ARTICLE

¹H, ¹⁵N and ¹³C assignments of yellow fluorescent protein (YFP) Venus

Shang-Te Danny Hsu · Caroline Behrens · Lisa D. Cabrita · Christopher M. Dobson

Abstract We present here the backbone and side-chain NMR assignments of YFP Venus, a 238-residue protein that emits yellow fluorescence in its native state. Venus is a variant of the green fluorescent protein (GFP), which has improved chromophore maturation and brightness, and the photochemistry and photophysics of which are insensitive to experimental conditions, such as the pH value and buffer content, making it a favourable biomarker.

Biological context

The green fluorescent protein (GFP) from the jelly fish *Aequorea Victoria* is one of the most commonly used biomarkers used in biological research leading to the Nobel Prize in Chemistry 2008 (Shaner et al. 2007; Tsien 1998). It features the spontaneous formation of a chromophore that involves an auto-catalytic cyclisation and subsequent oxidation of the backbone of residues 65–67 without the need for enzymes or added cofactors. This makes GFP an ideal fusion tag for a myriad of applications, such as monitoring protein expression, localisation, and mapping interactions in vivo and in vitro by using the fluorescence of GFP as a reporter. GFP consists of 238 residues that fold into an 11-stranded β-barrel which wraps around the central α-helix—where the chromophore is located and buried in the hydrophobic core—to form a β-can structure. Over the last decade, an array of GFP variants has been identified from natural sources (primarily from marine animals) or engineered additionally by mutagenesis. By altering the side-chain composition of the chromophore itself, i.e., residues 65–67, as well as that of the surrounding residues, the reported GFP variants cover a broad spectrum of fluorescence excitation and emission wavelengths ranging from blue to red, which makes it most versatile for various applications. While the wild type GFP requires a few hours for the chromophore to mature and its spectral properties are highly sensitive to experimental conditions, e.g., pH and salt contents, a yellow fluorescent protein (YFP), Venus, has been developed to yield rapid chromophore maturation, higher quantum yield, and insensitivity to environmental changes (Nagai et al. 2002). These desirable properties of Venus have made it an ideal fluorescence probe in for FRET (Shimozono and Miyawaki 2008) and single molecule studies (Yu et al. 2006).

The usefulness of GFP variants as biomarkers relies on a robust folding capacity of these proteins. A structure-based rational design is a key factor in the further improvement of the folding properties as well as the photochemistry and photophysics of GFP variants (Jackson et al. 2006). While more than 100 crystal structures of GFP variants have been deposited in the Protein Data Bank to date, including that of Venus (Rekas et al. 2002), NMR assignment data are, at present, only available at the backbone level for a single GFP variant, GFPuv (Georgescu et al. 2003; Khan et al. 2003). The availability of backbone assignments renders the identification of highly stable structural elements in GFPuv at a residue-specific level under native and denaturing conditions (Huang et al. 2007). Dynamics studies of the structural perturbations in GFPuv resulting from point mutations have also been reported (Seifert et al. 2003). Moreover, a very recent report that employed 1D ¹H NMR

S.-T. D. Hsu (✉) · C. Behrens · L. D. Cabrita · C. M. Dobson
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
e-mail: stdh2@cam.ac.uk

has identified intriguing indications of folding intermediate of a GFP variant by following the upfield shifted methyl proton resonances, although no side-chain assignments of GFP are currently available (Andrews et al. 2007).

We report here the backbone and side-chain ^1H, ^{15}N and ^{13}C assignments of YFP Venus with an emphasis on the I/L/V side-chain methyl groups. These assignments, in particular of the side-chain reporters, should enable us to obtain structural insights at an atomic resolution to delineate the complex folding processes and the common features within the GFP family.

Methods and experiments

The gene encoding YFP Venus was a kind gift from Prof. Atsushi Miyawaki at the Brain Science Institute, RIKEN, Japan, and was sub-cloned into a pET21 vector with a hexahistidine tag at the N-terminus. ^1H, ^{15}N (>95%), ^{13}C (>95%) and uniformly ^2H (~70%), ^{15}N (>95%), ^{13}C (>95%) labelled protein was expressed in the *E. coli* strain BL21 (DE3) and purified using a Ni-column (Ni-NTA

Superflow; Qiagen) followed by gel filtration chromatography (Superdex 75; GE). Selective ^{13}C labelling, [U-^{15}N, ^{13}C-aa], was achieved by addition of ^{13}C/^{15}N (>98%; Spectral Isotope) labelled amino acids to 0.5 l of ^{15}N-labelled minimal media (1 mM final concentration for each amino acid type) 20 min prior to IPTG induction. The cell cultures were harvested after 1.5 and 3 h of growth at 37°C; prolonged cell growth did not show significant scrambling of the supplemented amino acid types. Selective suppression of ^{15}N labelling, [U-^{15}N, \aa], was achieved by addition of unlabelled amino acids to ^{15}N-labelled minimal media following the same protocol as that for selective ^{13}C labelling. Six amino acid type-specific (un)labellings were made—[U-^{15}N, ^{13}C-I/F/V/L], [U-^{15}N, ^{13}C-V/L], [U-^{15}N, ^{13}C-F, \K], [U-^{15}N, \M\K\I\L], [U-^{15}N, \M\I] and [U-^{15}N, \K\L]—to facilitate residue-specific backbone assignments by comparing the ^{15}N–^1H projections of corresponding HNCO and ^{13}C coupled {^{15}N–^1H} TROSY spectra to identify the sequential (i − 1) and intraresidue correlations of the selectively (un)labelled amino acid types.

Purified protein was then concentrated to ca. 400 μM and buffer exchanged into Tris-HCl, HEPES, MES or

Fig. 1 Overlaid 2D {^{15}N–^1H} HSQC spectrum of Venus recorded at different pH values for optimisation of spectral quality. The spectra were recorded at 37°C using a 700 MHz spectrometer equipped with a cryogenic probe. Weak crosspeaks that are attributed to minor conformations of Gly residues (all 24 Gly have been unambiguously assigned) are indicated by red crosses. Aliased side-chain resonances which were only observed at lower pH values as a result of protonation are indicated by asterisks ($\delta(^1H) = 7$–8 ppm, $\delta(^{15}N) \sim 124$ ppm)
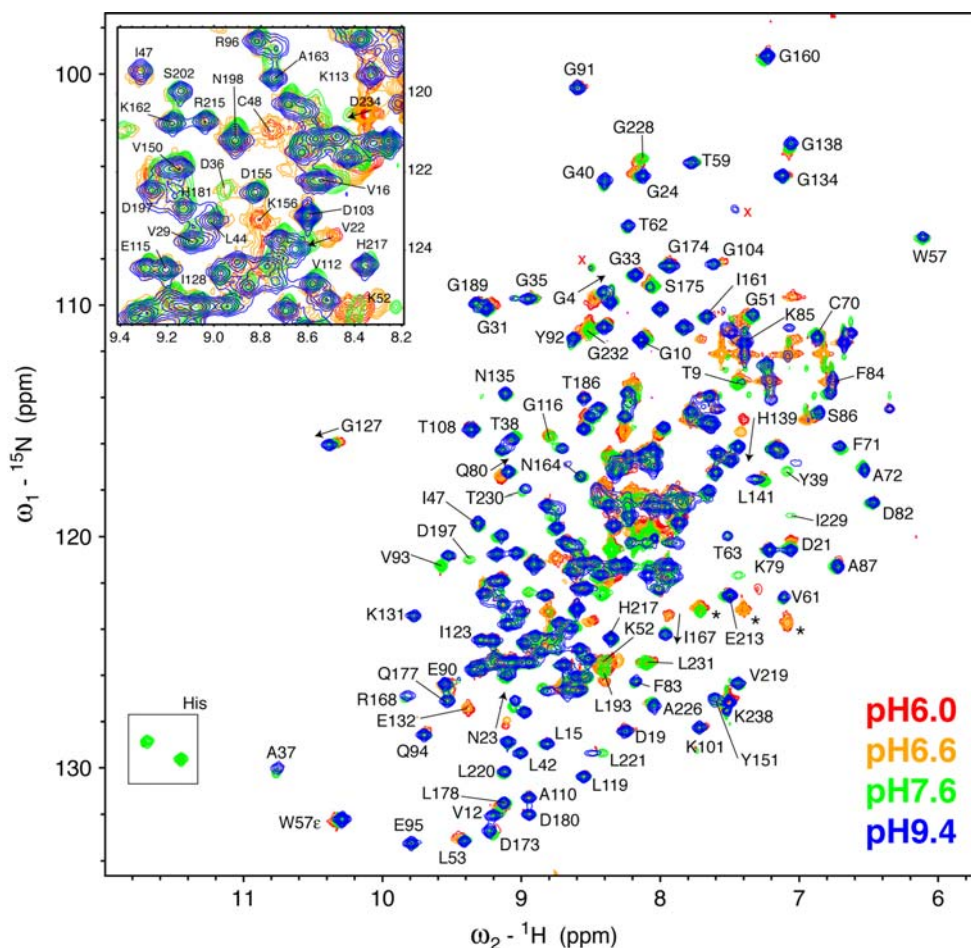
phosphate buffers to yield sample pH pf 6.0, 6.6, 7.6, 8.0, 9.6, in the presence of 0.002% NaN₃. For backbone assignments, TROSY versions of the triple resonance experiments (HNCA, HNCACB, intra-HNCACB, HNCO, HN(CA)CO) were recorded using a ²H(70%)/¹³C/¹⁵N labelled sample. Experiments were performed at 37°C on a Bruker Avance 700 MHz spectrometer equipped with a cryogenic triple resonance probe (Bruker BioSpin). The chemical shifts of individual spin systems (H^N, N, Cα, Cβ and C′) collected manually and the backbone resonance assignments were achieved iteratively through a combination of computer-aided automated assignment using the programme MARS (Jung and Zweckstetter 2004) and visual inspections. For side-chain resonance assignments, 3D ¹H–¹⁵N NOESY-HSQC, 3D ¹H–¹³C NOESY-HSQC, 3D HcCH-COSY, 3D HcCH-TOCY and 3D hCCH-COSY spectra were recorded using the [U-¹⁵N, ¹³C-I/V/L] labelled sample, and HBHA(CO)NH were also recorded using the ²H(70%)/¹³C/¹⁵N labelled sample, and the resonances were assigned manually. All NMR data were processed and analysed by TopSpin (Bruker BioSpin), NMRPipe (Delaglio et al. 1995) and Sparky (Goddard and Kneller) software packages.

Extent of assignment and data deposition

An array of experimental conditions including pH values (ranging from 6.0 to 9.6; Fig. 1) and temperatures (ranging from 20 to 47°C; data not shown) have been tested to obtain an optimal condition for the assignment purpose based on the spectral quality and the number of observed crosspeaks in the {¹⁵N–¹H} TROSY spectra. While some crosspeaks exhibit marked chemical shift changes and intensity changes under different conditions, the overall number of crosspeaks, however, is essentially the same throughout the wide range of tested conditions (Fig. 1). Overall, broader linewidths are observed at lower pH values and low temperatures (data not shown). In particular, the NMR sample of Venus begins to show visible signs of aggregation at pH 6.0, consistent with expectation of the loss of charges when close to the theoretical isoelectric point of 5.85, close to the pKa of the yellow fluorescence of Venus, 5.8, below which point the fluorescence intensity is lost rapidly (Rekas et al. 2002). The backbone assignment of Venus was therefore carried out using primarily data recorded at pH 7.6 and at 37°C; 3D HNCA, 3D HNCO and 3D ¹H–¹⁵N NOESY-HSQC, recorded at pH 6.6 and at

Fig. 2 Structural mapping of the extent of backbone assignment. **a** The crystal structure of Venus (PDB entry 1MYW) is shown in a ribbon representation with unassigned residues coloured in red and labelled with corresponding identities. The chromophore is shown as yellow spheres. **b** Definition of the dimer interface of Venus in the crystalline state. Residues that are involved in intermolecular contact between the monomers are shown in blue with mesh representing the solvent accessible surface. **c** The extent of backbone amide assignments. The primary sequence of Venus is shown with the same colouring scheme as in (**a**). In addition, proline residues are shaded to indicate the locations where sequential connectivities are interrupted
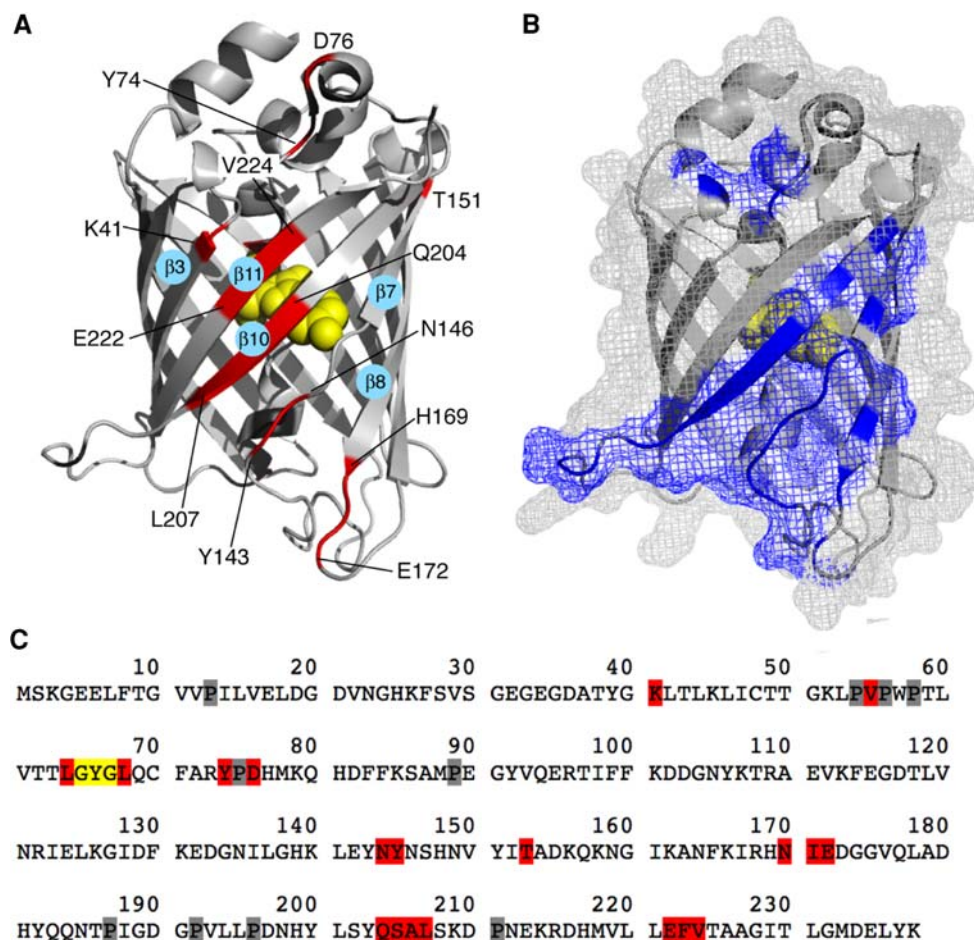
37°C, were used to identify or confirm additional assignments. The condition used here is the closest to those under which Venus is used as a spectroscopic probe in vivo and therefore most useful for future structural studies associated with the in vivo data.

Following a standard sequential assignment procedure, 90% of all assignable $^1H^N$–^{15}N pairs (201 out of 224) and 86% of all $^{13}C'$, $^{13}C\alpha$ and $^{13}C\beta$ (592 out of 690) resonances of Venus have been assigned (Fig. 1). The relatively low level of assignment is in part due to the limited number of backbone amide crosspeaks available for the assignment; essentially all the resolved crosspeaks in the {^{15}N–^1H} TROSY spectrum observed across a wide range of pH and temperature values have been assigned. Structural mapping of the unassigned residues revealed a cluster encompassing parts of strands 3, 7, 8, 10 and 11 (Fig. 2a), forming a surface that largely overlaps with the dimer interface with which Venus self-associates in the crystalline state (Rekas

et al. 2002) (Fig. 2b). Under NMR conditions, the hydrodynamic radius (R_h) derived through pulse field gradient diffusion measurements is ca. 6-fold larger than that of an immunoglobulin domain (14 kDa), while a factor of two is expected for a monomeric Venus in solution. The unexpectedly large R_h is in fact concentration-dependent, suggesting that Venus undergoes transient interactions in solution to form not only dimers but also higher order oligomers which contribute to the loss of resonances corresponding to the unassigned residues.

Note that triply labelled NMR samples (^2H/^{13}C/^{15}N) with partial deuteration ($\sim$70%) were prepared to alleviate the relaxation loss due to the moderate size of Venus (27 kDa), leading to a prolonged ^{15}N transverse relation time of 42.3 $\pm$ 5.9 ms, which is sufficiently long for most triple resonance experiments (unpublished data). The 70% partial deuteration was necessary in order to retain part of the highly protected amide groups whose hydrogen
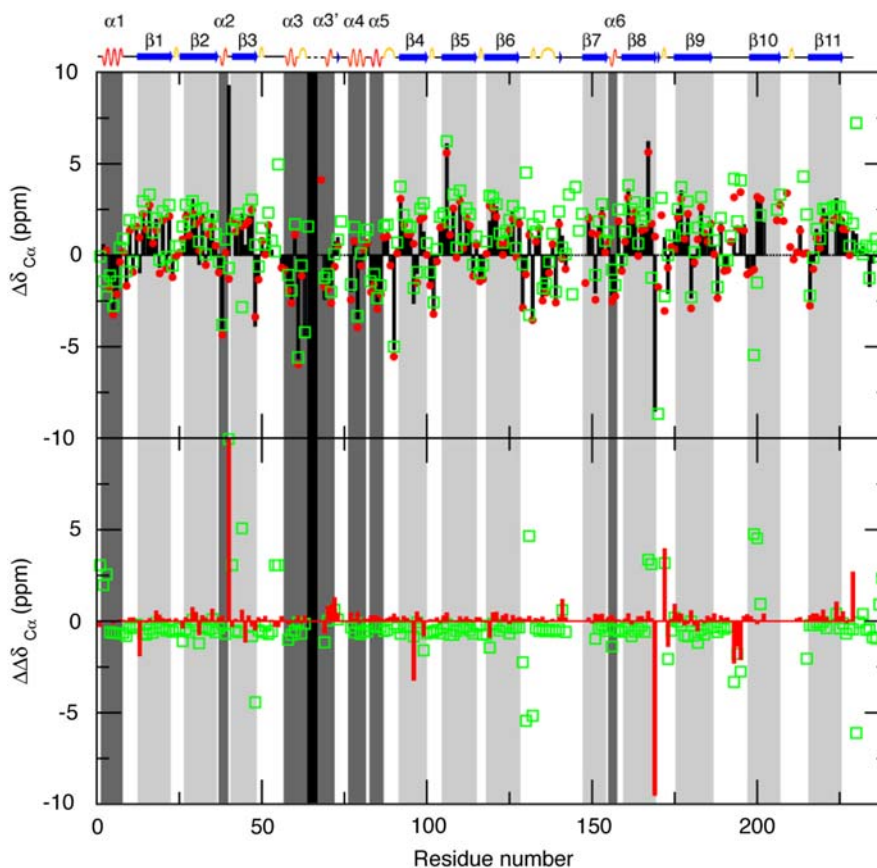

Fig. 3 Comparison of Cα chemical shifts of Venus and GFPuv. *Top panel* Secondary Cα chemical shifts of Venus (*black bar*) and those of GFPuv derived from BMRB entries 5666 (Khan et al. 2003) (*filled red circles*) and 5144 (Georgescu et al. 2003) (*open green squares*), $\Delta\delta_{C\alpha} = \delta_{C\alpha}^{YFP/GFP} - \delta_{C\alpha}^{random\ coil}$, after re-referencing using SHIFT-COR (Zhang et al. 2003). Sequence-dependent random coil chemical shifts were calculated as described previously (Schwarzinger et al. 2001). *Lower panel* Differences of secondary Cα chemical shifts between YFP and different GFPuv assignments, $\Delta\Delta\delta_{C\alpha} = \Delta\delta_{C\alpha}^{YFP} - \Delta\delta_{C\alpha}^{GFP}$. The sets of residues that form α-helices and β-strands in the crystal structure (PDB entry 1MYW) are indicated by *red wiggles* and *blue arrows* at the top of the diagram with labels that are consistent with the nomenclature used in the crystallographic study, and are shaded in *dark* and *light grey boxes*, respectively. Residues 65–67, which form the cyclised chromophore, are highlighted in *black* (Rekas et al. 2002)
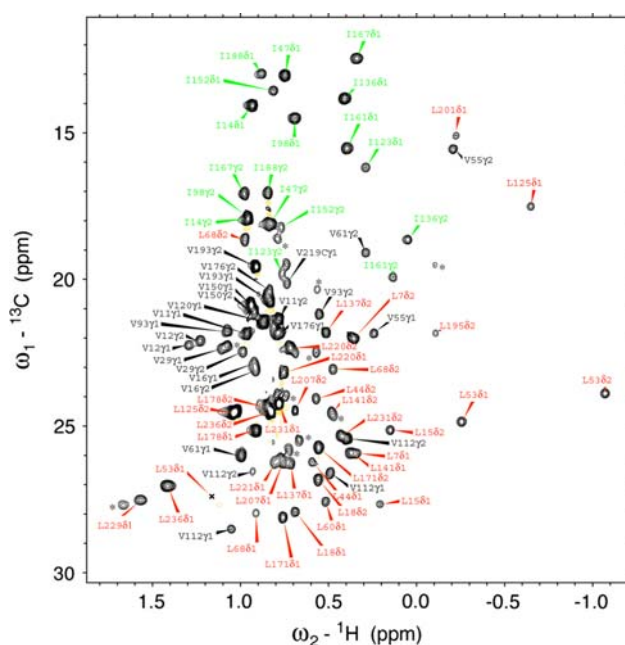
Fig. 4 Methyl region of the 2D {^{13}C–^{1}H} constant-time HSQC spectrum of Venus recorded at 37°C and at 700 MHz. The assignments for Ile, Leu and Val are coloured in green, red and black, respectively

exchange rates are exceeding low even in the presence of chemical denaturants or strong acids, and the refolding of Venus is not efficient (some have half lives of several months; unpublished data). Perdeuteration that requires back-exchange for the amide protons through cycles of unfolding and refolding procedures is therefore not desirable.

Although similar in sequence compositions (differs in 10 out of 238 residues), significant differences can be found in the 2D {^{15}N–^{1}H} HSQC spectra of Venus and GFPuv (not shown). Comparison of the Cα chemical shifts of Venus with respect to the two sets of previously reported values of GFPuv, which report secondary structure contents, revealed systematic offsets. We therefore re-referenced the deposited assignments, BMRB entries 5144 (Georgescu et al. 2003) and 5666 (Khan et al. 2003), using the online server SHIFTCOR (http://redpoll.pharmacy.ualberta.ca/shiftcor/ (Zhang et al. 2003)), and corrections of 3.08 and −0.30 ppm, respectively, were carried out to yield two similar overall profiles of the secondary Cα chemical shifts of GFPuv that are in general agreement with that of Venus. The differences of secondary Cα chemical shifts between YFP and GFPuv, $\Delta\Delta\delta_{C\alpha} = \Delta\delta_{C\alpha}^{YFP} - \Delta\delta_{C\alpha}^{GFP}$, are -0.26 ± 1.55 and 0.13 ± 1.17 ppm for BMRB entries 5144 and 5666, respectively (Fig. 3). The two assignments were obtained under very similar conditions, both in phosphate buffered saline (pH 7.0 for BMRB 5144 and pH 7.2 for BMRB 5666) and at 310 K. Close inspection into the

differences between the two, however, revealed marked deviations for BMRB entry 5144 from those of Venus (determined at pH 7.4), particularly in loop regions connecting individual β-strands, suggesting some degrees of pH-dependent conformational rearrangements in GFP. Common in these two GFPuv assignments, in relation with the observed deviations from those of Venus, are the Cα shifts of T38, H169, E172, P192 and L194, all of which are conserved and are located in the loop regions. Overall, the deviations of Cα chemical shifts between GFPuv and Venus suggest that the solution structures of the two GFP variants are highly similar and that marginal conformational differences occur only in the loop regions.

In addition to the backbone assignments, we assigned nine out of 10 alanine side-chain methyl groups through a combination of HBHA(CO)NH and 3D ^{1}H–^{15}N NOESY-HSQC. Additionally, 82% (84 out of 102) of all side-chain methyl groups in Ile, Val and Leu (most of the unassigned resonances are located in the loop regions which exhibit severe spectral overlaps) were assigned using a [U-^{15}N, ^{13}C-I/F/V/L] labelled sample (Fig. 4). Future studies are underway to investigate in detail the structural and dynamical characteristics of these side-chain methyl groups. The assigned chemical shifts have been deposited in the BMRB under accession number 15826.

References

Andrews BT, Schoenfish AR, Roy M, Waldo G, Jennings PA (2007) The rough energy landscape of superfolder GFP is linked to the chromophore. J Mol Biol 373:476–490. doi:10.1016/j.jmb.2007.07.071

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293. doi:10.1007/BF00197809

Georgescu J, Rehm T, Wiehler J, Steipe B, Holak TA (2003) Backbone HN, C, Ca and Cb assignment of the GFPuv mutant. J Biomol NMR 25:161–162. doi:10.1023/A:1022296413190

Goddard TD, Kneller DG Sparky 3. University of California, San Francisco

Huang JR, Craggs TD, Christodoulou J, Jackson SE (2007) Stable intermediate states and high energy barriers in the unfolding of GFP. J Mol Biol 370:356–371. doi:10.1016/j.jmb.2007.04.039

Jackson SE, Craggs TD, Huang JR (2006) Understanding the folding of GFP using biophysical techniques. Expert Rev Proteomics 3:545–559. doi:10.1586/14789450.3.5.545

Jung YS, Zweckstetter M (2004) Mars—robust automatic backbone assignment of proteins. J Biomol NMR 30:11–23. doi:10.1023/B:JNMR.0000042954.99056.ad

Khan F, Stott K, Jackson S (2003) ^{1}H, ^{15}N and ^{13}C backbone assignment of the green fluorescent protein (GFP). J Biomol NMR 26:281–282. doi:10.1023/A:1023817001154

Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. Nat Biotechnol 20:87–90. doi:10.1038/nbt0102-87

Rekas A, Alattia JR, Nagai T, Miyawaki A, Ikura M (2002) Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity. J Biol Chem 277:50573–50578. doi:10.1074/jbc.M209524200

Schwarzinger S, Kroon GJ, Foss TR, Chung J, Wright PE, Dyson HJ (2001) Sequence-dependent correction of random coil NMR chemical shifts. J Am Chem Soc 123:2970–2978. doi:10.1021/ja003760i

Seifert MH, Georgescu J, Ksiazek D, Smialowski P, Rehm T, Steipe B, Holak TA (2003) Backbone dynamics of green fluorescent protein and the effect of histidine 148 substitution. Biochemistry 42:2500–2512. doi:10.1021/bi026481b

Shaner NC, Patterson GH, Davidson MW (2007) Advances in fluorescent protein technology. J Cell Sci 120:4247–4260. doi:10.1242/jcs.005801

Shimozono S, Miyawaki A (2008) Engineering FRET constructs using CFP and YFP. Methods Cell Biol 85:381–393. doi:10.1016/S0091-679X(08)85016-9

Tsien RY (1998) The green fluorescent protein. Annu Rev Biochem 67:509–544. doi:10.1146/annurev.biochem.67.1.509

Yu J, Xiao J, Ren X, Lao K, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. Science 311:1600–1603. doi:10.1126/science.1119623

Zhang H, Neal S, Wishart DS (2003) RefDB: a database of uniformly referenced protein chemical shifts. J Biomol NMR 25:173–195. doi:10.1023/A:1022836027055