

# Tubulin Secondary Structure Analysis, Limited Proteolysis Sites, and Homology to FtsZ<sup>†</sup>

José M. de Pereda, Daniel Leynadier, Juan A. Evangelio, Pablo Chacón, and José M. Andreu\*

*Centro de Investigaciones Biológicas, CSIC, Velázquez 144, 28006 Madrid, Spain*

*Received June 7, 1996; Revised Manuscript Received September 6, 1996*<sup>⊗</sup>

**ABSTRACT:** The far-ultraviolet circular dichroism spectrum of the  $\alpha\beta$ -tubulin dimer analyzed by six different methods indicates an average content of approximately 33%  $\alpha$  helix, 21%  $\beta$  sheet, and 45% other secondary structure. Deconvolution of Fourier transform infrared spectra indicates 24% sheet, 37% (maximum) helix, and 38% (minimum) other structure. Separate alignments of 75  $\alpha$ -tubulin, 106  $\beta$ -tubulin, and 14  $\gamma$ -tubulin sequences and 12 sequences of the bacterial cell division protein FtsZ have been employed to predict their secondary structures with the multiple-sequence method PHD [Rost, B., & Sander, C. (1993a) *J. Mol. Biol.* 232, 584–599]. The predicted secondary structures average of 33%  $\alpha$  helix, 24%  $\beta$  sheet, and 43% loop for the  $\alpha\beta$  dimer. The predictions have been compared with sites of limited proteolysis by 12 proteases at the surfaces of the heterodimer and taxol-induced microtubules [de Pereda, J. M., & Andreu, J. M. (1996) *Biochemistry* 35, 14184–14202]. From 24 experimentally determined nicking sites, 18 are at predicted loops or at the extremes of secondary structure elements. Proteolysis zone A (including acetylatable Lys40 and probably Lys60 in  $\alpha$ -tubulin and Gly93 in  $\beta$ -tubulin) and proteolysis zone B (extending between residues 167 and 183 in both chains) are accessible in microtubules. Proteolysis zone C, between residues 278 and 295, becomes partially occluded in microtubules. The  $\alpha$ -tubulin nicking site Arg339-Ser340 is at a loop following a predicted  $\alpha$  helix in proteolysis zone D. This site is protected in taxol microtubules; however, a new tryptic site appears which is probably located at the N-terminal end of the same helix. Zone D also contains  $\beta$ -tubulin Cys354, which is accessible in microtubules. Proteolysis zone E includes the C-terminal hypervariable loops (10–20 residues) of each tubulin chain. These follow the two larger predicted helical zones (residues 372–395 and 405–432 in  $\beta$ -tubulin), which also are the longer conserved part of the  $\alpha$ - and  $\beta$ -tubulin sequences. Through combination of this with other biochemical information, a set of surface and distance constraints is proposed for the folding of  $\beta$ -tubulin. The FtsZ sequences are only 10–18% identical to the tubulin sequences. However, the predicted secondary structures show two clearly similar (85–87 and 51–78%) regions, at tubulin positions 95–175 and 305–350, corresponding to FtsZ 65–135 and 255–300, respectively. The first region is flanked by tubulin proteolysis zones A and B. It consists of a predicted loop1–helix–loop2–sheet–loop3–helix–loop4–sheet fold, which contains the motif (KR)GXXXXG (loop1), and the tubulin–FtsZ signature G-box motif (SAG)GGTG(SAT)G (loop3). A simple working model envisages loop1 and loop3 together at the nucleotide binding site, while loops 2 and 4 are at the surface of the protein, in agreement with proteolytic and antigenic accessibility results in tubulin. The model is compatible with studies of tubulin and FtsZ mutants. It is proposed that this region constitutes a common structural and evolutionary nucleus of tubulins and FtsZ which is different from typical GTPases.

The  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulins are eukaryotic proteins of about 450 residues with characteristic GTP binding motifs, which are different from typical GTPases (Sage et al., 1995; Bourne et al., 1991). The  $\alpha\beta$ -tubulin dimer ( $M_r = 100\,000$  Da) assembles dynamically forming microtubules, long hollow cylindrical polymers which are essential for cell division. The  $\alpha$ - and  $\beta$ -tubulin families consist of different isotypes, which are highly conserved, differentially expressed, and post-translationally modified (Little & Seehaus, 1988; Mandelkow & Mandelkow, 1995). The exchangeable GTP bound to the  $\beta$  subunit is hydrolyzed to GDP as a result of

assembly. Tubulin liganded to GDP is inactive for microtubule assembly and easily forms rings related to the curled protofilaments of depolymerizing microtubule ends (Melki et al., 1990; Díaz et al., 1994; Hyman & Karsenti, 1996).  $\gamma$ -Tubulin is essential for microtubule nucleation at the centrosome (Oakley, 1992; Zheng et al., 1995; Moritz et al., 1995). A highly divergent  $\gamma$ -tubulin gene is essential for cell division and microtubule organization in *Saccharomyces cerevisiae* (Sobel & Snyder, 1995).

FtsZ is a bacterial protein essential for cell division, which stays in the cytoplasm for most of the cell cycle (Plá et al., 1991) and localizes in the septum zone at the time of division (Bi & Lutkenhaus, 1991; Lutkenhaus, 1993). The members of the FtsZ family share several characteristic short sequences with tubulins (Mukherjee & Lutkenhaus, 1994), in particular the unique G-box motif (SAG)GGTG(SAT)G and the (inverted) P-loop (KR)GXXXXG, both of which have been related to the sequence motif GXXXXGK(ST) which is

<sup>†</sup> This work was supported in part by DGICYT Grant PB920007 and a predoctoral fellowship from Caja Madrid to J.M.d.P. D.L. held postdoctoral fellowships from MEC and the European Commission. J.A.E. and P.C. had predoctoral fellowships from Comunidad de Madrid and MEC, respectively.

\* Author to whom correspondence should be addressed. Fax: 34-1-5627518. E-mail: cibjm07@cc.csic.es.

<sup>⊗</sup> Abstract published in *Advance ACS Abstracts*, October 15, 1996.

involved in binding the phosphates of the nucleotide in GTPases (Bourne et al., 1991). However, the overall sequence homology between FtsZ and tubulins is low. The sequence of a divergent *ftsZ* gene from the archaeobacterium *Halobacterium salinarum* has recently been reported, which is more related to tubulins than the *ftsZ* genes from eubacteria (Margolin et al., 1996). FtsZ proteins bind and hydrolyze GTP in a protein concentration- and cation-dependent manner (RayChaudhuri & Park, 1992; de Boer et al., 1992; Mukherjee et al., 1993). GTP and magnesium induce FtsZ to polymerize into large structures (Mukherjee & Lutkenhaus, 1994; Bramhill & Thompson, 1994). Moreover, it has been shown that FtsZ can assemble into protofilament sheets which have intermolecular spacings very similar to those of the microtubule lattice and also into minirings which resemble those observed upon microtubule depolymerization (Erickson et al., 1996).

Taken together, the similarities between both protein families suggest that FtsZ may be a prokaryotic cytoskeletal protein homolog of tubulin [for reviews, see Erickson (1995) and Vicente and Errington (1996)] and support the hypotheses that ancient FtsZ might have evolved into tubulin and protofilament sheets into microtubules (Margolin et al., 1996; Erickson et al., 1996). It appears that both proteins might actually be much more similar than previously expected. Therefore, the high-resolution structure of one of them might give clues to the structure of the other; however, no success with any of both proteins has been reported to date. On the other hand, approximately 200 tubulin and about 12 FtsZ gene sequences from diverse organisms are presently known, and a vast amount of biochemical knowledge is available, particularly for  $\alpha\beta$ -tubulin (Ludueña et al., 1992; Monasterio et al., 1995). In the preceding paper (de Pereda & Andreu, 1996), we have reported the mapping of a number of surface peptide bonds of the  $\alpha\beta$ -tubulin dimer from mammalian brain and taxol-induced microtubules, which are cleaved by a panel of 12 different proteases. The purpose of the present study is twofold: first, to predict the secondary structure of tubulin in comparison with experimentally determined proteolysis sites and, second, to unravel possible similarities between the structures of tubulins and FtsZ. Since multiple-sequence comparisons may reveal evolutionary and significant structural information, we have employed multiple-sequence alignments to predict the secondary structures (Rost & Sander, 1993a,b) of tubulins and FtsZ. This has revealed unexpected similar predicted secondary structures in substantial parts of their sequences. This prediction is compared with the average secondary structure of the  $\alpha\beta$  dimer estimated with up-to-date circular dichroism and FTIR<sup>1</sup> analysis methods. The secondary structure prediction is combined with, and tested against, the experimental proteolysis results and other biochemical and genetic information; partial structural working models are proposed for the overall folding of the  $\beta$ -tubulin chain and for a common nucleus of tubulin and FtsZ which participates in GTP binding.

## MATERIALS AND METHODS

**Materials.** Tubulin was purified from calf brain, equilibrated in sodium phosphate buffer, and measured spectrophotometrically as in the preceding paper; other materials were as described (de Pereda & Andreu, 1996).

**Circular Dichroism.** CD<sup>1</sup> spectra of 20  $\mu$ M tubulin were acquired with a Jasco J-720 dichrograph employing 0.01 cm cells at 25 °C; four scans between 185 and 260 nm (bandwidth of 1 nm) were averaged. The instrument was calibrated between 180 and 350 nm with ammonium camphorsulfonate-*d*<sub>10</sub> (Katayama Chemical, Jasco) as a standard substance (de Tar, 1969). Ellipticity was calculated with a mean residue weight of 110.

The estimation of the percentage of each secondary structure by analyzing protein CD spectra depends on the method and the choice of standards. Therefore, the tubulin CD was analyzed employing six different methods presently available at our laboratory. The least-squares method of Yang et al. (1986) was employed to fit the experimental CD from 190 to 240 nm with a linear combination of four standard spectra of  $\alpha$  helix,  $\beta$  sheet, turn, and random coil components extracted from a data set of 15 proteins of known three-dimensional structure. The convex constraint analysis (CCA; Perczel et al., 1991, 1992a) uses no three-dimensional standards and was employed to extract singular components from a data set including the 195–240 nm CD spectra of 25 proteins plus tubulin; alternatively, the problem CD spectrum was analyzed as a linear combination of the five-reference curve set obtained by the convex constraint analysis of the 25 reference spectra, employing the Lincomb program (Perczel et al., 1992b). With the singular value decomposition method Varselect (Manavalan & Johnson, 1987), a variable-selection procedure was applied to the results obtained by systematically removing CD spectra from an initial basis set of 33 proteins. With the self-consistent Selcon method (Sreerama & Woody, 1993), the initial basis set was formed by the tubulin CD spectrum to be analyzed and all the spectra similar to it from a data set of 24 standard proteins, and an initial guess of the unknown secondary structure was made; the solution replaced the initial guess, and the procedure was repeated until convergence. The K2D program (Andrade et al., 1993) is a nonlinear method which was employed to evaluate the secondary structure of tubulin with a neural network approach and extract the secondary structure information from a data set of CD spectra of 24 standard proteins from 200 to 240 nm.

**Fourier Transform Infrared Spectroscopy.** Tubulin was equilibrated into D<sub>2</sub>O buffer by Sephadex G-25 gel chromatography (final concentrations, 20 and 50  $\mu$ M), and FTIR spectra were obtained (2000–1000  $\text{cm}^{-1}$  with a nominal resolution of 2  $\text{cm}^{-1}$ ; 100–500 interferograms) with Perkin-Elmer 1725 and 2000 interferometers equipped with MTC detectors, employing CaF<sub>2</sub> windows and 25–40  $\mu$ m spacers at 25 °C. The absorbance of the buffer and atmospheric water vapor were subtracted from the data; optionally, the contribution of the Arg, Asn, Gln, and Tyr side chains (Chirgazde et al., 1975) of tubulin were subtracted. Overlapping bands were resolved by Fourier self-deconvolution (Lorentzian line shape of 30  $\text{cm}^{-1}$  full width at half-height, with Gaussian line shape for apodization and a resolution enhancement factor of 1.8). Band assignment and secondary structure estimation by least-squares spectrum fitting were

<sup>1</sup> Abbreviations: CD, circular dichroism; FTIR, Fourier transform infrared spectroscopy; TR, trypsin; CH, chymotrypsin; EL, elastase; CL, clostripain; LC, protease lysine-C; TH, thermolysin; V8, protease V8 from *Staphylococcus aureus*; PA, papain; SB, subtilisin; K, proteinase K; DN, proteinase aspartic-N; BR, bromelain.

made as described (Goormaghtigh et al., 1990), considering known uncertainties (Goormaghtigh et al., 1994; Surewicz et al., 1993; Arrondo et al., 1993).

**Sequence Analysis and Prediction Methods.** The tubulin and FtsZ sequences available in the SWISS-PROT database (March 30, 1996) (Bairoch & Boeckmann, 1993) were used for the multiple-sequence alignments. The sequences in the  $\alpha$ -,  $\beta$ -,  $\gamma$ -tubulin and FtsZ families were separately aligned with the program ClustalW (version 1.6) (Thompson et al., 1994). The protein sequences of porcine brain  $\alpha$ -tubulin (TBA\_PIG) (Ponstingl et al., 1981) and  $\beta$ -tubulin (TBB\_PIG) (Krauhns et al., 1981) and the translated sequences of the human  $\gamma$ -tubulin (TBG\_HUMAN) (Zheng et al., 1991) and *Escherichia coli* FtsZ genes (FTSZ\_ECOLI) (Yi & Lutkenhaus, 1985) were chosen as guide sequences (first sequences in each alignment), and their numbering is used throughout. The conservation of sequences within each group was examined on these alignments. The positional sequence variation within each aligned family was measured as described (Sander & Schneider, 1991; the PHD server was employed). Note that the  $\alpha$ - and  $\beta$ -tubulin sequences outnumber by far the  $\gamma$ -tubulin and FtsZ ones, and therefore, the variation results are not directly comparable.

Predictions of secondary structure and solvent accessibility were obtained from the PHD server (Rost et al., 1994a). The PHD method (Rost & Sander, 1993a,b, 1994a,b) uses the information contained in multiple-sequence alignments as input to a neural network trained with a nonredundant protein structure data base. It has been reported to have the best predictive accuracy from only sequence information (Rost & Sander, 1995). For the tubulins and FtsZ predictions, we used the four aligned groups of sequences as separate inputs, employing all the available sequences in each case, in order to obtain maximum variation in the multiple alignments. The insertions in the guide sequences were removed. In order to check the stability of the secondary structure predictions, different PHD runs were made with alignments in which the number of sequences had been reduced by deleting the fragments and the more divergent sequences from each class. The results pointed out a low predictive variability, which affected mainly small elements, the extremes of longer ones, and the C termini of tubulin and FtsZ. The similarity of the four predicted secondary structures was measured per residue ( $Q_3$ ) and per overlapping segment (Sov) (Rost et al., 1994b) employing the PHDaccu tool. Predictions on the single-glycine sequences were made with the segment method SSP (Solovyev & Salamov, 1994). The local helical propensity of the tubulin and FtsZ polypeptide chains was evaluated with the program Agadir (Muñoz & Serrano, 1994), which is a statistical method based on experimental data of amino acid contributions to the stability of isolated helices. The amphipatic  $\alpha$  helical character of the sequences was evaluated with the hydrophobic moment method (Eisenberg et al., 1984; Peplot program, University of Wisconsin Genetic Computer Group, Inc.) and also with a graphical representation of the peptide chain on helical wheels.

## RESULTS

**Spectroscopic Analysis of the Average Secondary Structure of the  $\alpha\beta$ -Tubulin Dimer.** Figure 1 shows the far-ultraviolet circular dichroism spectrum of purified calf brain tubulin. This spectrum is comparable to those from earlier tubulin

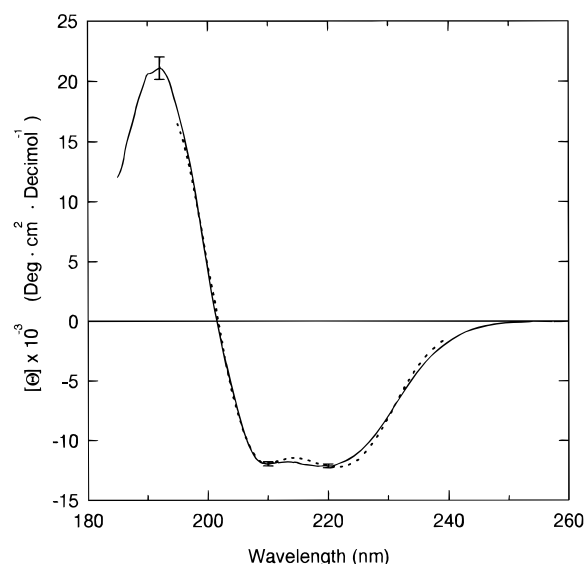


FIGURE 1: Circular dichroism of purified bovine brain tubulin in 10 mM sodium phosphate and 0.1 mM GTP of pH 7.0 and in the same buffer plus 6 mM  $MgCl_2$  at pH 6.7. The spectrum shown is the actual unsmoothed average of four samples, two in each buffer. The ellipticity values were  $21\,280 \pm 950$  (192 nm maximum),  $-12\,080 \pm 180$  (210 nm minimum), and  $-12\,240 \pm 170$  (220 nm minimum)  $\text{deg cm}^2 \text{dmol}^{-1}$ . The dashed line is a representative model spectrum generated with the CD analysis method CCA (Percezel et al., 1991, 1992a).

CD studies (Lee et al., 1978; Andreu & Timasheff, 1982; Andreu et al., 1986); however, the signal to noise ratio is higher, and the overall accuracy is better. Methods of analysis of secondary structure from CD are very useful for characterizing changes but have limited accuracy in absolute terms. The  $\alpha$  helical fraction is generally estimated with good accuracy, but the accuracy for  $\beta$  sheet and other components is very variable (Woody, 1995; Greenfield, 1996). The analysis of the secondary structure contents with six methods is presented in Table 1. It can be observed that the estimates from these different methods are not very divergent. The average over all methods suggests that tubulin contains ca. 33%  $\alpha$  helix, 21%  $\beta$  sheet, and 45% other structure, according to CD spectroscopy. Note that averaging in this case is a simplifying procedure, since the reliability of each method relative to the others has not been considered.

The amide I' Fourier transform infrared absorption spectrum of tubulin in  $D_2O$  buffer is shown in Figure 2. Tubulin CD spectra in  $H_2O$  and  $D_2O$  buffers were identical (not shown). The assignment of the components of the deconvoluted spectra is shown in Table 2. According to this method (Goormaghtigh et al., 1990), tubulin contains 24%  $\beta$  sheet, a maximum of 37%  $\alpha$  helix, and a minimum of 38% other structure, which is fully compatible with the CD analysis (Table 1). However, using an earlier band assignment table (Susi & Byler, 1986) gives 30% sheet, a maximum of 21% helix, and a minimum of 48% other structure (not shown). It is difficult to get an estimate of the accuracy of the FTIR-derived secondary structure. These methods have inherent limitations different from those of CD, including the extent of hydrogen isotope exchange and the assumption of identical absorption coefficients for the different secondary structures (Surewicz et al., 1993; Arrondo et al., 1993; Goormaghtigh et al., 1994). A conclusion combining the CD and FTIR analyses that is probably safe

Table 1: Tubulin Secondary Structure Content Estimated from Far-UV CD Spectra

method of analysis	helix	extended $\beta$ sheet			other		
		parallel	antiparallel	total sheet	$\beta$ -turn	unordered	total
Yang <sup>a</sup>	0.39			0.21	0.12	0.28	0.46
Lincomb <sup>d</sup>	0.20		0.20	0.20	0.29	0.31	0.60
CCA (five components) <sup>e</sup>	0.31			0.31	0.24	0.15	0.39
Varselct <sup>e</sup>	0.33	0.05	0.15	0.20	0.16	0.29	0.45
Selcon <sup>f</sup>	0.36	0.06	0.09	0.15	0.26	0.22	0.48
K2D <sup>b</sup>	0.41			0.21			0.38
average	0.33 $\pm$ 0.07			0.21 $\pm$ 0.05	0.21 $\pm$ 0.06	0.25 $\pm$ 0.06	0.45 $\pm$ 0.08

<sup>a</sup> Yang et al. (1986). <sup>b</sup> Andrade et al. (1993). <sup>c</sup> Perczel et al. (1991, 1992a). <sup>d</sup> Perczel et al. (1992b) (this method includes the contribution of the parallel  $\beta$  sheet with the  $\beta$  turn component; therefore, it may overestimate the  $\beta$  turn content at the expense of the  $\beta$  sheet content). <sup>e</sup> Manavalan and Johnson (1987). <sup>f</sup> Sreerama and Woody (1993).

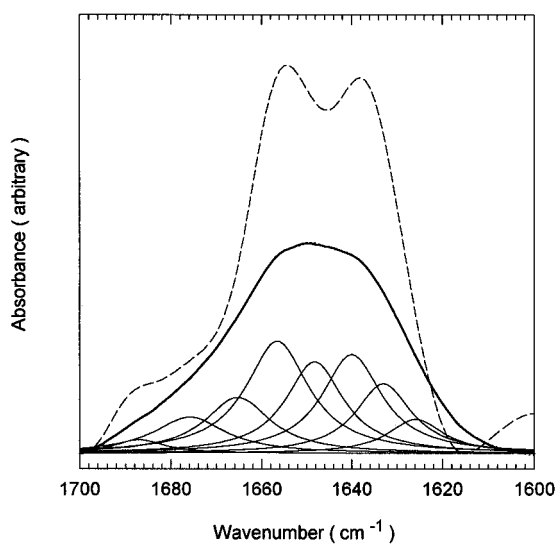


FIGURE 2: Fourier transform infrared spectrum of purified bovine brain tubulin equilibrated in 10 mM sodium phosphate and 0.1 mM GTP buffer in D<sub>2</sub>O at pH 7.4 and 25 °C (the solvent spectrum has been subtracted). This sample had been in D<sub>2</sub>O at 4 °C for 3 h before acquiring the spectrum. The areas under the amide I', amide II', and amide II bands were 3280, 2420, and 630 arbitrary units, respectively: thick line, experimental spectrum; dashed line, deconvoluted spectrum (resolution enhancement factor of 1.8); and thin lines, eight bands employed to fit the experimental spectrum (Goormaghtigh et al., 1990), whose sum is indistinguishable from it.

is that tubulin contains around 33%  $\alpha$  helix, 24%  $\beta$  sheet, and 42% other structure.

*Predictive Analysis of the Secondary Structures and Solvent Exposure of the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -Tubulin and FtsZ Sequences.* Figure 3 shows the alignment and the predicted secondary structures of the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulin and FtsZ families of proteins. Only one representative guide sequence is shown from 75, 106, 14, and 12 sequences respectively employed. In contrast to classical methods of secondary structure prediction using single sequences, for which the secondary structure potentials of most of the tubulin sequences are low (de la Viña et al., 1988), the multiple-sequence PHD method employed (Rost & Sander, 1995) gives high reliability indexes for a substantial part of the sequences. The secondary structures predicted with higher than 82% expected average accuracy according to this method are drawn in black, while those predicted with lower expected accuracy are drawn in gray in Figure 3 (note that the overall reported accuracy of the method is 72  $\pm$  9%). The SSP method (Solovyev & Salamov, 1994) applied to the single-guide sequences gave a generally compatible

Table 2: Tubulin Secondary Structure Content Estimated from FTIR Spectra

wavenumber (cm <sup>-1</sup> )	assignment <sup>a</sup>	fraction <sup>a</sup>
1685	$\beta$ sheet	0.03
1674	turn	0.09
1664	turn	0.11
1656	$\alpha$ helix <sup>b</sup>	0.21
1648	$\alpha$ helix/unordered <sup>c</sup>	0.16
1640	unordered	0.18
1632	$\beta$ sheet	0.13
1625	$\beta$ sheet	0.08
	total $\alpha$ helix (maximum)	0.37
	total $\beta$ sheet	0.24
	total other structure (minimum)	0.38

<sup>a</sup> Assignments of FTIR spectra in D<sub>2</sub>O (see Figure 2) were made according to Goormaghtigh et al. (1990, 1994) (see Materials and Methods). Numbers given are the average of two independent measurements with different instruments, subtracting and not subtracting the side chain absorption in each case (standard deviations were  $\leq$  0.01). Addition of 6 mM MgCl<sub>2</sub> to the tubulin samples nonsignificantly modified the results. <sup>b</sup> Note that any residual nondeuterated structure would also absorb in this zone. <sup>c</sup> This band was consistently observed by straight deconvolution of spectra with resolution enhancement factors above 2.

prediction of helical and strand segments (not shown). Most helices predicted by PHD exhibited an intrinsic helical potential detected in the guide sequences by the independent peptide helicity analysis program Agadir (Muñoz & Serrano, 1994); exceptions included the  $\beta$ -tubulin helices H11 and H13 and the FtsZ helices aligned with tubulin helices H5 and H6 in Figure 3 (data not shown). The prediction for buried (b) or exposed (e) residues is shown under the secondary structure scheme. The average secondary structure predicted for  $\alpha$ - and  $\beta$ -tubulin, shown in Table 3, is coincident with the CD-FTIR analysis of secondary structure (Tables 1 and 2). Predicting the secondary structure of three subfamilies and of a distant relative is to our knowledge a new application of the PHD method. The sequences within the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulin and FtsZ groups are 53–99, 72–99, 44–98, and 45–57% identical to their respective guide sequences. The  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulin families are typically 32–43% identical among them. The overall sequence homology of FtsZ to the tubulins is low (typically 10–18% identical residues in this alignment); however, homology is apparent in several short motifs [the putative phosphate and Mg<sup>2+</sup> binding motifs found in tubulin and FtsZ are underlined in Figure 3; see RayChaudhuri and Park (1994)].  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulin have similar predicted secondary structures, saving local differences, with levels of identity of 65–72% per residue, while the secondary structure predicted for the

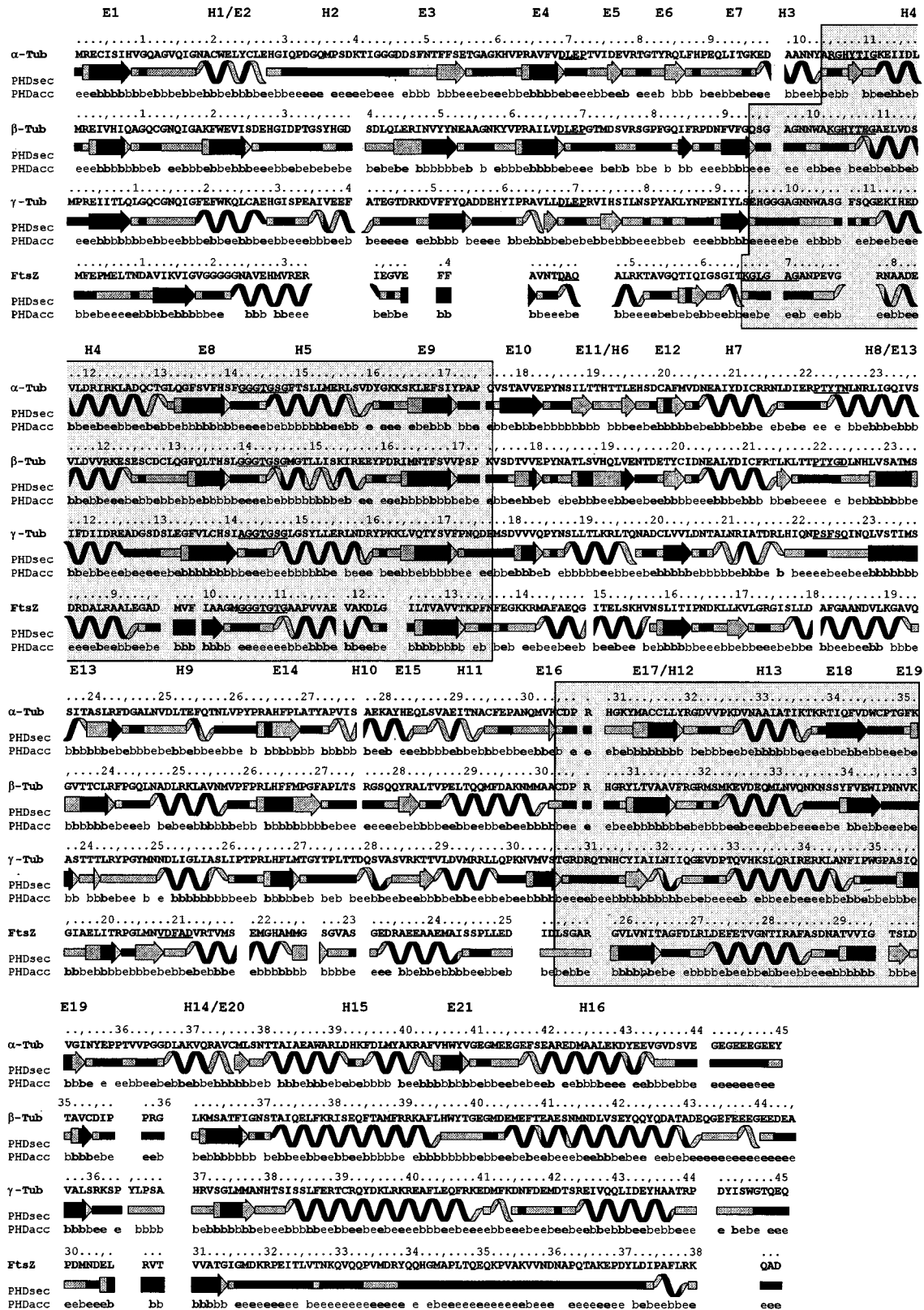


FIGURE 3: Secondary structure and solvent accessibility predictions and comparison of the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulin and FtsZ families of proteins. The residue number and the amino acid residues of the guide sequences are indicated. The separate predictions for each family were compared by aligning the four guide sequences. Note that the procedure introduces artificial gaps in the secondary structure elements. The characteristic sequence motifs of tubulin and FtsZ are underlined. The secondary structure prediction by the PHDsec method (Materials and Methods) is graphically displayed with helical symbols, arrows, and ribbons for  $\alpha$  helices, extended  $\beta$  sheet strands, and loops (other), respectively. The predictions with an expected accuracy under and over the 82% are in light gray and black, respectively. The alignment of the C termini of FtsZ was poor, and therefore, the extended loop prediction is probably not reliable. The secondary structure elements predicted for  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulin are named on the top of the first sequence, (H)  $\alpha$  helix and (E)  $\beta$  strand, by sequential order (note that not all the secondary structure elements are present in the three proteins). The three-state solvent accessibility prediction by the PHDacc method is shown below; e indicates exposed, b buried, and blank intermediate solvent accessibility. Predictions with an expected accuracy under and over 69% are in normal and bold type, respectively. The largest regions with similar secondary structure prediction for all  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulins and FtsZ are shaded with gray.

Table 3: Tubulin Secondary Structure Content Predicted by PHDsec

chain	helix	extended	loop
$\alpha$ -tubulin	0.36 (0.24) <sup>a</sup>	0.20 (0.10)	0.44 (0.23)
$\beta$ -tubulin	0.31 (0.22)	0.27 (0.17)	0.42 (0.23)
$\alpha\beta$ mean value	0.33 (0.23)	0.24 (0.14)	0.43 (0.23)

<sup>a</sup> The content of secondary structure expected to have an accuracy better than 82% is shown in parentheses.

FtsZ family is 44–46% identical with the predicted tubulin structures. Interestingly, the predicted secondary structure of FtsZ shows two zones very similar to the tubulins, which are shadowed in the figure. Most striking is the first zone, extending approximately from residue 65 to residue 135 in FtsZ and from residue 95 to 175 in tubulins. This consists of a strongly predicted  $\alpha$ - $\beta$ - $\alpha$ - $\beta$  secondary structure (loop-helix-loop-extended loop-helix-loop-extended loop) which contains a sequence resembling an inverted phosphate binding motif (the first loop) and the tubulin-FtsZ signature motif GGTG(SAT)G (the third loop). The sequence identity of this zone of FtsZ with that of tubulins is typically 19–31% in the alignment employed (and only 12–25% excluding the two motifs). However, the predicted secondary structure of this FtsZ zone is 85–87% identical residue per residue to the corresponding tubulin zones. If the comparison was between experimentally determined structures instead of predictions, the percent identity in this zone would clearly correspond to the average 88% in homologous protein structures (which is considered the goal in secondary structure prediction), far from the 35% coincidence among random protein pairs (Rost et al., 1994b). The predicted structural homology does not appear to result from locally similar sequences. Since the probability of predicting such a secondary structural arrangement in both

proteins by chance is expected to be very low, the predicted homology can be regarded as significant. The second zone shadowed in Figure 3 has a 9–17% sequence identity. Its predicted secondary structure in FtsZ is 62, 78, and 51% identical to those of  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulins, respectively. Note that for this zone the percentages of identical prediction for the tubulins are as follows:  $\alpha$ - $\beta$ , 80%;  $\gamma$ - $\alpha$ , 60%; and  $\gamma$ - $\beta$ , 66%; therefore, the local predictions for FtsZ and  $\beta$ -tubulin are nearly as close as those for  $\alpha$ - and  $\beta$ -tubulin.

**Proteolytic Sites at the Surface of the  $\alpha\beta$ -Tubulin Dimer and Microtubules.** Purified native tubulin, in the forms of  $\alpha\beta$  dimers or taxol-induced microtubules, was gently proteolyzed with a panel of 12 different proteases, and the cleavage points were mapped onto the  $\alpha$ - and  $\beta$ -tubulin sequences. The primary results, summarized in Figure 14 and Table 2 of the preceding paper (de Pereda & Andreu, 1996), show that the cleavage points cluster into five well-defined zones. The proteolysis sites of each individual proteolysis zone (A–E) are now presented on the  $\alpha$ - and  $\beta$ -tubulin sequences in Figures 4–8, respectively. The arrows marked S indicate sites which were similarly accessible in tubulin dimers and in taxol-induced microtubules, whereas sites marked P were comparatively protected from cleavage in the assembled form. The sites marked M were exclusively observed in microtubules. The maps of surface sequences of tubulin find their best use in combination with other structural information. The secondary structure prediction will be compared with this set of experimental constraints for each individual proteolysis zone in the Discussion, where structural working models will also be proposed.

## DISCUSSION

The majority of tubulin nicking sites are at predicted loops. The predicted secondary structural environments of the 24 mapped cleavage points of 12 proteases on the  $\alpha$ - and  $\beta$ -tubulin chains are as follows. Eleven protease cleavage points are at predicted loops. Five are at the extremes of

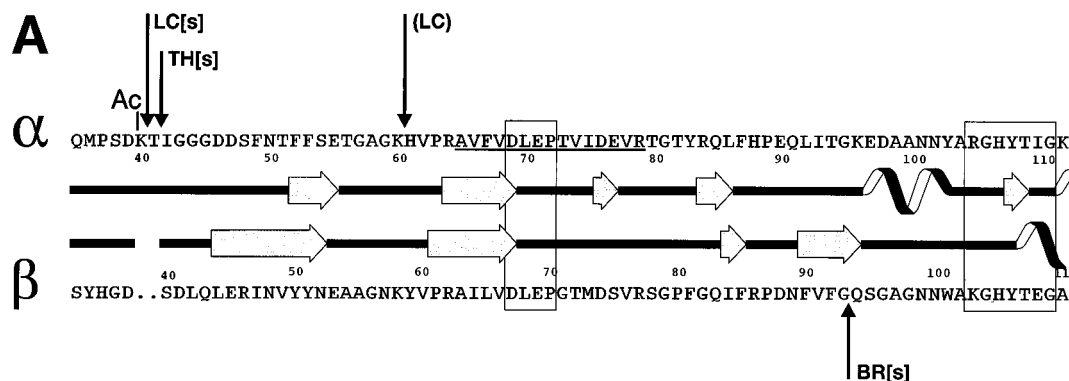


FIGURE 4: Zone A of preferential proteolysis of the  $\alpha\beta$ -tubulin dimer, shown onto the sequences of porcine brain  $\alpha$ - and  $\beta$ -tubulin. The predicted secondary structures (see Figure 3 for details) are indicated between both sequences. Nicking points which are accessible to proteases and the taxol-induced microtubules are marked by [s]. Ac indicates the post-translational acetylation position in  $\alpha$ -tubulin. The boxes mark the putative guanine binding motif and the reversed phosphate binding motif in  $\beta$ -tubulin and the corresponding sequence in the  $\alpha$  chain. A buried strand of extended  $\beta$  sheet is predicted between cleaved loops at position 40 and 60 (see also Figure 3). Another predicted buried  $\beta$  strand follows and overlaps the conserved sequence DLEP around position 70, which was suggested to be possibly equivalent to a guanine binding motif (Linse & Mandelkow, 1988). Two photoaffinity labeling studies of the exchangeable GTP binding site detected the  $\beta$ -tubulin peptides 63–72 and 63–77 as major labeled fragments (Kim et al., 1987; Linse & Mandelkow, 1988). However, more recent studies have also identified, far apart from this zone,  $\beta$ -tubulin Cys12 and the fragment 3–19 as primary sites of GTP incorporation (Shivanna et al., 1993; Jarayam & Haley, 1994). The yeast  $\beta$ -tubulin mutant D67N has been reported not to bind GTP, whereas the mutant E69Q binds GTP-Mg<sup>2+</sup> in a manner similar to that of the wild-type protein (Farr & Sternlicht, 1992). However, the multiple  $\beta$ -tubulin mutants D67A/E69A, D67A/E69A/G71W, and D74A/R77A have no effect on the yeast phenotype (Reijo et al., 1994). The  $\alpha$ -tubulin epitope mapped for the monoclonal antibody TU01 (Grim et al., 1987; Draber et al., 1989; sequence underlined in Figure 4) becomes occluded by taxol-induced microtubule assembly, and the phage display antibody V $\alpha$ NT1, whose epitope is probably at positions ~80–102, binds only to the minus microtubule ends (Fan et al., 1996).

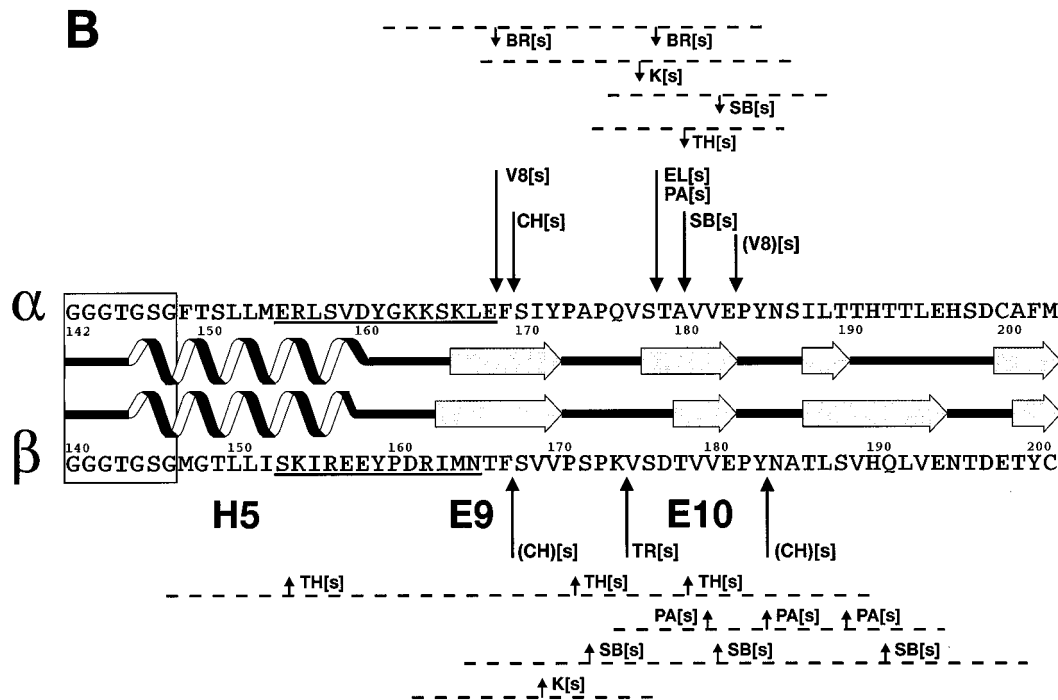


FIGURE 5: Zone B of preferential proteolysis of tubulin in comparison to the secondary structure predictions. This display is similar to Figure 4. Arrows indicate the sequenced cleavage positions, while positions assigned from protease specificities have the protease initials in parentheses. Positions of approximately located nicking points are indicated by the shorter arrows and dashed lines in separate rows. The box marks the characteristic Gly cluster motif. Sequences recognized by the site-directed antibodies anti $\alpha$ (155–168) and anti $\beta$ (153–165) (Arévalo et al., 1990) are underlined.

predicted extended sheet strands. Four are at predicted sheet strands, while two are at the extremes of predicted helices and two within predicted helices. The majority of nicking points are located at zones with high predictive accuracy (Figure 3) and correlate to sequence variability and/or hydrophilicity maxima (not shown). While any predictive method is bound to give a certain proportion of erroneously predicted elements, we consider the prediction compatible with comparison to experimentally determined proteolysis points. We tend to give credibility to strongly predicted secondary structure elements, particularly  $\alpha$  helices.

*Proteolysis Zones A are Nonequivalent in  $\alpha$ - and  $\beta$ -Tubulin and Are Accessible in Microtubules.* As indicated in Figure 4, the cleavage points by proteases lysine-C and thermolysin in zone A of  $\alpha$ -tubulin are adjacent to acetylatable Lys40, in a large predicted surface loop which is accessible in microtubules. This is also one of the more variable zones among the  $\alpha$ -tubulin sequences.  $\beta$ -Tubulin has a compatible prediction but no equivalent Lys (and few thermolysin cleavable residues) in this zone. A minor lysine-C cleavage of  $\alpha$ -tubulin probably maps at Lys60, a smaller predicted loop. Bromelain cleaves  $\beta$ -tubulin at Gly93-Gln94, at the junction of a short predicted extended strand with a loop which at the other end contains the reversed phosphate binding motif KGHYTEG(A) [note that the corresponding sequence in  $\alpha$ -tubulin is RGHYTIG(K), and it is not well-detected in  $\gamma$ -tubulin]. This nicking point is accessible in taxol-induced microtubules. A minor cleavage by lysine-C (not shown in Figure 4) takes place at an uncertain lysine residue, either Lys58 (corresponding to a similar cleavage in the  $\alpha$  chain), Lys103 (at the reversed P-loop), Lys122 (at the C-terminal end of a following helix; see Figure 3); or Lys154 (close to proteolysis zone B).

*Tubulin Proteolysis Zone B Extends between Residues 167 and 183 and Is at the Surface of Microtubules.* This zone

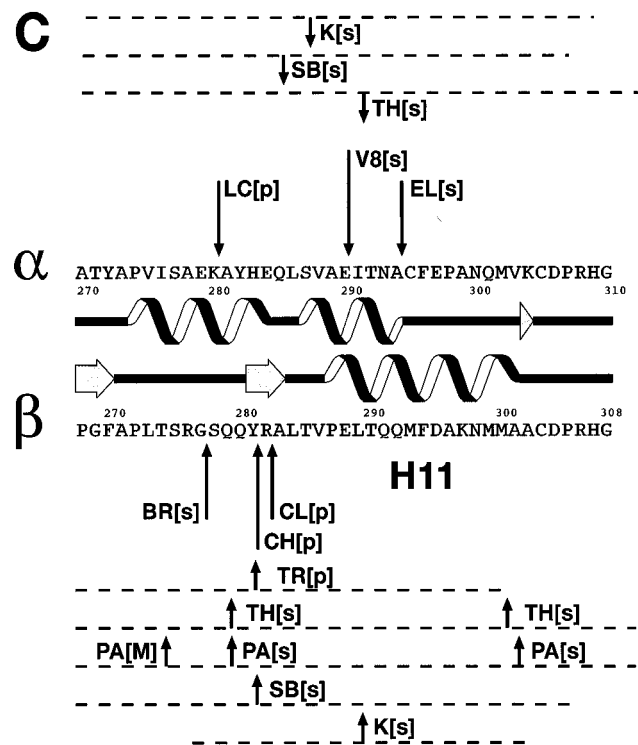


FIGURE 6: Zone C of preferential proteolysis of tubulin and secondary structure prediction, displayed in a manner similar to that of Figure 4. Nicking points which become unaccessible in taxol-induced microtubules are indicated by [p], while cleavages which are only detected in microtubules are indicated by [M].

is adjacent to sequences  $\alpha$ (155–168) and  $\beta$ (153–165), which are accessible to specific antibodies in the tubulin heterodimer (Arévalo et al., 1990) and contain a hydrophilic predicted loop around position 160. As shown in Figure 5, the clearly predicted  $\alpha$  helix H5 overlaps the characteristic glycine cluster (phosphate binding loop) motif GGGTGSG

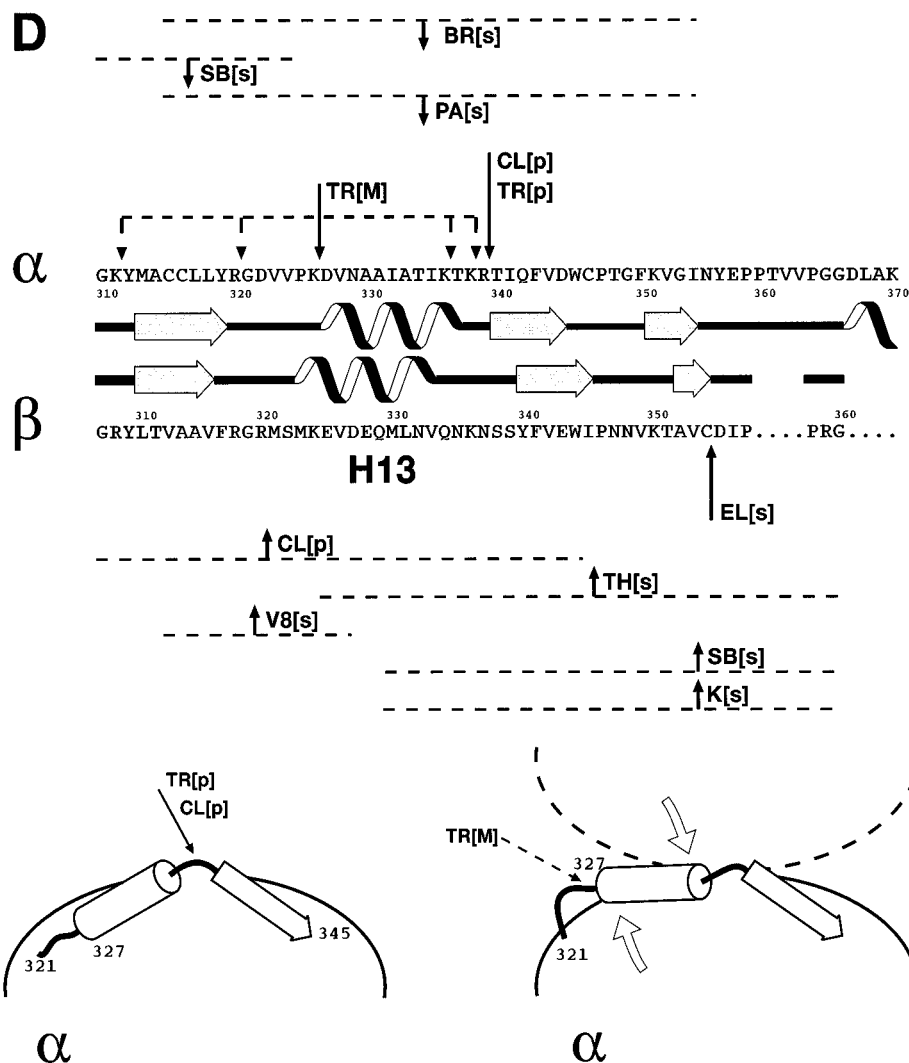


FIGURE 7: (Top) Zone D of preferential proteolysis of tubulin and secondary structure predictions, displayed in a manner similar to that of Figures 4 and 6. Note that trypsin produces a single new cleavage in microtubule-assembled  $\alpha$ -tubulin, indicated by TR[M], whose more probable location is indicated by a solid arrow, and the other possible locations are indicated by dashed arrows [see de Pereda and Andreu (1996)]. (Bottom) Tentative model shown to illustrate the protection of the clostripain/trypsin cleavage point and the exposure of the new tryptic cleavage point in proteolysis zone D of  $\alpha$ -tubulin upon microtubule assembly. This particular model envisages these changes simply as a movement of a helical segment (indicated by the large arrows) induced by contact with another tubulin molecule in the microtubule lattice (dashed lines).

(marked in the figure) and these antigenic sequences. The accessibility of the later zones to antibodies supports the notion that helix H5 may be oriented with its N terminus toward a nucleotide phosphate binding pocket and its C terminus toward the predicted loop at the surface. Protease V8 and chymotrypsin cleave the predicted  $\beta$  strand E9 (residues 165–170 approximately). Trypsin, elastase, and papain cleave the following loop and the amino end of the next short predicted  $\beta$  strand E10, while subtilisin cleaves in the middle of this element. The following predicted loop is again cleaved by protease V8 and chymotrypsin. Each of the 24 cleavage sites detected in zone B in the tubulin heterodimer was similarly exposed at the surface of taxol-induced microtubules.

*Tubulin Proteolysis Zone C between Residues 278 and 295 is Partially Occluded in Microtubules.* As shown in Figure 6, the protease lysine-C cleaves  $\alpha$ -tubulin at Lys280-Ala281, at the hydrophilic C-terminal part of a segment predicted as helix. However, this segment has a loop prediction in  $\beta$ -tubulin, and it is cleaved by bromelain at the corresponding site, Gly277-Ser278. Chymotrypsin and clostripain cleave  $\beta$ -tubulin at Tyr281-Arg282 and Arg282-Ala283, respec-

tively, predicted as a short extended strand or loop in both subunits. Downstream from this loop comes the predicted helix H11, which is shorter and weaker in  $\alpha$ -tubulin and is cleaved at residues situated at the same helical face by protease V8 (Glu290-Ile291) and elastase (Ala294-Cys295). Some of the proteolysis points of this relatively small zone C are protected in microtubules, while others are accessible, which indicates location at the microtubule surface. Accepting the head to tail orientation of tubulin  $\alpha\beta$  dimers along the protofilaments of microtubules and assuming that upon assembly both monomers are subject to similar lateral proteolytic accessibility changes suggest that the protected residues in zone C are far from the longitudinal contact sites at the ends of the heterodimer but may be near the lateral contact sites between tubulin molecules at the inter-protofilament grooves. The detailed accessibility pattern of the nicking points of zone C is complex, since (i) cleavage points in nearby sequence positions and (ii) the corresponding positions of the  $\alpha$ - and  $\beta$ -tubulin sequences cleaved by LysC and bromelain both show opposite behavior.

*Proteolysis of Tubulin Zone D Detects a Conformational Change in Microtubules.* The results of limited proteolysis



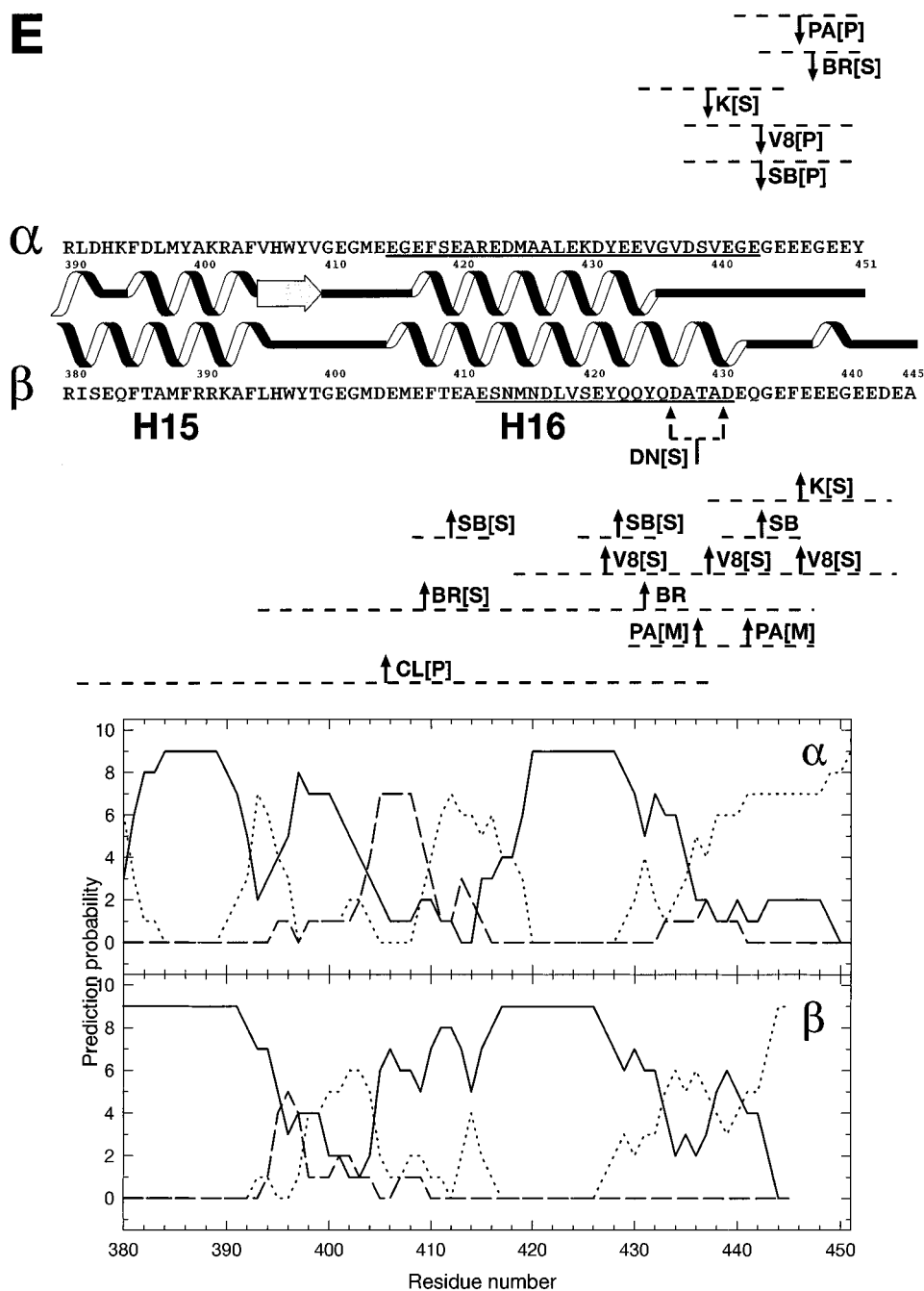


FIGURE 8: (Top) Proteolysis zone E at the carboxy termini of  $\alpha$ - and  $\beta$ -tubulin, in comparison to their secondary structure predictions, displayed as in Figure 7. (Bottom) PHD secondary structure prediction potentials of these  $\alpha$ - and  $\beta$ -tubulin zones: solid line,  $\alpha$  helix potential; dashed line,  $\beta$  strand potential; and dotted line, loop potential. These profiles are shown to facilitate detailed comparison with the proteolysis data in the top panel. Note, for example, the clustering of the approximately located cleavage points at both sides of the main helical maximum which defines the predicted helix H16 in  $\beta$ -tubulin.

are compared to the secondary structure prediction in the top panel of Figure 7. The largest predicted structural element in this zone is helix H13. Clostripain and trypsin cleave  $\alpha$ -tubulin at Arg339-Thr340, which is at a predicted loop on the C-terminal side of H13 and is well-protected by microtubule assembly. A corresponding cleavage in  $\beta$ -tubulin has not been observed. A new point of tryptic attack appears in  $\alpha$ -tubulin when it is assembled in taxol-induced microtubules, which is not located on H13, but probably at its amino-terminal loop. This is a clear indication of a tubulin structural change induced by assembly or taxol binding. If the secondary structural elements in this zone were to remain unchanged, a simple possible model would be that the loop C terminal to helix H13 is occluded by the intermolecular contacts, this helix acts as a lever (without

making extensive contact), and its N-terminal loop is pushed out into the solvent, zone D being located at the outer part of some interdimer contact zone in the microtubule lattice (see the hypothetical model scheme in the bottom panel of Figure 7). An alternate explanation would be the sequential cleavage of the N-terminal loop of H13 followed by Arg339-Thr340 only in the dimer; however, this is inconsistent with the fragment pattern observed (de Pereda & Andreu, 1996) and implies a similar contact in microtubules. Elastase cleaves  $\beta$ -tubulin at Cys354-Asp355, at the C-terminal end of a small weakly predicted  $\beta$  strand, and this nicking point is accessible in microtubules.

*Proteolysis Zone E Includes the C-Terminal 10–20 Residues of Each Tubulin Chain, Following Two Large Conserved Helical Zones.* As shown in the top panel of

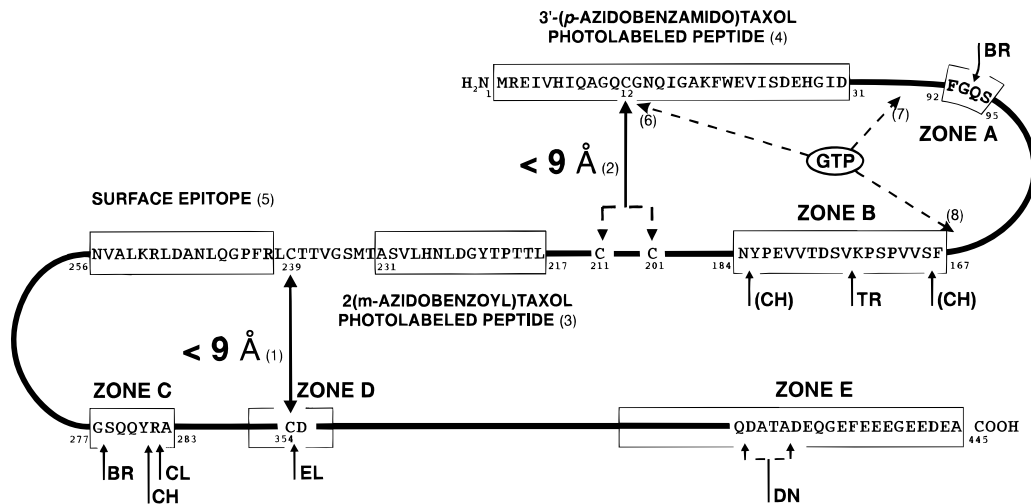


FIGURE 9: Scheme of surface and distance constraints for  $\beta$ -tubulin sequences separate in the polypeptide chain. The elastase accessible residue Cys354 (zone D) is maximally 0.9 nm from Cys239 (1; Little & Ludueña, 1985). Cys239 is close upstream from the heterodimer surface epitope 241–256, which becomes occluded by assembly (5, Arévalo et al., 1990), and nearly downstream is the peptide 217–231 which is photolabeled by 2-(*m*-azidobenzoyl)taxol (3; Rao et al., 1995). One of the Cys residues (201 or 211) is maximally 0.9 nm from Cys12 (2; Little & Ludueña, 1987). Cys12 is also a site of GTP photolabeling (6; Shivanna et al., 1993) and is within the peptide 1–31 that is photolabeled by 3'-(*p*-azidobenzamido)taxol (4; Rao et al., 1994). Other nucleotide-photolabeled peptides are 63–77 (7; Kim et al., 1987; Linse & Mandelkow, 1988) and 155–174 (8; Hesse et al., 1987). The positions of surface proteolysis zones A–E (described in detail in Figures 4–8) are indicated in the scheme. The region between proteolytic zones A and B is analyzed in Figure 10.

Figure 8, roughly the last 10 residues of  $\alpha$ -tubulin constitute an extremely acidic tail which is cleaved by subtilisin, protease V8, proteinase K, bromelain, and papain. The corresponding 15 or 19 residues in  $\beta$ -tubulin are selectively removed by proteinase Asp-N, and the zone is less specifically cleaved by the other proteases, including the multiple cleavage by subtilisin, protease V8, and bromelain and the microtubule cleavages by papain. It is difficult to exactly map proteolysis points in this zone. These are the more variable zones of the  $\alpha$ - and  $\beta$ -tubulin sequences, where most of the variable sequences of the tubulin isoforms and the known post-translational modifications cluster [see for example Redecker et al. (1994)]. Zone E is in general less accessible and more protected from limited proteolysis by microtubule assembly in  $\alpha$ -tubulin than in  $\beta$ -tubulin. This is consistent with the proposed locations at the longitudinal contact zones and the interface between subunits, respectively [see discussion in de Pereda and Andreu (1996)]. The extreme C-terminal zone has a mixed loop-helical potential in the case of  $\beta$ -tubulin. This is shown by the secondary structure potentials in the bottom panel of Figure 8. As originally pointed out by Ponstingl et al. (1979), the coil to helix transition may depend on the microenvironment and on interactions with other proteins. It is interesting that two microtubule-specific papain nicking sites map to this zone of  $\beta$ -tubulin. Next to these zones, the strongly predicted  $\alpha$  helix H16 extends for 17–25 residues upstream. This helix is not cleaved in  $\alpha$ -tubulin under the present conditions, while bromelain, subtilisin, and clostripain cleave  $\beta$ -tubulin probably at the amino-terminal part of H16. This helical zone may actually be comprised of several shorter helical segments instead of a single rigid helix, since its total length (4.3 nm) would be relatively large (in comparison with the dimensions of the tubulin monomer) to be located at the microtubule surface. This helical zone is amphipathic, suggesting potential interaction with other structural elements of tubulin or of microtubule binding proteins. Following upstream, there is a predicted loop and a short  $\beta$  strand which connect to the strongly helical segment H15, the other longest helix

predicted in tubulin, already detected by earlier analyses (Ponstingl et al., 1979; de la Viña et al., 1988). Interestingly, both H15 and H16, close to the highly variable C extremes, constitute the longest conserved zones of the  $\alpha$ - and  $\beta$ -tubulin sequences, which suggests important structural roles for them.

*Surface Mapping Combined with Other Biochemical Information.* The results of limited proteolysis of tubulin complement existing biochemical information. It is known that  $\beta$ -tubulin Cys354 can be cross-linked to Cys239 by a bivalent sulfhydryl reagent with a maximal effective length of 0.9 nm and that the cross-link is inhibited by colchicine (Little & Ludueña, 1985). The cleavage of the Cys354–Asp355 bond by elastase maps Cys354 to the surface of microtubules and therefore Cys239 to a relative maximal depth of 0.9 nm. In fact, Cys239 is close to the  $\beta$ -tubulin sequence 241–256, which is accessible to specific antibodies in the surface of the heterodimer and becomes occluded in microtubules (Arévalo et al., 1990). As an example, Figure 9 presents a scheme of these distance and surface constraints and other features expected to apply to the three-dimensional folding of the  $\beta$ -tubulin chain.

*A Model for the Region between Proteolysis Zones A and B: A Common Nucleus of Tubulins and FtsZ Which Participates in GTP Binding.* This region includes the inverted P-loop and G-box motifs characteristic of tubulin and FtsZ (see Figures 4 and 5). When the predicted secondary structures of the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulin and FtsZ protein families are compared, a significant homology is suggested (see Figure 3 and Results). This consists of a loop–helix–loop–strand–loop–helix–loop–strand–loop fold with a high predictive probability. A local model of the supersecondary structure of this tubulin zone is shown in Figure 10A. The model is the result of putting together the secondary structure prediction with the nucleotide binding motifs and the accessibility to proteases and antibodies. Surface zones A and B define between them a proteolytically resistant, compact region, extending from residue 94 to residue 167 in  $\beta$ -tubulin, with the exception of a minor

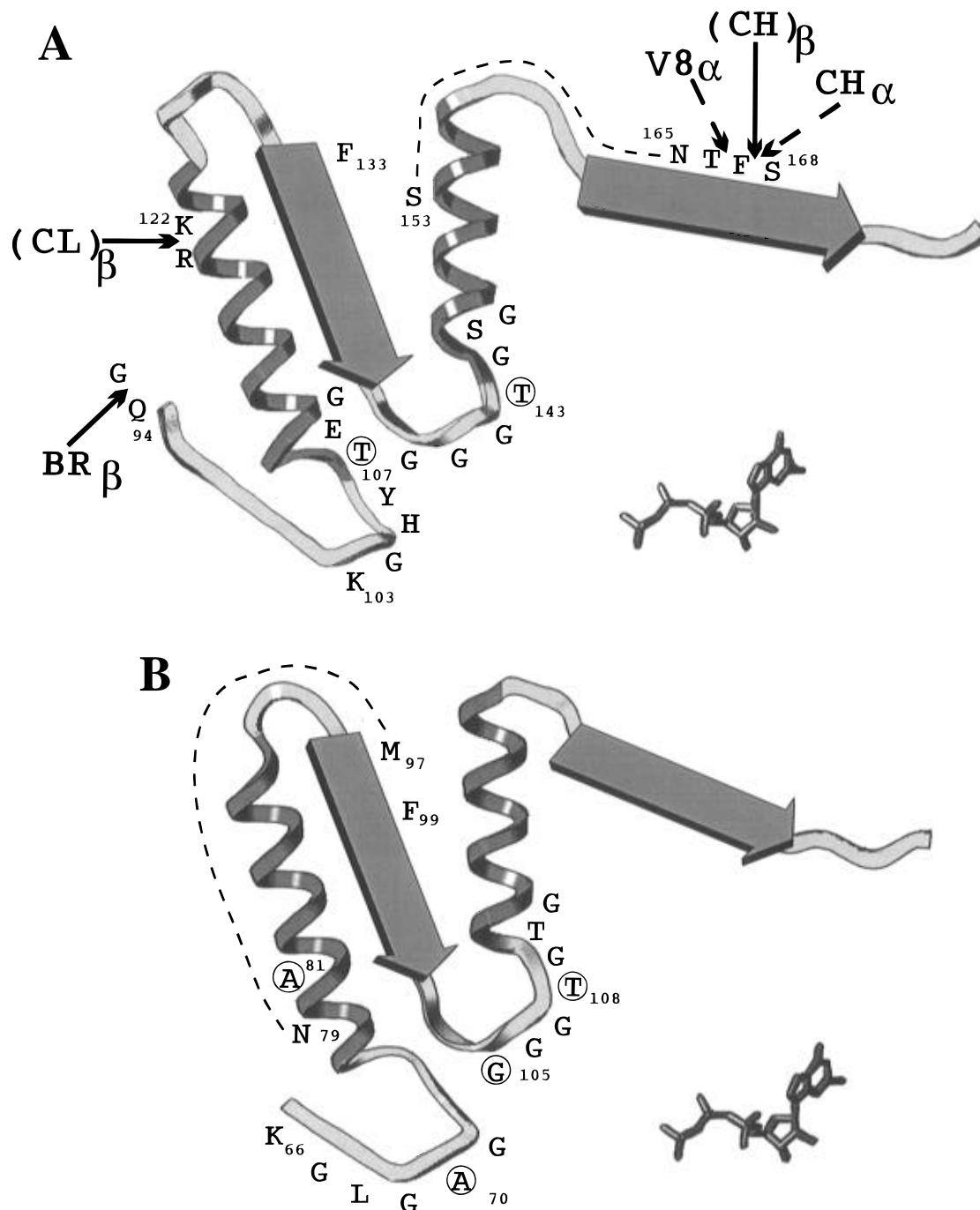


FIGURE 10: Local working model of a predicted homologous nucleus of tubulins and FtsZ. Panel A illustrates the model for  $\beta$ -tubulin (in which the chymotrypsin and protease V8 nicking points on  $\alpha$ -tubulin have been added at their homologous positions), while panel B illustrates the model for FtsZ. The surface of both proteins is to the upper left of the drawings, while the inside is to the lower right. The dashed lines indicate the mapping of epitopes of monoclonal antibodies. Residues encircled are important for GTP hydrolysis from studies with mutants (see text). A GTP molecule is represented as scaled to facilitate comparison.

cleavage by clostripain probably located at Arg121-Lys122 [see Figure 14 in de Pereda and Andreu (1996)]. The surface loop following the site of bromelain cleavage, Gly93-Gln94 (zone A), extends toward the inverted P-loop inside the protein, continuing with helix H4, which is clearly amphipatic (not shown) and has its C terminus probably oriented toward the surface of the protein so that it can be cleaved by clostripain in the vicinity of a hydrophilic loop. The tubulin sequences in this loop are highly conserved; the consensus sequence is (DE)X(CS)XXL(QE)GF, extending from residue 125 to Phe133 in  $\beta$ -tubulin [monospecific antibodies to peptides homologous to the  $\alpha$ (120–131) and  $\beta$ (119–129) sequences, containing this predicted loop, unfortunately failed

to recognize native or denatured tubulin; Arévalo et al., 1990]. After this loop, the potential buried  $\beta$  strand E8 leads to the G-box loop (note that this type of predicted strand may also be a helix; Barton, 1995). Helix H4 and predicted strand E8 may interact so that the amino terminus of the former is near the carboxy terminus of the latter, and the inverted P-loop and the G-box loop are close to each other. The model (Figure 10) is limited to a two-dimensional representation; that is, it does not define the positions of the other secondary structure elements relative to each other. Following the G-box, helix H5 connects again with a surface loop of the protein. This is detected by the reactivity of sequences  $\alpha$ (155–168) and  $\beta$ (153–165) with their mono-

specific antibodies (Arévalo et al., 1990). A zone predicted as  $\beta$  strand follows, which in  $\alpha$ -tubulin is cleaved by chymotrypsin and protease V8, and continues with a loop and the rest of the proteolysis zone B. Since the sequence  $\beta$ (153–165) (Arévalo et al., 1990) and region B are accessible in microtubules (de Pereda & Andreu, 1996) and the predicted length of helix H5 is 1.7–2.1 nm, the nucleotide binding site should be roughly in the outer half of the microtubule wall (Andreu et al., 1992). An antibody to the sequence  $\beta$ (154–165), including the end of helix H5 and the loop, has been reported to inhibit nucleotide exchange and tubulin assembly (Hesse et al., 1987). Photoaffinity labeling with GTP resulted in the identification of one labeled fragment corresponding to the  $\beta$ (155–162) sequence (Hesse et al., 1987).

The model is compatible with results from tubulin mutants. Mutant proteins T107G, T107K, and T107W at the inverted P-loop have modified GTPase activities, consistent with the notion that the  $\beta$ -tubulin sequence 103–109, KGHYTEG, may be functionally equivalent to the typical phosphate binding site of GTPases, GXXXXGK, though its orientation is inverted (Davis et al., 1994). The mutant  $\beta$ -tubulin T143G at the G-box has reduced GTPase activity [C. Dougherty and K. Farrell, unpublished observations cited by Sage et al. (1995)]. The multiple yeast  $\beta$ -tubulin mutants I152F/R156K/E157A/E158A, K154A/R156A, and E157A/D158A which are located in predicted helix H5 are lethal; however, the double mutant E125A/D128A, which maps to the predicted loop between helix H4 and strand E9, has no phenotypic effect (Reijo et al., 1994). The corresponding local model for FtsZ, based on the homology of predicted secondary structure with tubulin, is shown in Figure 10B. Less biochemical and genetic information is available for FtsZ. The epitope of one monoclonal antibody to FtsZ maps to sequence 79–97 in the native protein (Voskuil et al., 1994). This comprises a predicted helix and the loop 94–96, corresponding to tubulin helix H4 and its carboxy-terminal loop. The residue Phe99 in the following  $\beta$  sheet strand is highly conserved in the FtsZ and tubulin families ( $\beta$ -tubulin Phe133). The mutation A70T (FtsZ1) in the reversed P-loop strongly reduces the GTPase activity, though the mutated protein binds GTP to an extent similar to that of wild-type FtsZ (Dai et al., 1994). The mutant protein F268C (FtsZ114) binds and hydrolyzes GTP, while the double mutant A81V/F268C (FtsZ100) exhibits reduced GTP binding and hydrolysis (Dai et al., 1994), suggesting a role for Ala81, which is at the predicted helix 78–93, corresponding to tubulin helix H4. Bacteria bearing the mutation G105S (FtsZ84) in the G-box only can divide at the permissive temperature, and the *in vitro* GTPase activity of FtsZ surprisingly transforms into ATPase in FtsZ84 (Ray-Chaudhuri & Park, 1994). The mutation T108A (FtsZ3) is lethal (Bi & Lutkenhaus, 1990), and the mutated protein does not bind GTP (Mukherjee et al., 1993).

This new model (Figure 10) is clearly different from typical GTPases, although some partial structural similarities can be found. The phosphate binding motif GXXXXGK (Bourne et al., 1991) is preceded by a  $\beta$  sheet strand and followed by an  $\alpha$  helix which goes from the GTP binding site toward the solvent in elongation factor Tu (la Cour et al., 1985; Kjeldgaard et al., 1993), ras p21 (Pai et al., 1989; Brunger et al., 1990; Tong et al., 1991), and transducin (Noel et al., 1993). A similar situation can also be found in

ATPases such as in the RecA protein (Story & Steitz, 1992). In the model (Figure 10); the tubulin–FtsZ G-box motif is similarly placed between strand and helix; however, its sequence is different. The conserved Phe residue of tubulin and FtsZ ( $\beta$ -tubulin Phe133, FtsZ Phe99) might be equivalent to a conserved Phe residue in the Ras family whose aromatic ring is close to the guanine base of the bound nucleotide, even if it is separated from the P-loop by a helix of ca. 10 residues (Valencia et al., 1991). On the other hand, the “inverted P-loop” of  $\beta$ -tubulin and FtsZ shows more motif similarity to an inverted GXXXXGK sequence than to the G-box. However, it is difficult to understand how a sequence may be substituted in the structure by its inverted equivalent. The predicted structural environment of this tubulin–FtsZ motif appears to be unrelated to GTPases.

This model does not intend to include the complete GTP binding site or all the possible structurally homologous regions of tubulin and FtsZ (see Figure 3 and zone D in Figure 7) but only a predictable nucleus (Barton, 1995). The model predicts a characteristic fold present in the tubulin and FtsZ families of proteins, which may constitute an ancestral core of these atypical GTPases. We propose critical use of these (Figures 9 and 10) and other similar schemes to assist in the construction of three-dimensional models and in the design of new experiments with both proteins.

## ACKNOWLEDGMENT

We thank Dr. G. Fasman, Dr. C. Johnson, and Dr. R. Woody for making their CD analysis algorithms available to us, Dr. E. Goormaghtigh for his FTIR analysis software, and Drs. V. Muñoz and L. Serrano for Agadir and Dr. M. Andrade for help with sequence analysis. We are indebted to Dr. M. Gasset, Drs. P. Rodríguez and O. Monasterio, and Dr. J. Villalain for FTIR spectroscopy and to Dr. J. L. Arrondo for discussion. We appreciate critical comments on the manuscript by Dr. M. Vicente and Dr. A. Valencia.

## SUPPORTING INFORMATION AVAILABLE

Two tables and four figures depicting the data and graphical representations of the analysis and predictions of the tubulin and FtsZ sequences (7 pages). Ordering information is given on any current masthead page. The sequence alignments employed are available from the authors.

## REFERENCES

- Andrade, M. A., Chacón, P., Merelo, J. J., & Morán, F. (1993) *Protein Eng.* 6, 383–390.
- Andreu, J. M., & Timasheff, S. N. (1982) *Biochemistry* 21, 534–543.
- Andreu, J. M., De la Torre, J., & Carrascosa, J. L. (1986) *Biochemistry* 25, 5230–5239.
- Andreu, J. M., Bordas, J., Diaz, J. F., García de Ancos, J., Gil, R., Medrano, F. J., Nogales, E., Pantos, E., & Towns-Andrews, E. (1992) *J. Mol. Biol.* 226, 169–184.
- Arévalo, M. A., Nieto, J. M., Andreu, D., & Andreu, J. M. (1990) *J. Mol. Biol.* 214, 105–120.
- Arrondo, J. L., Muga, A., Castresana, J., & Goñi, F. (1993) *Prog. Biophys. Mol. Biol.* 59, 23–56.
- Bairoch, A., & Boeckmann, B. (1993) *Nucleic Acids Res.* 21, 3093–3096.
- Barton, G. J. (1995) *Curr. Opt. Struct. Biol.* 5, 372–376.
- Bi, E., & Lutkenhaus, J. (1990) *J. Bacteriol.* 172, 5602–5609.
- Bi, E., & Lutkenhaus, J. (1991) *Nature* 354, 161–164.
- Bourne, H. R., Sanders, D. A., & McCormick, F. (1991) *Nature* 349, 117–127.

- Bramhill, D., & Thompson, C. M. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 5813–5817.
- Brünger, A. T., Milburn, M. V., Tong, L., deVos, A. M., Jankarik, J., Yamaizumi, Z., Nishimura, S., Ohtsuka, E., & Kim, S. H. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 4849–4853.
- Chizgarde, Y. N., Fedorov, O. V., & Trushina, N. P. (1975) *Biopolymers* 14, 679–694.
- Dai, K., Mukherjee, A., Xu, Y., & Lutkenhaus, J. (1994) *J. Bacteriol.* 175, 130–136.
- Davis, A., Sage, C. R., Dougherty, C. A., & Farrell, K. W. (1994) *Science* 264, 839–842.
- de Boer, P., Crossley, R., & Rothfield, J. (1992) *Nature* 359, 254–256.
- de la Viña, S., Andreu, D., Medrano, F. J., Nieto, J. M., & Andreu, J. M. (1988) *Biochemistry* 27, 5352–5365.
- de Pereda, J. M., & Andreu, J. M. (1996) *Biochemistry* 35, 14184–14202.
- de Tar, D. F. (1969) *Anal. Chem.* 41, 1406–1408.
- Díaz, J. F., Pantos, E., Bordas, J., & Andreu, J. M. (1994) *J. Mol. Biol.* 238, 214–225.
- Dráber, P., Bráberová, E., Linhartová, I., & Viklicky, V. (1989) *J. Cell Sci.* 92, 519–528.
- Eisenberg, D., Weiss, R. M., & Terwilliger, T. C. (1984) *Proc. Natl. Acad. Sci. U.S.A.* 81, 140–144.
- Erickson, H. P. (1995) *Cell* 80, 367–370.
- Erickson, H. P., Taylor, D. W., Taylor, K. A., & Bramhill, D. (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93, 519–523.
- Fan, J., Griffiths, A. D., Lockhart, A., Cross, R. A., & Amos, L. A. (1996) *J. Mol. Biol.* (in press).
- Farr, G. W., & Sternlicht, H. (1992) *J. Mol. Biol.* 227, 307–321.
- Goormaghtigh, E., Cabiaux, V., & Ruyschaert, J. M. (1990) *Eur. J. Biochem.* 193, 409–420.
- Goormaghtigh, E., Cabiaux, V., & Ruyschaert, J. M. (1994) *Subcell. Biochem.* 23, 329–450.
- Greenfield, N. J. (1996) *Anal. Biochem.* 235, 1–10.
- Grim, M., Breitling, F., & Little, M. (1987) *Biochim. Biophys. Acta* 914, 83–88.
- Hesse, J., Thierauf, M., & Ponstingl, H. (1987) *J. Biol. Chem.* 262, 15472–15475.
- Hyman, A. A., & Karsenti, E. (1996) *Cell* 84, 401–410.
- Jarayam, B., & Haley, B. E. (1994) *J. Biol. Chem.* 269, 3233–3242.
- Kim, H., Ponstingl, H., & Haley, B. E. (1987) *Fed. Proc.* 46, 2229.
- Kjelgaard, M., Nissen, P., Thirup, S., & Nyborg, J. (1993) *Structure* 1, 35–50.
- Kraus, E., Little, M., Kempf, T., Hofer-Warbinek, R., Ade, W., & Ponstingl, H. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78, 4156–4160.
- la Cour, T. F. M., Nyborg, J., Thirup, S., & Clark, B. F. C. (1985) *EMBO J.* 4, 2385–2388.
- Lee, J. C., Corfman, D., Frigon, R. P., & Timasheff, S. N. (1978) *Arch. Biochem. Biophys.* 185, 4–14.
- Linse, K., & Mandelkow, E. M. (1988) *J. Biol. Chem.* 263, 15205–15210.
- Little, M., & Ludueña, R. F. (1985) *EMBO J.* 4, 51–56.
- Little, M., & Ludueña, R. F. (1987) *Biochim. Biophys. Acta* 912, 28–33.
- Little, M., & Seehaus, T. (1988) *Comp. Biochem. Physiol.* 90B, 655–670.
- Ludueña, R. F., Banerjee, A., & Khan, I. A. (1992) *Curr. Opin. Cell Biol.* 4, 53–57.
- Lutkenhaus, J. (1993) *Mol. Microbiol.* 9, 403–409.
- Manavalan, P., & Johnson, W. C. (1987) *Anal. Biochem.* 167, 76–85.
- Mandelkow, E., & Mandelkow, E. M. (1995) *Curr. Opin. Cell Biol.* 7, 72–81.
- Margolin, W., Wang, R., & Kumar, M. (1996) *J. Bacteriol.* 178, 1320–1327.
- Melki, R., Carlier, M. F., Pantaloni, D., & Timasheff, S. N. (1990) *Biochemistry* 29, 9143–9152.
- Monasterio, O., Andreu, J. M., & Lagos, R. (1995) *Comments Mol. Cell. Biophys.* 8, 273–306.
- Moritz, M., Braunfeld, M. B., Sedat, J. W., Alberts, B., & Agard, D. A. (1995) *Nature* 378, 638–640.
- Mukherjee, A., & Lutkenhaus, J. (1994) *J. Bacteriol.* 176, 2754–2758.
- Mukherjee, A., Dai, K., & Lutkenhaus, J. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90, 1053–1057.
- Muñoz, V., & Serrano, L. (1994) *Nat. Struct. Biol.* 1, 399–409.
- Noel, J. P., Hamm, H. E., & Sigler, P. B. (1993) *Nature* 366, 654–662.
- Oakley, B. R. (1992) *Trends Cell Biol.* 2, 1–5.
- Pai, E. F., Kabsch, W., Krengel, U., Holmes, K., John, J., & Wittinghofer, A. (1989) *Nature* 341, 209–214.
- Perczel, A., Hollósi, M., Tusnády, G., & Fasman, G. D. (1991) *Protein Eng.* 4, 669–679.
- Perczel, A., Park, K., & Fasman, G. D. (1992a) *Anal. Biochem.* 203, 83–93.
- Perczel, A., Park, K., & Fasman, G. D. (1992b) *Proteins* 13, 57–69.
- Plá, J., Sanchez, M., Palacios, P., Vicente, M., & Aldea, M. (1991) *Mol. Microbiol.* 5, 1681–1686.
- Ponstingl, H., Little, M., Kraus, E., & Kempf, T. (1979) *Nature* 282, 423–424.
- Ponstingl, H., Kraus, E., Little, M., & Kempf, T. (1981) *Proc. Natl. Acad. Sci. U.S.A.* 78, 2757–2761.
- Rao, S., Krauss, N. E., Heerding, J. M., Swindell, C. S., Ringel, I., Orr, G. A., & Horuntz, S. B. (1994) *J. Biol. Chem.* 269, 3132–3134.
- Rao, S., Orr, G. A., Chaudhary, A. G., Kingston, D. G. I., & Horuntz, S. B. (1995) *J. Biol. Chem.* 270, 20235–20238.
- RayChaudhuri, D., & Park, J. T. (1992) *Nature* 359, 251–254.
- RayChaudhuri, D., & Park, J. T. (1994) *J. Biol. Chem.* 269, 22941–22944.
- Redecker, V., Levilliers, N., Schmitter, J. M., Le Caer, J. P., Rossier, J., Adoutte, A., & Bré, M. H. (1994) *Science* 266, 1688–1691.
- Reijo, R. A., Cooper, E. M., Beagle, G. J., & Huffaker, T. C. (1994) *Mol. Biol. Cell* 5, 29–43.
- Rost, B., & Sander, C. (1993a) *J. Mol. Biol.* 232, 584–599.
- Rost, B., & Sander, C. (1993b) *Proc. Natl. Acad. Sci. U.S.A.* 90, 7558–7562.
- Rost, B., & Sander, C. (1994a) *Proteins* 19, 55–72.
- Rost, B., & Sander, C. (1994b) *Proteins* 20, 216–226.
- Rost, B., & Sander, C. (1995) *Proteins* 23, 295–300.
- Rost, B., Sander, C., & Schneider, R. (1994a) *Comput. Appl. Biosci.* 10, 53–60.
- Rost, B., Sander, C., & Schneider, R. (1994b) *J. Mol. Biol.* 235, 13–26.
- Sage, C. R., Dougherty, C. A., Burns, R. G., Wilson, L., & Farrell, K. W. (1995) *Biochemistry* 34, 7409–7419.
- Sander, C., & Schneider, R. (1991) *Proteins* 9, 56–68.
- Shivanna, B. D., Mejillano, M. R., Williams, T. D., & Himes, R. H. (1993) *J. Biol. Chem.* 268, 127–132.
- Sobel, S. G., & Snyder, M. (1995) *J. Cell Biol.* 131, 1775–1788.
- Solovyyev, V. V., & Salamov, A. A. (1994) *Comput. Appl. Biosci.* 10, 661–669.
- Sreerama, N., & Woody, R. W. (1993) *Anal. Biochem.* 209, 32–44.
- Story, R. M., & Steitz, T. A. (1992) *Nature* 355, 374–376.
- Surewicz, W. K., Mantsch, H. H., & Chapman, D. (1993) *Biochemistry* 32, 389–394.
- Susi, H., & Byler, D. M. (1986) *Methods Enzymol.* 130, 290–311.
- Thompson, J. D., Higgins, D. G., & Gibson, J. (1994) *Nucleic Acids Res.* 22, 4673–4680.
- Tong, L. A., de Vos, A. M., Milburn, M. W., & Kim, S. H. (1991) *J. Mol. Biol.* 217, 503–516.
- Valencia, A., Chardin, P., Wittinghofer, A., & Sander, C. (1991) *Biochemistry* 30, 4637–4648.
- Vicente, M., & Errington, J. (1996) *Mol. Microbiol.* 20, 1–7.
- Voskuil, J. L. A., Westerbeek, C. A. M., Wu, C., Kolk, A. H. J., & Nanninga, N. (1994) *J. Bacteriol.* 176, 1886–1893.
- Woody, R. W. (1995) *Methods Enzymol.* 246, 34–71.
- Yang, J. T., Chuen-Shang, C. W., & Martinez, H. M. (1986) *Methods Enzymol.* 130, 208–269.
- Yi, Q. M., & Lutkenhaus, J. (1985) *Gene* 36, 241–247.
- Zheng, Y., Jung, M. K., & Oakley, B. R. (1991) *Cell* 65, 817–823.
- Zheng, Y., Wong, M. L., Alberts, B., & Mitchison, T. (1995) *Nature* 378, 578–583.