sum of the "normal" atomic scattering factor (f^0) and a "complex anomalous" correction term having real (f') and imaginary (f") components:

$$f = f^0 + f' + f'' \qquad \text{(eq. 2.7)}$$

The change of f with wavelength is known as dispersion effect. The dispersion component of the anomalous scattering is f'. The magnitude of f" is proportional to the absorption coefficient. When a narrow beam of monochromatic radiation passes through a thickness (t) of the crystal, the emergent intensity (I) is related to the incident intensity (I_0) by:

$$I = I_0 e^{-\mu t} \qquad \text{(eq. 2.8)}$$

where μ is the linear absorption coefficient. Anomalous scattering has been used for many years as a source of phase information in macromolecular crystallography. Wavelength-tuneable synchrotron radiation allows the real component (f') to be used as well, providing the opportunity for direct phasing through combination of the orthogonal effects of f' and f". To exploit the f' dispersive component, measurements must be made at multiple wavelengths. Both the real and the imaginary parts of the anomalous scattering vary with wavelength,

especially in the region of the absorption edge of the anomalous scatterer. The differences in X-ray diffraction pattern at different wavelengths can be used to calculate estimates of the phases of the protein by a method equivalent to MIR. The optimal wavelengths for MAD experiments using scatterers correspond to a minimum f' (inflection point of the absorption edge) and to a maximum f'', and a point above the edge at which f' is different from its minimum value and f'' is still large. Data collection on frozen crystals reinforces the advantages of multiwavelength data collection because there is virtually no radiation damage. Particular care must be taken also in the scaling of the data sets and in the calculation of the absorption effects at different wavelengths.

Direct methods are the common way to determine phases in small molecule crystallography, but due to the high number of atoms per assymetric unit and the limited resolution that is obtained from most protein crystals, this approach has rarely been successful for large biomolecules. Most recently the number of protein structures solved by direct methods is increasing and the development is promising, but small molecule size, high resolution and the good data quality are still a prerequisite.

The introduction presented here was adapted from Branden & Tooze (1999). For more details about crystallization procedures please refer to McPherson (1999) and for the basis of X-ray crystallography to Drenth (1994) or Giacovazzo (1992).

2.2 NMR Spectroscopy

2.2.1 General Background

Nuclear magnetic resonance spectroscopy plays an important role in structural investigations complementing X-ray crystallography for small and medium size proteins (below 30 kDa) (Montelione et al., 2000; Prestegard et al., 2001). NMR can deliver strong results in several areas of structural biochemistry. It is the basis for a wide range of experiments and can be used to determine structure-function relationships (Shuker et al., 1996), to find binding partners with their specific binding sites (Stoll et al., 2001), to investigate dynamics of proteins, to distinguish multiple conformations (Mühlhahn et al., 1998), to compare apo and holo forms of proteins and map the binding sites of their cofactors (Wijesinha-Bettoni et al., 2001) or to determine pKa values of ionizable groups (Fielding, L., 2000). A series of spectra taken under different conditions may be used to monitor aggregation and formation of amyloid fibrils (Zurdo et al., 2001), to determine K_D values of binding partners (Shuker et al., 1996), or to track hydrogen exchange with real time NMR in

proteins dissolved in D_2O (Canet et al., 2002). The ability to detect weak ligands binding to target molecules has made NMR also increasingly important in drug discovery (Diercks et al., 2001; Pellecchia et al., 2002). The unique power of NMR lies in its capability to semi-quantitatively estimate unstructured regions of the polypeptide chain in the otherwise partially folded protein and to identify proteins that are heterogeneous because of aggregation or other conformational effects. For full NMR structure investigations samples of 200 – 400 µl with a protein concentration of 0.5-1.0 mM are required. This corresponds to about 10 – 15 mg/ml of the protein which is the concentration usually used for crystallographic screening. Spectra can also be recorded in up to 1 mM Tris buffer. An important advantage of NMR is that the sample is not destroyed by the measurement, so it is possible to continue with crystallization attempts after NMR characterization. NMR can serve as a powerful tool for screening proteins for sample conditions that favor crystallization or which are candidates for the NMR solution structure determination.

When protein molecules are placed in a strong magnetic field, the spin of their atoms aligns along the field. This equilibrium alignment can be changed to an excited state by applying radio frequency (RF) pulses to the sample. When the nuclei of the protein molecule revert to their equilibrium state, they emit RF radiation that can be measured. The exact frequency of the emitted radiation from each nucleus depends on the molecular environment of the nucleus and is different for each atom, unless they are chemically equivalent and have the same molecular environment. These different frequencies are called the chemical shift of the respective atom. The nature, duration, and combination of applied RF pulses can be varied enormously and different molecular properties of the sample can be probed by selecting the appropriate combination of pulses.

2.2.2 Application of NMR in structural proteomics.

A simple one-dimensional proton experiment, the most basic spectrum in NMR spectroscopy that can be acquired in a short time (usually not longer than a few minutes) for samples as dilute as 0.01 mM contains already a great amount of information. For example Figure 2.7 shows the discrimination of unfolded and folded proteins.

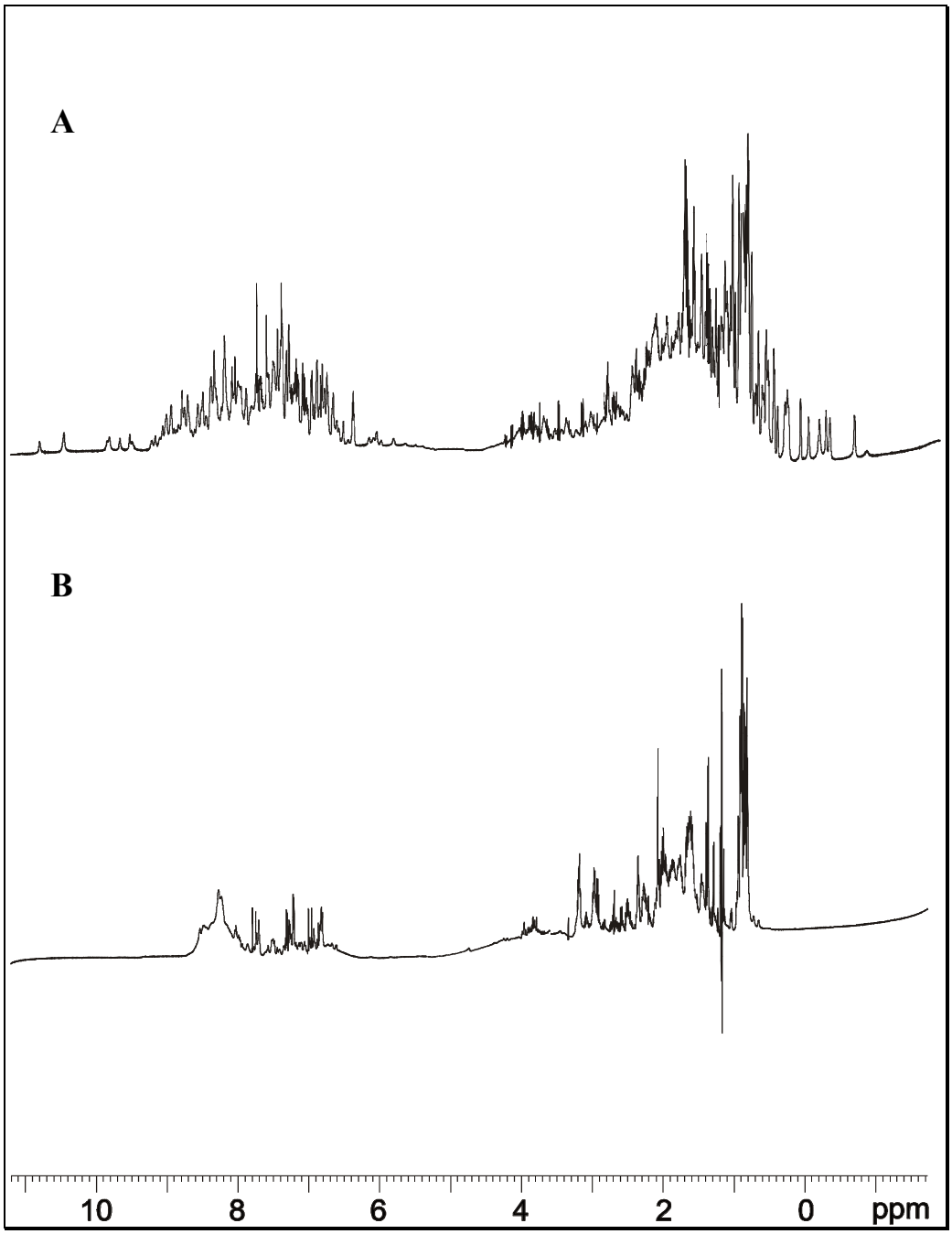

Fig. 2.7 *Characterization of protein structures by one-dimensional NMR spectroscopy.* ***A.*** *A typical one-dimensional proton NMR spectrum of a folded protein with signal dispersion downfield (left) of 8.5 ppm and upfield (to the right) of 1 ppm. The spectrum shows the N-terminal 176 residue domain of the adenylyl cyclase assiociated protein (CAP) at pH 7.3.* ***B.*** *An unfolded protein sample. Strong signals appear around 8.3 ppm, the region characteristic for amine groups in random coil conformation. No signal dispersion is visible below approximately 8.5 ppm. Also to the right of the strong methyl peak at 0.8 ppm no further signals show up. The sample is an unfolded domain of the IGF binding protein 4 (IGFBP-4, residues 147-229).*

The appearance of intensities at shifts near 8.3 ppm is an indicator for a disordered protein (Wüthrich, 1986), as this region is characteristic of backbone amides in random coil configuration. On the other hand signal dispersion beyond 8.5 ppm proves proteins to be folded. A typical intensity pattern of a folded protein is shown in the Fig. 2.7 A. Following the same argument, in the aliphatic region of the spectrum between 1.0 and and –1.0 ppm a large signal dispersion versus a steep flank of the dominant peaks at approximately 1 ppm separates a structured from an unfolded protein (Fig. 2.7; A and B, respectively). On the other hand distinguishing between a protein that is only partially foled and a mixture of folded and unfolded proteins is difficult with NMR without having additional information from e.g. gel filtration or other biochemical methods. While the signal dispersion of the resonances is generally connected to folding, aggregation can be detected by observing the linewidth of the signals. The NMR signal from larger molecules will decay much faster than that from smaller ones (Abragam, 1961). This in turn will produce broader lines for the resonances of larger molecules. Thus the linewidths of the signals in any NMR spectrum are correlated to the size of the molecule. Using the line width from known monomeric proteins of a given size as a reference, observation of the line width in a one-dimensional spectrum will also yield information on the molecular weight and aggregation of the molecule under investigation. Furthermore, attempts to prevent aggregation by e.g. dilution of the sample, addition of mild detergents as CHAPS or lowering pH value can thus be monitored by NMR to find optimal sample conditions (Kalus et al., 1998; Anglister et al., 1993; Edwards et al., 2000). While the extent of folding is crucial both for X-ray crystallography and NMR, aggregation is not. Actually some proteins that yield rather poor quality NMR spectra due to aggregation or low solubility might give excellent crystals (Kalus et al., 1997; Baumgartner et al., 1998). Thus sample conditions that are optimal for crystallography might not necessarily be optimal for NMR spectroscopy and vice versa. One-dimensional spectra may additionally provide information on α-helical or β-strand structures in a protein. The C^{α} protons in a helix display few resonances in the region between 5 and 6 ppm, while those in a β-sheet resonate in this region (Wishart et al., 1991). The use of one-dimensional spectra provides also information about purity of the sample. Any small compounds, be it peptides or other impurities will be readily detected by its sharp resonances.

2.2.3 Two-dimensional NMR

Two-dimensional spectra provide improvement of resolution and for this reason they are frequently used for screening and binding studies. In homonuclear two-dimensional spectra the diagonal corresponds to a normal one-dimensional NMR spectrum. The peaks off the diagonal result from different "magnetic" interactions between atoms. A Correlation Spectroscopy (COSY) experiment gives peaks between hydrogen atoms that are covalently connected through one or two other atoms. A Nuclear Overhauser Effect (NOE) spectrum, on the other hand, gives peaks between pairs of hydrogen atoms that are close together in space even if they are quite distant in the primary amino acid sequence.

One of the simplest and most powerful among the heteronuclear two-dimensional experiments is the Heteronuclear Single-Quantum Coherence (HSQC) experiment. This spectrum is the first step in any structure elucidation as it maps the backbone amide groups of a protein according to their proton and nitrogen frequencies. A whole set of three-dimensional spectra later used to assign the NMR signals to their respective amino acid residue is based on the HSQC experiment. For this kind of spectrum ^{15}N-labeled protein samples are required. The HSQC shows one peak for every proton bound directly to a nitrogen atom and thus exactly one signal per residue in the protein (apart from proline which is devoid of proton bound nitrogen and some additional side chain signals which can easily be identified). The positions of the peaks are indicative of structured or disordered proteins in the same way as described for one-dimensional spectrum. In case of an unfolded protein all signals cluster in a characteristic "blob" around a ^{1}H frequency of 8.3 ppm with little signal dispersion in both dimensions. In a spectrum of structured protein, the peaks show large signal dispersion (Fig. 2.8 A and B, respectively).

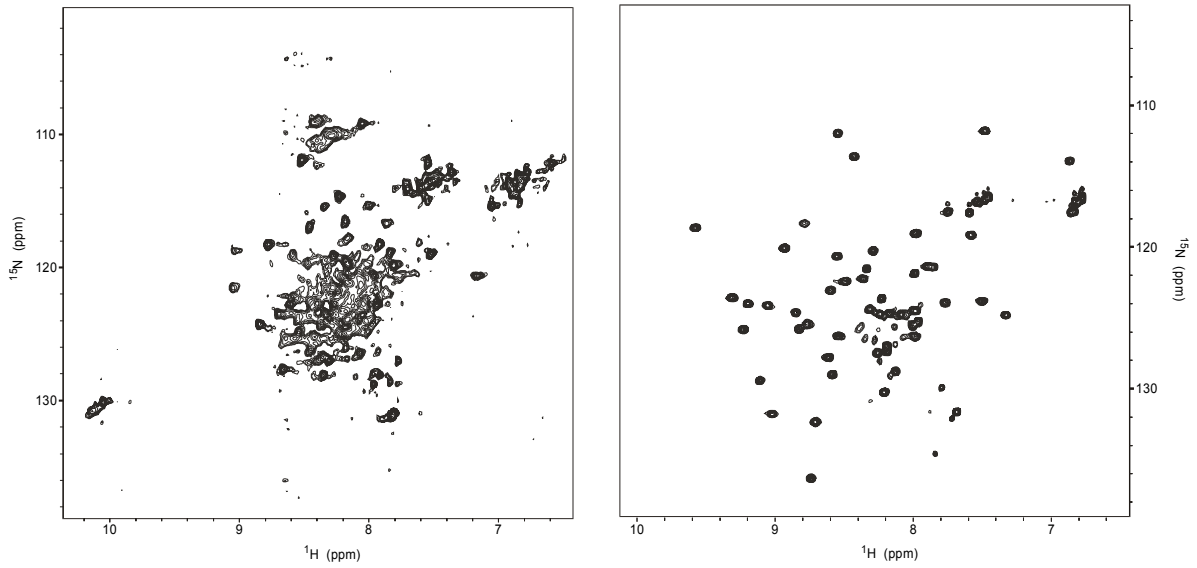

Fig. 2.8 *^{15}N-HSQC spectra of unfolded and folded proteins. The left panel shows a ^{15}N-HSQC spectrum of partially unstructured protein fragment of 80 amino acid residues. All signals cluster around a ^{1}H frequency of 8.3 ppm. Also the signal dispersion in the ^{15}N dimension is limited. The broad unresolved signals in the middle of the spectrum indicate either aggregation in the sample or conformational heterogeneity on an ms-µs timescale (both cases being unfavorable for NMR studies). The signal at 10 ppm is not diagnostic for a folded protein, but stems from the sidechain amide group of a tryptophan residue. The right panel shows the spectrum of a folded, 55 residue long construct of the IGFBP-5 protein. The peaks show large signal dispersion in both dimensions.*

Thus if the peaks are assigned their respective sequential position in the polypeptide chain, disordered regions may be identified. As the number of signals in the HSQC spectrum corresponds approximately to the number of residues in the protein under investigation, conformational heterogeneity can easily be detected by a surplus of peaks. To optimize sample conditions, pH titrations, titration with cofactors or other molecules or variation of temperature may be performed while repeatedly recording spectra. For example Zhang & Forman-Kay showed that low temperatures and neutral pH stabilize the folded state of an SH3 domain, while high temperatures and neutral pH tend to favor the unfolded state (Zhang & Forman-Kay, 1995). On the other hand low pH has also been reported to prevent aggregation (Anglister et al., 1993).

The main advantage of ^{15}N-HSQC spectra in screening for ligands is the ability to detect the binding of small, weakly bound ligands to ^{15}N-labeled target proteins. Another advantage of NMR is the ability to rapidly determine the different binding site locations of the

fragments, which is critical for interpreting the structure activity relationship and for guiding the synthesis of linked compounds. The method called Structure-Activity Relationships by NMR (SAR) employs this information (Shuker et al., 1996). In this method small organic molecules that bind to proximal subsites of a protein are identified, optimized, and linked together to produce high-affinity ligands. The first step of the method is screening a library of low molecular weight compounds to identify molecules that bind to the protein. Binding is determined by the observation of ^{15}N- or ^{1}H-amide chemical shift changes in the HSQC spectrum. From an analysis of the chemical shift changes, the approximate location of the ligand binding site can be defined. Screening further compounds may then identify a second ligand. When two "lead" fragments have been selected, their location and orientation in the ternary complex are determined experimentally either by NMR spectroscopy or by x-ray crystallography. Finally, on the basis of this structural information, compounds are synthesized in which the two fragments are linked together with the goal of producing a high affinity ligand (Shuker et al., 1996). The SAR by NMR method is limited by the solubility of compounds at milimolar concentrations and is relevant only to small biomolecules (MW < 30kDa) that can be obtained in large quantities. NMR is very beneficial and fast in screening for structure activity relationships and interactions. This has made NMR also increasingly important and a powerful tool not only in structural proteomics but also in drug discovery.

For a better understanding of NMR and all aspects of modern NMR techniques the following books are recommended:

- Cavanagh, Fairbrother, PalmerIII &Skelton (1996) Protein NMR Spectroscopy. Principles and Practice, Academic Press, New York
- Wüthrich (1994) NMR of Proteins and Nucleic Acids, John Wiley & Sons, New York

3. Materials and Laboratory Methods

3.1 Materials

All chemicals used in the work were supplied from Merck (Darmstadt, FRG) or Sigma (Deisenhofen, FRG) unless otherwise indicated.

Strains

Escherichia coli: DH5α, BL21, BL21(DE3), BL21(DE3)pLYS(E)

Plasmids for Protein Expression

GFP	**RFP**
pRsetB	pt7RFPav

Plasmids for protein overexpression in *E.coli* were a kindly gift from Dr. Boris Steipe

CAP

PT7-7 and pQE80L (QIAGEN, FRG)

Enzymes and Other Proteins

- hen egg white lysozyme
- RNaseA
- DNaseI
- factor Xa (NEB, FRG)
- thrombin (Sigma, FRG)

Molecular Weight Marker for SDS-PAGE Electrophoresis (NEB, FRG):

protein	**source**	**Apparent MW (Da)**
Maltose-binding protein-β-galactosidase	*E.coli*	175 000
Maltose-binding protein-paramyosin	*E.coli*	83 000

Glutamic dehydrogenase	Bovine liver	62 000
Aldolase	Rabbit muscle	47 500
Triosephosphate isomerase	Rabbit muscle	32 500
β-lactoglobulin A	Bovine milk	25 000
lysozyme	Chicken egg white	16 500
aprotinin	Bovine lung	6 500

Other Chemicals

Antibiotics:

- Ampicillin, sodium salt
- Chloramphenicol
- Kanamycin, monosulfate

Protease Inhibitors:

- Complete Protease Inhibitors Cocktail (Roche, FRG)
- Pefablock® SC (Roche, FRG)
- Phenylmethylsulfonyl fluorid (PMSF) (Roche, FRG)

Isotopically Enriched Chemicals:

- Deuterium oxide, D_2O 99%, 99.99% (Campro Scientific, Berlin, FRG)
- Deuterium oxide, D_2O 99%, 99.9% (Campro Scientific, Berlin, FRG)
- Deuterium oxide, D_2O 95%, 95% (Campro Scientific, Berlin, FRG)
- ^{13}C-L-Glucose, 99,9% (Campro Scientific, Berlin, FRG)
- ^{15}N-Ammonium chloride, NH_4Cl 99.9% (Campro Scientific, Berlin, FRG)
- ^{15}N-L-Alanine, 99,9% (Campro Scientific, Berlin, FRG)
- ^{15}N-L-Glycine, 99,9% (Campro Scientific, Berlin, FRG)
- ^{15}N-L-Isoleucine, 99,9% (Campro Scientific, Berlin, FRG)
- ^{15}N-L-Leucine, 99,9% (Campro Scientific, Berlin, FRG)
- ^{15}N-L-Lysine, 99,9% (Campro Scientific, Berlin, FRG)
- ^{15}N-L-Methionine, 99,9% (Campro Scientific, Berlin, FRG)
- ^{15}N-L-Phenyloalanine, 99,9% (Campro Scientific, Berlin, FRG)
- ^{15}N-L-Valine, 99,9% (Campro Scientific, Berlin, FRG)

- ^{15}N-L-Tyrosine, 99,9% (Campro Scientific, Berlin, FRG)

Standard Chemicals:

- Acetic acid
- Acrylamide
- L-Arginine
- Ammonium chloride, NH_4Cl
- Ammonium persulfate, APS
- Biotin
- Boric acid, H_3BO_3
- Calcium chloride, $CaCl_2$
- Citric acid
- Cobalt (II) chloride, $CoCl_2$
- Coomassie Brillant Blue R-250
- Copper (II) chloride, $CuCl_2$
- Dimethylsulfoxide, DMSO
- Dipotassium hydrogenphosphate, K_2HPO_4
- Disodium hydrogenphosphate, Na_2HPO_4
- Dithiothreitol, DTT
- Ethanol
- Ethylendiamintetraacetic acid, disodium salt, EDTA
- Formaldehyde
- Ferrous citrate
- D-Glucose
- Glutardialdehyde
- L-Glutathione, oxidized, GSSG
- L-Glutathione, reduced, GSH
- Glycerine
- Glycine
- Guanidine hydrochloride
- Hydrochloric acid, HCl
- Imidazole
- Isopropanol

- Isopropyl-β-D-thiogalactopyranoside, IPTG
- Magnesium chloride, $MgCl_2$
- Magnesium sulfate, $MgSO_4$
- Manganese (II) chloride, $MnCl_2$
- β-Mercaptoethanol, β-ME
- Methanol
- N,N'-Methylenbisacrylamide
- Nonidet P-40, NP-40
- Potassium chloride, KCl
- Potassium dihydrogenphosphate, KH_2PO_4
- Silver nitrate, $AgNO_3$
- Sodium acetate
- Sodium azide, NaN_3
- Sodium carbonate, Na_2CO_3
- Sodium chloride, NaCl
- Sodium dihydrogenphosphate, NaH_2PO_4
- Sodium dodecylsulphate, SDS
- Sodium hydrogencarbonate, $NaHCO_3$
- Sodium hydroxide, NaOH
- Sodium molybdate, Na_2MoO_4
- Sodium thiosulfate
- N,N,N',N'-Tetramethylenethylendiamine, TEMED
- Thiamin
- Tricine
- Trifluoroethanol, TFE
- Tris-(hydroxymethyl)-aminomethane, TRIS
- Triton X-100
- Tryptone
- Yeast Extract
- Zinc acetate, $Zn(Ac)_2$

Buffers and Media

All buffers, stock solutions and media, if not mentioned here, were prepared exactly as described in Sambrook & Russell (2001).

LB Medium:

Tryptone	10 g/l
Yeast Extract	5 g/l
NaCl	5 g/l

For the preparation of agar plates the medium was supplemented with 15 g agar. Antibiotics were added after the medium has been cooled to 50°C.

Minimal Medium (MM) for Uniform Enrichment with ^{15}N and/or ^{13}C

Stock solutions:

1. thiamin, 1%
2. antibiotics
3. $MgSO_4$, 1M
4. Zn-EDTA solution:

EDTA	5 mg/ml
$Zn(Ac)_2$	8.4 mg/ml

   Dissolved separately in small water volumes, then mixed together.
5. trace elements solution:

H_3BO_3	2.5 g/l
$CoCl_2*H_2O$	2.0 g/l
$CuCl_2*H_2O$	1.13 g/l
$MnCl_2*2H_2O$	9.8 g/l
$Na_2MoO_4*2H_2O$	2.0 g/l

   If difficult to dissolve, pH was lowered with citric acid or HCl.
6. glucose, 5 g/25 ml, separately autoclaved (for uniform enrichment with ^{13}C 2 g/25 ml)

For 1 liter medium:

1. mixture was prepared:

NaCl	0.5 g

trace elements solution	1.3 ml
citric acid monohydrate	1 g
ferrous citrate	36 mg (dissolved in 120µl conc. HCl, heated)
KH_2PO_4	4.02 g
K_2HPO_4 $3H_2O$	7.82 g
Zn-EDTA solution	1 ml
NH_4Cl or $^{15}NH_4Cl$	1 g

2. pH was adjusted to 7.0 with NaOH
3. the mixture was autoclaved
4. 25 ml separately autoclaved glucose was added
5. other compounds were added (previously sterile filtered):

thiamin	560 µl
antibiotics	(half of the usual amount for the LB-medium)
$MgSO_4$, 1M	2 ml

Medium for Selectively Enrichment Amino Acids

1. a mixture was prepared:

Ala, Glu, Gln, Arg, Gly	400 mg
Asp, Met	255 mg
Cytosine, Guanosine, Uracil	125 mg
Asn, Ile, Val, Leu, His, Lys, Pro, Thr, Tyr	100 mg
Phe, Thymidine, Thymine	50 mg
CaCl2	10 mg
NaAc	2 g
KH2PO4	10 g
Citric acid	1 g
Trace elements solution	1 ml
Ferrous citrate (preparation see protocol for MM)	36 mg
Zn-EDTA	1 ml
Ser	1.6 g
NH_4C	1 g

2. the pH was adjusted to 7.4 with 1M NaOH,
3. the mixture was autoclaved,

4. after autoclaving the following sterile filtered solutions were added:

Cys, Trp, nicotinic acid	50 mg
Biotin	0.1 mg
MgSO4	2 ml
Thiamin	560 µl
Glucose	20 ml
Antibiotics	(2x of the usual amount for the LB-medium)

Concentrations of stock solutions' were like for MM.

IPTG stock solution:

IPTG was dissolved in water (2.38 g/10 ml) to the end concentration of 1 M. The stock solution was sterile filtered and stored in aliquots at –20°C until used. The stock solution was diluted 1:1000 when added to the medium, unless otherwise indicated.

Ampicillin stock solution:

Ampicillin was dissolved in water (1 g/10 ml) to the end concentration of 100 mg/ml. The stock solution was sterile filtered and stored in aliquots at –20°C until used. The stock solution was diluted 1:1000 when added to the medium.

Canamycin stock solution:

Canamycin was dissolved in water (0.5 g/10 ml) to the end concentration of 50 mg/ml. The stock solution was sterile filtered and stored in aliquots at –20°C until used. The stock solution was diluted 1:1000 when added to the medium.

Chloramphenicol stock solution:

Chloramphenicol was dissolved in ethanol (0.5 g/10 ml) to the end concentration of 50 mg/ml. The stock solution was sterile filtered and stored in aliquots at –20°C until used. The stock solution was diluted 1:1000 when added to the medium.

Crystallization Buffer:

5 mM tris-HCl, pH 8.0	2.42 g/4l
50 mM NaCl	11.68 g

Phosphate-Buffered Saline (PBS) Buffer:

10 mM $Na_2HPO_4 \cdot 2H_2O$, pH 7.3	1.78 g/l
1.8 mM KH_2PO_4	1.36 g
140 mM NaCl	8.18 g
2.7 mM KCl	0.2 g
0.05 % NaN_3	0.5 g

Laboratory Equipment

Consumables:

Centripreps YM3, YM10	Amicon, Witten, FRG
Dialysis tubing Spectra/Por MW 3500, 10000	Roth, Kleinfeld, Hannover, FRG
Falcon tubes, 15 ml, 50 ml	Becton Dickinson, Heidelberg, FRG
Maxi-Prep, Plasmid Isolation Kit	Qiagen, FRG
NMR-tubes, 5 mm	Wilmad, Buena, NJ, USA
Parafilm	American National, Canada
Pipette tips 10 µl, 200 µl,1000 µl	Gilson, Villiers-le Bel, France
Plastic disposable pipettes 1 ml, 5 ml, 10 ml, 25 ml	Falcon, FRG
Reaction cups 0.4 ml, 1.5 ml, 2 ml	Eppendorf, FRG
Sterile filters Millex 0.22 µm, 0.45 µm	Millipore, Molsheim, FRG
Syringes 1 ml, 2 ml, 10 ml, 20 ml, 60ml	Braun, Melsungen FRG
Ultrafiltration membranes YM3, YM10	Amicon, Witten, FRG

Chromatography equipment, columns and media:

ÄKTA explorer 10	Amersham Pharmacia, Freiburg, FRG
Peristaltic pump P-1	Amersham Pharmacia, Freiburg, FRG
Fraction collector RediFrac	Amersham Pharmacia, Freiburg, FRG
Recorder REC-1	Amersham Pharmacia, Freiburg, FRG
UV flow through detector UV-1	Amersham Pharmacia, Freiburg, FRG
BioloLogic LP System	Biorad, FRG
HiLoad 16/60 Superdex S30pg, S200pg	Amersham Pharmacia, Freiburg, FRG
HiLoad 26/60 Superdex S75pg	Amersham Pharmacia, Freiburg, FRG
HiLoad 10/30 Superdex S75pg	Amersham Pharmacia, Freiburg, FRG
Mono Q HR 5/5, 10/10	Amersham Pharmacia, Freiburg, FRG

Mono S HR 5/5, 10/10	Amersham Pharmacia, Freiburg, FRG
NiNTA-agarose	QIAGEN, FRG
Buthyl Sepharose 4 FF	Amersham Pharmacia, Freiburg, FRG
Q-Sepharose FF	Amersham Pharmacia, Freiburg, FRG
SP-Sepharose FF	Amersham Pharmacia, Freiburg, FRG
Glutathione Sepharose	Amersham Pharmacia, Freiburg, FRG

Miscellaneous:

Autoclave	Bachofer, Reutlingen, FRG
Balances PE 1600, AE 163	Mettler, FRG
Centrifuge Avanti J-30I	Beckman, USA
Centrifuge Microfuge R	Beckman, USA
Centrifuge 3K15	Sigma, FRG
Centrifuge 5414	Eppendorf, FRG
Chambers for SDS PAGE and Western blotting	MPI für Biochemie, FRG
Ice machine Scotsman AF 30	Frimont, Bettolino di Pogliano, Italy
MARresearch image plates, mar345	MARresearch, Hamburg, FRG
Magnetic stirrer Heidolph M2000	Bachofer, Reutlingen, FRG
NMR-spectrometer DRX500, DRX600	Bruker, Rheinstetten, FRG
pH-meter pHM83	Radiometer, Copenhagen, Denmark
Pipettes 2.5µl, 10µl, 20µl, 200µl, 1000µl	Eppendorf, FRG
Quarz cuvettes QS	Hellma, FRG
Shaker	Adolf-Kühner AG, Switzerland
Spectrophotometer	Amersham Pharmacia, Freiburg, FRG
Ultrafiltration cells, 10ml, 50ml, 200ml	Amicon, Witten, FRG
Vortex	Cenco, FRG
X-ray generator RU2000, 45kV, 120mA	Rigaku, Tokyo, Japan

3.2 Molecular Biology Techniques

All employed molecular biology protocols, if not mentioned here, were used exactly as described in Sambrook & Russell (2001).

Electroporation

Protocol for Electrocompetent Cells:

1. Bacteria were streaked on an LB agar plate, and incubated at 37°C overnight.
2. 50 ml of LB medium in a 250 ml flask were inoculated with a single colony from the LB plate and incubated at 37°C with shaking (200 rpm) overnight.
3. 1 l of LB medium in a 3 l flask was inoculated with the 50 ml overnight culture. The culture was grown in a shaking incubator (200 rpm) at 37°C until the OD_{600} was between 0.5 – 0.6 (approximately 2 hours).
4. The culture was transferred to the two chilled, sterile 500 ml centrifuge bottles and incubated on ice for 30 min. Thereafter centrifugation followed at 2000G for 15min. at 0 – 4°C.
5. The supernatant was decanted, and bottles put back on ice. The cell pellet in each bottle was resuspended in approximately 500 ml of cold (0 – 4°C) sterile water, and subsequently centrifuged as before.
6. The cells in each bottle were washed again with 250 ml of cold sterile water, and centrifuged.
7. The cell pellet in each bottle was then resuspended in 20 ml of cold sterile 10% glycerol and transferred to a chilled, sterile, 50ml centrifuge tube. Centrifugation followed at 4000G for 15 min. at 0 – 4°C.
8. The 10% glycerol was decanted and the pellet resuspended for the second time in 1 ml cold sterile 10% glycerol.
9. Using a pre-chilled pipette the cell suspension was aliquoted (40 µl) to pre-chilled 1.5 ml tubes and frozen immediately in liquid nitrogen. The aliquots were kept at –80°C ready for use.

Transformation of the Electrocompetent Bacteria:

1 µl of a plasmid DNA solution in water was mixed together with the 40 µl aliquot of electrocompetent bacteria and put between the electrodes of a 0.1 cm electroporation cuvette (Biorad, FRG). The cuvette was then put into the electroporator (Stratagene, FRG), and a pulse of 1660 V was applied. The value of the time constant was observed (usually 3-5 ms). The mixture was then washed out from between the electrodes with 1ml of the sterile pre-warmed (37°C) LB medium (without antibiotics), transferred to a sterile 1.5ml tube and

shaked (800rpm) at 37°C. After 1 hour the cells were streaked on a LB agar plate with an appropriate antibiotic.

Bacterial Cultures

Bacterial Culture in LB medium:

1. 50 ml LB were inoculated with a fresh single bacterial colony and incubated overnight at 37°C with vigorous shaking (280rpm) in a 100 ml flask.
2. Pre-warmed 1 l LB medium in 3 l flask was inoculated with 10ml of the overnight culture, supplemented with appropriate antibiotica, and incubated at 37°C with shaking (150 rpm) until the OD_{600} reached the 0.7 value.
3. Induction by IPTG addition followed. 0.1-1 mM IPTG concentration was usually used. The cells were then grown until the expected OD was reached.

Bacterial Culture in MM:

1. 2 ml LB were inoculated with a single colony, and shaked (150 rpm) overday in a 15 ml falcon tube at 37°C.
2. 20 ml MM were inoculated with 50 µl the overday culture, and shaked (280 rpm) overnight in 100 ml flask at 37°C.
3. 1 l MM was inoculated with 20 ml of the overnight culture (1:50), and shaked (150 rpm) in a 3 l flask, until the expected optical density was reached.

3.3 Tools of Biochemistry

All biochemical methods that are not mentioned here were performed exactly according to Sambrook & Russell (2001) or Coligan et al. (1995).

SDS Polyacrylamide Gel Electrophoresis (SDS PAGE)

The glycine SDS PAGE was performed exactly as described in Sambrook & Russell (2001). For small proteins or peptides, however, the tricine SDS PAGE is better suitable, as it has better resolution in low molecular weight range. The tricine SDS PAGE was adapted from Schagger & von Jagow (1987).

Tricine SDS PAGE with urea:

Stock solutions:

1. buffer A

3M tris	181.71 g/500 ml
0.3% SDS	1.5 g
0.05% NaN_3	0.25 g
pH adjusted to 8.45 with HCl	

2. buffer B

acrylamide	24 g/50 ml
bis-acrylamide	0.75 g

3. buffer C

acrylamide	23.25 g/50 ml
bis-acrylamide	1.5 g

4. buffer D

ammonium persulfate	10 %

5. 6 M urea 36.04 g/100 ml

The gels were prepared in chambers for 9 gels. Therefore the solutions volumes below are given always per 9 gels.

1. **stacking gel** (upper, poured as last, total volume 30 ml)

buffer A	7.5 ml
buffer B	2.5 ml
water	20 ml
buffer D	300 µl
TEMED	30 µl

2. **spacer gel** (middle, poured as second, length 1 cm, total volume 33 ml)

buffer A	10 ml
buffer B	6 ml
water	14 ml
6 M urea	3 ml
buffer D	200 µl
TEMED	20 µl

3. **separating gel** (lower, poured as first, length 4-5 cm, total volume 62 ml)

buffer A	20 ml

buffer C	20 ml
water	13 ml
6 M urea	9 ml
buffer D	400 µl
TEMED	40 µl

Buffer D was prepared always freshly before use. Buffer D and TEMED were added immediately before the gels were poured. Spacer gel was poured immediately after pouring the separating gel so that they could mix together creating a polyacrylamide concentration gradient. Different voltage was used for distinct gels:

stacking gel:	25-30 V
spacer gel:	50 V
separating gel:	75-80 V

Different buffers were used for anode and cathode:

1. anode buffer (+)

0.2 M tris-HCl, pH 8.9	24.22 g/l

2. cathode buffer (-)

0.1 M tris-HCl, pH 8.25	12.2 g/l
0.1 M tricine	17.9 g/l
0.1 % SDS	

Staining of Proteins

Staining of proteins with Coomassie-Blue and with Ponceau-Red was performed like as in Sambrook & Russell (2001). The silver staining, however, was slightly modified as described below:

Silver Staining:

Stock solutions:

1. solution 1

   300 ml ethanol
   150 ml acetic acid
   water up to 1000 ml

2. solution 2

41 g sodium acetate
250 ml ethanol
water up to 1000 ml
immediately before use add: 0.1 g/50 ml sodium thiosulfate
250 µl/50 ml glutardialdehyde

3. solution 3
   1 g silver nitrate
   water up to 1000 ml
   immediately before use add: 15 µl/50 ml formaldehyde
4. solution 4
   25 g sodium carbonate
   water up to 1000 ml
   pH adjusted to 11.5 with carbonate/hydrogencarbonate
   immediately before use add: 15 µl/50 ml formaldehyde
5. solution 5
   18.6 g EDTA
   water up to 1000 ml

The gels were stained in the following manner:

solution	**washing time**
1	1 h
2	1-12 h (overnight)
water	3x10 min
3	30 min
4	-
5	-

Determination of Protein Concentration

The concentration of proteins in solutions was estimated with the assistance of the Bradford reagent (BioRad; Bradford, 1976). 10 µl of the protein solution (or 1 µl, if the protein solution is concentrated) to be measured was added to 625 µl of the BioRad-reagent working solution (the working solution was prepared by 1:5 dilution of BioRad-reagent stock solution in the PBS buffer or water, stored in the fridge). Then the mixture was diluted with 400 µl water. After thoroughly mixing the sample, the OD_{595} was measured. As a reference a

similar mixture was prepared with 10 µl water instead of the protein solution. OD was subsequently converted into the protein concentration on the basis of a BSA calibration curve.

NMR Samples Preparation

If not otherwise indicated, the samples for NMR spectroscopy were concentrated and dialyzed against PBS buffer. Typically, the sample concentration varied from 0.3 to 1.0 mM. Before measuring, the sample was centrifuged in order to sediment aggregates and other macroscopic particles. 450 µl of the protein solution were mixed with 50 µl of D_2O (5-10%) and transferred to an NMR sample tube.

Crystallization Procedure

Protein samples for crystallization trials were prepared in the following manner. A more than 95% pure protein sample was concentrated and purified by gel filtration (HiLoad Superdex column S75, S30 or S200) in a low salt crystallization buffer. Collected fractions were pooled and concentrated using the Amicon concentrating cell until the expected protein concentration was reached (5-100 mg/ml). The membrane was washed several times in the crystallization buffer before use. Subsequently, the sample was filtered through the 0.22 µm Millipore filter, equilibrated previously twice with the crystallization buffer. The samples were kept on ice or at RT. Reservoirs were filled with 500 µl of the reservoir buffer. Hanging or sitting drop techniques were employed.

4. Investigations on the Fluorescent Protein Family

4.1 ^{1}H, ^{15}N, ^{13}C NMR Spectroscopy of GFPuv

4.1.1 Biological background

Green fluorescent proteins not only provide powerful tools for monitoring gene expression, protein movement and protein interactions (Chalfie et al., 1994; Tsien, 1998) but also exhibit complex photophysical behaviour as observed by ensemble (Chattoraj et al., 1996; Lossau et al., 1996; Brejc et al., 1997) and single molecule measurements (Dickson et al., 1997; Garcia-Parajo et al., 2000; Zumbusch & Jung, 2000). Up to now the family of GFP-like proteins comprises 27 cloned and spectroscopically characterized proteins (Labas et al., 2002) of which eighteen high-resolution crystal structures are available (http://www.rcbs.org). The overall structure of GFP consists of an 11-stranded β-barrel with a central helix that carries the chromophore (Ormoe et al., 1996; Yang et al., 1996). The X-ray diffraction studies and a variety of physicochemical methods highlight an apparent exceptional stability of the GFP fold in which the chromophore lies rigidly inside the conformationally inflexible GFP molecule (Striker et al., 1999). The stability of GFP against temperature, denaturants and proteases is very high (Tsien, 1998; Ward, 1981). However, even the earliest absorption spectroscopy studies by Ward et al. indicated some conformational flexibility in the chromophore-binding site (Ward et al., 1982). More recently, several fluorescence, X-ray crystallographic, and computational studies tried to address the contributions of various fast dynamic processes to the fluorescence spectra of GFPs. At room temperature, denatured GFP or model chromophores in solution show no significant fluorescence. However, when frozen as ethanol glass at 77K, such compounds become highly fluorescent (Niwa et al., 1996; Bell et al., 2000). Inside the protein environment the chromophore is very sensitive to mutations of surrounding residues (Kummer et al., 1998; 2000). Competition of radiationless photoisomerization and fluorescence emission has been suggested to provide an explanation for the differences in quantum yield of fluorescence in various GFP mutants (Kummer et al., 1998; Voityuket et al., 1998). It was proposed that any increase in the degrees of motional freedom has to be avoided when designing new GFP variants (Voityuket et al., 1998). For photoisomerization pathways the rotations around either of the two bridging bonds between the two rings of the chromophore and a so-called hula-twist have been suggested, but it was

not possible to determine the timescales of conformational changes quantitatively by molecular dynamics simulations (Weber et al., 1999; Chen et al., 2001).

NMR has been suggested as a tool for elucidating the dynamics of chromophore formation, water accessibility of the chromophore and conformational flexibility in GFP (Bulina et al., 2002; Prendergast, 1999). In contrast to many spectroscopic techniques, NMR spectroscopy provides a large frequency range for studying dynamical processes from picosecond to second timescales and even longer at atomic resolutions. For example, motional and thermodynamical information for backbone amides in proteins can be obtained from measurements of ^{15}N relaxation rates in ^{15}N labeled proteins (Haupts et al., 1998; Spyracopoulos & Sykes, 2001). Hydrogen-deuterium exchange experiments allow for a characterization of conformational fluctuations in secondary structure elements (Kay, 1998). NMR studies on GFPs turned out to be a challenge due to the strong tendency of GFPs to aggregate (Ward, 1981). This study investigates the dynamical properties of the GFP β-barrel that plays an essential role in protecting the chromophore from environmental influences. The characterization of the backbone flexibility of GFP by ^{15}N NMR relaxation measurements allowed the determination of the residue specific spectral densities of motions on the picosecond to nanosecond timescale. To identify residues that may play a role for the structure and dynamics of the β-barrel, sequences of several GFP variants and the perlecan-binding G2 fragment of nidogen-1 (Dempsey, 2001) are compared by structure specific and sequence specific alignment. To investigate the influence of histidine 148 on backbone dynamics the mutation His 148 Gly was introduced in GFPuv. ^{1}H-^{15}N NMR correlation spectra show the presence of a second conformation in the mutant protein.

4.1.2 Sample Preparation

^{15}N, ^{15}N/99%-^{2}H, ^{15}N/^{13}C, and ^{15}N/^{13}C/70%-^{2}H labeled samples of GFPuv were produced as described by Georgescu et al., (2000 and 2002). The samples were dissolved in phosphate buffered saline (PBS) at pH 7.0.

To produce the GFPuv His148Gly mutant the QuikChange site-directed mutagenesis kit (Stratagene) was used. Mutagenesis primers (MWG-DNA) were designed to introduce the specific experimental mutation. After PCR, reaction products were digested (37°C, 1 h) by Dpn1 restriction enzyme (10 U/l) and transformed into *E. Coli* XL1_Blue supercompetent cells. Transformation plates were incubated at 37°C for 16 h. DNAs from 8 different colonies were sequenced and 7 of them had the desired mutation. After this His148Gly GFPuv was

expressed in *Escherichia Coli*. Expression was controlled by the T7 promotor and provided by the gene for β-lactamase. After induction with 0.5 mM isopropylthiogalactoside (IPTG) His6-tagged GFP was overexpressed in BL21(DE3) cells in 1 liter LB-medium for 6 h at 37°C. The cell pellet was resuspended in 50 ml lysis buffer (300 mM NaCl, 50 mM Na_2HPO_4, pH 8.0) with 1 mg DNAse, 1 mg RNAse, 1 mg $MgCl_2$ and 0.1 M phenylmethylsulfonyl fluoride (PMSF) and then sonicated with Micro-Tip and centrifugated for 45 minutes at 80.000 rpm in a Beckmann centrifuge. The protein was purified by Ni^{2+} affinity chromatography. Ni-NTA resin (Qiagen) was incubated with the supernatant for 1 hour at 4°C, poured into a column and subsequently washed with 100 ml of a lysis buffer. The protein was eluted from the Ni-NTA resin with 20 ml elution buffer (300 mM NaCl, 50 mM Na_2HPO_4, 500 mM imidazole, pH 8.0). Then the protein was dialysed against PBS (115 mM NaCl, 8 mM KH_2PO_4, 16 mM Na_2HPO_4, pH 7.0) and concentrated to 15-20 mg/ml using Centricon 10 (Milipore). After additional gel filtration with a Superdex 75 column (Pharmacia), the sample was concentrated again to a concentration of 20-25 mg/ml. The purity of the sample was checked by mass spectroscopy. Two independent samples were produced to ensure the reproducibility of the effects observed by NMR spectroscopy.

4.1.3 Optical Spectroscopy

UV-absorption spectra of 10 μM GFPuv and GFPuv His148Gly in PBS buffer were recorded on a Perkin-Elmer UV/VIS spectrophotometer at room temperature. The spectra were scaled using a known molar extinction coefficient of GFP (ε_M = 21000 M^{-1} cm^{-1} at λ = 277 nm). Optical excitation and emission spectra were measured on a Perkin-Elmer LS50B spectrometer at RT. The protein concentration was adjusted to 50 nM. In order to acquire the excitation spectra, the emission at 510 nm was measured. To record the emission spectra, the samples were excited at 396 nm. In both cases a slit of 2.5 nm was used.

4.1.4 NMR Spectroscopy

All heteronuclear ^1H-^{15}N NMR experiments were performed on either Bruker DRX500, DRX600, or DMX750 spectrometers equipped with 5 mm ^1H-^{13}C-^{15}N triple resonance TXI probeheads, including triple-axis gradients (DRX600, DMX750) or z-axis gradients (DRX500). The spectra were recorded at 310K with a sweepwidth of 17.5 ppm for ^1H and 42 ppm for ^{15}N. Assignment of the backbone resonances was accomplished with triple

resonance spectra including HNCA (Cavanagh et al., 1996), TROSY-HNCA (Salzmann et al., 1998), HN(CO)CA, HNCO, CBCA(CO)NH (Cavanagh et al., 1996). A HBHA(CBCACO)NH (Cavanagh et al., 1996), spectrum was also recorded on a Bruker Cryoprobe™ DRX 600 spectrometer. In addition, HSQC (Mori et al., 1995) spectra of selectively ^{15}N labeled Gly, Phe, Tyr, Met, Ala, Leu, Lys, Ile, and Val and inverse ^{15}N Asn, Thr, His, and Asp labeled samples of GFPuv were recorded. 2D NOESY (τ_m=160ms) spectra and 3D NOESY-HSQC (τ_m=160ms) supplemented the NMR data. Where appropriate, the spectra were recorded on 70% or 99% deuterated samples of GFPuv. The program Sparky (Goddard et al., 1990) and in-house software CCNMR (Cieslar et al., 1990) were used for analysing the spectra. Details of the NMR data acquisition and resonance assignment are described in Georgescu et al. (2000 and 2002).

A pH titration and a HD exchange (Dempsey, 1998) experiment were conducted based on ^{1}H-^{15}N HSQC spectra (Mori et al., 1995). For the titration the pH of a ^{15}N labeled sample of GFPuv in PBS buffer was reduced from pH 7.7 to pH 6.1 in series of 0.1 by adding the appropriate amount of 1M H_3PO_4. For each step a HSQC spectrum was recorded. For the HD exchange experiment the ^{15}N labeled sample of GFPuv was lyophilized and dissolved in 99.99% D_2O. Then a series of 16 HSQC spectra were recorded. The measurement time for one spectrum was 1 hour. In the following month one spectrum per week was recorded. Finally spectra after 2, 3 and 5 months were recorded, between measurements the protein was stored at +4°C.

4.1.5 NMR diffusion measurements

The translational diffusion coefficient (Dingley et al., 1995) of GFPuv at 310 K was measured by a modified Watergate-^{1}H-^{1}D-sequence (Sklenar et al., 1993) with strong, sine-shaped gradients centred on the selective Watergate pulse. The gradients were calibrated by 1D NMR imaging of a 3.5 mm Teflon plug in a standard 5 mm NMR tube filled with $CuSO_4$ doped water. The signal attenuation I/I_0 due to translational diffusion is given by:

$$I = I_0 e^{-bD},$$

with D being the translational diffusion coefficient. The corresponding b-value for sine-shaped gradients was calculated using the relationship (Seifert, 2002)

$$b = \gamma^2 G^2 \frac{\delta^2}{\pi^2}\left(4\Delta - \delta\right),$$

where *G* represents the maximum strength of the gradient. The timing of the gradients before and after the selective Watergate pulse was set to Δ = 17 ms and δ = 15 ms. The apparent mass of the protein can be calculated from the diffusion coefficient according to (Cantor & Schimmel, 1980):

$$M = \left(\frac{kT}{6\pi\nu FD} \right)^3 \frac{4\pi N_A}{3\left(\overline{V}_2 + \kappa\overline{V}_1\right)} .$$

For the apparent mass calculation the following parameters (T = 310K) were used: the protein form factor F = 1.03, hydration of the protein κ = 0.34, viscosity of water $\nu = 6.91 \cdot 10^{-4}$ Ns/m^2, specific volume of the protein $\overline{V}_2 = 0.731 \cdot 10^{-3}$ m^3/kg, and specific volume of water $\overline{V}_1 = 1.00669 \cdot 10^{-3}$ m^3/kg. k and N_A represent the Boltzmann constant and Avagadro's number, respectively.

4.1.6 Sequence Alignment

As can been seen in Fig. 4.1, *Aquorea victoria* GFP/GFPuv, *Discosoma striata* DsRed, *Anemona sulcata* asCP and the G2 fragment of mouse nidogen-1 share several conserved amino acids despite their different origin and function. The low rmsd values for the structures of GFP, DsRed, and nidogen-1 indicate a high degree of structural similarity despite a low degree of sequence homology (e.g. only 10% for nidogen-1 in comparison to GFP (Kummer et al., 1998). Residues: Gly20, Phe27, Gly40, Leu60, Gly104, Gly127, and Pro196 (GFP numbering) are conserved in all proteins under consideration. Most of these similarities correlate well with the start or end of secondary structure elements suggesting a key role especially of the conserved glycine residues for the overall GFP fold. In addition, Phe71 in GFP, nidogen-1 and asCP is changed to a tyrosine in DsRed. Phe100 in GFP, DsRed, and nidogen-1 is mutated to tyrosine in asCP. His148 in GFP is changed into a tyrosine in nidogen-1. Furthermore, several residues are conserved between DsRed, nidogen-1 and asCP, including Gln110, Ser113, Phe92, Glu146, Thr149-Glu150, and Val160.

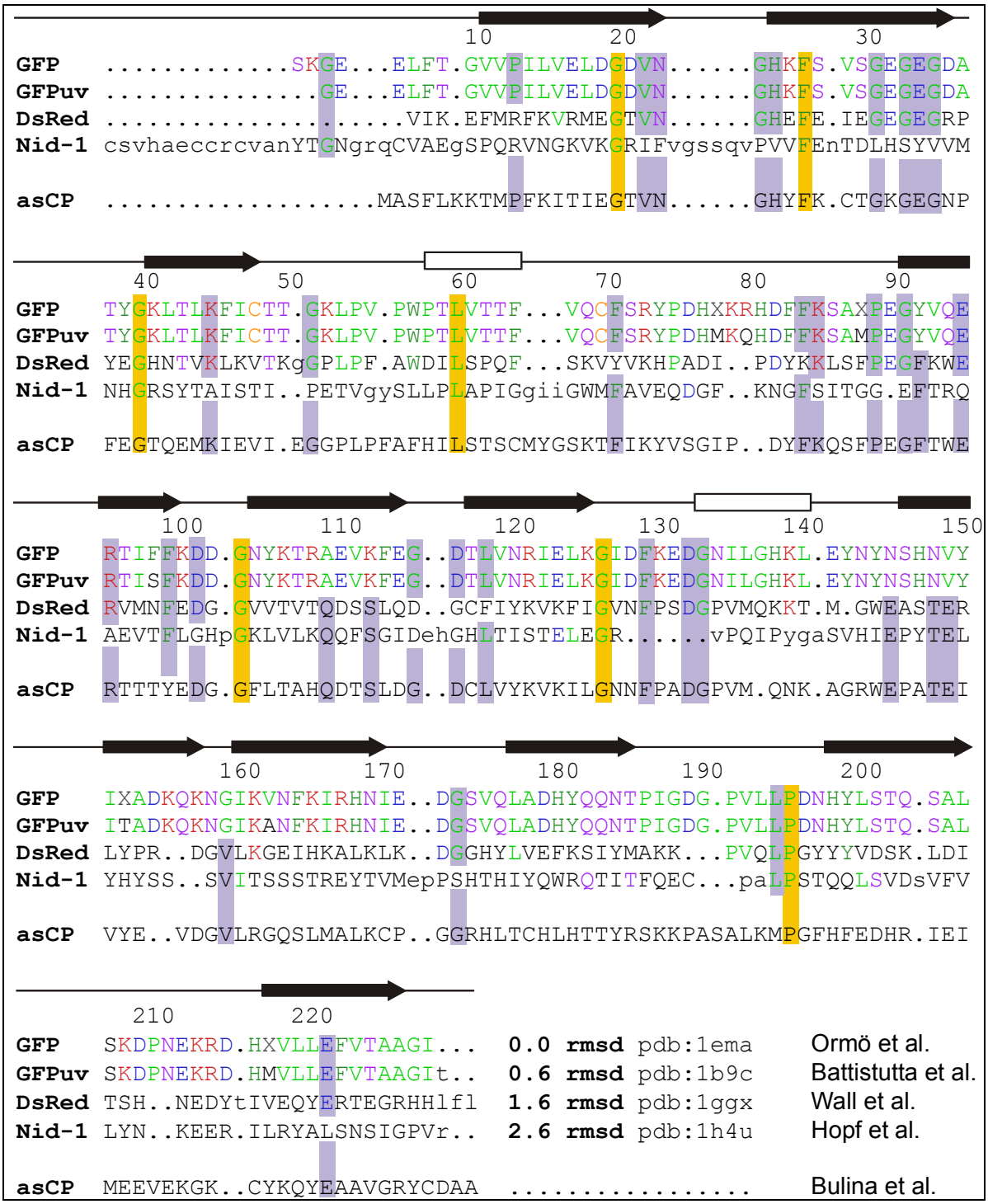

Fig. 4.1 *The sequence alignment of GFP (PDB: 1ema), GFPuv (PDB: 1b9c), DsRed (PDB: 1ggx), Nidogen-1 (PDB: 1h4u) and asCP (no structural data available) is shown. The alignment of GFP, GFPuv, DsRed and Nidogen-1 is based on a FSSP/DALI database search Holm & Sander (1996) The sequence of asCP was aligned according to Bulina et al., (2002). The backbone rmsd values are given in Å relative to GFP. The amino acids conserved in all five proteins are marked with yellow bars. Blue bars highlight amino acids conserved to a smaller degree.*

Fig. 4.2 shows a stereoplot of the GFP crystal structure highlighting the residues conserved in all proteins under consideration.

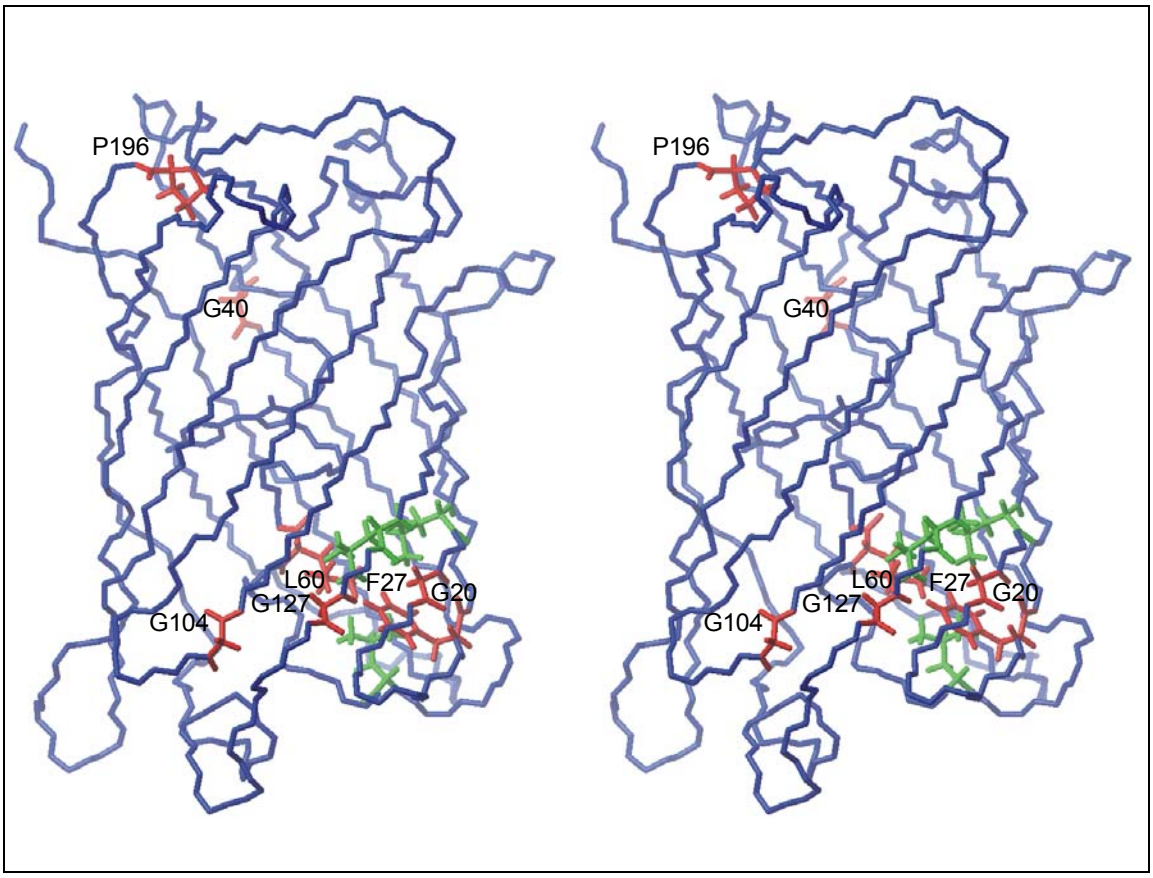

Fig. 4.2 *Stereoplot of the crystal structure of GFPuv (PDB: 1b9c) with the residues conserved in GFP, DsRed, asCP, and the G3 fragment of nidogen-1 depicted in red. Residues L18, L53, and L125 that form the hydrophobic pocket with L60 and F27 are shown in green.*

4.1.7 Translational Diffusion

Figures 4.3A and 4.3B illustrate the result of the translational diffusion NMR experiment. Considering the uncertainties of the gradient calibration a diffusion constant of D = (120 ± 10) $\mu m^2/s$ was measured. Using the specific model this results in an apparent mass of GFPuv of (45 ± 10) kDa under the conditions used in NMR experiments. This is significantly larger than 27 kDa of the monomeric GFPuv, but still smaller than 54 kDa expected for a dimer suggesting that GFPuv in solution exists as mixture of monomers and

dimers. This is in agreement with the known dimer dissociation constant K_D = 0.1 mM of GFPs as determined by analytical ultracentrifugation (Philips, 1998). Fig. 4.3C illustrates the GFP dimer interface as observed by X-ray crystallography.

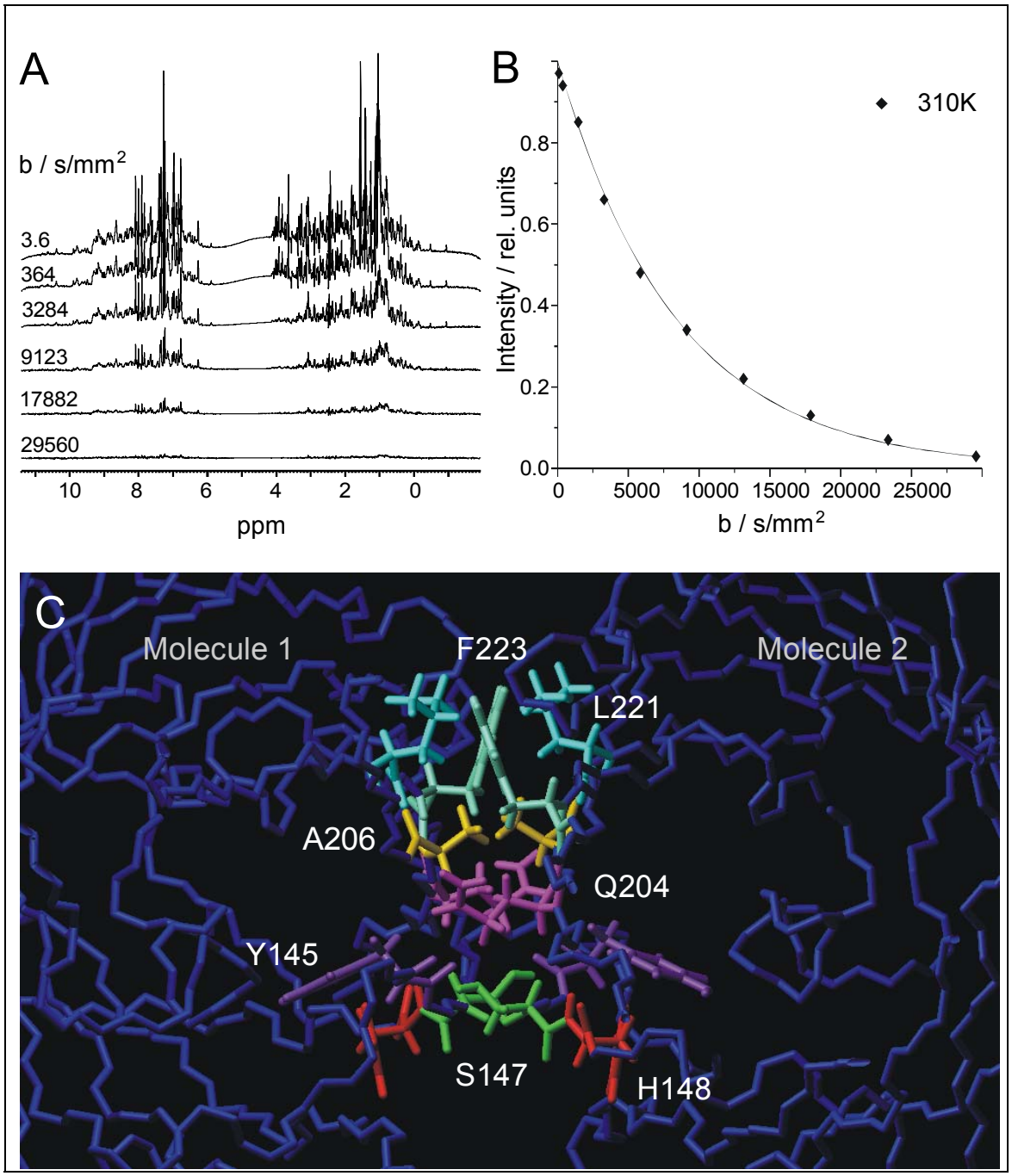

Fig. 4.3 *NMR measurement of the translational diffusion of GFPuv (**A.** and **B.**).*
***C.** The dimerization interface of GFP*

4.1.8 NMR Assignment

80% of the backbone amide groups of GFPuv were assigned by standard triple resonance method (see Fig. 4.4). The assignment was complicated by a substantial amount of missing interresidue connectivities in the triple resonance NMR spectra. The assignment of the central α-helix and of β-sheets 7, 8, and 10 could not be completed. These difficulties can be explained by the necessity of a neutral pH, the tendency of GFP to aggregate in solution, and the presence of conformational exchange on μs - ms time scales – as described below – that lead to line broadening.

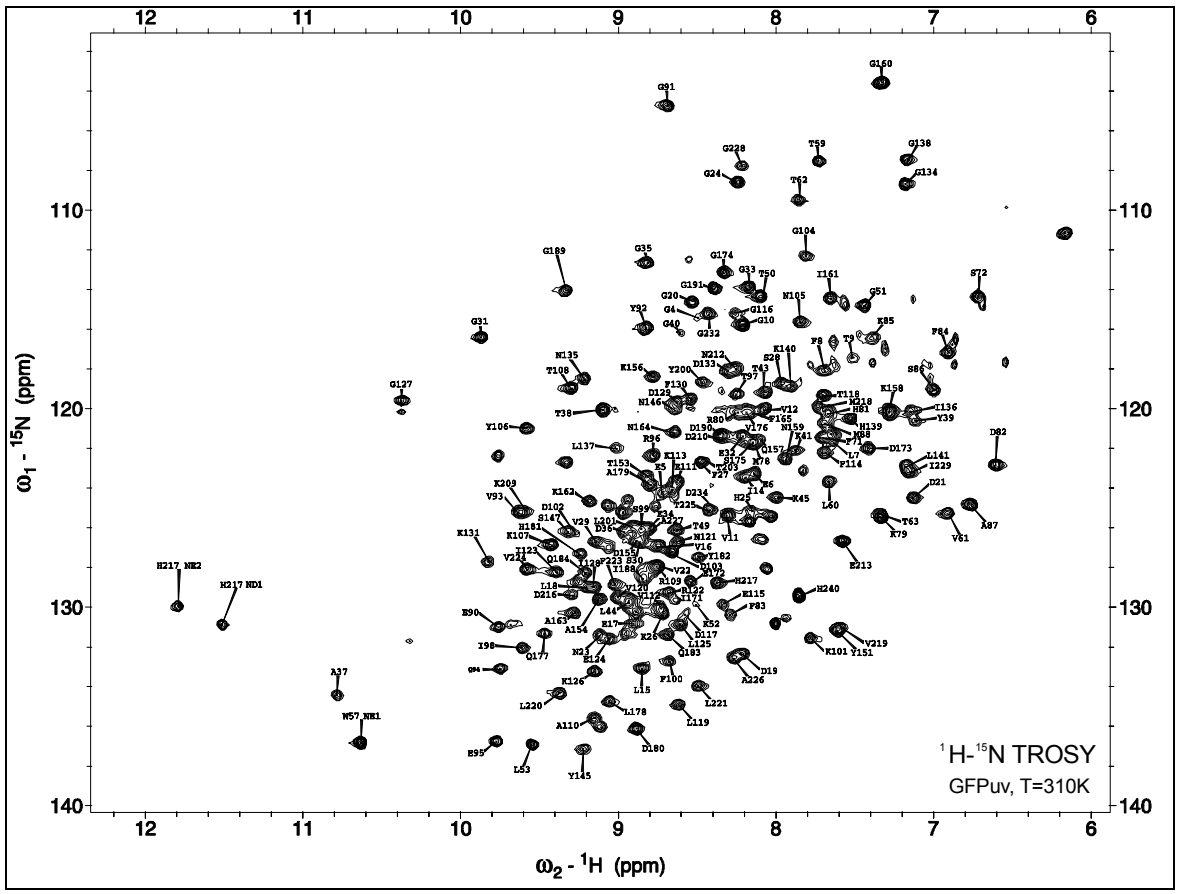

Fig. 4.4 *A 2D ^{1}H-^{15}N NMR TROSY spectrum of GFPuv at 310K. The labels indicate the sequence-specific resonance assignment.*

4.1.9 Mutant His148Gly

The UV absorption and excitation spectra of GFPuv and GFPuv mutant His148Gly show that the mutation does not dramatically affect the absorption maximum of the chromophore's A state at 398 nm (Fig. 4.5). The intensity of peak A in the absorption spectrum and the emission peak at 508 nm is reduced to 82% in the mutant compared to

GFPuv. In contrast, peak B in the absorption spectrum is reduced to 26% in the mutant protein. Peaks A and B are usually assigned to the neutral and anionic state of the chromophore, respectively. The emission at 510 nm is reduced to 80% in the mutant protein, which is agreement with the reduced absorption of state A.

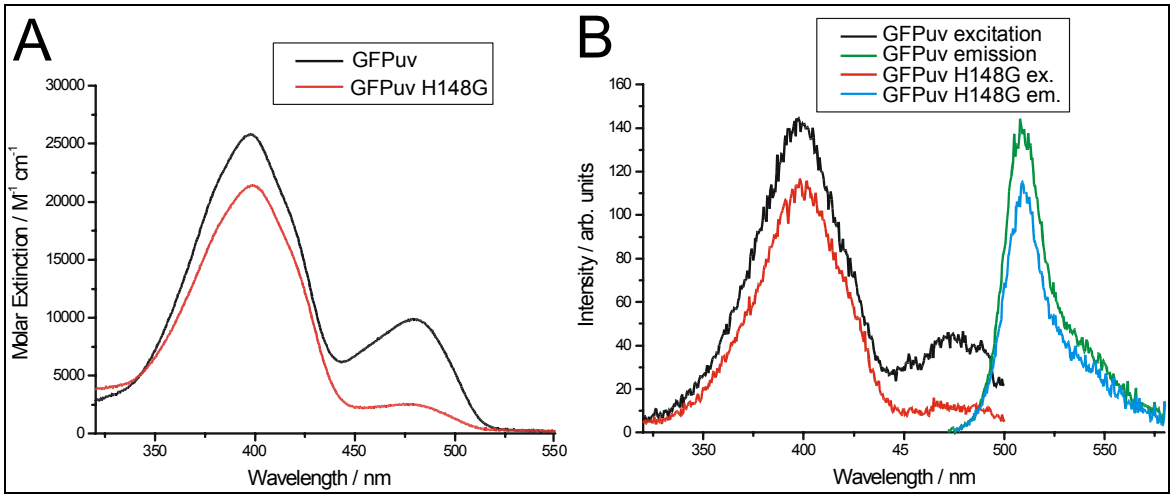

Fig. 4.5 ***A.*** *UV/VIS absorption spectra* ***B.*** *Excitation and emission spectra of GFPuv and GFPuv mutant H148G.*

Mutation His148Gly leads to the appearance of double peaks for several residues in the HSQC spectrum of GFPuv His148Gly (Fig. 4.6).

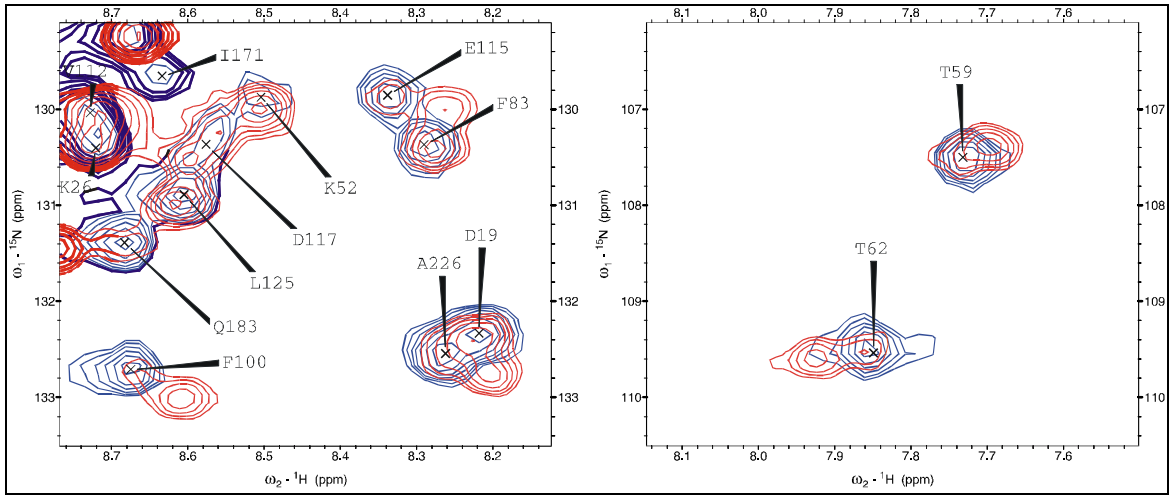

Fig. 4.6 *Two regions from the ^{1}H-^{15}N HSQC spectrum of GFPuv (blue) and GFPuv His148Gly (red) are shown exemplifying the appearance of double peaks in the mutant His148Gly for several backbone amide groups.*

The peak volume ratio of minor and major peaks was on average 0.45:1 at 310K. Minor peaks corresponded to the original resonances in the spectrum of GFPuv. The small difference in chemical shift of major and minor peaks of approximately 30 - 60 Hz indicates a timescale of exchange at the order of 0.01 s to 0.1 s. At a temperature of 290K the minor peaks seem to be even more pronounced although broadening of the resonances due to a higher solvent viscosity complicated the detection of several resonances. These temperature dependent double peaks provide an indication for a slow exchange processes affecting the backbone structure of the mutant. The following residues with known assignment are affected by these exchange processes: Thr43, Thr49, Thr59, Thr62, Ser72, Phe83, Thr97, Ile98, Phe100, Phe130, Lys158, Leu178, Tyr182, Gln183, Thr203, His217, Val224 and Ala226. In addition the heteronuclear NOE values of the minor confirmation are significantly reduced for the following residues (data not shown): Gly33, Glu34, Gly40, Lys41, Phe84, Phe100, Gly116, Asn164 and Phe165. The heteronuclear NOE values of the major conformation are comparable to those obtained for GFPuv. Figure 4.7 maps the location of residues that were assigned and unambiguously affected by exchange on the crystal structure of GFP.

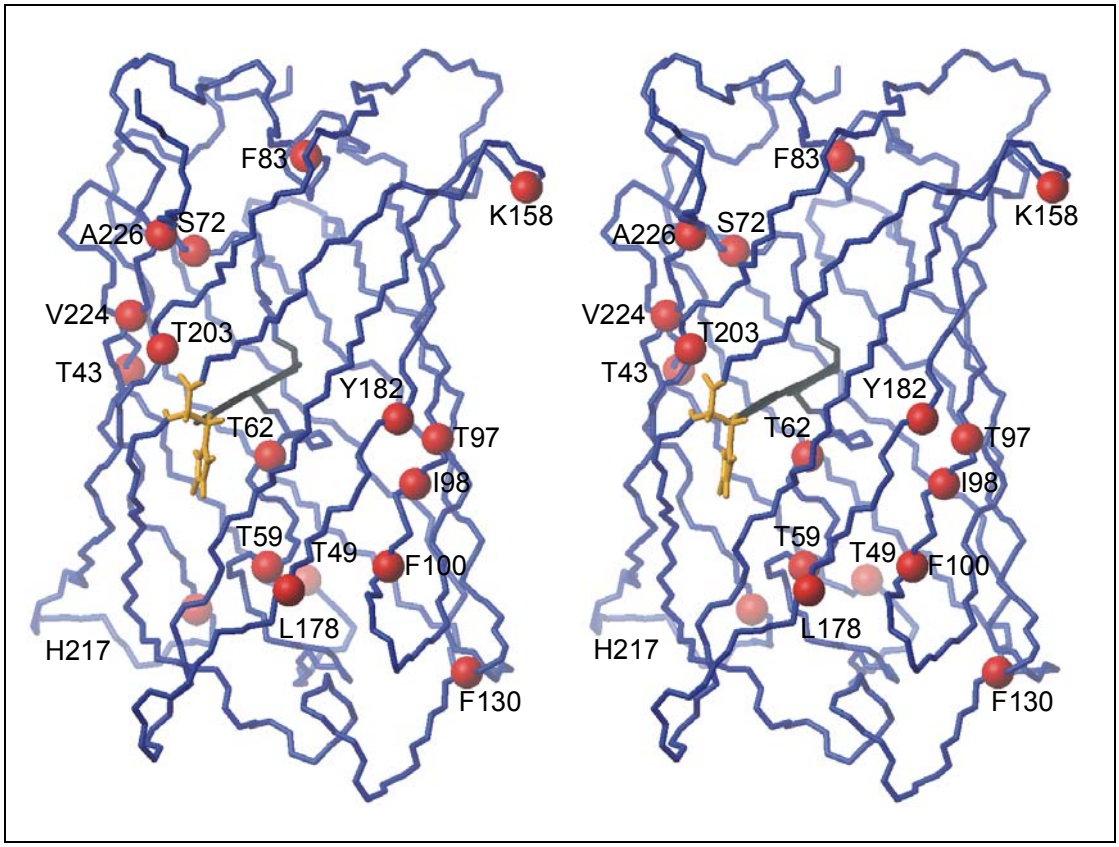

Fig. 4.7 *Dual conformation in GFPuv His 148 Gly. The affected amide nitrogens are shown as red balls with respect to the GFP structure. The chromophore and H148 are depicted in gray and orange, respectively.*

These residues cluster in the central α-helix (as far as assigned) and β-strand 4, however, the effect is not limited to residues near His148 and many residues distributed all over the β-barrel are affected.

4.1.10 Discussion

Comparison of GFP-like proteins from several organisms is useful to reveal residues that are crucial for the GFP stability. Four glycines, one phenylalanine, one leucine, and one proline residue are conserved among all species. The conserved amino acids Phe27 and Leu60 together with Leu125 (in asCP: Ile, in Dsred: Phe) form a hydrophobic core within the barrel (see Fig. 4.2) that may play a role in positioning the central α-helix during protein folding: In this process the β-strands have to wrap around the central helix to close the barrel. It is usually assumed that protein folding is driven by the thermodynamic advantage inherent in the sequestration of hydrophobic residues in the interior of the protein away from the solvent (Prendergast, 1999). However, in the core of GFP, i.e. around the chromophore, there is a surprising number of polar residues (Tsien, 1998). Therefore other nucleation points for protein folding are likely to exist. Hydrophobic residues Phe27 and Leu60 are conserved in all proteins mentioned here suggesting that these residues might be important for the stable protein structure, presumably during the folding process.

In addition, conserved glycine residues may also play an essential role for the GFP folding. It has been shown that the Thr22Gly mutation in *Drosophila* drk protein, which restores the otherwise highly conserved glycine residue in the diverging β-turn connecting β-strand 3 and 4 of SH3 domains, significantly stabilizes the drk SH3 structure (Mok et al., 2001). Conserved glycines with a strong correlation at the start or end of β-strands have been reported for the large subunit of ribulose biphosphate carboxylase/oxygenase from *Anacystis nidulans* (Cheng et al., 1998). It was proposed that glycines might serve to be hinges enabling the movement of loops. Such hinges may also be important for the GFP β-barrel.

Several other residues are conserved in only three or four of the compared species. Based on an alignment of GFP and the nidogen-1 G2 fragment it has been speculated that both proteins derive from common ancestor (Hopf et al., 2001).

The apparent protein mass of GFPuv of approximately 45 kDa determined from translational diffusion indicates the presence of a mixture of monomers (27 kDa) and dimers (54 kDa) in solution. The extent of self-association in the GFPuv sample is obvious when comparing the overall rotational correlation of GFP (238 residues, τ_c = 22 ns) to the one of e.g. maltose binding protein (MBP (Gardner et al., 1998), 370 residues, τ_c = 16 ns), which was measured at the same temperature as GFP. It was proposed that the mutations present in GFPuv, Phe99Ser, Met153Thr, and Val163Ala, reduce the aggregation tendency (Fukuda et al., 2000). Our results clearly show that these mutations cannot prevent self-association of GFPuv under the conditions needed for NMR spectroscopy, i.e. the high protein concentration of approximately 1 mM. Additionally, under these conditions one-dimensional ^{1}H NMR spectra (see Fig. 4.8) of GFPuv and Clontech EGFP (Phe64Leu, Ser65Thr) show no significant difference in ^{1}H line width between GFPuv and EGFP.

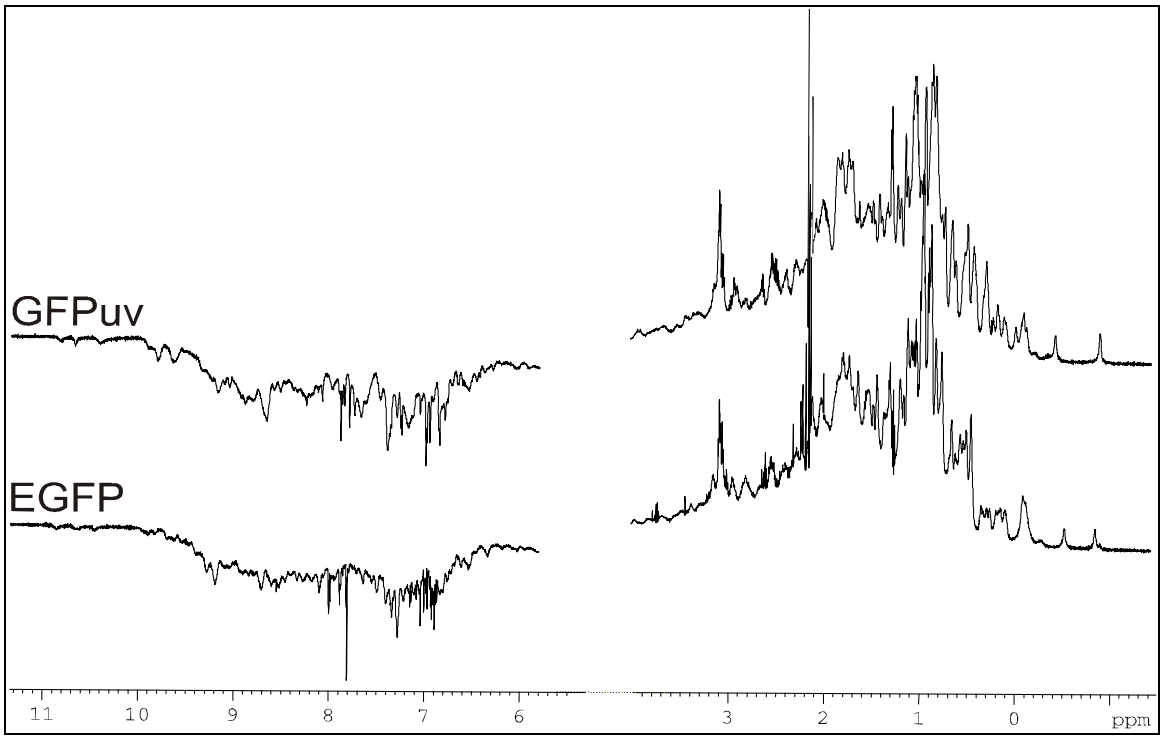

Fig. 4.8 *^{1}H NMR spectra of GFPuv and EGFP (Clontech) recorded at a temperature of 310 K and a proton frequency of 600 MHz.*

Additional mutations like Ala206Lys, Leu221Lys, and Phe223Arg (see Zacharias et al., (2002) and Fig. 4.3C) are known to reduce hydrophobic interactions and might therefore allow for production of NMR samples with less aggregation. It is not yet clear if these hydrophobic interactions play a physiological role for any kind of membrane association or

interaction with the light emitting protein aequorin. It is known that other GFPs, like *Renilla* GFP, are obligate dimers (Tsien, 1998) whereas DsRed forms a tetramer (Zacharias et al., 2002 and paragraph 4.3). In general, the self-association of GFPs is a relatively conserved feature. Recently it has been speculated that a relationship between oligomerization and habitat temperature may exist (Zacharias et al., 2002). Hydrophobic interactions may also be relevant for the interaction of GFP and the light-emitting protein aequorin. Remarkably, the structurally very similar (2.6 Å rmsd compared to GFP) nidogen-1 is supposed to play a role in membrane assembly (Hopf et al., 2001). Therefore further studies should address the structural and phylogenetic relationship of GFPs to proteins like nidogen-1 and, for example, β-barrel membrane proteins OmpX, OmpA (Fernandez et al., 2001) and others (Schulz, 2000).

Figure 4.9 maps experimental findings on the GFP secondary structure. With 80% of the backbone amides being assigned, some important conclusions can be drawn from the relaxation measurements (data not shown): (i) the overall structure of GFP is rigid on the picosecond to nanosecond time scale. This is in agreement with molecular dynamics simulation, which predicted a RMS deviation of only 0.9 Å during a 1 ns period (Helms et al., 1999). (ii) A significant variation of the η/R_2 values can be observed in β-sheets 4, 5, and 6 and especially in the first α-helix where the η/R_2 values are lower at the beginning and at the end of the sheets compared to the center of the sheets. This indicates motions on the μs – ms timescale in the turns connecting the sheets. These features are not observed on the ps – ns timescale in the spectral density function $J(\omega)$ at least for the β-sheets. (iii) One side of the protein, corresponding to β-strands 7, 8, 9, and 10, shows increased HD exchange rates. Increased HD exchange rates in β-sheets are usually interpreted as the result of hydrogen bonds being broken by conformational fluctuations on the μs - ms time scale (Dempsey, 2001). The first two observations may not be unexpected, but the last issue needs further discussion.

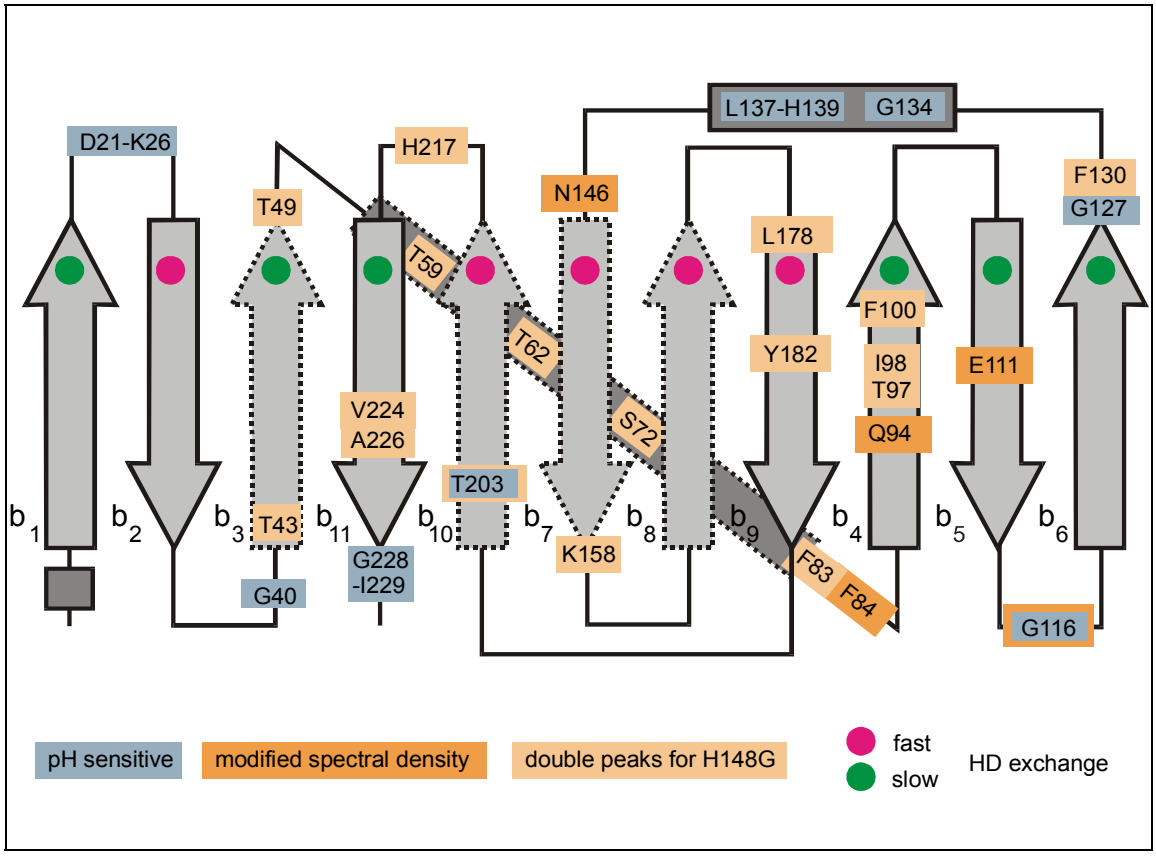

Fig.4.9 *The observed dynamic properties of GFPuv are mapped on the GFP fold. β-sheets and α-helices are drawn as arrows and boxes, respectively. Dotted arrows indicate β-sheets with incomplete resonance assignment.*

In crystal structures of GFPs a reduced inter-strand hydrogen-bonding network (see Fig. 4.10A and 4.10B) can be observed for the β-strands 7, 8, and 10. This may be caused either by dynamic conformational processes or by inefficient sidechain packing inside the GFP β-barrel, which leads to diverging β-strands and therefore larger inter-strand distances. For β-barrel folds purely geometrical considerations showed that efficient side chain packing and low conformational energies are related to the overall geometry of the cylinder which is characterized by the number of β-sheets *n* and the "shear number", *S*, a measure of the stagger of the strands (McLachlan, 1979; Murzin et al., 1994). Crystal structures of GFP-like proteins show that these proteins are characterized by $n = 11$ and $S > 20$, a mean slope of the strands with respect to the long axis of the protein of $\alpha = 41°$ and a mean radius of the barrel of approximately 12 Å. In contrast to the proteins studied by McLachlan, (1979) and Murzin et al. (1994), GFP-like proteins contain a central α-helix. However, the proposed condition for

efficient side-chain packing, and therefore low-energy β-barrels, $S = 2n$ (Murzin et al., 1994) is fulfilled to good approximation. Therefore inefficient side chain packing is unlikely to be the origin of the disturbed hydrogen-bonding network.

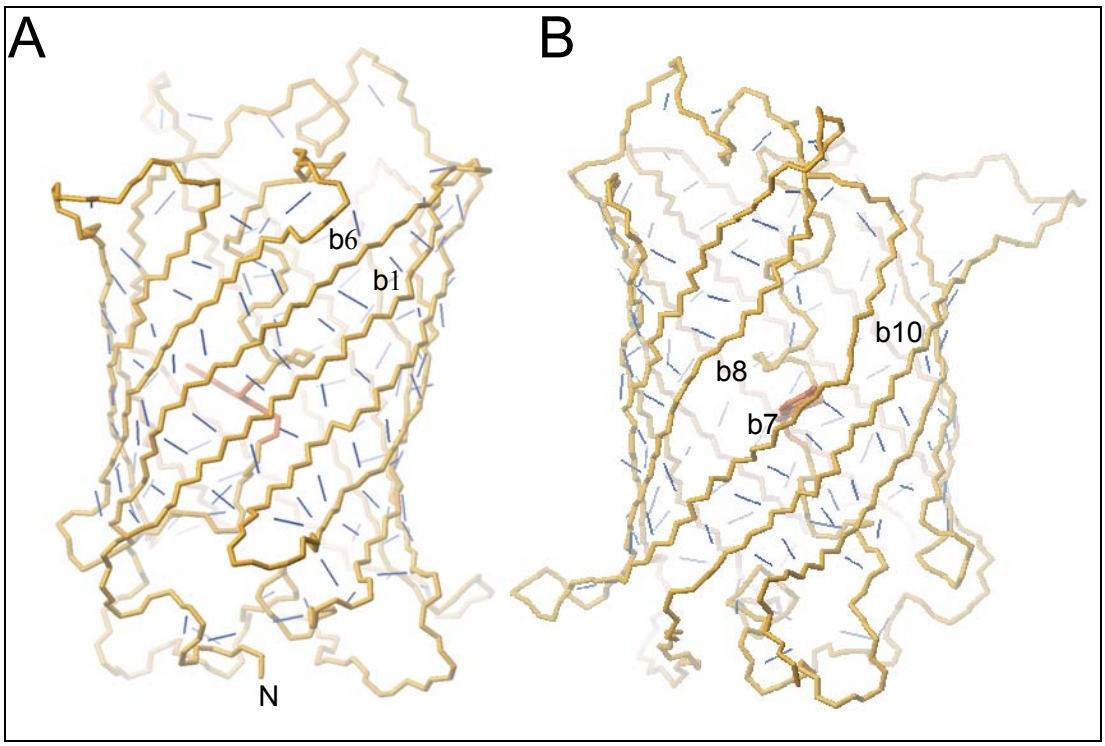

Fig. 4.10 *The hydrogen-bonding network (blue lines) on opposite sides of the GFP β-barrel is depicted in A. and B. based on the crystal structure coordinates from PBD entry 1b9c (Battistutta et al., 2000).*

Several indications underline the importance of dynamical processes around β-strand 7: Firstly, according to molecular dynamics simulations conformational fluctuations occur in the cleft between β-sheets 7 and 8, mainly connected to flipping of the Arg168 sidechain (Helms et al., 1999) at least on a nanosecond timescale. Secondly, ^{19}F NMR studies on the cyan variant of GFP showed that His148 is likely to be involved in a slow exchange process (Seifert et al., 2002). Finally, it has been speculated that conformational fluctuations are the origin of the pH sensitivity of the GFP chromophore (Haupts et al., 1998). For example, mutations at the positions 147, 149, 164-168, 202, and 220 (located in β-strands 7, 8, 10, and 11) have been shown to increase pH-sensitivity of the GFP chromophore (Miesenbock et al., 1998). This effect is supposed to be caused by an increased conformational flexibility around the chromophore. Therefore a dynamic process, e.g. an exchange between different backbone conformations, is most likely the origin of the increased hydrogen exchange rates, which may also explain the difficult resonance assignment in this part of the protein.

In summary, the observed HD rates indicate a higher degree of conformational fluctuations that disturb the inter-strand hydrogen-bonding network around β-strand 7. NMR methods exist to measure hydrogen bonds directly, e.g. via the $^{1}J_{NC}$ coupling constants, but the high molecular weight and aggregation tendency of GFP precludes the practical usuage of such sequences (Juranic et al., 1995 and 1996). From the point of view of fluorescence the observed conformational flexibility is not optimal for a high quantum yield. For example, such motions may allow fluorescence quenchers like oxygen to enter the β-barrel. In addition, the backbone flexibility may also affect the chromophore-stabilizing network and therefore increase the rate of photoisomerization. To further optimize fluorescence properties of GFPs by reducing conformational flexibility, clarification is needed whether this conformational flexibility is necessary for protein folding or if suitable mutations can reduce such structural instabilities. However, since the β-strand 7 is interrupted around residue 148 in most GFP structures and surprisingly also in the nidogen-1 structure where Tyr540 corresponds to His148, the presence of a predetermined breaking point at this point is suggested. It has been speculated that the closure of the cleft between β-strands 7 and 8 is one of the last events in the formation of the β-barrel upon folding of the protein (Helms et al., 1999). It is therefore more likely that the mentioned flexibility is an intrinsic property of the GFP fold that might result from its folding.

To investigate the relationship between structural properties of β-strands 7, 8, and 10, and GFP function (i.e. fluorescence), histidine 148 located in β-sheet 7 was mutated to glycine. The most apparent difference in the optical properties of GFPuv His148Gly compared to native GFPuv is the strong reduction of absorption at 478 nm, referred as band B arising from the anionic state of the chromophore. In addition, results here show that this mutation leads to the appearance of an allosteric slow exchange processes in the protein backbone. Therefore this mutation destabilizes the overall structure of the protein, leading either to an increase in the amplitude of low-frequency motions or a shift in time scale which makes the motions visible in the HSQC spectrum. The effect observed in GFPuv His148Gly has to clearly distinguish from the slow exchange process observed in CFP (Seifert et al., 2002), which takes place on a much shorter timescale (1.3 ms) compared to GFPuv His148Gly ($k_{ex} \approx 0.01$ - 0.1 s).

The most interesting question that arises from the disscused observations is whether a mechanistic link can be drawn from structural dynamics derived from NMR data to optical properties. First of all, since light absorption takes place on the femtosecond time scale – much faster than any structural process in proteins – UV absorption spectra display only an

instantaneous average of the immediate chromophore environment (Lakowicz, 1999). Secondly, for the interpretation of fluorescence data it is always necessary to infer the physiochemical behavior of the environment from measured photophysical parameters on the assumption that the photophysical responses are understood *a priori*, i.e. a plausible model of the chromophore and its environment is available (Moncrieffe et al., 2000). For GFP a model of photophysics has been derived from crystal structures in combination with time-resolved fluorescence measurements (Tsien, 1998; Prendergast, 1999). In the case of GFPuv His148Gly the main question is whether the observed strong reduction in absorption of the anionic state B – whereas the fluorescence emission is nearly unchanged – is related to the structural imhomogenity as seen in NMR spectra. The creation of the non-equilibrium excited I* state – which is supposed to give rise to the fluorescence emission – by deprotonation of the chromophore's phenolic oxygen seems to be quite unaffected by the mutation of His148Gly. However, the anionic state's B $\rightarrow$ B* transition is significantly less prominent which may be attributed to a shifted ratio of neutral to anionic states in the chromophore's thermodynamic equilibrium. So, it seems that the substitution of His148 with glycine and thereby removing a major part of the internal hydrogen bonding network affects only the state of chromophore protonation in equilibrium. The access of solvent molecules through the structural fluctuations in the β-barrel alters the immediate chromophore environment but still allows a fast and efficient creation of the fluorescent state by chromophore deprotonation. Further studies including time-resolved fluorescence spectroscopy and NMR relaxation measurements of the involved sidechains have to be conducted to unambiguously determine the pathways of proton transfer in this mutant.

4.2 ^{19}F NMR Spectroscopy of EGFP, ECFP and EYFP

4.2.1 Biological background

As mentioned already in *Chapter 2*, in contrast to many spectroscopic techniques, nuclear magnetic resonance (NMR) spectroscopy provides a large frequency range for studying dynamical processes from picosecond to second timescales and even longer at atomic resolutions. For example, motional and thermodynamical information for backbone amides in proteins can be obtained from measurements of ^{15}N relaxation rates in ^{15}N labeled proteins (Spyracopouloset al., 2001; Kay, 1998). Dynamic NMR spectroscopy (Sandström, 1982) is able to reveal populations and rate constants of exchange processes in proteins.

Further information, e.g. on the energy landscape of proteins, may be revealed by ^{17}O and ^{2}H NMR relaxation measurements (Denisov, 1996). Side-chain dynamics can be investigated by measurement of, for example, ^{13}C relaxation of methyl groups (Sandström, 1982; Palmer et al., 1999; Mulder et al., 2001) or ^{19}F relaxation (Hull & Sykes, 1974, 1975a, 1975b) in fluorine labeled aromatic rings. ^{19}F NMR spectroscopy offers the advantage over ^{1}H, ^{15}N and ^{13}C NMR that only specifically labeled sites give rise to an NMR signal which circumvents in many cases the assignment of highly crowded ^{1}H, ^{15}N and ^{13}C NMR spectra. In addition many proteins of known structure, which are too large (M > 30 kDa) for conventional multidimensional NMR studies, fall within the range of molecular weights accessible to ^{19}F NMR (M < 100 kDa) (Danielson & Falke, 1996). In addition it is usually assumed that fluorine substitution has negligible steric effects on the protein structure (Danielson & Falke, 1996), although indications exist that it may influence catalytic behaviour (Dominguez et al., 2001) and cause subtle differences in sidechain and backbone conformation (Xiao et al., 1998). Fluorine NMR has been used for ligand binding (Rong, et al., 1997) and folding studies (Bann, et al., 2002; Ropson & Frieden, 1992) and for probing conformational (Danielson & Falke, 1996; Ropson & Frieden, 1992, Kwon et al., 2002) and light-induced structural changes (Klein-Seetharaman et al., 1999).

Here the first quantitative ^{19}F NMR characterization of the conformational dynamics in the cyan fluorescent protein (CFP), a GFP mutant in which tyrosine 66 is replaced with a tryptophan is presented. The incorporation of different fluoro-tryptophans (i.e. 4-, 5-, and 6-fluoro-Trp) results in direct post-translational integration of the indole moiety and concomitantly ^{19}F nucleus into the CFP chromophore providing an almost ideal reporter of its dynamics. Using this approach it was possible to demonstrate that either the CFP chromophore itself or residues in its vicinity interconvert between multiple conformations on the time scale of milliseconds. These slow motions should therefore be considered in the interpretation of structural and spectroscopic properties of the ground states of GFPs.

4.2.2 Protein Expression and Purification

The incorporation of fluorinated analogues 4-fluoro-Trp (4FW), 5-fluoro-Trp (5FW) and 6-fluoro-Trp (6FW) into Clontech GFPuv (F99S, M153T, V163A), Clontech EGFP (F64L, S65T) and Clontech ECFP (F64L, S65T, Y66W, N146I, M153T, V163A) was achieved in *E. coli* Trp-auxotrophic strain ATCC49980 (Minks et al., 1999) using the selective pressure incorporation (SPI) method developed in our lab (Minks et al., 2000). The

protein expression host *E. coli* ATCC49980 was routinely co-transformed with two plasmids: ampicillin resistant pQE-30-PP4 harbouring the GFP and CFP gene sequence under control of the T5 promoter, and kanamycin resistant pREP4 containing a repressor gene *lac*I^q. The native and fluorinated proteins included a polyhistidine tag and were purified using Ni-chelate columns. The purity of the recombinant proteins was checked by SDS-PAGE. The quantitative replacement of the native Trp-residues by its non-canonical analogs was routinely confirmed by electrospray mass spectrometric analyses (ESI-MS, API III, Sciex Perkin Elmer) as described earlier (Minks et al., 1999). Before recording NMR spectra the proteins were dialysed against bi-distilled water and the samples were concentrated to 5-10 mg/ml. 10% D_2O was added to allow the field locking of the NMR spectrometer. All chemicals were purchased from Sigma or Aldrich unless stated otherwise.

4.2.3 Optical Spectroscopy

UV-absorption spectra of the proteins in the phosphate buffered saline (PBS: 115 mM NaCl, 8 mM KH_2PO_4, 16 mM Na_2HPO_4, pH 7.3) were recorded on a Perkin-Elmer Lambda 17 UV/VIS spectrophotometer. Molar extinction coefficients (ε_M) for native and fluorinated proteins were determined at room temperature (20 °C) according to the procedure described by Mach et al. (1992). Native and 6FW-EGFP have essentially the same absorbance (λ_{max}= 277 nm: ε_M = 21000 $M^{-1}cm^{-1}$; λ_{max}= 488 nm: ε_M = 38000 $M^{-1}cm^{-1}$). Molar extinction coefficients (ε_M) for native and substituted ECFP variants were different: native ECFP: λ_{max}= 278 nm: ε_M = 25000 $M^{-1}cm^{-1}$; λ_{max}= 434 nm: ε_M = 25500 $M^{-1}cm^{-1}$; 6FW-ECFP: λ_{max}= 280 nm: ε_M = 23300 $M^{-1}cm^{-1}$; λ_{max}= 430 nm: ε_M = 22900 $M^{-1}cm^{-1}$. Fluorescence spectra were recorded on a Perkin-Elmer spectrometer (LS50B) equipped with digital software. Protein samples (0.25 μM; slit 2.5 nm) were excited at 488 nm (EGFP) or at 450 nm (ECFP) and the emission spectra were recorded in the 490 – 540 nm (EGFP) or 460 – 540 nm (ECFP) range. Emission and excitation spectra of 6FW ECFP were measured at temperatures of 288K, 293K, 298K, 303K, 308K, 313K, 318K, and 323K.

4.2.4 NMR Spectroscopy

^{19}F NMR measurements were carried out on a Bruker DRX500 spectrometer equipped with a dual ^{1}H-^{19}F probehead. The fluorine Larmor frequency was 470 MHz. All measurements were performed at a temperature of 303 K with exception of six spectra taken

at temperatures between 293 K and 318 K. Due to the negative NOE enhancement factor of fluorine nuclei in proteins no proton decoupling was used. All ^{19}F NMR spectra were referenced to an external reference consisting of 10 mM TFA in 90% H_2O, 10% D_2O at a temperature of 303 K. For a typical one-dimensional ^{19}F NMR spectrum of 14 kHz sweep-width up to 16k scans with an interscan delay of 1 s were recorded. Baseline distortions caused by a relatively long prescan delay were corrected by manual baseline fitting.

The ^{1}H-^{19}F J-coupling constants in 4-, 5- and 6-fluoro-tryptophan were measured in ^{19}F NMR spectra of pure tryptophan samples. To evaluate the effect of J-coupling on the results of lineshape fitting experimentally, spectra of 6FW CFP with and without proton decoupling were recorded. Despite of a strong reduction of signal intensity upon decoupling no difference in lineshape of the W66 resonances was visible (data not shown). Therefore it was concluded that the calculated values for the rate constants are not affected by J-coupling.

pH titrations of the 4FW and 6FW labeled CFP sample were carried out by dissolving the protein in 50 mM borate buffer at pH 9.1 with 100 mM NaCl and then lowering the pH by adding the appropriate amount of 1% H_3PO_4 for each step. ^{19}F spectra at pH 9.1, 8.1 and 7.1 were recorded. In order to record a ^{19}F NMR spectrum of unfolded 6FW CFP, urea was added and heating to 95°C for 10 minutes was applied. To analyse the influence of protein concentration on ^{19}F NMR spectra a sample of 6FW ECFP was diluted in 4 steps to one sixteenth of the original concentration. The ^{19}F NMR spectra of the reference and of the diluted samples were recorded with 8k scans (reference, 8 mg/ml, 0.3 mM monomeric), 16k scans (4 mg/ml, 0.16 mM), 32k scans (2 mg/ml, 0.08 mM), 64k scans (1 mg/ml, 0.04 mM) and 128k scans (0.5 mg/ml, 0.02 mM). The effect of UV irradiation on 6FW ECFP was checked by recording ^{19}F NMR spectra after irradiating the NMR sampl with a 450 W Xe-lamp (~ 5 mW light power at 400 nm) for 4 hours.

4.2.5 Modelling of the CFP Structure

To evaluate the possible structural origins of exchange the structure of CFP was modelled based on the crystal structure coordinates of 4-amino-tryptophan CFP (Bae, personal communication, see also (Budisa et al., 2002)). In this CFP variant the two tryptophans (W57 and W66) are replaced by the non-canonical amino acid 4-amino-tryptophan. In order to construct the CFP model the 4-amino group of the tryptophan (W66) involved in chromophore formation was removed and a short run of energy minimization was performed using the modelling software package SYBYL (Tripos, Inc., St. Louis, MO). No

fluorine atom was modelled in the chromophore since the replacement of a hydrogen atom with a fluorine atom is supposed to have negligible steric effect (Danielson et al., 1996). To reveal gap regions around the chromophore the program SURFNET (Laskowski, 1995) was used. SURFNET generates gaps by fitting spheres between all pairs of atoms and computing three-dimensional density maps.

4.2.6 UV and Fluorescence Spectroscopy

As expected, the UV absorption profile as well as fluorescence spectra of EGFP (F64L, S65T) were not affected by fluorination of W57, which is placed about 15 Å apart from the chromophore. In fact, the single absorption peak at 488 nm of both EGFP variants as well as the emission maximum at 510 nm clearly indicate the anionic state of the chromophore (Fig. 4.11 A). Conversely, native ECFP (F64L, S65T, Y66W, N146I, M153T, V163A) exhibits two chromophore absorption bands (434 and 452 nm) and two distinct fluorescence emission maxima (476 and 500 nm). Fluorination of W57 and the chromophore indole induced 4 nm blue shift in the absorption maxima of 6FW-ECFP (430, 448 nm) as well as a decrease in the absorbance intensity by about 10 %. These effects are even more pronounced in fluorescence properties of 6FW-ECFP: while emission maxima are negligibly blue shifted by about 2 nm, the fluorescence intensity is lowered by about 50 % compared to the native ECFP (Fig. 4.11 B). It should be however noted that basic features of spectral behaviour of ECFP (i.e. the presence of two spectral bands in UV and fluorescence) are not changed upon fluorination. A detailed study on the effects of fluorination on the optical properties of GFP variants is reported by Budisa et al., (2002).

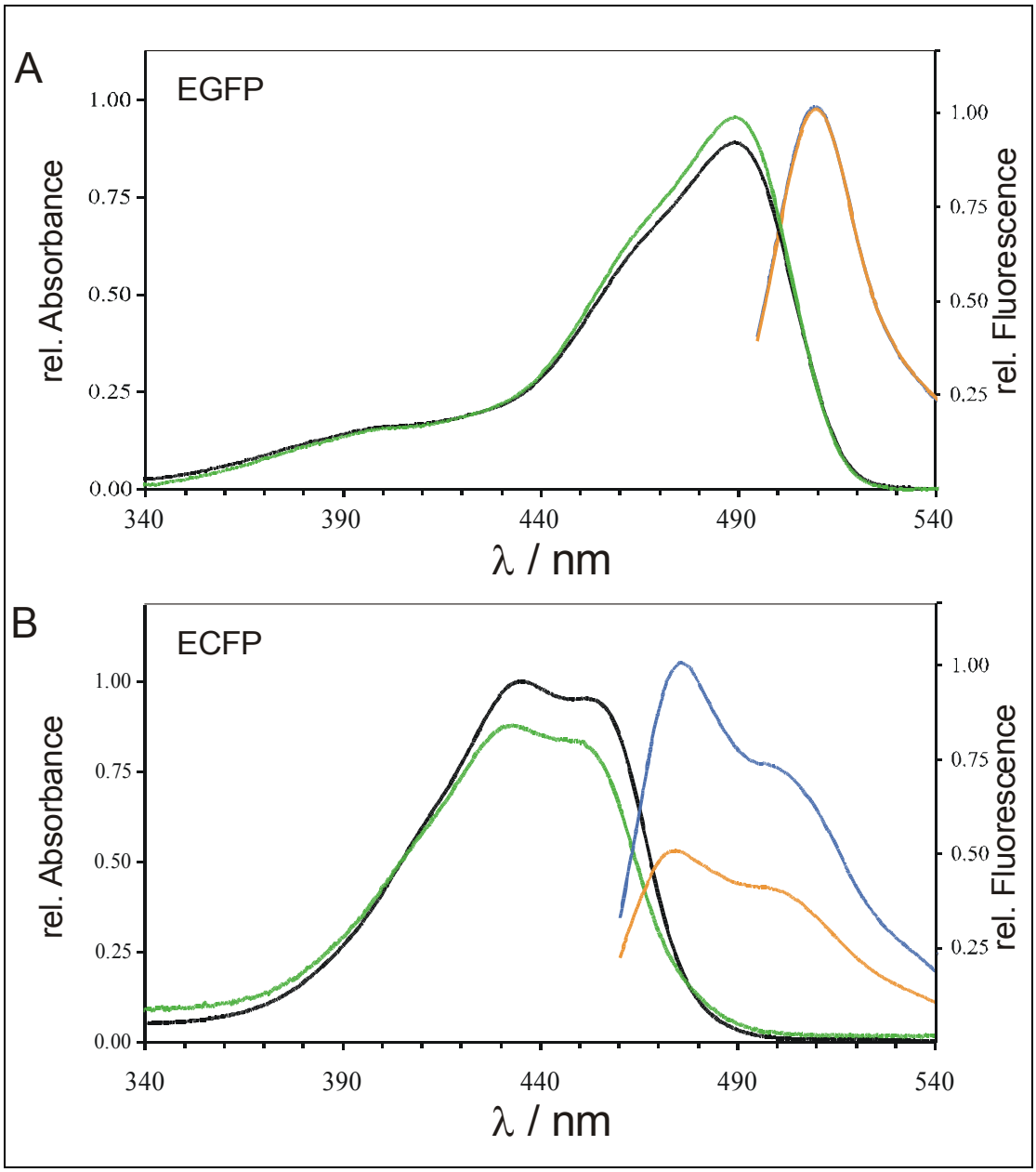

Fig. 4.11 *UV absorption and fluorescence emission spectra of EGFP (**A.** black and blue), 6FW EGFP (A, green and orange), ECFP (**B.** black and blue) and 6FW ECFP (**B**, green and orange) were recorded at neutral pH. Note that the absorption band around 395 nm characteristic for wild type GFP is suppressed in EGFP and its fluorinated counterpart. Only one absorption band at 488 nm and emission maxima at 510 nm are characteristic for the anionic state of chromophore in EGFP. On the other hand the UV-absorbance and fluorescence spectra for both native and fluorinated ECFP are characterised by two spectroscopically distinguishable states.*

4.2.7 NMR Assignment

The ^{19}F NMR resonance of free 4-fluoro-tryptophan is split into a quartet ($^{3}J_{FH}$ = 12 Hz, $^{5}J_{FH}$ = 4.6 Hz), whereas 5- and 6-fluoro-tryptophan show a sextet structure ($^{3}J_{FH}$ = 11 Hz, $^{5}J_{FH}$ = 5.5 Hz) (data not shown). These splittings are obscured in the broad protein signals when these tryptophans are incorporated into EGFP and ECFP proteins. The ^{19}F NMR spectra of ECFP and EGFP labeled with 4-, 5-, or 6-fluoro-tryptophan exhibited four and two resonances, respectively (Fig. 4.12). CFP contains only two tryptophan residues, W57 and W66, whereas GFP has only one tryptophan, W57. This clearly indicates that each of the two tryptophans in CFP exists in two states. The more and less populated states of W57 and W66 are marked with subscript A and B, respectively (Fig. 4.12). Since one of the CFP tryptophans, W66, is incorporated into the CFP chromophore during protein folding by cyclization of the backbone, two states exist within the chromophore. The resonances of 4FW and 6FW ECFP can be unambiguously assigned by comparison with the EGFP spectra. The assignment of the resonances of 5FW ECFP is complicated by the fact that a second peak is not visible in the 5FW EGFP spectrum. Since W66 and W57 are incorporated in the protein in a 1:1 ratio, the sum of integrals for both states of each residue must give the same value. The ratio of the integrals of states A and B for W57 and W66 are consistent within the 4FW and 5FW CFP samples ($W66_A/W66_B$ = 0.3 and $W57_A/W57_B$ = 0.2). 6FW CFP shows integral ratios of $W66_A/W66_B$ = 0.6 and $W57_A/W57_B$ = 0.6. The difference in chemical shift for the two states of W66 is 2.6 ppm in 4FW and 6FW ECFP. The ^{19}F NMR spectra of 4FW and 6FW GFPuv did not show double peaks (data not shown) indicating a different timescale of exchange.

The appearance of two peaks for each tryptophan in the spectra of ECFP and EGFP and the different broadening of those peaks provided a first indication for the existence of slow exchange processes between different states of the tryptophan sidechains. This is especially important since it gives a strong hint that a slow exchange process exists within the chromophore of CFP.

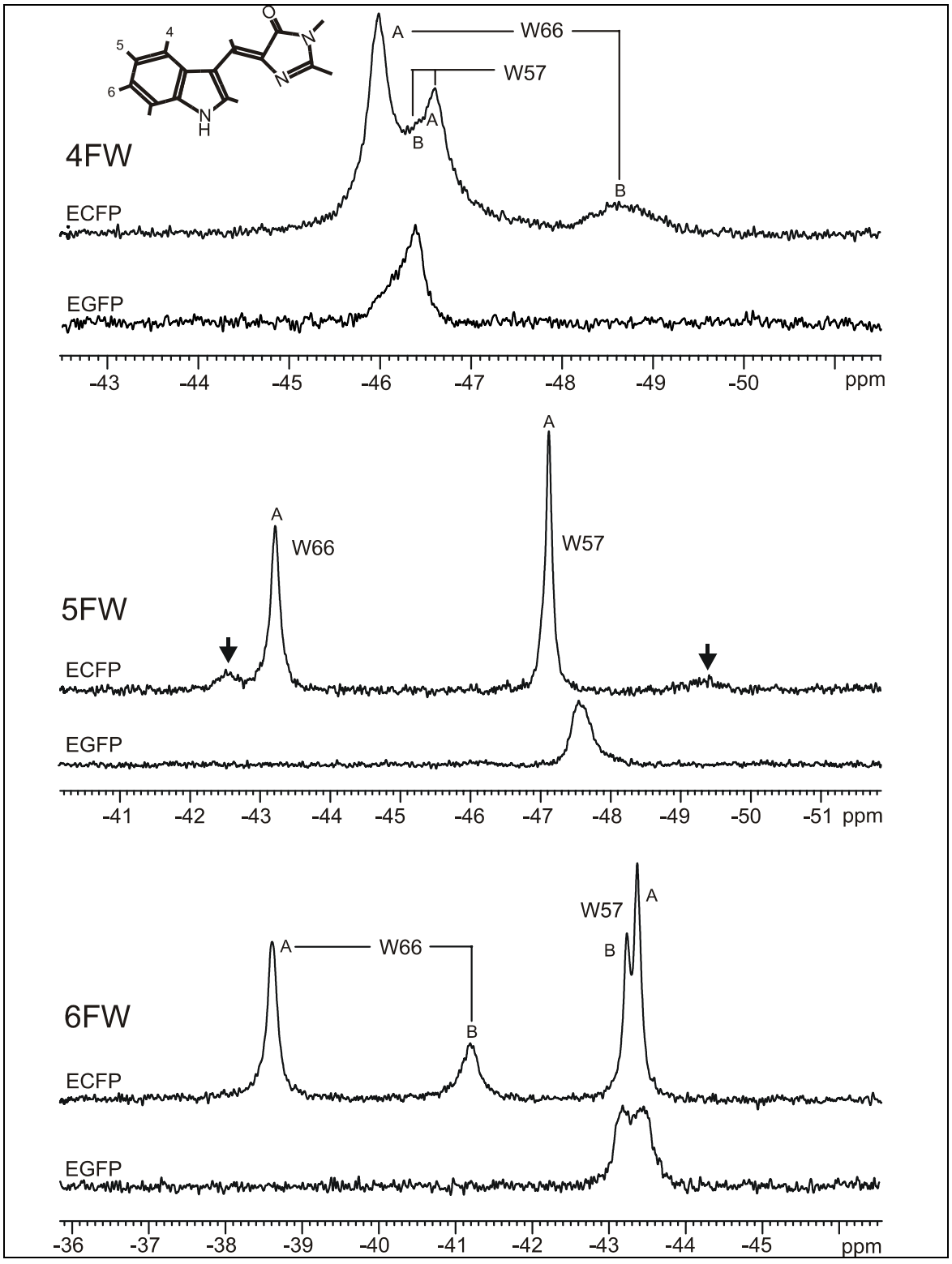

Fig. 4.12 *^{19}F NMR spectra of ECFP and EGFP labeled with 4FW, 5FW and 6FW: The chromophore of CFP and the fluorinated sites are shown in the upper left corner. The spectra were recorded at a temperature of 303K without proton decoupling and referenced relative to TFA. An exponential line broadening of 3 Hz was used to increase signal-to-noise ratio. The*

arrows in the spectrum of 5FW ECFP indicate two resonances, which were assigned by comparison of peak integrals (see text).

4.2.8 Thermodynamic Analysis

The temperature dependence of ^{19}F NMR spectra of 6FW ECFP provides additional evidence for the existence of slow exchange processes (data not shown). These dynamic NMR measurements show that the exchange rate and state populations change with temperature. The mean lifetime of the W66 states was found to be $\tau = 1/k_{ex} = 1.2$ ms to 1.4 ms for temperatures in the range of 293 K to 318 K. The rate of interconversion from $W66_A$ to $W66_B$ (k_A) decreases with temperature, whereas the rate of interconversion from $W66_B$ to $W66_A$ (k_B) increases exponentially with temperature.

4.2.9 Influence of Denaturation, pH, Protein Concentration and Irradiation with UV Light

To clarify the nature of the two states of ECFP found by NMR, the effect of denaturation was examined (Fig. 4.13). The fluorine resonances experience a pronounced upfield shift of 3 to 5 ppm upon protein unfolding. This can be attributed to the change of environment from the hydrophobic core of GFP to the hydrophilic solution. The assignment of the resonances of the unfolded protein was based on the chemical difference of W57 and W66. One resonance has exactly the chemical shift of pure 6-fluoro-tryptophan in solution. This resonance is attributed to W57. The indole of W66 is bound to an imidazolinone ring and should therefore exhibit a different chemical shift for the fluorine atom at the 6 position. Therefore the resonance at -43.4 ppm is assigned to W66. Despite unfolding the W57 shows two distinct resonances. This is attributed to proline *cis* and *trans* conformations in the P56-W57-P58 sequence. In the vicinity of the W66 resonance a small peak is detectable whose origin however could not be established without being too speculative. The resonances of unfolded CFP are significantly narrower compared to folded CFP indicating higher flexibility and the absence of exchange broadening. The “ground state” $W66_B$ in folded CFP is always shifted upfield compared to the “excited state” $W66_A$. But due to several significant contributions to ^{19}F chemical shift (Lau et al., 2000) this observation cannot be attributed to a specific origin at the moment.

pH titration with 6FW and 4FW ECFP (data not shown) showed no significant changes for W66 resonances but a line narrowing can be observed for the W57 resonances. Protein concentration has no effect on the two states observed by ^{19}F NMR in the 6FW ECFP sample (see Fig. 4.14). The dilution spectra show that the two states of W66 and W57 cannot be attributed to monomer-dimer states of ECFP. No significant changes in ^{19}F NMR spectra of 6FW ECFP after irradiation with UV light for four hours could be detected.

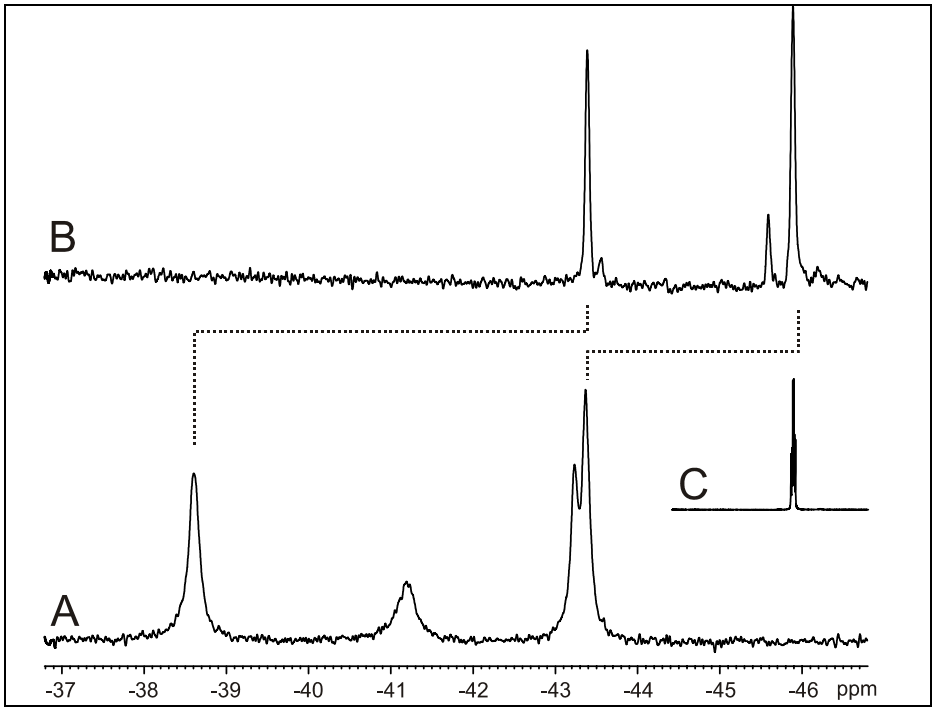

Fig. 4.13 *^{19}F NMR spectra of folded 6FW ECFP (**A**), unfolded 6FW ECFP (**B**) and 50mM 6FW (**C**). Even unfolded CFP clearly shows double peaks for the W57 resonance. For the W66 only a very small second peak is observed.*

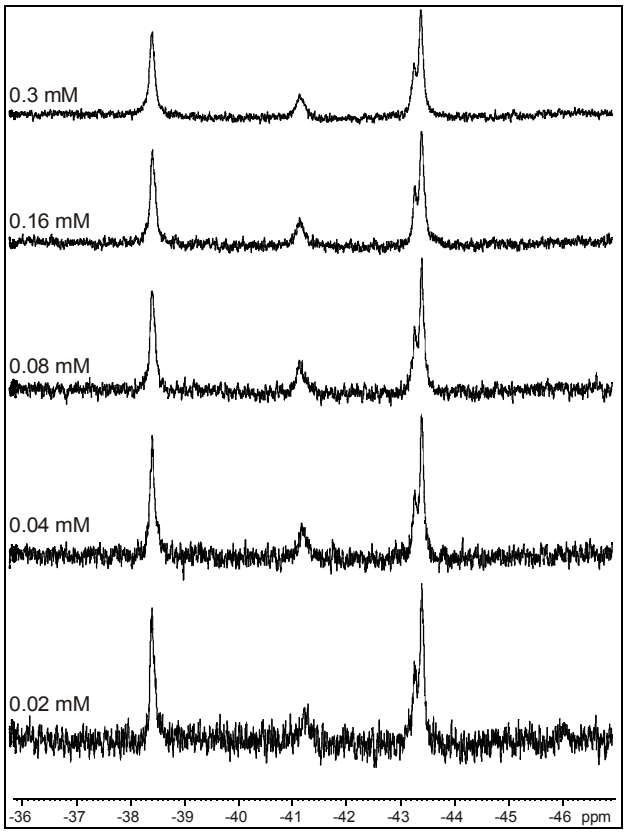

Fig. 4.14 *Concentration dependence in ^{19}F NMR spectra of 6FW ECFP: protein concentration does not affect the exchanging states in ECFP significantly. This excludes dimerization as source of the observed states. The concentration is given as the total concentration of monomeric protein (27 kDa). The decrease in signal-to-noise ratio results from the linear increase of signal intensity with the number of scans per spectrum but a quadratic increase of noise level.*

4.2.10 Discussion

In the last three decades ^{19}F-NMR studies have taken advantage of high sensitivity of ^{19}F chemical shifts to the changes in local environments including van der Waals packing interactions and local electrostatic fields with no background signals from solvent. In most cases, crystallographic studies confirmed that structures of fluorinated proteins are indistinguishable from those of wild-type protein forms (Minks et al., 2000). This makes ^{19}F NMR a powerful technique as it allows to probe subtle conformational changes and to reveal intrinsic dynamic features of the proteins. Aromatic residues are especially suitable targets for fluorination because they are normally less likely to be involved in internal motions than smaller amino acids, thus facilitating chemical shift calculations. In this context, the residue

specific labelling of EGFP and ECFP with fluorine containing tryptophan analogues offers a good model for NMR studies as well as the opportunity to study possible effects of fluorination on the dynamic properties of protein chromophore.

Presented spectroscopic investigations of substituted proteins at neutral pH clearly indicate that incorporation of electron-withdrawing fluorine in the environment (i.e. W57) of the EGFP chromophore does not affect absorbance and steady-state fluorescence properties of this molecule (Fig. 4.11). On the other hand, additional fluorination of W66 of ECFP provides direct integration of the fluorine nucleus into the chromophore. Therefore, the limited blue shift in both fluorescence and absorbance as well as the dramatic decrease in fluorescence intensity upon fluorination can be attributed to this H→F substitution. However, it should be kept in mind that the general behaviour of ECFP is not substantially changed by fluorination since two states of chromophore seen as absorption bands at around 430 nm and 450 nm as well as in fluorescence are preserved (Fig 4.11). Thus these spectral properties represent relatively static properties of the GFP chromophore whose dynamic behaviour can certainly be better dissected using ^{19}F NMR.

Since light absorption takes place on a femtosecond timescale – much faster than any structural process in proteins - the information on structural dynamics of proteins obtained by UV absorption spectroscopy is limited to an instantaneous average of the immediate environment of the absorbing moiety on this timescale (Laskowicz, 1999). Compared to the UV spectroscopy NMR is able to characterize motions on a much broader range of timescales. Each of the ECFP residues W57 and W66 exhibits two ^{19}F NMR resonances, which is a clear sign for slow exchange processes. For both tryptophans, W66 and W57, the very low f = NOE - 1 values exclude the existence of pronounced motions on the picosecond to nanosecond timescale. Excessive motions on timescales shorter than picoseconds are also excluded by the small values of T_1. ^{19}F linewidths indicate the existence of motions on a time scale of nanoseconds and longer. The temperature dependence and denaturation behaviour observed in the ^{19}F NMR spectra of 6FW ECFP provides direct evidence for the existence of slow exchange processes. The mean lifetime of the W66 states was found to be 1.2 ms to 1.4 ms for temperatures in the range of 293 K to 318 K. Slow conformational exchange between two ground states of the chromophore is also in-line with the existence of two UV absorption and fluorescence peaks for CFP and their temperature dependence. The pH titration suggests that the two exchanging states of the chromophore in ECFP cannot be attributed to different protonation states of the indole nitrogen. This leads to the conclusion that the two states of the

CFP chromophore are formed in a different way compared to GFP where the protonation state of the phenolic oxygen determines the state of the chromophore.

Both tryptophans in ECFP show double peaks. Therefore, it could be thought first that this double peak behaviour may be general for tryptophans, but the ^{19}F NMR spectrum of GFPuv showed only a single peak for W57. The more probable explanation of double signals for both tryptophans in the NMR spectrum of ECFP is that the signals of W57 originate from different processes than those of W66, as the resonances of W57 can be easily explained by the exchange between *cis* and *trans* proline conformation (Wüthrich, 1994) in the P56-W57-P58 sequence. Whereas an exchange process of different origin causes the resonances of the chromophore tryptophan W66. In order to determine the exchange process the steric interactions in the chromophore cavity of CFP has to be examined for which a crystal structure would be desirable. Since attempts to crystallize fluorinated CFP were not yet successful the structure CFP was modelled based on X-ray data from the closely related 4-amino-tryptophan CFP (Budisa et al., 2002). It is well known that even minor substitutions in the core of proteins can cause substantial structural rearrangements therefore the model may give only a first impression on the steric interactions going on in CFP; nevertheless, this provides substantial aid for interpreting the NMR data.

In principle, three possibilities for the exchange process within CFP exist that may explain experimental data: (i) the two states correspond to the monomer-dimer states of CFP, (ii) they are related to the exchanging ground states of the chromophore itself or (iii) sidechains in the vicinity of the chromophore exchanges between two conformations.

The explanation that the two states are simply the monomer and dimer of the protein is the least plausible interpretation. Although GFPs are known to dimerize, the dissociation constant of wild-type GFP is approximately K_D = 0.1 mM as measured by analytical ultracentrifugation (Phillips, 1998; Ward et al., 1998). The K_D indicates weak binding for the dimer formation; in addition the symmetric dimers seen in the X-ray structures are rather loose without any interleaved monomer units (Yang et al., 1996) so that even if GFP formed a tight complex the immediate atomic environment, which to the first approximation is responsible for the difference in the chemical shift, would be almost identical for a given nucleus in the monomer and dimer unit. This is in strong contrast to the large difference in ^{19}F chemical shift ($\Delta\delta$ = 2.6 ppm) between the two states, $W66_A$ and $W66_B$. An example of NMR characterization of a tight nM range dimer, with interleaved monomers, is provided by Rous sarcoma virus protease (Schatz et al., 2001). Even in this case, a single set of NMR peaks was

observed in the ^{1}H-^{15}N correlation spectrum, indicating the monomer-dimer equilibrium to be in fast exchange with respect to the NMR time scale.

Aromatic ring flipping is a well-known phenomenon found in proteins and to evaluate the possibility of the CFP chromophore being directly involved in such an exchange process, an examination of the known GFP structures is useful. In GFP, the chromophore is stabilized by a hydrogen bond network, which mainly involves its phenolic hydroxyl group and H148 and T203. Despite tight packing in the chromophore cavity it was shown by molecular dynamic simulations that concerted rotations around the τ and φ angles, so-called hula-twists, are a possible pathway of photoisomerization of the GFP chromophore since electronic excitation reduces the energy barrier for these rotations (Weber et al., 1999; Chen et al., 2001). Without electronic excitation the hula-twist will occur only with a reduced rate.

The model of CFP structure predicts an overall structure similar to the well-known GFP structures, as expected. In CFP, a rotation of the chromophore around the φ axis is very unlikely since this would cause a rotation of the long axis of the indole, which is hindered by surrounding residues. But a rotation around the τ axis in concert with an in-plane rotation of the imidazolinone ring appears to be possible (Fig. 4.15), since both rotamers τ=0° and τ=180° fit into the same chromophore cavity.

Weber et al. showed that the *cis* and *trans* isomers of the same protonation state in GFP are similar in their heats of formation (< 5 kJ/mol) (Weber et al., 1999). Applied to ECFP, this is in agreement with the small energy difference determined by NMR. Weber et al. also calculated the energy barriers for bond rotations in the ground state of the GFP chromophore (Weber et al., 1999). Most of them are in the range of 100 - 200 kJ/mol, except e.g. the φ rotation in the neutral and anionic state or the τ rotation in a zwitterionic state, which have lower energy barriers. Since the free energy of activation for the exchange process was estimated to be in the range of 57-63 kJ/mol, the rotation around one of the bonds connecting the two rings of the chromophore appears to be feasible for the indole-imidazolinone chromophore of CFP. One might argue that in CFP a hydrogen bond between the indole amide and T203 may exist. But the breaking of a hydrogen bond introduces an energy barrier of typically 20 kJ/mol, which is only a fraction of the 60 kJ/mol estimated as the height of the overall energy barrier, therefore not prohibitive for the proposed exchange process.

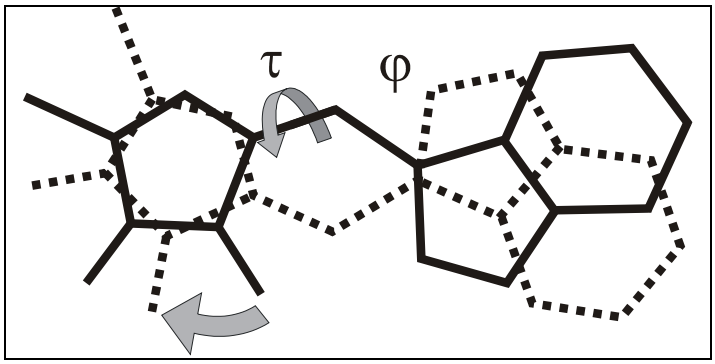

Fig. 4.15 *The possible rotamer conformations of the CFP chromophore are illustrated. A rotation around τ is accompanied by an in-plane rotation of the imidazolinone ring.*

Another possible reason for slow motion processes is the exchange between different conformations of neighbouring sidechains. In GFP, two residues, H148 and T203, are closest to the phenolic ring of the chromophore. In most of the crystal structures of GFPs, the β-barrel is somewhat perturbed around H148 with residues 144 to 150 not being hydrogen bonded to the adjacent backbone residues 165 to 170 (Wachter et al., 1998). This suggests that H148 has an increased flexibility compared to other residues within the β-barrel. The comparison of the CFP model structure with the GFP structure suggests that it is sterically feasible for H148 in CFP to adopt two conformations: one with the H148 sidechain oriented towards the chromophore like in GFP and another one with H148 pointing towards the solvent (see Fig. 4.16).

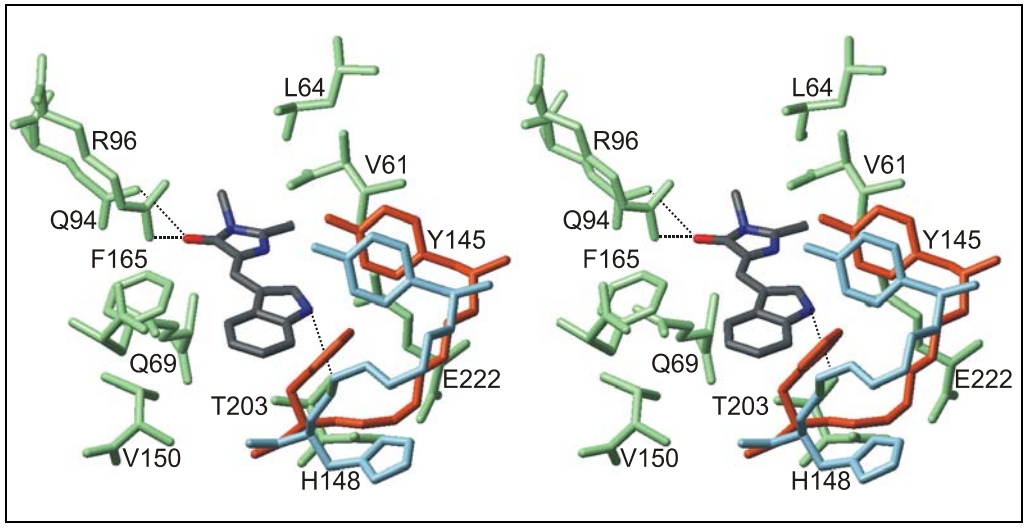

Fig. 4.16 *Stereo view of the chromophore environment in the CFP model structure: Residues surrounding the chromophore are shown in green. The chromophore itself is shown as a stick representation; the carbon, nitrogen and oxygen atoms are depicted in black, blue and red,*

respectively. The comparison of the conformation around residue H148 for GFP (in this case taken from PDB entry 1emm (Palm et al., 1997)) and the CFP model is shown in orange and cyan, respectively. Close inspection of both structures suggests that the conformation present in GFP may also be adopted in CFP. Hydrogen bonds are depicted as dotted lines.

The different conformations are expected to introduce distortions in the backbone conformation of the adjoining residues. The hydrogen bond between H148 and the chromophore is removed in CFP, at least in one conformation. If the histidine sidechain flips from a chromophore-neighbouring conformation to a solvent-exposed conformation, the chemical shift of the fluorine atoms at the 4-, 5- or 6-position of the indole will be affected by the change in electronic environment. A flipped H148 sidechain opens a channel to the solvent and additionally gives the chromophore a larger degree of motional freedom. But the NMR pH-titration clearly shows that the chromophore is not significantly affected by pH changes. Either surrounding hydrophobic residues hinder the entry of water molecules or the imino nitrogen is involved in a hydrogen bond and therefore does not titrate in the measured range of pH. Assuming that the solvent exposed conformation of H148 is higher in energy, the increased disorder associated with H148 flipping from its position in the protein core to a more flexible, solvent-exposed position may account for the observed enthalpy-entropy compensation. Taken together the exchange of H148 between two conformational states is the most probable explanation for our NMR observations, but more X-ray studies are necessary to provide direct evidence.

For applications in molecular biology a higher quantum yield of CFP fluorescence is desirable. To achieve an enhancement of the fluorescence the influence of the slow exchange process has to be reduced. Therefore replacement of H148 by an amino acid that (i) is not able to get in contact with the chromophore and (ii) stabilizes the backbone conformation should enhance the fluorescence of CFP. If the chromophore itself exchanges between different conformations, it has to be stabilized. For example, the mutation of T203 in CFP to an aromatic amino acid may lead to the formation of a stacked π-electron system like in YFP, which is likely to reduce slow exchange motions within the chromophore and therefore may increase the quantum yield of fluorescence.

4.3 Crystal Structure of Red Fluorescence Protein (DsRed, drFP, RFP) from the coral *Discosoma striata*

4.3.1 Biological background

Matz et al. (1999) identified a protein based on the similarity of its sequence to that of the „green fluorescent protein“ but which glowed at a bright red. This protein is called the red fluorescent protein (DsRed, drFP583, RFP, names used equivalently) and is responsible for the red coloration around the oral disk of a coral of the *Discosoma striata* genus. DsRed, is a 28-kDa polypeptide and has the same chromophore as GFP. Mature DsRed has an emission maximum of 583 nm and an absorbance maximum of 558 nm but for immature DsRed emission is at 500 nm and excitation at 475 nm. DsRed shows also slow kinetics of maturation proceeding via the green fluorescent protein (GFP)-like green intermediate to the final red species. Maturation takes from hours to days and may be incomplete. In solution the protein forms tetramers and two of these may weakly associate into an octamer. They have an extremely similar fold to that of the *Aequora victoria* GFP. Each monomer consists of an 11-stranded β barrel with a coaxial helix. It is formed from an internal tripeptide Gln-Tyr-Gly (residues 66-68). However amino acid sequence homology is low (only about 23 %), several amino acids in the immediate vicinity of the chromophore are strictly conserved e. g., Glu-215 and Arg-95 in RFP corresponding to Glu-222 and Arg-96 in GFP. These are probably essential for chromophore formation. In the electron density of the crystal, which was solved with a 2.8 Å resolution an open form of the chromophore was observed before maturation. This is, as published by Yarbrough et al. (2001), because of a mixture of different forms of the chromophore. Immature green species contain a *cis*-peptide bond between residues Phe-65 and Gln-66 and display two predominant conformations of the side chain of Ser-69. As a general mechanism, the rate-limiting step in generating red fluorescence may be deprotonation of the α carbon of Gln-66 by Ser-69, generating a reactive carbanion, or attack by dioxygen via a superoxide radical anion intermediate. The oxidation reaction would then proceed via a hydroperoxide adduct and result in the release of hydrogen peroxide on completion. The geometry associated with the α-carbon of Gln-66 was not restrained during refinement, and all four of the chromophores of the teramer clearly adopted a planar geometry that is consistent with the sp^2 hybridization of this center. The β-carbon and main-chain nitrogen of Gln-66 and the carbonyl carbon of Phe-65 are coplanar with the rest of the chromophore. Gross et al. (1999), based on theoretical calculations, predicted that despite its

apparent conjugation with the rest of the chromophore the C=O bond of Phe-65 is almost perpendicular to the plane of the chromophore. Yarbrough suggested also that Glu-215 plays two roles in the maturation chemistry of DsRed: it is important in the formation of the green intermediate and the stereochemical environment of Glu-215 is crucial for the formation of both the green- and red-emitting species. For maintaining Glu-215 as an anion, Lys-70 would be important, consistent with the failure of the Lys70Met mutant to mature beyond the green emitting state. In GFP Ser-69 is replaced by Val-68; Glu-222 is positioned differently with respect to the chromophore, and the 64-65 peptide bond is *trans*, all of which would prevent further maturation in case of GFP.

The new structural results suggest that red fluorescence appears to be a result of an extremely subtle interplay of a more highly evolved and optimised assemblage of active side residues than is the case for GFP.

4.3.2 Protein Expression and Purification

DsRed was expressed in *Escherichia coli*. Expression was controlled by the T7 promotor (and resistance against ampicillin) and provided by the gene for β-lactamase. His_6-tagged DsRed was overexpressed using the plasmid pt7RFPav (3500 bp) in BL21DE3 cells in 1 liter of LB-medium after being induced with 0.5 mM isopropylthiogalactoside (IPTG) for 6 h at 37°C. Cell paste was resuspended in 50 ml lysis buffer (300 mM NaCl, 50 mM Na_2HPO_4, pH 8.0) with 1mg DNAse, 1 mg RNAse and 1 mg $MgCl_2$ and 0.1 mM phenylmethylsulfonyl fluoride (PMSF) and then sonicated with Micro-Tip and centrifuged for 45 min at 80 000 rpm in a Beckman centrifuge. The protein was purified by Ni^{2+} affinity chromatography over Ni-NTA agarose (Qiagen). For the binding to the resin the supernatant was incubated with Ni-NTA agarose for 1h at 4 °C, poured onto a column and subsequently washed with 100 ml lysis buffer. The protein was eluted from the Ni-NTA resin with 20 ml elution buffer (300 mM NaCl, 50 mM $Na_2HPO_{4,n}$ 500 mM imidazole), pH 8.0. Then the protein was dialysed against the PBS buffer (115 mM NaCl, 8mM KH_2PO_4, 20 mM Na_2HPO_4, pH 7.0). DsRed was concentrated to 15-20 mg ml^{-1} using a Centricon 10 (Millipore). Then gel filtration was carried out using a Superdex 75 column (Amersham Biosciences). After this the sample was again concentrated up to concentration 20-25 mg ml^{-1} and flash frozen in liquid nitrogen and stored at –80 °C in small aliquots for crystallization trials.

4.3.3 Crystallization

Crystals were grown in 1-2 weeks at 4 °C by hanging drop vapor diffusion against 20 % PEG 8000, 0.1 M sodium cacodylate (pH 6.5) and 0.2 magnesium acetate tetrahydrate. Drops were 2 µl of protein solution and 2 µl of well solution, macroseeding was carried out 24 h after wells were set up and larger crystals could be produced after few days (Fig. 4.17).

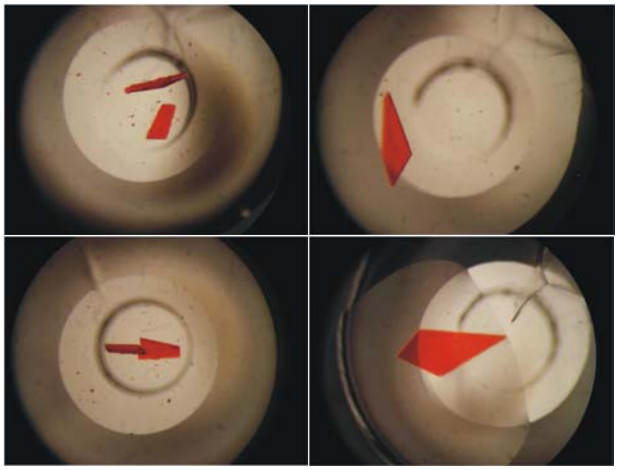

Fig. 4.17 *Crystals of DsRed, the picture is done without polarization filter.*

Crystals that reached a maximum size (0.4 × 0.3 × 0.5 mm) were harvested using the reservoir solution and mounted in a thin-walled glass capillary tube. A complete data set up to 2.8 Å was collected at room temperature "in-house" using a 300 mm Mar Research image plate detector. The crystal belonged to space group $P2_1$ with cell constants a = 55.8 Å, b = 124.6 Å, c = 62.3 Å, β = 94.5˚ and contained four molecules per asymmetric unit. Diffraction data (see Table 4.1) was evaluated, scaled and reduced with DENZO-SCALEPACK (Otwinowski, 1993).

4.3.4 Structure Determination

The crystal structure was solved by the molecular replacement with CNS (Brünger et al., 1998), using the structure of *Discosoma striata* DsRed (PDB accession code 1ggx) as a search model. The diffraction data from 20 Å to 2.8 Å was used to calculate the rotation and translation functions. The correlation factor and the R-factor of the best solution, after rigid body refinement, was 25.5 % and 51.3 %, respectively. The model building was performed by iterative rounds in MAIN (Turk, 1992) and refined using CNS (Brünger et al., 1998). After several cycles of model building, conjugate gradient minimization and simulated annealing, a structure with good stereochemistry was obtained. A Ramachandran plot calculated using the

program PROCHECK (Laskowski et al., 1993) shows that all the residues fall into the most favoured 89.5%) or additionally allowed regions (10.5%)(Fig. 4.18)

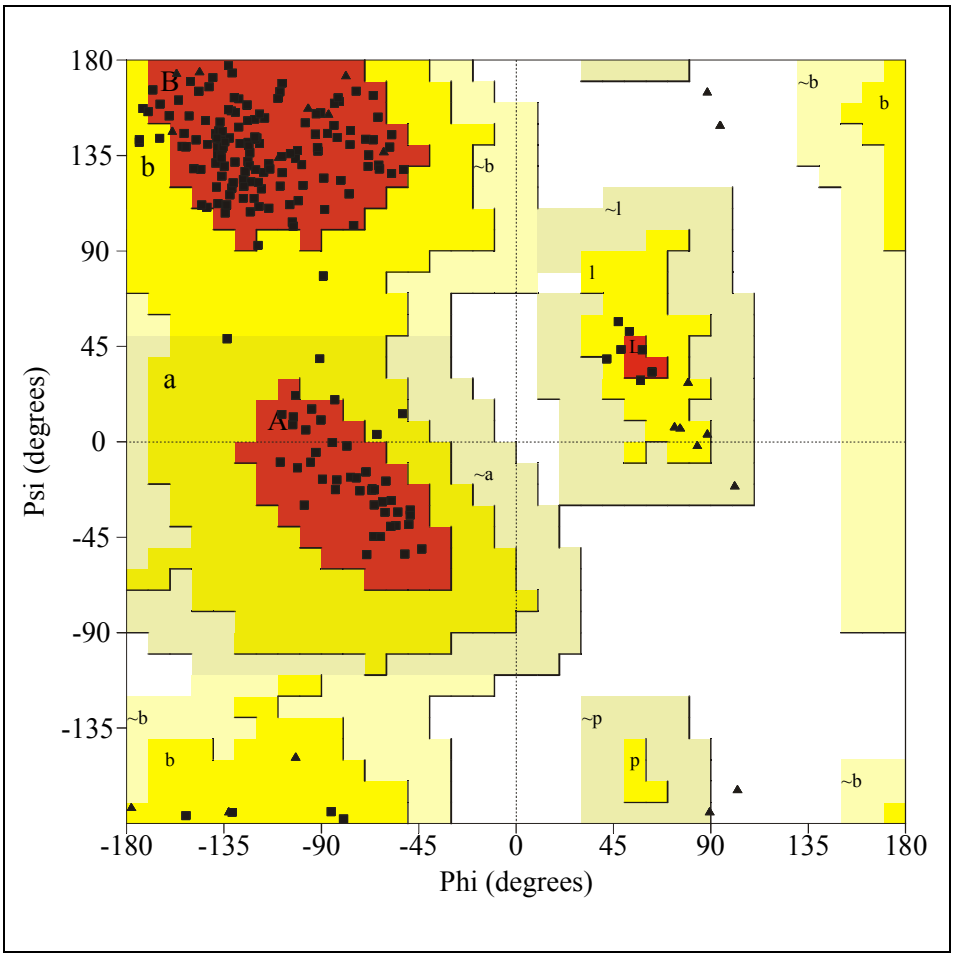

Fig. 4.18 *The Ramachandran plot for the refined DsRed structure.*

Table 4.1 *X-ray Data and Refinement Statistics*

Space group	$P2_1$
Crystal cell constants	
a	55.8 Å
b	124,6 Å
c	62,3 Å
β	94.5 °
Limiting resolution (Å)	2.8
Total number of reflections measured	280341
Unique reflections (20 - 2.8 Å)	70236
R_{merge}[1] (20 – 2.8 Å)	10,2 %

R_{merge}[1] (2.9 – 2.8 Å)	36,6 %
Completeness (%),(20 – 2.8 Å)	91,3
Completeness (%),(2.9 – 2.8 Å)	92,8
Solvent molecules	418
Reflection used for refinement	65789
Sigma cut-off for refinement	2.0
Resolution range used	20 - 2.8 Å
R-value[2], overall	19.01 %
R_{free}[3]	26,7 %
R.m.s. standard deviations:	
Bond length	0.02 Å
Bond angles	1.65 °

R_{merge}[1]= $[\Sigma_h\Sigma_i \mid I(h,i) - \langle I(h)\rangle \mid / \Sigma_h\Sigma_i \mid I(h,i)] \times 100$, where I(h,i) is the intensity value of the i-th measurement of h and <I(h)> is the corresponding mean value of h for all i measurements of h. The summation is over all measurements.

R-value[2] = $(\Sigma \mid Fo - Fc \mid / \Sigma Fo) \times 100$

R_{free}[3] was calculated randomly omitting 7% of the observed reflections from refinement and R-factor calculation.

4.3.5 Description of the Structure

DsRed is an 11-stranded β-can with a central α-helix and α-helikal caps at the barrel ends, nearly identical in topology to the homologous GFP (see Fig. 4.19). Four non-crystallographically related molecules in the *P1* crystals form a tightly packed tetramer with orthogonal 222 symmetry through two extensive protein interactions surfaces. As is the case in GFP, the DsRed chromophore rests in the middle of α-helix that runs through the center of the β-can and clearly shares much of the chemical mechanism of formation of the chromophore with GFP, namely autocatalytic cyclization of residues Gln 66, Tyr 67 and Gly 68 and dehydrogenation of the Cα-Cβ bond of Tyr 67. The majority of the backbone structural variations between DsRed and GFP occur in loop regions that form the ends of the β-can structure, but structural changes are only minor, as can be seen from the overlay of the two models (Fig.4.19)

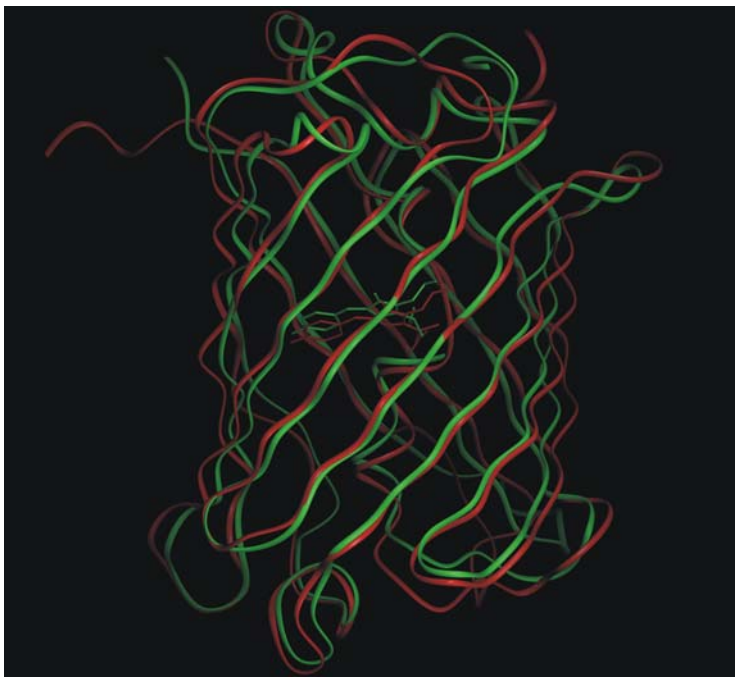

Fig. 4.19 *Superposition of structures of GFPuv and DsRed. The fold of both proteins is very similar despite low (~ 23 %) amino acid sequence homology.*

The DsRed has a more compact hydrophobic loop (residues 201-209), three residues shorter, which packs tightly against the protein core. This is because of an equivalent to Pro-211 is missing, which in avGFP forced the loop into an extended conformation. In the case of avGFP there is a long loop composed primarily of charged and polar residues extending far from the body of the protein. A number of concerted changes in the structure of DsRed force main chain residues 197-220 and 140-145 closer to the chromophore. For both proteins in their highly symmetric structures even some irregularities were conserved. A good example is the region of the seventh β-strand centred on residue Ser-146 in DsRed. In avGFP the corresponding residue is His-148. The side-chain of His-148 occupies the space between strands and partially satisfies the backbone hydrogen bonding potential in this region, however it is not required for a bulge formation (Wachter et al., 1998). In DsRed, the bulge is less pronounced and strands 7 and 8 form hydrogen bonds on either side of residues 145-148. The reestablishment of hydrogen bonding on the carboxyl-terminal side of residue 148 moves the end of strand 7 closer to strand 8 and away from strand 10 and opens a cleft in the surface of the molecule. This phenomenom takes place also in avGFP but is not as strongly pronounced.

4.3.6 Oligomerization

Oligomerization of DsRed occurs through the two protein interfaces, which differ dramatically in chemical character, suggesting the possibility of fluorescence resonance transfer (FRET) between chromophores within the tetramer. This was confirmed by time-resolved anisotropy experiments, which indicated a rapid phase depolarization in DsRed. The physical basis for FRET is the dipole-dipole coupling between the donor and acceptor chromophores. The probability of the energy transfer depends on i) the spectral overlap between donor emission and acceptor absorbances, ii) relative angular displacement and iii) distance between the respective dipoles.
Figure 4.20 shows the relative orientation of the chromophores within the DsRed tetramer and Table 4.2 the distances between them.

Table 4.2 *Distances between chromophores within tetramer*

Chromophore pairs	A–B C-D	A-C B-D	A –D B-C
Distance (Å)	22	38	43
Angular orientation (˚)	21	47	41

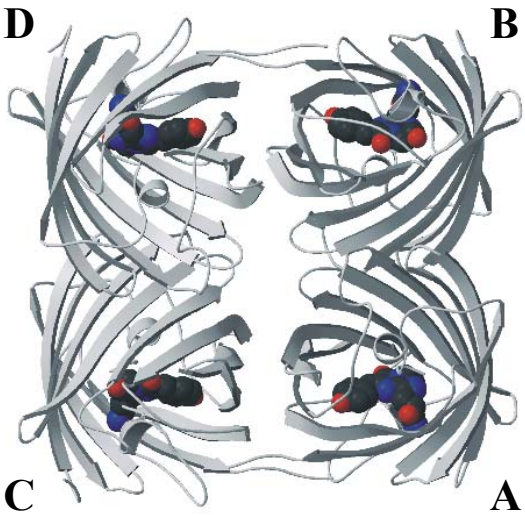

Fig. 4.20 *Ribbon diagram of the DsRed tetramer. Monomers are labelled with letters A-D and atoms in the chromophore (residues Gln 66, Tyr 67 an Gly 68) and the main chain cis peptide bond between Phe 65 and Gly 68 are shown as van der Waals spheres.*

Orientation and distances between chromophores within the tetramer suggest that chromophores A-B and C-D are the most probable FRET pairs.

In contrast to GFP, DsRed is found only in tetrameric and octameric states in solution (see figure 4.21):

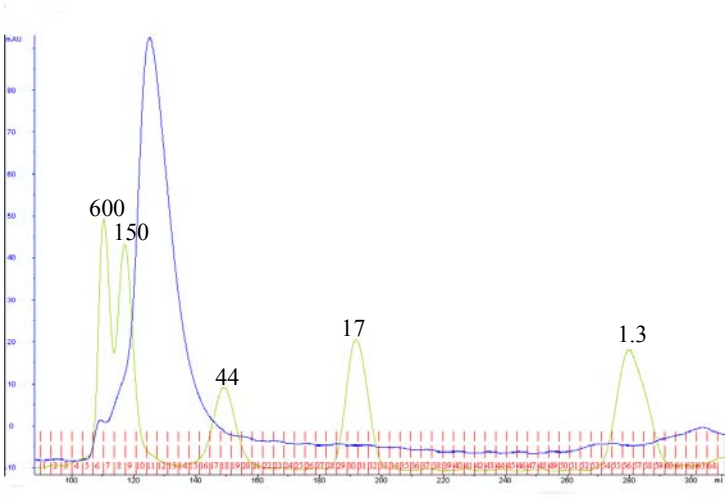

Fig. 4.21 *Size exclusion chromatography of the DsRed. Purified DsRed (20 mg/ml in PBS buffer) was eluted with 5 mM Tris, 50 mM NaCl, pH 8.0 (blue line). Molecular weight for calibration of proteins (green line) is given.*

The monomers are arranged as dimers of dimers with the A and C chains making close contacts via their surfaces and the A and B chains interacting via large buried patches. 25 % of the monomer surfaces is not solvent accessible in the tetramer. The A-B interface consists of hydrophobic interactions between small side chains and a few salt-brigges are also present. Interaction in the A-C interface consist largely of salt bridges and hydrogen bonds. There are several interesting and unusual features in the A-C and B-D interfaces e.g. the carboxy terminus of the A monomer embraces the C monomer forming a “clasp” about the local 2-fold axis. Much of the A-C interface involves the bungle region of the fold (which contacts the chromophore), suggesting that tetramer formation may be important for correct folding and establishment of the chromophore environment (Yarbrough et al., 2001).

4.3.7 The DsRed Chromophore Structure and Future Perspectives

The chromophore of GFP arises from cyclization of residues Ser 65 and Gly 67 and dehydratation to produce the imidazolinone intermediate (Heim et al., 1994; Cubitt et al., 1995). In DsRed cyclization of the backbone between Gln 66 and Gly 68 gives rise to the imidazoline ring which is planar to the phenolic ring of Tyr 67 (Fig. 4.22 A and B)

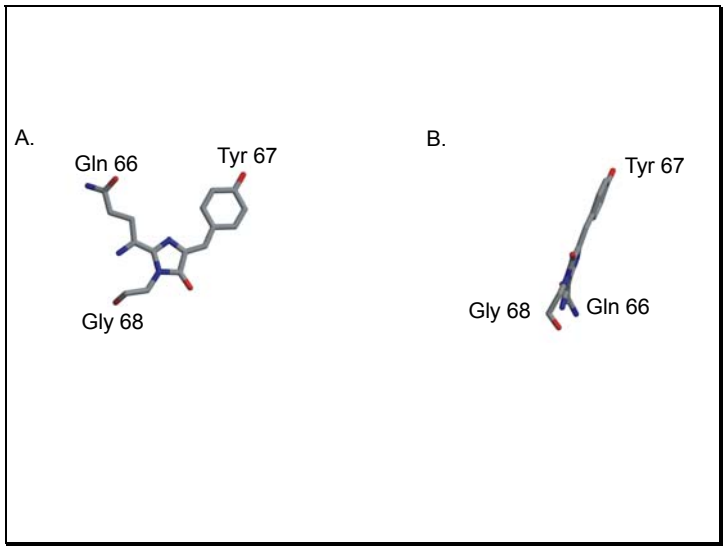

Fig. 4.22 *The chromophore structure of DsRed in different views to illustrate the planar architecture and cis peptide bonds*

Gln 66 is observed to be planar (originally sp^3 hybridization and tetrahedral configuration) and sp^2 hybridized, consistent with creation of the double bond between Cα and N of position 66. In addition, residues Phe 65 and Gln 66 of the DsRed chromophore are connected by an unexpected *cis* peptide bond. Cis peptide bonds occur rarely in proteins because of steric repulsion between side chains in that conformation. Rare examples of non-proline *cis* peptide bonds are preceded by β-strand secondary structures and are close to active sites in protein structures e.g. in (α/β) barrels where side chain contacts between the two residues are sharing the *cis* bond. Modelling of a *trans* peptide bond between Phe 65 and Gln 66 is inconsistent with the crystallographic data and will not allow the chomophore to be planar. It is clear that the DsRed protein environment stablilizes the formation of this unique structure and three factors seems to play role in it. They are :

1. the bond disrupts the middle of the α-helix within the β-can and there is no steric hindrance between the Phe 65 and Gln 66 side chains;
2. electron delocalization of Gln 66 main chain atoms into the chromophore molecular orbital;
3. all potential hydrogen bonds are satisfied upon *trans-cis* isomerization.

Wall et al (2000) showed that peptide bond isomerization can be also a key step in the maturation of the red fluorescence in DsRed. This process in DsRed is slower than in case of GFP and takes from hours to days. Rotation around Phe 65 – Gln 66 bond is energetically disfavored because of their partial double bond character. It is a result of delocalization that

creates a polar resonance species with planar geometry. Because of energetic interactions, the *cis* configuration can be stalilized relative to the *trans* in DsRed and the rate of achieving this geometry is probably slow given that the bond starts in the *trans* state upon folding of the protein.

In refined electron densities, an immature, not-planar conformation of the chromophore was seen. Similar data were published by Yabrough et al., (2001) and recently many other papers described kinetics of maturation of variants of DsRed (Verkhusha et al., 2001; He et al., 2002; Sacchetti et al., 2002; Terskikh et al., 2002; Bevis and Glick, 2002; Campbell et al., 2002). For most biotechnological applications slow maturation and oligomerization of DsRed are useless. This was one of the reasons why so many systematic mutagenesis studies were carried out to find out which residues could make maturation faster (Bevis and Glick, 2002) and stabilize the protein in monomeric states (Campbell et al., 2002). This research resulted in novel variants of DsRed with improved fluorescent properties e.g. mRFP1 – monomeric DsRed which confer greater tissue penetration and spectral separation from autofluorescence and other fluorescent proteins.

4.4 Conclusions

Studies presented here show that GFPuv exists in dynamic dimer-monomer equilibrium in solution at the high concentrations used for NMR spectroscopy. The β-barrel structure of GFPuv is inflexible on the nanosecond to picosecond timescale with only the connecting turns being flexible on the microsecond to millisecond timescale. The mutational study on His148 reveals a predetermined breakpoint within the otherwise regular β-barrel. The mutation His148Gly leads to a slow exchange process between two conformations of the β-barrel. A sequence alignment of related proteins from different organisms suggests a major role of glycine residues for the GFP fold.

The concerted use of ^{19}F NMR relaxation and dynamic NMR provides direct evidence for the existence of a slow exchange process between two states on the millisecond timescale (1.3 ms) including or near the chromophore of CFP. The structural origin of the exchanging states has not yet been fully clarified, but it seems to be clear that either the chromophore itself or more likely the sidechain of His 148 exchanges between two conformations. The differences in thermodynamic properties found for the different fluorination sites show that the introduction of fluorine into the protein has a small but observable effect on dynamic parameters. Therefore fluorine cannot be considered purely as a non-interacting “reporter

atom" in ^{19}F NMR studies. However, since the introduction of fluorine does not alter the timescales of dynamic processes, useful information can be extracted by comparing proteins with different fluorine-labeled sites. Presented results also demonstrate the use of an "atomic mutation" methodology to investigate proteins that contain aromatic amino acids. The knowledge of dynamics of ECFP opens several possibilities for manipulating the chromophore of CFP in order to improve the fluorescence quantum yield of CFPs. This may lead to the development of new, optimised variants of GFPs.

DsRed has a three-dimensional structure that is very similar to GFP, despite a rather low amino acid sequence homology (roughly 23%). The monomer consists of an 11-stranded β barrel with a coaxial helix. The chromophore is formed by cyclization of the carbon chain between –Glu-Tyr-Gly- (residues 66-68) and subsequent dehydrogenation of the Cα-Cβ bond of Tyr 67. DsRed is an ideal partner of GFP in fluorescence resonance energy transfer experiments. The protein proves to be stable under harsh pH conditions and its extremely resistant to photobleaching. DsRed has a high quantum yield, is photostable and can be used with laser-based confocal microscopes and flow cytometers. These characteristics make DsRed an ideal candidate for fluorescence imaging, and special for multicolour experiments involving GFP and its variants. But unfortunately wild-type DsRed has also several drawbacks; they are: the slow and possibly incomplete maturation for the red fluorescence and its tendency to oligomerize even at dilute concentration. If DsRed is fused to another protein, tetramerization of the DsRed domain can perturb the function and localization of the protein under investigation because of massive aggregation. By overcoming these problems (by site directed mutagenesis and optimization of different variants) DsRed, together with GFP, may become an invaluable tool in pure and applied biological research.

5. Structure of the N-Terminal Domain of the Adenylyl Cyclase-Associated Protein (CAP) from *Dictyostelium discoideum*

5.1 Biological Background

Cyclase associated proteins (CAPs) are highly conserved, ubiquitous, bifunctional proteins that modulate the actin-based cytoskeleton and play a role in RAS signaling (Hubberstey and Mottilo, 2002). CAP was first isolated from *Saccharomyces cerevisiae* as a 70 kDa component of the adenylyl cyclase complex (Cyr1p) (Field et al., 1990, Freeman et al., 1995). The CAP homologue from *Dictyostelium discoideum* is involved in the microfilament reorganisation at anterior and posterior plasma membrane regions during directed cell movement, it exhibits approximately 39% identity and 61% similarity to CAPs from human and yeast, and shows like all CAPs sequenced so far a similar length and the common domain organization (Gottwald et al., 1996). The amino-terminal domain of *Dictyostelium* CAP encompasses residues 1-215, and is followed by a proline rich region (residues 216-255) and the carboxy-terminal domain (residues 256-464) (Fig. 5.1).

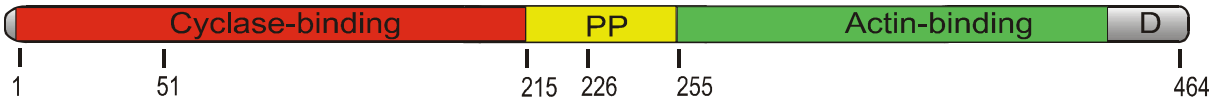

Fig. 5.1. *The domain organization of CAPs. CAP has a highly conserved domain structure. At the amino terminus, CAP has an adenylyl cyclase binding domain separated by a polyproline domain (PP) from the carboxy terminus with an oligomerization domain (D). A fragment of the N-terminal domain consisting of residues 51-226 (CAP-N) was used for structure determination.*

The amino-terminal region contains the adenylyl cyclase binding site (Nishida et al., 1998), the proline rich middle domain interacts with proteins containing SH3 domains, and finally the carboxyl terminus is necessary and sufficient for both protein dimerization and actin binding (Freeman et al., 1995, Gottwald et al., 1996; Wesp et al., 1997,). Despite its ubiquitous distribution and its involvement in signaling pathways there are only a few studies on the structure of the cyclase-associated protein under way (Hofmann et al., 2002; Roswarski et al.). This study focuses on the structure of the CAP N-terminus, which is apparently responsible for the interaction with the adenylyl cyclase and is of importance for the Ras

signaling pathways. The N-terminal domain shows an all-helical bundle structure. Including the yet unpublished data of the C-terminal domain and biochemical data about the oligomerization of the protein, we propose a model that suggests the formation of a ring-like multimer.

5.2 N-terminal Domain Constructs of CAP

The original recombinant construct of residues 1-226 of *Dictyostelium* CAP harboured the complete N-terminal domain, and was checked for its folding properties by NMR spectroscopy. NMR spectroscopy is a powerful technique for studying folded-unfolded equilibria in macromolecules because an unfolded protein shows a small dispersion of chemical shifts of amino acid residues (Wüthrich, 1986; Zhang and Forman-Kay, 1995, Rehm et al., 2002). This is particularly true for amide resonances that are clustered at around 8.3 ppm in the proton spectrum of an unfolded protein. NMR spectra of the N-terminal domain revealed a diagnostic distribution of resonance lines at 8.3 ppm, which together with the resonance intensity pattern of the up-field shifted aliphatic resonances (from 0.3 ppm to 0.2 ppm) versus the intensity to the bulk aliphatic resonance lines at 0.8 ppm allowed to estimate that about 15% of the domain was in a random coil conformation (Fig. 5.2). Spectra taken after the sample was left at room temperature for two weeks, together with subsequent amino acid analysis and mass spectrometry, revealed cleavage of the first 50 N-terminal residues of this domain. Based on these results, we constructed and expressed a 176-residue segment from residue 51-226, designated here as CAP-N, which proved to be folded and stable for at least several months at 4ºC

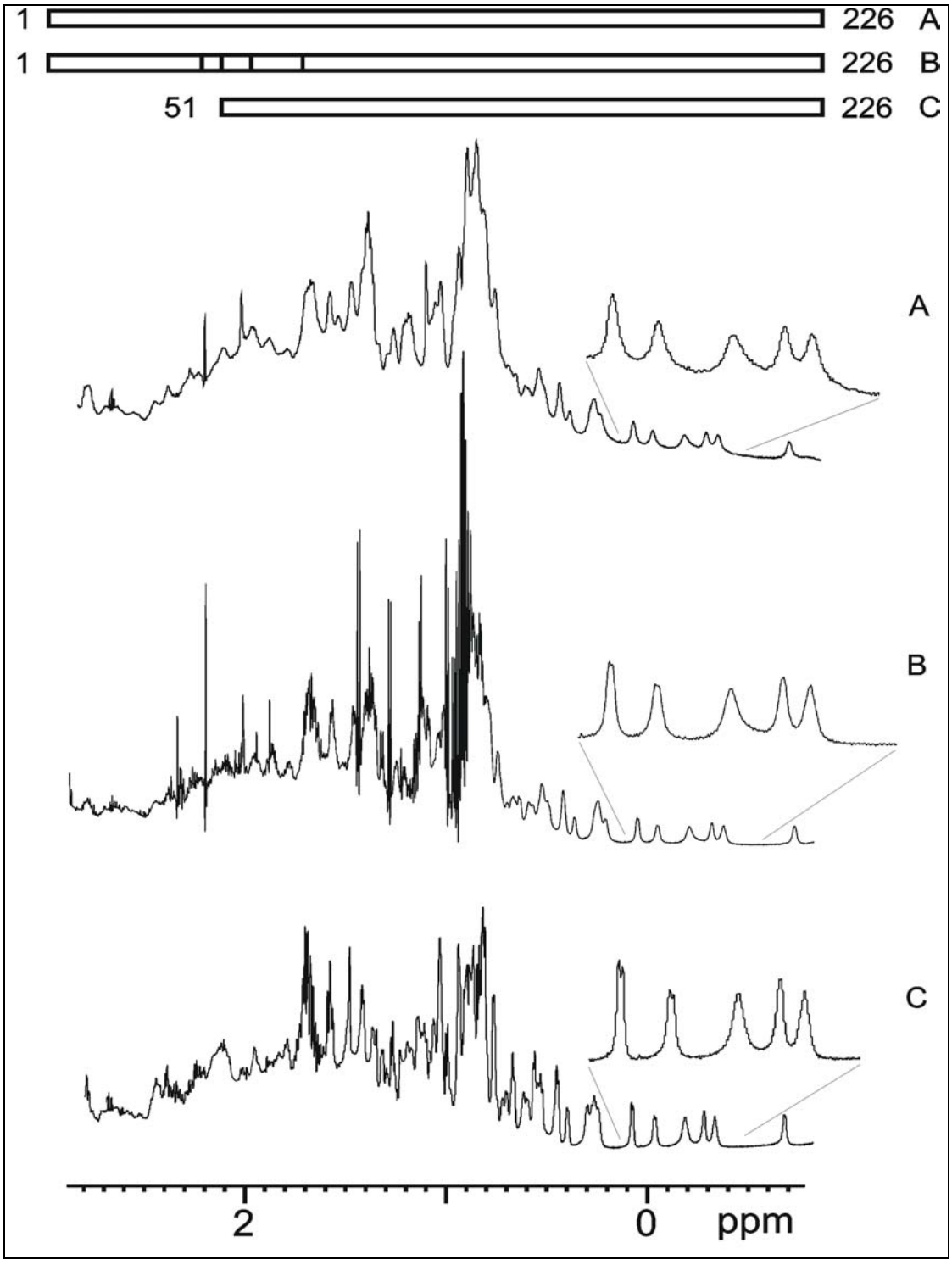

Fig. 5.2 *The aliphatic region of one-dimensional spectra of CAP. A The N-terminal 226 residue construct as indicated in the scheme above. B. A mixture of several constructs of different length (residues 1-226, 44-226, 51-226, 56–226). The overlap leads to broader lines compared to spectrum A). Short peptides give rise to sharp signals around 1 ppm. C. The stable core of the protein (residues 51-226) only. The sharp signals from impurities are removed and the linewidth is substancialy improved compared to spectrum A.*

5.3 Cloning, Expression and Crystallization of N-terminal CAP

To amplify the N-terminal domain of 176 residues of CAP (CAP-N) we employed the polymerase chain reaction (PCR) method. The PCR product was subcloned into the pT7-7 expression vector using the *Nde*I and *Bam*HI restriction sites (Tabor, 1990). *E. coli* BL21 cells harboring the plasmid were grown at 37°C to an OD_{600} of 0.6-0.8. To induce protein expression, IPTG was added to a final concentration of 0.5 mM and cells were further incubated over night. After lysis and centrifugation at 100.000 × g the supernatant was purified using DE52 (Whatman) anion exchange, phosphocellulose (P11, Whatman) cation exchange, hydroxylapatite (Bio-Rad) and Superdex 75 prep grade columns (Amersham Biosciences). Crystallization was carried out with the sitting drop vapour diffusion method by mixing equal volumes of protein and reservoir solution (30% PEG8000, 0.2 M $MgCl_2$, 10 mM β-mercaptoethanol and 0.1 M MES pH 6.1). Crystals of two different morphologies appeared in the same drop after 2-3 weeks and grew to a final size of ~0.2 × 0.2 × 0.1 mm (Fig. 5.3)

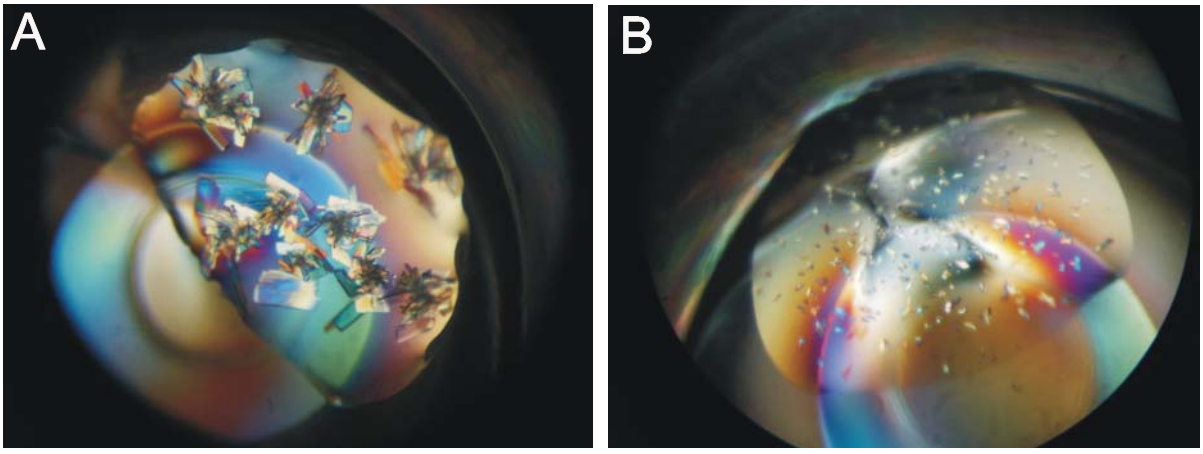

Fig. 5.3 *Crystals of N-terminal domain of CAP of two different morphologies. They grew from 30% PEG8000, 0.2 M $MgCl_2$, 10 mM β-mercaptoethanol and 0.1 M MES pH 6.1 at 4°C*
The colours are due to manipulation of the polarization filter in the microscope.

The crystals belong to the space group *P1* (with unit cell dimensions: a=37.52 Å, b=42.13 Å, c=53.74 Å, α=97.42°, β=105.09° and γ=97.15°) or to the space group $P2_1$ (with unit cell dimensions: a=50.19 Å, b=30.80 Å, c=52.87 Å, β=110.73°). The *P1* crystal contains one dimer per asymmetric unit whereas the monoclinic crystal form contains one monomer. A

native data set was obtained from plunge-frozen crystals to a resolution of 1.4 Å. For derivatization, the crystals were soaked for 15 min in a solution containing 3 mM of K_3IrCl_6. The crystals were then mounted in cryo-loops directly from the soaking solution and flash-frozen in the cold nitrogen stream. Annealing of the crystals by repeated freezing and thawing significantly reduced the mosaicity of the diffraction pattern (Fig. 5.4).

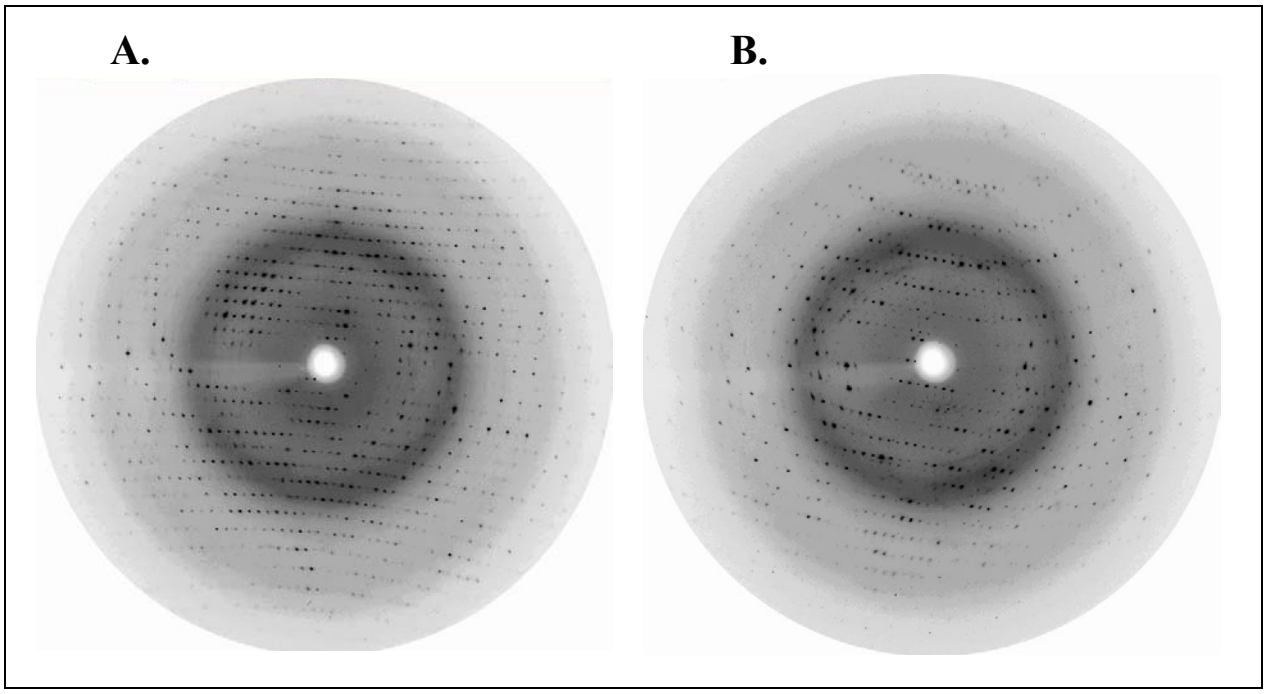

Fig. 5.4 *Diffraction pattern of N-terminal domain of CAP crystals.* ***A****. native data* ***B.*** *K_3IrCl_6 derivative The resolution at the edge of the images is 1.4 Å and 1.7 Å, respectively.*

X-ray data were collected at the wiggler beamline BW6 at DORIS (DESY, Hamburg, Germany) (Blume et al., 2001) using the 156-mm MarCCD detector (Mar-USA, Evantson, IL) at 100K and were processed and scaled with Denzo/Scalepack (Otwinowski and Minor, 1997). The data statistics are summarized in Table 5.1.

Table 5.1. *Data collection and phasing statistics*

	Native	**Iridium**	**Native**
Data collection			
Space group	*P*1	*P*1	$P2_1$
Cell constants (Å)	a=37.52 α=97.42		a= 50.19
	b=42.13 β=105.09		b=30.80 β=110.73
	c=53.74 γ=97.15		c=52.87
Resolution range (Å)	20-1.4	20-1.7	20-2.3
Wavelength (Å)	0.9760	1.105	0.9176
Observed reflections	3602228	208923	46028
Unique reflections	58737	34069	6804
Whole range			
Completeness (%)	93.7	94.0	97.9
[a]R_{merge}	3.2	4.3	5.2
$I\sigma(I)$	15.3	13.8	68.6
Last shell			
Resolution range (Å)	1.42-1.40	1.74-1.70	2.34-2.40
Completeness (%)	92.9	93.1	98.9
R_{merge}	12.8	21.0	10.7
$I\sigma(I)$	4.6	2.8	48.5
[b]Phasing power		2.3	
[c]FOM		0.63	
R_{culis}		0.61	
Number of sites		5	

[a] $R_{merge} = \frac{\Sigma(I - \langle I \rangle)}{\Sigma I}$; I is the intensity of an individual measurement and <I> the corresponding mean value.

[b] Phasing power = rms ($|F_H|/E$); $|F_H|$ is the heavy atom structure factor amplitude and E the residual lack of closure.

[c] Figure of merit; FOM $= \frac{|F(hkl)best|}{|F(hkl)|}$; $|F(hkl)|$ is the amplitude of an individual structure factor amplitude, and $|F(hkl)_{best}|$ is the best estimate for this amplitude.

5.4 Structure Determination and Refinement

The structure was determined by a combination of single isomorphous replacement with anomalous scattering (SIRAS) on the triclinic crystal. The phasing procedure was carried out with programs of the CCP4 program suite (Collaborative Computational Project No. 4) and CNS (Brünger, 1998). Anomalous difference Patterson syntheses were calculated from the peak (λ=1.105nm) data set. Refinement of heavy atom parameters and phase calculations from native and derivative data sets were performed with CNS, MLPHARE and Sharp. Solvent flattening and histogram matching were performed with the program DM, assuming a solvent content of 40%. The solvent flattened electron density map showed molecular bounders and allowed the recognition of few secondary structural elements. The map quality was further improved by a non-crystallographic symmetry averaging of the density using the program MAIN (Turk, D. 1992). The resulting electron density map was of good quality and enabled the incorporation of a complete model with the exception of 6 C-terminal residues. The initial model had a crystallographic R-factor of over 40% for all reflections in the resolution range of 10-2.0 Å. Real space electron density averaging was performed with MAIN in combination with CCP4 routines. The model was refined using the program CNS and corrected by using the interactive 3D graphics of the program MAIN. The final model has a R crystallographic factor of 18,4% and a free R factor of 21%.

The monoclinic crystal form was substituted from the Patterson search technique using the model coordinates from the *P1* crystal. The model was refined to a R factor of 19.6% and a free R factor of 24%. Searches for structurally related proteins were carried out with the DALI server. The refinement statistics data are summarized in Table 5.2.

Table 5.2 *Refinement statistics*

	Dimer	**Monomer**
No. of reflections	3602228	3201134
Resolution (Å)	20 - 1.4	20 - 1.7
R-factor (%)	18.4	19.6
R_{free} (%)	21	24

Content of asymmetric unit		
No. of protein molecules	2	1
No. of protein residues/atoms	352/6936	176/3468
No. of solvent atoms	698	257
Other atoms or molecules	Mg	MES
Ramachandran plot		
No. of residues in core regions	155	157
No. of residues in disallowed regions	0	0

5.5 Structure Description

Initial crystallization attempts using CAP-N yielded only poorly diffracting crystals. When proteins were crystallized immediately after gel filtration and always in the presence of β-mercaptoethanol, crystals were obtained that diffracted to a better than 2.5 Å resolution on the laboratory X-ray source, and better than 2.0 Å on the synchrotron. CAP-N crystallized both as a dimer (Fig. 5.5) and a monomer (Fig. 5.6) in the crystallographic asymmetric unit. The dimer and monomer crystals had different morphologies and diffracted to 1.4 and 1.7 Å resolution in two different symmetry groups, *P1* and $P2_1$, respectively. The structure of the *P1* crystal was solved by SIRAS of iridium-soaked crystals at a resolution of 1.4 Å and refined to an R-value of 18.4 %. The final model comprises the dimer molecule of CAP-N, a magnesium atom and 698 water molecules, and displays good stereochemistry. The two molecules are related by a non-crystallographic 2-fold axis.

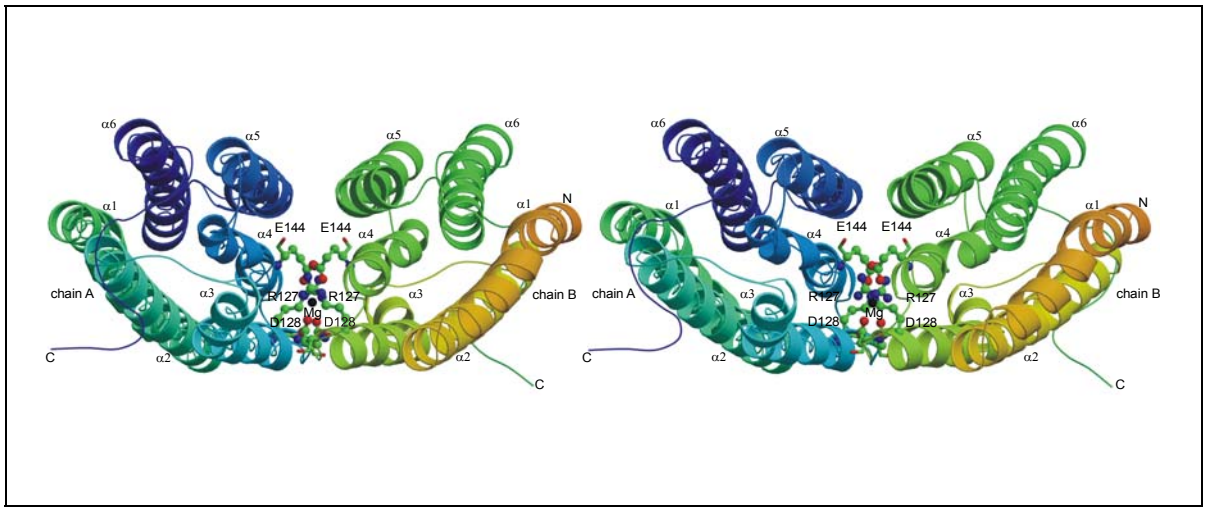

Fig. 5.5 *Ribbon presentation of the CAP-N dimer molecule with a magnesium ion and six residues involved in its coordination. Chain A is coloured in a blue-to-dark violet and chain B is coloured in an orange-to-green gradient. The dimer is shown from the bottom (in correlation to the monomer) and rotated 90° to the right. Shown carbon atoms are green, oxygen red, nitrogen blue and the magnesium ion is black.*

The monomer crystals consist of one CAP-N molecule, one MES molecule and 257 water molecules (Fig. 5.6). The dimensions of the monomer are ~33 Å × 52 Å × 30 Å. The complete dimer has an approximate size of 64 Å × 54 Å × 30 Å. Comparison of the monomer structure with that of the dimer showed that they are essentially identical. We use the higher resolution structure of the dimeric form of CAP-N for describing the details of the CAP-N fold.

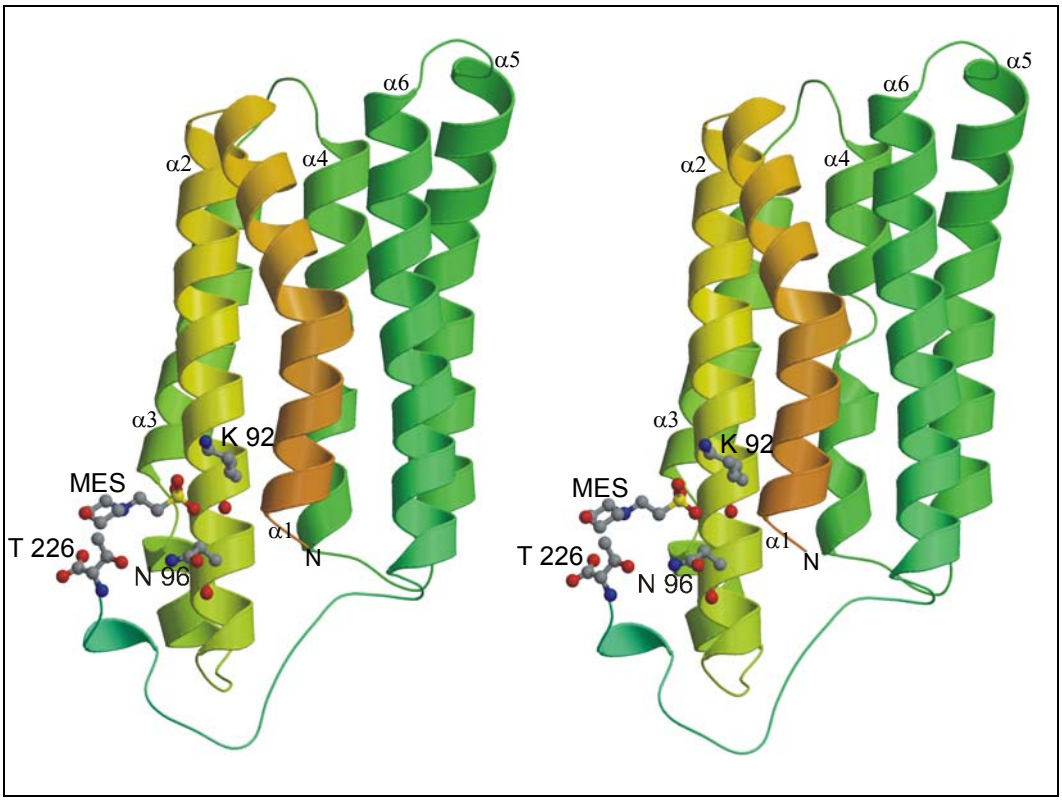

Fig. 5.6 *Stereo ribbon presentation of the monomer molecule. Colours correspond to that one from chain B of the dimer molecule. The side chains of the residues involved in specific interaction and coordination of the MES molecule and the MES molecule itself are shown in ball-and-stick presentation. The colours of the ball-and-stick are as follows: carbon, grey; nitrogen, blue; oxygen, red; sulphur, yellow.*

5.6 Overall Fold

The overall fold of CAP-N is presented in Fig. 5.5, 5.6 and the secondary structure elements found in the dimeric and monomeric forms are given in Table 5.3.

Table 5.3 *Secondary structure nomenclature*

Helix number	Dimer	Monomer
α1	52 – 73	52 – 73
α2	75 – 99	75 – 100
α3a	107 – 112	107 – 111
α3b	114 – 128	114 – 128

α4a	137 – 143	136 – 143
α4b	145 – 153	146 – 153
α5	158 – 180	158 – 180
α6	185 – 207	185 - 208

The CAP-N structure is built by an α-helix bundle composed of six antiparallel helices. Helix α1 is formed by residues 52 to 73 and runs approximately antiparallel to helix α2, which is the longest helix of the bundle (24 amino acids, residues 75 to 99). The two next helices α3 and α4 are distorted and subdivided into α3a (residues 107 to 112), α3b (residues 114 to 128), α4a (137 to 143) and α4b (145 to 153). In α3 one turn is distorted at residue 113; this residue has φ- and ψ-values of –95.4º and 7.2º, respectively. For α4, φ and ψ for the two amino acids 144 and 145 do not fit the ideal helix conformation (φ, ψ: -81.2º, -7.1º; -118.5º, 1.2º, respectively), and the ψ angles for several residues in α4b are between the ideal values for α- and 3_{10}-helices (the average φ and ψ are –60º and -20º, respectively). Helix α5 consists of residues 158 to 180; helix α6 (from 185 to 207) is arranged antiparallel to α1. A helical turn is also present at the C-terminus of the molecule but does not participate in the formation of the bundle fold. A total of 69% of residues of CAP-N are involved in α-helical structures. Helices α3 and α4 are connected by a long nine-residue loop, which is stabilized by hydrogen bonding interactions with Lys 125 (helix α3b), Leu 139, Ser 140 and Ala 141 (helix α4a). The longest loop is near the C-terminal part of the molecule and contains residues from 208 to 226. Because this long loop is stabilized only by three hydrogen bonds (with residues Lys 102, Tyr 204 and Ile 205) it may constitute a mobile element of the structure. An interesting feature of the monomer structure is the presence of a MES molecule, which is non-covalently bound to residues Lys 92, Asn 96 and Thr 226 (Fig. 5.6).

In the dimer, two monomers interact through helices 3 and 4 from both chains. This generates a four-helix interaction with a monomer/monomer interface that has mainly a hydrophobic character. The residues involved in a metal binding site that holds a hexa-hydrated Mg^{2+} are however exclusively hydrophilic (Fig. 5.7).

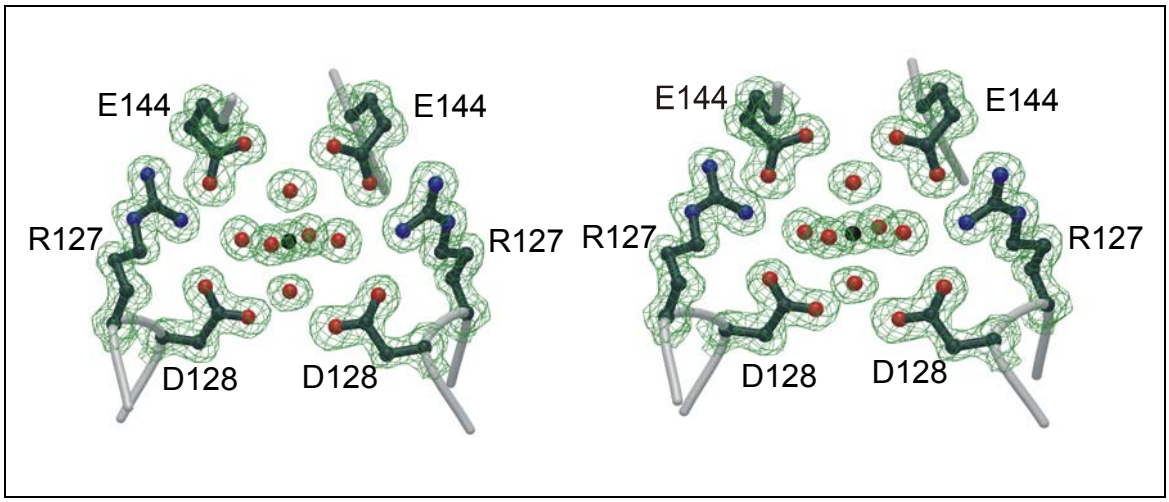

Fig 5.7 *Stereo view of a representative part of electron density map around the magnesium binding site. The 2|Fo| - |Fc| electron density map (1.4 Å) resolution, contoured at 1 σ above the mean) corresponds to residues 127, 128 and 144 from both chains of the dimer and six coordinating water molecules. The colours in the protein are as follows: carbon atoms, dark grey; nitrogens, blue; oxygens, red and magnesium, black. Orientation of the molecule is like in Fig. 5.5.*

The temperature factors, coordination characteristics and distances (about 2 Å) leave no doubt that the coordination is provided by a magnesium ion, as Mg^{2+} was present at a high concentration in the crystallization medium. Arg 127, Asp 128 and Glu 144 from each monomer are at a hydrogen bonding distance from water molecules surrounding the metal ion at an average distance of 2.1 Å. A coordination sphere of a positively charged amino acid, namely Arg 127, compensates for this accumulation of negative charges from Asp 128 and Glu 144. In addition, hydrogen bonds are formed between Asp 128 and Gly 124, and Nε from the side chain of Glu 144 contributes to the formation of another hydrogen bond with Ser 140.

5.7 Structural Comparisons

The DALI algorithm found seven proteins with high similarity scores to the CAP-N structure (Table 5.4). The fold of the 14-3-3 proteins (PDB ID; 1QJA) is the closest to the CAP-N structure (z-score = 7.7, Table 5.4). Figure 5.8 shows a structural alignment of CAP-N and the 14-3-3-protein. All known structures of different 14-3-3 proteins contain an 11 α-helix bundle with an up-down-up-down right-handed topology.

Table 5.4 *DALI statistics*

Name	PDB ID	z-score	r.m.s.d.	RES-ALI	RES-ALL	Identity (%)
14-3-3ζ	1QJA	7.7	3.8	128	217	7
Malt domain IV	1HZ4	5.4	3.3	96	366	11
Threonin phosphatase	1A17	5.0	4.6	87	159	10
CO-Chaperone	1FPO	4.4	3.6	83	171	10
Leptin	1AX8	4.4	2.7	80	130	9
Kerosteroid Isomerase	1E2A	4.2	3.4	78	102	10
Crystalin	1AUW	4.1	10.2	140	447	9
Arfaptin	1I4D	4.0	3.6	81	188	12

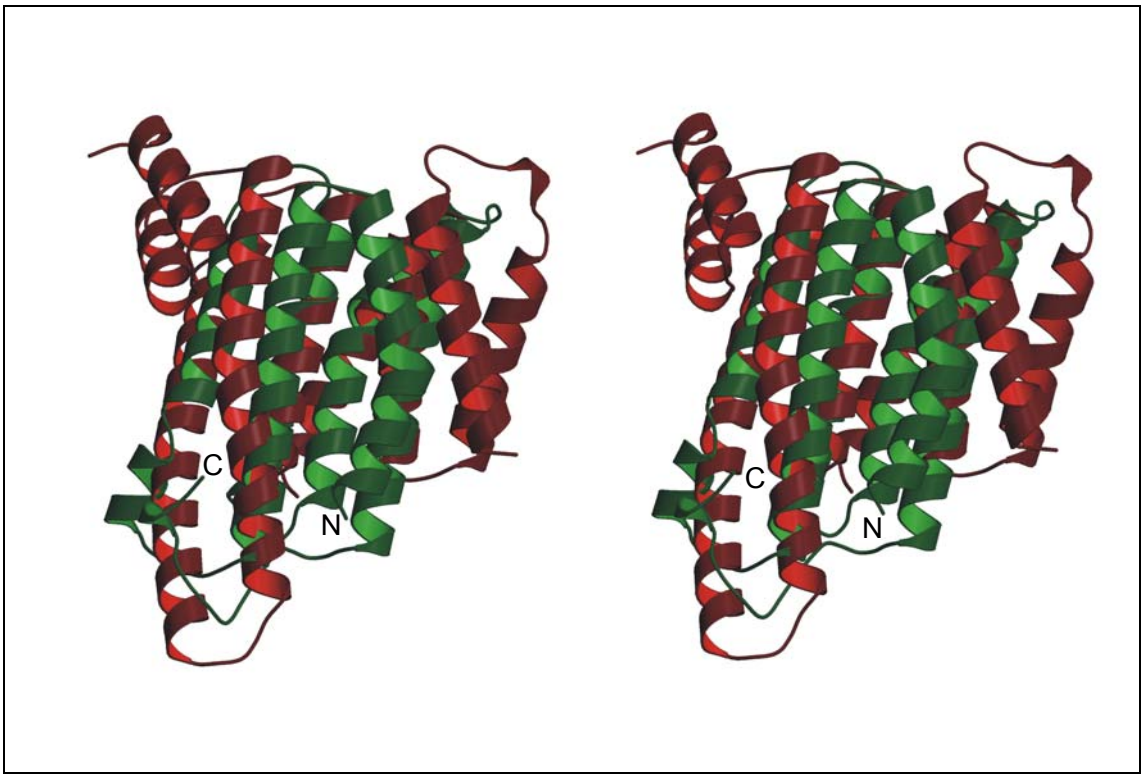

Fig. 5.8 *Superposition of CAP-N and 14-3-3. Orientation of CAP-N and superimposed 14-3-3 protein is like in Fig. 5.6.*

5.8 CAP in *Dictyostelium* is a Multimer

During fractionation of cellular extracts from *D. discoideum* by gel filtration chromatography it was noted that CAP did not elute at the expected position for a monomeric protein of 50 kDa but rather as a multimer. Bacterially expressed CAP exhibited a similar behaviour and eluted from a gel filtration column in front of catalase, which was added for control (Fig. 5.9; upper gel). Catalase, a 60 kDa protein, forms multimers of 240 kDa. From this it appears that CAP is at least a hexamer in solution. Multimerization of CAP was further confirmed by chemical crosslinking studies using EDC. CAP could be crosslinked very efficiently as shown by disappearance of the monomer and appearance of multimers, which were detected after immunoblotting and probing with monoclonal antibodies (Fig. 5.9;lower gel).

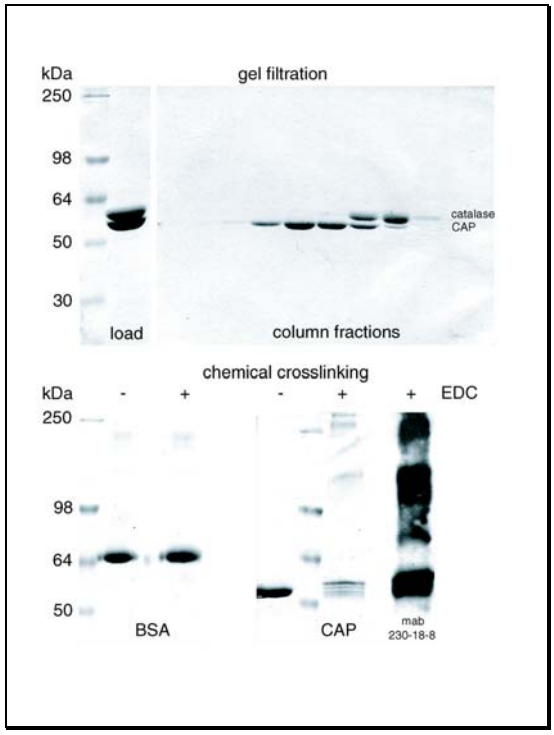

Fig. 5.9 *Multimerization of CAP. Gel filtration (upper gel) and chemical crosslinking (lower gel) show that CAP is a multimer. Upper gel: 150 µg of recombinant CAP were mixed with 100 g of catalase (240 kDa) as an internal standard, and the mixture was loaded onto a Superose 12 column (Amersham Biosciences). The load is shown in the first lane, followed by fractions #22 to #30. CAP eluted as a sharp peak at a position that corresponds to a molecular mass of about 370 kDa. Lower gel.: Bovine serum albumin (control) and CAP were treated with EDC for 3 min and further analyzed by SDS-PAGE. Crosslinked multimers of*

CAP can be detected in Coomassie Blue stained gels and after incubation with the monoclonal antibody 230-18-8 (Gottwald et al., 1996).

5.9 Discussion

The N-terminal domain of CAP is exclusively composed of α-helices connected by irregular loops of 5-12 residues; the helices are arranged into a six-helix bundle which is connected in the complete protein to the C-terminal domain through a proline-serine rich linker. Interestingly, the structure of the C-terminal domain of *S. cerevisae* CAP has been solved recently (Roswarski, R.A., to be published, PDB ID; 1K4Z). In contrast to our N-terminal domain structure, the C-terminus of CAP is solely built by parallel β-strands that form a right-handed β-helix of 6 turns. The β-helix itself forms a homodimer with the two β-structures arranged in an antiparallel orientation.

It is clear from these structural data that CAP domains have a preponderance for dimerization and that this feature should be important for the biological activities of CAP. In the N-terminal domain a magnesium ion seems to play a crucial role in dimerization. Its intermolecular environment consists of the Arg 127, Asp 128 and Glu 144 residues and their symmetric copies are arranged in a tetrahedral fashion. The complexity of this interaction and the high degree of conservation of the above mentioned amino acids in CAP sequences from other organisms (Fig. 5.10) argue strongly that this dimer represents the physiologically active form of CAP-N (Fig. 5.5).

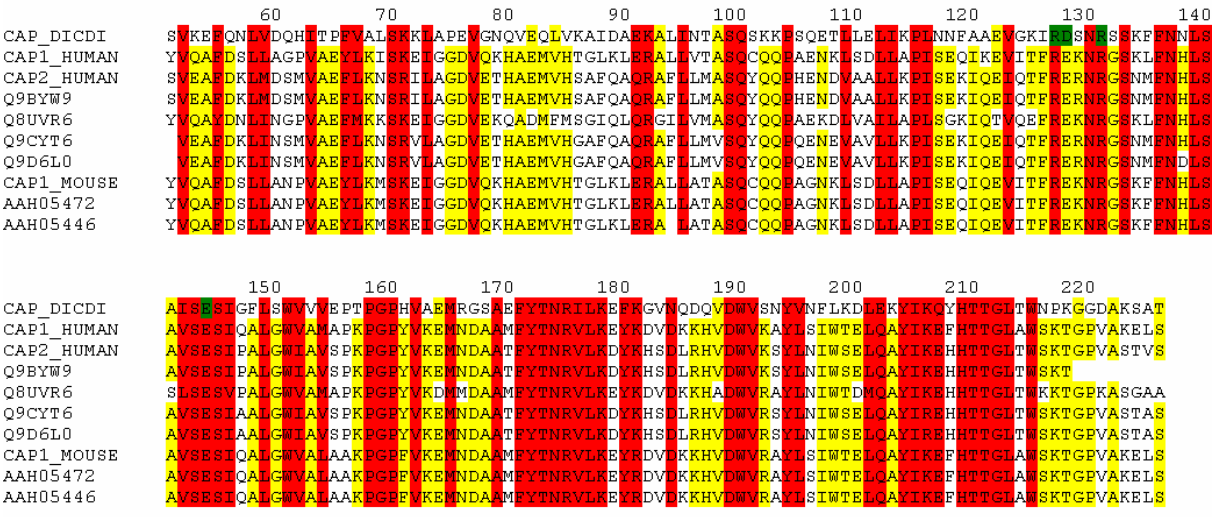

Fig. 5.10 *Structural conservation of CAP proteins. Sequence alignment of eight proteins from various organisms. The alignment was initially performed with ClustalW (Thompson, 1994) and subsequently manually modified. Conserved or similar residues in all 8 proteins are*

boxed in red, residues similar or conserved in more then 4 proteins are shaded in yellow and residues, which coordinated the Mg ion in the dimer molecule are green.

The dimer interface covers ~7.3 % of the accessible surface of the monomer. This relatively small value and the fact that most residues participating in the interface are hydrophilic, suggests that the driving force for dimerization stems from magnesium binding. In the absence of Mg^{2+} ions, negatively charged residues from both monomer molecules accumulate without compensation and, as a consequence, the relative position of the two domains may be destabilized. In the monomer molecule of the monoclinic crystal we could observe the MES molecule non-covalently bound to the protein moiety (Fig. 5.6). MES is located at the C-terminus of the longest loop of CAP-N and at the linker between N and C-terminal regions, which is flexible because of weak stabilization by hydrogen bonds.

Structures of both the N- and C-terminal domains together with NMR information on flexible fragments in CAP and biochemical data from gel filtration and chemical crosslinking enable us to model the structure of the entire CAP molecule with a high degree of confidence. We assume that CAP exists as hexamer or even octamer in its native form. It is interesting to note that the cyclase and actin binding sites are located on positions that are structurally independent from each other.

The fold of CAP-N is similar to that of the core of the 14-3-3 proteins although CAP-N has only 7% amino acid identity to the human homologue. 14-3-3 proteins constitute a highly conserved family of homo- and heterodimeric molecules present in abundance in all eukaryotic cells and have been implicated as key regulators of signal transduction events (Aitken, 1995, 1996; Fu et al., 2000). It was therefore interesting to find that Zhou et al. (2000) observed the interaction of the N-terminal CAP domain from *Lenzinus edodes* with the 14-3-3 proteins. The interaction between the N-terminal domain of CAP and 14-3-3 proteins, studied by the two hybrid system and co-precipitation experiments, was shown to be strong, specific and conserved in *S. pombe*, although it appears weaker than between the *L. edodes* homologues in the two hybrid system. Co-precipitation using fusion proteins produced in *E. coli* confirmed the interaction of the N-terminal region of *L. edodes* CAP and 14-3-3. However, this interaction was restricted to the N-terminal domain and not observed for full length CAP either in *L. edodes* or in *S. pombe*. The authors speculated that the C-terminal region of CAP might mask the 14-3-3 binding site in the N-terminal domain. We have studied the binding of a human full length 14-3-3σ with CAP-N and the long variant of the N-

terminal domain of CAP (residues 1-215) by NMR and found no evidence for any interaction between these two proteins.

The modes of dimerization of CAP-N and 14-3-3 proteins are different. In CAP-N, helices 3, 4, and 5 participate in the dimer interaction, whereas in 14-3-3 two short N-terminal helices, not present in CAP-N, interact with helices 2 and 3 (in the CAP-N numbering system).

5.10 Biological Implications

The cyclase associated protein CAP has at least two distinguishable functions: the N-terminus plays a role in RAS signaling, the C-terminus binds to G-actin and thus alters the dynamic rearrangements of the microfilament system. The structure of the N-terminus is an all-helix bundle, which crystallizes as monomer or a dimer. Interestingly, the dimer is formed by a Mg^{2+} ion and the orientation of the two bundles suggests an intermolecular dimerization. Because the C-terminus forms dimers as well, CAP might adopt a ring-like hexamer with the adenylyl cyclase binding domains in the middle and the actin-binding domain at the outer surface of the ring. The typical topology of adenylyl cyclases in mammalian cells contains 12 transmembrane domains. In *D. discoideum* proteins of nearly identical topology have been identified but they are distinct because they have a guanylyl and not an adenylyl cyclase activity (Roelofs et al., 2001). It remains to be shown whether *D. discoideum* CAP can associate also with guanylyl cyclases, but considering the bulky transmembrane areas of those cyclases, a ring-like oligomer of CAP subunits could very well fit into this complex connecting the cyclase to the actin cytoskeleton.

6. Summary

The work in this thesis has been carried out in the Department of Structural Research at the Max Planck Institute for Biochemistry from October 1999 to November 2002. Two projects were undertaken: structural and dynamic investigations on green fluorescent-like proteins and the X-ray structure determination of the N-terminal domain of the adenylyl cyclase associated protein (CAP) from *Dictyostelium discoideum*.

The nuclear magnetic resonance (NMR) characterization of the fluorescent proteins the green fluorescent protein (GFP) and cyan fluorescent protein (CFP) paved the way for investigations of their dynamic properties. GFP and its mutants are well-known to be extremely valuable tools in molecular biology. ^{1}H-^{15}N NMR spectra of the His148Gly GFP mutant indicate that the mutation affects the conformational stability of GFP, we detected two conformations of the protein in a slow exchange on the NMR time scale. Also the ^{19}F NMR studies of variants of GFP and CFP labeled with fluorinated tryptophans, supported by temperature-, concentration- and folding-dependent experiments provided direct evidence for the existence of a slow exchange process between two different conformational states of CFP. The possibility to generate a series of single atom exchanges ("atomic mutations") like H→F in this study offers a useful approach for characterizing and quantifying dynamic processes in proteins by NMR.

The structure of the red fluorescent protein (DsRed) is also described with the emphasis on the conformational properties of the chromophore. DsRed, a 28-kDa polypeptide, displays two new properties of the fluorescent protein family: dramatically red-shifted spectra and oligomerization to form tetramers. DsRed also shows slow kinetics of maturation proceding via a GFP - like green intermediate to the final red species. The crystal structure of DsRed was solved by molecular replacement using the previously determined structure of *Discosoma* corals DsRed as a search model. DsRed is an 11-stranded β-can with a central α-helix and an α-helical cap on the barrel ends, nearly identical in topology to the homologous GFP.

Structural investigations on the N-terminal domain of CAP (CAP-N) from *D. discoideum* yielded the crystal structure of this protein. CAPs are widely distributed and highly conserved actin binding proteins that regulate actin remodeling in response to cellular signals. Two functional binding sites are located on the opposite sides of the amino acid sequence. The N-terminus of CAPs plays a role in Ras signaling, the C-terminus binds to G-actin and thus alters the dynamic rearrangements of the microfilament system. The crystal

structure of CAP-N was determined by a combination of single isomorphous replacement with anomalous scattering (SIRAS) of iridium-soaked crystals. The overall structure consists of an α-helix bundle composed of six antiparallel helices. The N-terminal domain of the *D. discoideum* CAP is found in monomeric and dimeric states. The dimer is formed by an Mg^{2+} ion and the orientation of the two bundles, together with gel filtration and cross-linking experiments, suggests an overall model for an octameric state of the whole CAP molecule. The presented structure is the first for an N-terminal domain of any CAP and plays a crucial role in the understanding of specific functions for these domains.

7. Zusammenfassung

Die hier beschriebene praktische Arbeit umfasste zwei verschiedene Projekte: Die strukturellen und dynamischen Untersuchungen dreier verschiedener fluoreszierender Proteine und die Aufklärung der Röntgenstruktur der N-terminalen Domäne von Adenylyl Cyclase-assoziiertem Protein (CAP-N) aus *Dictyostelium discoideum*.

Die Charakterisierung der fluoreszierenden Proteine, des Grün fluoreszierenden Proteins (GFP) und des Cyan fluoreszierenden Proteins (CFP) mit Hilfe der Kernresonanzspektroskopie (NMR) haben den Weg für die Untersuchungen ihrer dynamischen Eigenschaften geebnet. GFP und seine Mutanten sind äußerst wertvolle Werkzeug in der Molekularbiologie. Die Studien mit Hilfe der ^{15}N NMR Spektroskopie an GFP und seiner Mutante His148Gly zeigten eine deutliche Flexibilität in ihrer Konformation. Die ^{1}H-^{15}N NMR Spektren deuten darauf hin, dass die Mutation His148Gly die Stabilität beeinflusst. Aus den Ergebnissen kann man auf das Vorhandensein zweier Konformationen schließen, die sich auf einer für NMR messbaren Zeitskala ineinander umwandeln. Die ^{19}F NMR Studien an Mutanten von GFP und CFP, die mit fluoriertem Tryptophan markiert waren, ermöglichen den Nachweis von langsamen Bewegungen auf molekularer Ebene. Neben den Dynamikstudien wurden auch Experimente zur Abhängigkeit von Temperatur, Konzentration und Faltung unter der Anwendung der NMR Spektroskopie an CFP durchgeführt. Das Ergebnis bietet einen direkten Beweis für die Existenz eines langsamen Austauschprozesses zwischen zwei verschiedenen Konformationen des Proteins. Die Möglichkeit des gezielten Ersetzens einzelner Atome wie H→F bietet einen nützlichen Ansatz für die Charakterisierung und Quantifizierung dynamischer Prozesse in Proteinen mit Hilfe der NMR Spektroskopie. Darüber hinaus wird die Struktur des rot fluoreszierenden Proteins beschrieben und einige interessante Besonderheiten der Chromophorkonformationen diskutiert. DsRed, ein 28-kDa großes Polypeptid, zeigt zwei neue Eigenschaften der Familie von fluoreszierenden Proteinen: Eine deutliche rot Verschiebung in den Fluoreszenz Spektren und die Bildung von Tetrameren. DsRed besitzt auch eine langsame Kinetik des Reifungsprozesses über eine GFP-ähnliche grüne Zwischenstufe zur endgültigen roten Spezies. Die Kristallstruktur des DsRed wurde mit Hilfe von Isomorphem Ersatz unter Verwendung der früher ermittelten Struktur des DsRed aus der Koralle *Discosoma striata* als Suchmodell bestimmt. Das DsRed bildet ein 11-strängiges β-Faß mit einer zentral liegenden α-Helix. Die Topologie ist damit fast identisch mit der Topologie des homologen GFP. Die langsame Reifung und Oligomerisierung des DsRed in Lösung führte zu Untersuchungen von monomer vorliegenden Mutanten, die einen

schnelleren Reifungsprozess besitzen. Heutzutage entwickelt sich das DsRed zu einem nützlichen Hilfsmittel in der Molekularbiologie.

Im zweiten Teil der vorliegenden Dissertation wurde die Kristallstruktur des CAP-N aus *Dictyostelium discoideum* bestimmt. Die Cyclase-assoziierten Proteine sind weit verbreitete und hoch konservierte Aktin bindende Proteine, die die Neuordnung des Aktins als Antwort auf zellulare Signale regulieren. Die funktionellen Bindestellen sind an entgegengesetzten Positionen in der Aminosäuresequenz zu finden. Die N-terminale Domäne des CAPs spielt eine Rolle im Ras-Signalweg, die C-terminale Domäne bindet an G-Aktin und bewirkt die dynamische Neuausrichtung des Mikrofilamentssystems. Die Kristallstruktur des CAP-N wurde mit Hilfe einer Kombination von einzelner isomorpher Ersetzung mit anomaler Beugung an Kristallen, die mit Iridium getränkt sind, aufgeklärt. Die Struktur besteht aus einem α-helikalen Bündel aus 6 antiparallelen Helices. Die N-terminale Domäne des *Dictyosteliums* CAP liegt in monomerer und dimerer Form vor. Die Dimere werden durch ein Magnesium Ion stabilisiert. Die Orientierung der beiden Bündel legt die Vermutung nahe, dass das CAP-Molekül als Octamer vorliegt. Dieses Modell wird durch Gelfiltration und Cross-linking Experimente unterstützt. Die vorliegende Struktur ist die erste einer N-terminalen Domäne eines CAPs überhaupt und spielt eine entscheidende Rolle im Verständnis der spezifische Funktion dieser Domäne.

8. References

Abragam A.: *Principles of Nuclear Mechanism,* Oxford University Press, (1961)

Aitken A.: 14-3-3 proteins on the MAP, *TIBS* **20** (1995) 95-97

Aitken A.: 14-3-3- and its possible role in co-ordinating multiplesignaling pathways, *Trends Cell Bioll.* **6** (1996) 341-347

Anglister J., Grzesiek S., Ren H., Klee C.B. & Bax A.: Isotope-edited multidimensional NMR of calcineurin-B in the presence of the non-deuterated detergent Chaps, *J. Biomol. NMR* **3** (1993) 121-126

Bann J.G., Pinkner J., Hultgren S.J. & Frieden C.: Real-time and equilibrium ^{19}F-NMR studies reveal the role of domain-domain interactions in the folding of the chaperone PapD, *Proc. Natl. Acad. Sci. USA* **99** (2002) 709-714

Baumgartner R., Fernandez-Catalan C., Winoto A., Huber R., Engh R.A. & Holak T.A.: Structure of human cyclin-dependent kinase inhibitor p19^{INK4d}: comparison to known ankyrin-repeat-containing structures and implications for the dysfunction of tumor suppressor p16^{INK4d}, *Structure* **6** (1998) 1279-1290

Battistutta R., Negro A. & Zanotti G.: Crystal structure and refolding properties of the mutant F99S/M153T/V163A of the green fluorescent protein from coral, *Prot. Struct. Func. Gen.* **41** (2000) 429-437

Bell A.F., He X., Wachter R.M. & Tonge P.J.: Probing the ground state structure of the fluorescent protein chromophore using Raman spectroscopy, *Biochemistry*, **39** (2000) 4423-4431

Bevis B.J. & Glick B. S.: Rapidly maturing variants of the *Discosoma* red fluorescent protein DsRed, *Nature Biotech.* **20** (2002) 83-87

Blume H., Boesecke P., Bourenkov G.P., Kosciesza D. & Bartunik H.D.: The protein crystallography beamline BW6 at DORIS – automatic operation and high-throughput data collection, *Nuclear Instr. Meth. A* **467** (2001) 1358-1362

Bragg W.L. & Bragg W.H.: The diffraction of X-rays by crystals, *Proc. Roy Soc. Sect. A* **88** (1913) 428-438

Branden C. & Tooze, J. *Introduction to Protein Structure*. Garland Publ., New York, (1999)

Brejc K., Sixma T.K., Kitts P.A., Kain S.R., Tsien R.Y., Ormo M. & Remington S.J. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein, *Proc. Natl. Acad. Sci. U.S.A.* **94** (1997) 2306-2311

Brünger A.T., Adams P.D., Clore M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang, J. Kuszewski J., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T. & Warren G.L.: Crystallography and NMR System – A new software suite for macromolecular structure determination, *Acta Crystallogr. D* **54** (1998) 905-921

Budisa N., Bae J., Rubini M., Weyher E., Wenger W., Azim M.K., Moroder L. & Huber R.: (2003) *Manuscript in preparation*

Bulina M.E., Chudakov D.M., Mudrik N.N. & Lukyanov K.A.: Interconversion of Anthozoa GFP-like fluorescent and non-fluorescent proteins by mutagenesis, *BMC Biochemistry* **3** (2002) 7

Campbell R.E., Tour O., Palmer A.E., Steinbach P.A., Baird G.S., Zacharias D.A. & Tsien R.Y.: A monomeric red fluorescent protein, *Proc. Natl. Acad. Sci. U.S.A.* **99** (2002) 7877-7882

Canet D. Last A.M., Tito P., Sunde M., Spencer A., Archer D.B., Redfield C., Robinson C.V. & and Dobson C.M.: Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme, *Nat. Struct. Biol.* **9** (2002) 308-315

Cantor C. & Schimmel P. R.: *Biophysical chemistry part II: Techniques for the study of biological structure and function*. W.H. Freeman and Company, San Francisco, (1980)

Cavanagh J., Fairbrother A.G., PalmerIII & Skelton, N.J.: *Protein NMR Spectroscopy. Principles and Practice*, Academic Press, New York, (1996)

Chalfie M., Tu Y., Euskirchen G., Ward W.W. & Prasher D.C.: Green fluorescent protein as a marker for gene expression, *Science* **263** (1994) 802-805

Chattoraj M., King B.A., Bublitz G.U. & Boxer S.G.: Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer, *Proc. Natl. Acad. Sci. USA* **93** (1996) 8362-8367

Chen M.C., Lambert C.R., Urgitis J.D. & Zimmer M.: Photoizomerization f green fluorescent protein and the dimensions of the chromophore cavity, *Chem. Phys*. **270** (2001) 157-164

Cheng Z.Q. & McFadden B.A.: A study of conserved in-loop and out-of-loop glycine residues in the large subunit of ribulose bisphosphate carboxylase/oxygenase by directed mutagenesis, *Prot. Eng.* **11** (1998) 457-465

Cieslar C., Holak T. & Oschkinat H.: A program for the evaluation of 3d spectra applied to the sequential assignment of PBTI Utilizing 3D TOSCY-NOSY, *J. Magn. Reson.* **87** (1990) 400-407

Collaborative Computational Project No. 4 The CCP4 Suite: Programs for Protein Crystallography. *Acta Crysallogr. D* **50** (1994) 760-763

Cotton M., Tian C., Busath D.D., Shirts R.B. & Cross T.A.: Modulating dipoles for structure-function correlations in the gramicidin A channel, *Biochemistry* **38** (1999) 9185-9197

Cubitt A.B., Heim R., Adams S.R., Boyd A.E., Gross L.A. & Tsien R.Y.: Understanding, improving and using green fluorescent proteins, *Trends Biochem Sci.* **20** (1995) 448-55

Danielson M.A. & Falke J.J.: Use of ^{19}F NMR to probe protein structure and conformational changes, *Annu. Rev. Biophys. Biomol. Struct.* **25** (1996) 163-195

Dempsey C.E.: Hydrogen exchange in peptides and proteins using NMR spectroscopy, *Prog. Nucl. Magn. Reson. Spec.* **39** (2001) 135-170

Denisov V.P., Peters J., Hörlein H.D. & Halle B.: Using buried water molecules to explore the energy landscape of proteins, *Nature Struct. Biol*. **3** (1996) 505-509

Dierck T., Coles M. & Kessler H.: Applications of NMR in drug discovery, *Curr. Opin. Chem. Biol*. **5** (2001) 285-291

Dickson R., Cubitt A., Tsien R. & Woerner W.: On/off blinking and switching of single molecules of green fluorescent protein, *Nature* **388** (1997) 355-358

Dingley A.J., Mackay J.P, Chapman B.E., Morris M.B., Kuchel P.W., Hambly B.D. & King G.F.: Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: application to myosin light chain2. *J. Biomol. NMR* **6** (1995) 321-328

Dominguez M.A., Thornton K.C., Melendez M.G. & Dupureur C.M.: Differential effects of isomeric incorporation of fluorophenylalanines into Pvull endonuclease, *Prot. Struct. Funct. Gen.* **45** (2001) 55-61

Drenth J.: *Principles of protein X-ray crystallography*, Springer-Verlag, New York (1994)

Dunitz J.D.: Win some, los some : entalphy-entopy compensation in weak intermolecular interactions, *Chem. Biol.* **2** (1995) 709-712

Edwards A. Arrowsmith C.H., Christendat D., Dharamsi A., Friesen J.D., Greenblatt J.F. & Vedadi M.: Protein production: feeding the crystallographers and NMR spectroscopists, *Nat. Struct. Biol. Struct. genomics suppl*. (2000)

Esnouf R.M.: Further additions to MolScript version 1.4, including reading and contouring of electron density maps, *Acta Crystallog. D* **55** (1999) 938-940

Ewald P. P.: Das „reziproke“ Gitter in der Strukturtheorie, *Z. Kristallogr*. **56** (1921) 129-156

Field J., Vojtek A., Ballester R., Bolger G., Colicelli J., Ferguson K., Gest J., Katoka T., Michaeli T. & Powers S.: Cloning and characterization of CAP, the *S. cerevisiae* gene encoding the 70 kDa adenylyl cyclase-associated protein, *Cell* **61** (1990) 319-327

Frauenfelder H., Sligar S.G. & Wolynes P.G.: The energy landscape and motions of proteins, *Science* **254** (1991) 1598-1603

Freeman N., Chen Z., Horenstein J., Weber A. & Field J.: An actin monomer binding activity localizes to the carboxyl half of the *Saccharomyces cerevisiae* cyclase associated protein, *J. Biol. Chem.* **270** (1995) 5680-5685

Fernandez C., Hilty C., Bonjour S., Adeishvili K., Pervushin K. & Wüthrich K.: Solution NMR studies of the integral membrane proteins OmpX and OmpA from Escherichia coli, *FEBS Lett.* **504** (2001) 173-178

Fielding L.: Determination of the association constants (Ka) from solution NMR data, *Tetrahedon* **56** (2000) 6151-6170

Fu H., Subramanian R.R. & Masters S.C.: 14-3-3 Proteins: structure, function and regulation, *Annu. Rev. Pharmacol. Toxicol.* **40** (2000) 617-47

Garcia-Parajo M., Segers-Nolten G., Veerman J., Greve J. & von Hulst N.: Real-time light-driven dynamics of the fluorescence emission in single green fluorescent protein molecules, *Proc. Natl. Acad. Sci. U.S.A.* **97** (2000) 7237-7242

Gardner K.H., Zhang X.C., Gehring K. & Kay L.E. *J. Am. Chem. Soc.* **120** (1998) 11738-11748

Georgescu J.: *NMR-Untersuchungen und Strukturbestimmungen von Biomolekülen: Linkerpolypeptid $L_C^{7,8}$, Interleukin-16, Grün fluoreszierendes Protein (GFPuv)*, PhD thesis, TU Munich, Faculty of Chemistry (2000)

Georgescu J., Ksiazek D., Seifert M.H.J., Rehm T., Steipe B., Holak T.A. & Reuter, W.: (2002**)** *manuscript in preparation*

Giacovazzo C., Monaco H.L., Viterbo D., Scordari F., Gilli G., Zanotti M. & Catti M. (1992) *Fundamentals of crystallography*, Oxford University Press, Oxford

Goddard T.D. & Kneller D.G., *SPARKY 3*, University of California, San Francisco, (2001)

Gottwald U., Brokamp R., Karakesisoglou I., Schleicher M. & Noegel A.: Identification of a cyclase-associated protein (CAP) homologue in *Dictyostelium discoideum* and characterization of its interaction with actin, *Mol. Biol. Cell* **7** (1996) 261-272

Gross L.A., Baird G.S., Hoffman R.C., Baldridge K.K. & Tsien R.Y.: The structure of the chromophore within DsRed, a red fluorescent protein from coral, *Proc Natl Acad Sci U S A.* **22 (**2000) 11990-11995

Harker D.: The determination of the phases of the structure factors on non-centrosymmetric crystals by the method of double isomorphous replacement, *Acta Cryst.* **9** (1956) 1-9

Haupts U., Maiti S., Schwille P. & Webb W.W. Dynamics of fluorescence fluctuations in green fluorescent protein observed by fluorescence correlation spectroscopy, *Proc. Natl. Acad. Sci. U.S.A.* **95** (1998) 13573-13578

He X., Bell A.F. & Tonge P.J.: Synthesis and spectroscopic studies of model red fluorescent protein chromophores, *Org Lett.* **9** (2002) 1523-1526

Heim R. & Tsien R.Y.: Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer, *Curr Biol.* **2** (1996) 178-182

Helms V., Straatsma T.P. & McCammon J.A.: Internal Dynamics of Green Fluorescent Protein *J. Phys. Chem.* **103** (1999) 3263-3269

Hendricson W.A., Smith J.L., Phizackerley R.P. & Merrit E.A.: Crystallographic structure analysis of lamprey hemoglobin from anomalous disperssion of synchlotron radiation, *Proteins* **4** (1988) 77-78

Hofmann A., Hess S., Noegel A.A., Schleicher M. & Wlodawer A.: Crystallization of cyclase-associated protein from Dictyostelium discoideum, *Acta Crysallogr. D* **58** (2002) 1858-1861

Hofmeister F.: On the understanding of the effects of salts, *Arch. Exp. Pathol. Pharmacol.* **24** (1888) 247-260

Holm L. & Sander C.: Mapping the protein universe, *Science* **273** (1996) 595-603

Hopf M., Göhring W., Ries A., Timpl R. & Hohenester E.: Crystal structure and mutational analysis of a perlecan-binding fragment of nidigen –1, *Nature Struct. Biol***. 8** (2001) 634-640

Hoppe W.: Die Faltmolekülmethode: eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannten Molekülstrukturen, *Acta Cryst.* **10** (1957) 750-751

Huber R.: Die automatisierte Faltmolekülmethode, *Acta Cryst.* **19** (1965) 353-356

Huberstey A.V. & Mottillo E.P.: Cyclase associated proteins: CAPacity for linking signal transduction and actin polimerization, *FASEB J.* **490** (2002) 487-499

Hull W.E. & Sykes B.D.: ^{19}F nuclear magnetic resonance relaxation times and molecular motion of the individual fluorotyrosines, *Biochemistry* **13** (1974) 3431-3437

Hull W.E. & Sykes B.D.: Fluorotyrosine alkaline phosphatase: internal mobility of individual tyrosines and the role of chemical shift anisotropy as a ^{19}F nuclear spin relaxation mechanism in proteins, *J.Mol. Biol.* **98** (1975a) 121-153

Hull W.E. & Sykes B.D.: Dipolar nuclear spin relaxation of fluorine-19 in multispin systems. Application to fluorine-19 labelled proteins, *J. Chem. Phys.* **63** (1975b) 867-880

Juranic N., Ilich P. K. & Macura S.: Hydrogen bonding networks in proteins as revealed by the amide 1JNC' coupling constant, *J. Am. Chem. Soc.* **117** (1995) 405-410

Juranic N., Likic V.A., Prendergast F.G. & Macura S.: Protein-solvent Hydrogen bonding studied by NMR 1JNC' coupling constant determination and molecular dynamics simulations, *J. Am. Chem. Soc.* **118** (1996) 7859-7860

Kalus W., Baumgartner R., Renner C. Noegel A., Chan F.K.M., Winoto A. & Holak T.A.: NMR structural characterization of the CDK inhibitor $p19^{INK4d}$, *FEBS* **401** (1997) 127-132

Kalus W., Zweckstetter M., Renner C., Sanchez Y., Georgescu J., Groll M., Demuth D., Schumacher R., Dony C., Lang K., and Holak T.A.: Structure of the IGF-binding domain of the insulin-like growth factor-binding protein-5 (IGFBP-5): implications for IGF and IGF-I receptor interactions, *EMBO J.* **17** (1998) 6558-6572

Kay L.E.: Protein dynamics from NMR. *Nature Struct. Biol*., *NMR supplement* (1998) 513-517

Klein-Seetharaman J., Getmanova E.V., Loewen M.C., Reeves P.J. & Khorana H.G.: NMR spectroscopy in studies in light induced structural changes in mammalian rhodopsin: applicability of solution ^{19}F NMR, *Proc. Natl. Acad. Sci. USA* **96** (1999) 13744-13749

Korchuganov D.S., Nolde S.B., Reibarkh V.Y., Schulga A.A., Ermolyuk Y.S., Kirpichnikov M.P. & Arseniev A.S. NMR study of monomer-dimer equilibrium of barstar in solution, *J. Am. Chem. Soc.* **123** (2001) 2068-2069

Kraulis P.J.: MOLSCRIPT: a program to produceboth detaliled and schematic plots of protein structures, *J. Appl. Crystallogr*. **24** (1991) 946-950

Kummer A.D., Kompa C., Lossau H., Pöllinger-Dammer F., Michel-Beyerle M.E., Silva C., Bylina E., Coleman W., Yang M. & Youvan D.: Dramatic reduction in fluorescence quantum yield in mutants of green fluorescent protein due to fast internal conversion, *Chem. Phys.* **237** (1998) 183-193

Kummer A.D., Wiehler J., Rehaber H., Kompa C., Steipe B. & Michel-Beyerle M.E.: Effect of threonine 203 replacements on excited-state dynamics and fluorescence properties of the green fluorescent protein, *J. Phys. Chem. B.* **104** (2000) 4791-4798

Kwon K., Jiang Y.L., Song F. & Stivers J. T.: ^{19}F NMR Studies of vaccina type IB topoisomerase, *J. Biol. Chem.* **277** (2002) 353-358

Labas Y.A., Gurskaya N.G., Yanushevich Y.G., Fradkov A.F., Lukyanov K.A. & Matz M.V. Diversity and evolution of the green fluorescent protein family, *Proc. Natl. Acad. Sci. U.S.A.* **99** (2002) 4256-4261

Lakowicz J.R. *Principles of fluorescence spectroscopy*, 2nd edition; Kluwer Academic: New York, (1999)

Laskowski R.A.: SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions, *Mol. Graphics* **13** (1995) 323-330

Lau E.Y. & Gerig J.T.: Origins of fluorine NMR chemical shifts in fluorine-containing proteins, *J. Am. Chem. Soc.* **122** (2000) 4408-4417

Lossau H., Kummer A., Heinnecke R., Pöllinger-Dammer F., Kompa C., Bieser G., Jonsson T., Silvia C., Yang M., Youvan D. & Michel-Beyerle M.E.: Time-resolved spectroscopy of wild-type and mutant green fluorescent proteins reveals excited state deprotonation consistent with fluorophore-protein interactions, *Chem.Phys.*, **213** (1996) 1-16

Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G., Markelov M.L. & Lukyanov S.A. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat Biotechnol.* **17 (**1999) 969-73

McConnell H.M.: Reaction rates by nuclear magnetic resonance, *J. Chem. Phys.* **28** (1958) 430

McLachlan A.D.: Gene duplications in the structural evolution of chymotrypsin, *J. Mol. Biol.* **128** (1979) 49-79

McPherson A.: *Crystallization of Biological Macromolecules*, CSHL Press, New York, (1999)

Merritt E.A. & Bacon D.J.: Raster3D: photorealistic molecular graphics, *Methods Enzymol.* **277** (1997) 505-524

Miesenböck G., De Angelis, D.A. & Rothman J.E.: Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins, *Nature* **394** (1998) 192-195

Minks C., Huber R., Moroder L. & Budisa N.: Atomic mutations at the single tryptophan residue of human recombinant annexin V: effects on structure, stability and activity, *Biochemistry* **38** (1999) 10649-10659

Minks C., Alefelder S., Moroder L., Huber R. & Budisa N.: Towards new protein engineering: In vivo building and folding of protein shuttles for drug delivery and targeting by the selective pressure incorporation (SPI) method, *Tetrahedron* **56** (2000) 9431-9442

Mok Y.K., Elisseeva E.L., Davidson A.R. & Forman-Kay J.D.: Dramatic stabilization of an SH3 domain by a single substitution: roles of the folded and unfolded states, *J. Mol. Biol.* **307** (2001) 913-928

Moncrieffe M.C., Juranic N., Kemple M.D., Potter J.D., Macura S. & Prendergast F.G.: Structure-fluorescence correlations in a single tryptophan mutant of carp parvalbumin: solution structure, backbone and side-chain dynamics, *J Mol Biol.* **17** (2000) 147-63

Montelione G.T., Zheng D., Huang Y.J., Gunsalus K.C. & Szyperski T.: Protein NMR spectroscopy in structural genomics, *Nat. Struct. Biol.* **7** (2000) 982-985

Mori S.C., Abeygunawardana C., Johnson M.O. & van Zijl P.C.: Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation, *J. Magn. Reson. B*. **108** (1995) 94-98

Mulder F.A.A., Mittermaier A., Hon B., Dahlquist F.W. & Kay L.E.: Studying excited states of proteins by NMR spectroscopy, *Nature Struct. Biol.* **8** (2001) 932-935

Murzin A.G., Lesk A.M. & Chothia C.: Principles determining the structure of beta-sheet barrels in proteins. I. A theoretical analysis, *J. Mol. Biol.* **236** (1994) 1369-1400

Mühlhahn P., Zweckstetter M., Georgescu J., Ciosto C., Renner C., Lanzendorfer M., Lang K., Ambrosius D., Baier M., Kurth R. & Holak T.A.: Structure of interleukin 16 resembles a PDZ domain with an occluded peptide binding site, *Nat. Struct. Biol.* **5** (1998) 682-686

Nishida Y., Shima F., Sen H., Tanaka Y., Yanagihara C., Yamawaki-Kataoka Y., Kariya K. & Kataoka T.: Coiled-coil interaction of N-terminal 36 residues of cyclase associated protein with adenylyl cyclase is sufficient for its function in Saccharomyces cerevisiae Ras pathway*, J.Biol. Chem.* **43** (1998) 28019-24

Niwa H., Inouye S., Hirano T., Matsuno T., Kojima S., Kubota M., Ohashi M. & Tsuji F.I.: Chemical neture of the light emitter of the Aequorea green fluorescent protein, *Proc. Natl. Acad. Sci. USA*. **93** (1996) 13617-13622

Ormoe M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y. & Remington S.J.: Crystal structure of the Aequorea victoria green fluorescent proteins, *Science* **273** (1996) 1392-1395

Otwinowski Z. & Minor W.: Processing of X-ray diffraction data collected in oscillation mode, *Methods Enzymol.* **276** (1997) 307-326

Palm G.J., Zdanov A., Gaitanaris G.A., Stauber R., Pavlakis G.N. & Wlodawer, A.: The structural basis for spectral variations in green fluorescent protein, *Nat. Struct. Biol.* **4** (1997) 361-365

Palmer III A.G., Wright P.E. & Rance M.: Measurement of relaxation time constants for methyl groups by proton-detected hteronuclear NMR spectroscopy, *Chem. Phys. Lett.* **185** (1991) 41-46

Pellecchia M., Sem D.S. & Wüthrich K.: NMR in drug discovery, *Nat. Rev. Dr. Disc.* **11** (2002) 211-219

Phillips G.N.: *Green Fluorescent Protein*, Wiley-Liss, New York, (1998)

Prendergast F.G.: Biophysics of the green fluorescent protein, *Methods Cell Biol.* **58** (1999) 1-18

Prestegard J.H., Valafar H., Glushka J. & Tian F.: Nuclear Magnetic Resonance in the Era of Structural Genomics, *Biochemistry* **40** (2001) 8677-8685

Rehm T., Huber R. & T.A.: Holak Application of NMR in structural proteomics: Screening for proteins amenable to structural analysis, *Structure*. **10** (2002) 1013-1018

Renner C. & Holak T.A.: NMR N-15 relaxation of the insulin-like growth factor (IGF)-binding domain of IGF binding protein-5 (IGFBP-5) determined free in solution and in complex with IGF-II, *Eur. J. Biochem.* **268** (2001) 1058-1065

Roelofs J., Meima M., Schap P. & Van Haastert P.J.: The Dictyostelium homologue of mammalian soluble adenylyl-cyclase encodes a guanylyl cyclase, *EMBO J.* **16** (2001) 4341-8

Rong D., Lin C.-L.S., d'Avignon D.A., Lovey A.J., Rosenberger M. & Li E.: ^{19}F NMR studies of retinal transfer between cellular retinal binding proteins and phospholipid vesicles, *FEBS Lett.* **402** (1997) 116-120

Ropson I.J. & Frieden C.: Dynamic NMR spectral analysis and protein folding : Identification of a highly populated folding intermediate of rat intestinal fatty acid-binding protein by ^{19}F NMR, *Proc. Natl. Acad. Sci. USA* **89** (1992) 7222-7226

Roswarski D. A., Fedorov A. A., Dodatko T. & Almo S. C.: Crystal structure of the actin-binding domain of cyclase associated protein (CAP) from *Saccharomyces Cerevisiae, to be published*

Sacchetti A, Subramaniam V, Jovin TM & Alberti S.: Oligomerization of DsRed is required for the generation of a functional red fluorescent chromophore, *FEBS Lett.* **14** (2002) 13-9

Sali A.: 100,000 protein structures for the biologist, *Nat. Struct. Biol.* **5** (1998) 1029-1032

Salzmann M., Pervushin K., Wider G., Senn H. & Wüthrich K.: TROSY in tripleresonance experiments: New perspectives for sequential NMR assignment of large proteins, *Proc. Natl. Acad. Sci. U.S.A.* **95** (1998) 13585-13590

Sambrook J. & Russell D.W.: *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York (2001)

Samuel T., Weber O., Rauch P., Verdoodt B., Eppel J., T., McShea A., Hermeking H. & Funk J. O.: The G2/M regulator 14-3-3 sigma prevents apoptosis through sequestration of Bax, *J Biol Chem* **276** (2001) 45201-45206

Sandström J. *Dynamic NMR Spectroscopy,* Academic Press: London, (1982)

Schagger H. & von Jagow G.: Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa, *Anal. Biochem*. **166** (1987) 368-379

Schatz G.W., Reinking J., Zippin J., Nicholson L.K & Vogt V.M. Importance of the N terminus of rous sarcoma virus protease for structure and enzymatic function, *J. Virology* **75** (2001) 4761-4770

Schulz G.E.: β-Barrel membrane proteins. *Curr. Opin. Struct. Biol.* **10** (2000) 443-447

Seifert M.H.J., Ksiazek D., Azim M.K., Smialowski P., Budisa N. & Holak T.A. Slow exchange in the chromophore of a green fluorescent protein variant, *J. Am. Chem. Soc.* **124** (2002) 7932-7942

Shuker S.B., Hajduk P.J., Meadows R.P. & and Fesik S.W.: Discovering high-affinity ligands for proteins: SAR by NMR, *Science* **274** (1996) 1531-1534

Sklenar V., Piotto M., Leppik R. & Saudek V.: Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, *J. Magn. Reson*. **102** (1993) 241-245

Spyracopoulos L. & Sykes B.D.: Thermodynamic insights into proteins from NMR spin relaxation studies, *Curr. Opin. Struct. Biol*. **11** (2001) 555-559

Striker G., Subramaniam V., Seidel C.A.M. & Volkmer A.: Photochromicity and fluorescence lifetimes of green fluorescent protein, *J. Phys. Chem. B* **103** (1999) 8612-8617

Spyracopoulos L. & Sykes B.D.:Thermodynamic insights into proteins from NMR spin relaxation studies, *Curr. Opin. Struct. Biol*. **11** (2001) 555-559

Stoll R., Renner C., Zweckstetter M., Brüggert M., Ambrosius D., Palme S., Engh R.A., Golob M., Breibach I., Buettner R., et al.: The extracellular human melanoma inhibitory activity (MIA) Protein adopts an SH3 domain-like fold, *EMBO J.* **20** (2001) 340-349

Tabor S.: Expression using the T7 RNA polymerase/promoter system, *Current protocols in Molecular Biology* **16** (1990) 2.1-2.11

Terskikh A.V., Fradkov A.F., Zaraisky A.G., Kajava A.V. & Angres B.: Analysis of DsRed Mutants, *J. Biol. Chem.* **10** (2002) 7633-7636

Thompson J.D., Higgins D.G. & Gibson T.J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucl. Acids Res.* **22** (1994) 4673-4680

Tsien R.Y.: The green fluorescent protein, *Annu. Rev. Biochem.* **67** (1998) 509-544

Turk D.: *Weiterentwicklung eines Programms für Molekülgraphik und Elektronendichte-Manipulation und seine Anwendung auf verschiedene Protein-Strukturaufklärungen.* PhD thesis, Technische Univ. München, (1992)

Verkhusha V.V., Akovbian N.A., Efremenko E.N., Varfolomeyev S.D. & Vrzheshch P.V.: Kinetic analysis of maturation and denaturation of DsRed, a coral-derived red fluorescent protein, *Biochem.* **66** (2001) 1342-1351

Voityuk A.A., Michel-Bayerle M.E., Rösch N.: Structure and rotation barriers for ground and excited states of the isolated chromophore of the green fluorescent protein *Chem. Phys. Lett*. **1998**, 296, 269-276

Wachter R.M., Elsiger M.A., Kallio K., Hanson G.T. & Remington S.J.: Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein, *Structure* **6**, (1998) 1267-1277

Wall M.A., Socolich M. & Ranganathan R.: The structural basis for red fluorescence in the tetrameric GFP homolog DsRed, *Nature Struct. Biol.* **7** (2000) 1133-1138

Ward W.W.: In *Bioluminescence and Chemiluminescence: Basic Chemistry and Analytical Applications*; Eds. DeLuca, M.A., McElroy, W.D., pp. 235-242, Academic Press, New York, (1981)

Ward W.W., Prentice H.J., Roth A.F., Cody C.W. & Reeves S. C.: Spectral perturbations of the Aequorea green-fluorescent protein, *Photochem. Photobiol.* **35** (1982) 803-808

Weber W., Helms V., McCammon J.A. & Langhoff P.W.: Shedding light on the dark and weakly fluorescent states of green fluorescent proteins, *Proc. Natl. Acad. Sci. USA* **96** (1999) 6177-6182

Wesp A., Hicke L., Palecek J., Lombardi R., Aust T., Munn A.L. & Riezman H.: End4p/Sla2p interacts with actin-associated proteins for endocytosis in Saccharomyces cerevisiae, *Mol. Biol. Cell* **11** (1997) 2291-2306

Wijesinha-Bettoni R., Dobson C.M. & Redfield C.: Comparison of the Structural and Dynamical Properties of Holo and Apo Bovine α-Lactalbumin by NMR Spectroscopy, *J. Mol. Biol*. **307** (2001) 885-898

Wills Z., Emerson M., Rusch J., Bikoff J., Baum B., Perrimon N. & Van Vactor D.: A *Drosophila* homolog of cyclase associated proteins collaborates with the Abl tyrosine kinase to control midline axon pathfinding, *Neuron* **36** (2002) 611-622

Wishart D.S., Sykes B.D. & Richards F.M.: Relationship between nuclear magnetic shift and protein secondary structure, *J. Mol. Biol.* **222** (1991) 311-333

Wüthrich K. *NMR of proteins and nucleic acids,* John Wiley & Sons: New York, (1994)

Xiao G., Parsons J.F., Tesh K., Armstrong R.N. & Gilliland G.L.: Conformational changes in the crystal structure of rat glutathione transferase M1-1 with global substitution of 3-fluorotyrosine for tyrosine, *J. Mol. Biol*. **281** (1998) 323-339

Yarbrough D., Wachter R.M., Kallio K., Matz M.V. & Remington S.J.: Refined crystal structure of DsRed, a red fluorescent protein from coral, at 2.0 Å resolution, *Proc. Natl. Acad. Sci. USA* **98** (2001) 462-467

Yang F., Moss L.G. & Phillips Jr. G. N.: Molecular structure of the green fluorescent protein,*Nature Biotech.* **14** (1996) 1246-1251

Zacharias D.A., Violin J.D., Newton A.C. & Tsien R.Y.: Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells, *Science* **296** (2002) 913-916

Zhang O. & Forman-Kay J.D.: Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer*, Biochemistry*, **34** (1995) 6784-6794

Zhou G.L., Yamamoto T., Ozoe F., Yano D., Tanaka K., Matsuda H & Kawamuka M.: Identification of 14-3-3 protein from *Lenzinus edodes* that interacts with CAP (adenylyl cyclase-associated protein), and conservation of this interaction in fission yeast, *Biosci. Biotechnol. Biochem*. **64** (2000) 149-159

Zumbusch A. & Jung G.: Single Molecule Spectroscopy of the Green Fluorescent Protein: A Critical Assessment, *Single Mol.* **4** (2000) 261-270

Zurdo J., Guijarro J.I., Jimenez J.L., Saibil H.R. & Dobson C.M.: Dependence on solution conditions of aggregation and amyloid formation by an SH3 domain, *J. Mol. Biol.* **311** (2001) 325-340

9. Appendix: Abbreviations and Symbols

- Å — Ångstrøm (10^{-10}m)
- aa — amino acid
- ATP — adenosine triphosphate
- 1D — one-dimensional
- APS — ammonium peroxodisulfate
- bHLH — basic region and helix-loop-helix region
- bp — base pair
- BSA — bovine serum albumin
- cDNA — complimentary DNA
- CAP — adenylyl cyclase-associated protein
- CFP — cyan fluorescent protein
- COSY — correlation spectroscopy
- δ — chemical shift
- ΔG^{act} — activation energy
- ΔG_0 — reaction free energy
- ΔH_0 — reaction enthalpy
- $\Delta\nu$ — NMR linewidth
- $\Delta\sigma$ — anisotropy of chemical shift tensor
- ΔS_0 — reaction entropy
- Da — Dalton (g mol^{-1})
- DHFR — dihydrofolate reductase
- DMSO — dimethylsulfoxide
- DNA — deoxyribonucleic acid
- DNaseI — deoxyribonuclease I
- DTT — Dithiothreitol
- ECFP — enhanced CFP
- EDTA — ethylenediamine tetraacetic acid
- EGFP — enhanced GFP
- f — NOE enhancement factor
- FID — free induction decay

- 4FW 4-fluoro-tryptophan
- 5FW 5-fluoro-tryptophan
- 6FW 6-fluoro-tryptophan
- G gravity (9.81 m s^{-2})
- GFP green fluorescent protein
- GSH reduced glutathione
- GSSG oxidized glutathione
- GST glutathione S-transferase
- HSQC heteronuclear single quantum coherence
- Hz Hertz
- {I}-S NOE nuclear Overhauser effect on nucleus S by saturating nucleus I
- IPTG isopropyl-β-thiogalactopyranoside
- J spectral density of rotational diffusion
- $^3J_{FH}$ three-bond J-coupling constant
- $^5J_{FH}$ five-bond J-coupling constant
- k_A, k_B off-rates of states A and B
- LB Luria-Broth medium
- M mol l^{-1}
- MAD multiwavelength anomalous diffraction
- MIR multiple isomorphous replacement
- MM minimal medium
- MW molecular weight
- η asymmetry of chemical shift tensor
- NiNTA nickel-nitrilotriacetic acid
- NLS nuclear localization signal
- NMR nuclear magnetic resonance
- NOE nuclear Overhauser effect
- NOESY nuclear Overhauser enhancement spectroscopy
- OD optical density
- p_A, p_B populations of states A and B
- PAGE polyacrylamide gel electrophoresis
- PBS phosphate-buffered saline
- POL DNA polymerase α
- ppm parts per million

- R universal gas constant
- RMSD root mean square deviation
- RNaseA ribonuclease A
- S^2 generalized order parameter
- SAR structure-activity relationship
- SDS sodium dodecyl sulfate
- T_1 longitudinal relaxation time
- T_2 transversal relaxation time
- TEMED N,N,N',N'-tetramethylethylendiamine
- TFA trifluoroacetic acid
- TK thymidine kinase
- TOCSY total correlation spectroscopy
- TS thymidylate synthase
- τ_i correlation time of internal motions
- τ_C overall rotational correlation time
- YFP yellow fluorescent protein
- ω Larmor frequency

Amino acids and nucleotides are abbreviated according to either one or three letter IUPAC code.

Acknowledgements

I would like to thank everyone who contributed to this work.

In particular to Professor Luis Moroder for being my Doktorvater.

To Professor Robert Huber for the opportunity of working in his department and providing me with excellent conditions.

To my supervisor Doctor Tad A. Holak for his many scientific contributions, guidance and interest in my work.

To the crystallography masters, Dr. Hans-Georg Beisel and Dr. Hans Brandstetter for showing me the power of X-ray crystallography.

To the NMR team for their help, advice and good working atmosphere.

I am grateful to Markus Seifert for his excellent cooperation in various projects and his good advices.

I am indebted to Mariusz Kamionka for our many discussions which provided me with help and motivation.

I would like to thank Till for his friendship, strong support, proofreading of this thesis and his optimism.

Finally, I would like to acknowledge my family and Yuri for all the love, care and support during all these years.