

# Study of the Interactions Between a Proline-Rich Protein and a Flavan-3-ol by NMR: Residual Structures in the Natively Unfolded Protein Provides Anchorage Points for the Ligands

Christine Pascal,<sup>1</sup> Franck Paté,<sup>1</sup> Véronique Cheynier,<sup>1</sup> Marc-André Delsuc<sup>2,3,4</sup>

<sup>1</sup> INRA-UMR Sciences Pour l'Oenologie (SPO), 2 place Pierre Viala, Montpellier F34060, France

<sup>2</sup> INSERM, U554, Montpellier F-34090, France

<sup>3</sup> Université de Montpellier 1 et 2, CNRS, UMR 5048, Centre de Biochimie Structurale, Montpellier F34090, France

<sup>4</sup> IGBMC, Biomolecular NMR Group, INSERM, U596, CNRS, UMR7104, Illkirch F-67400, France

Received 28 October 2008; revised 11 April 2009; accepted 14 April 2009

Published online 28 April 2009 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bip.21221

## ABSTRACT:

Astringency is one of the major organoleptic properties of food and beverages that are made from plants, such as tea, chocolate, beer, or red wine. This sensation is thought to be due to interactions between tannins and salivary proline-rich proteins, which are natively unfolded proteins. A human salivary proline-rich protein, namely IB-5, was produced by the recombinant method. Its interactions with a model tannin, epigallocatechin gallate (EGCG), the major flavan-3-ol in green tea, were studied here. Circular dichroism experiments showed that IB-5 presents residual structures (PPII helices) when the ionic strength is close to that in saliva. In the presence of these residual structures, IB-5 undergoes an increase in structural content upon binding to EGCG. NMR data corroborated the presence of preformed structural elements within the protein prior to binding and a partial assignment was proposed, showing partial structuration. TOCSY experiments showed that amino acids that are involved in PPII helices are more likely to interact with EGCG than those in random coil regions, as if they were

anchorage points for the ligand. The signal from IB-5 in the DOSY NMR spectrum revealed an increase in polydispersity upon addition of EGCG while the mean hydrodynamic radius remained unchanged. This strongly suggests the formation of IB-5/EGCG aggregates. © 2009 Wiley Periodicals, Inc. *Biopolymers* 91: 745–756, 2009.  
**Keywords:** astringency; EGCG; epigallocatechin gallate; flavan-3-ol; IB-5; proline-rich protein; PRP; natively unfolded protein; preformed structural elements; PPII helices; CD; NMR; TOCSY; DOSY; <sup>1</sup>H-<sup>13</sup>C HSQC; NOESY; ROESY

This article was originally published online as an accepted preprint. The “Published Online” date corresponds to the preprint version. You can request a copy of the preprint by emailing the *Biopolymers* editorial office at [biopolymers@wiley.com](mailto:biopolymers@wiley.com)

## INTRODUCTION

Tannins, considered to be part of the plant defense system against environmental stressors,<sup>1–3</sup> are found in a wide variety of food and beverages. They have a range of effects on animals, including growth rate depression and inhibition of digestive enzymes.<sup>4</sup> These biological effects are related to the ability of tannins to interact with proteins and sometimes to cause them to precipitate.<sup>5</sup> These interactions also substantially influence the taste properties and palatability of many plant products.

Correspondence to: Marc-André Delsuc; e-mail: [delsuc@igbmc.fr](mailto:delsuc@igbmc.fr)

Contract grant sponsor: ANR (project Protanin)

© 2009 Wiley Periodicals, Inc.

Our interest is focused on tannin interactions with salivary proline-rich proteins, which are thought to be responsible for astringency<sup>3,6–8</sup> and to be part of animal and human defense mechanisms against dietary tannins.<sup>4,9–11</sup>

Proline-rich proteins constitute about two thirds of human parotid salivary proteins.<sup>12,13</sup> They are usually divided into glycosylated, acidic, and basic types, which have widely different functions.<sup>14</sup> Acidic PRPs are reported to bind calcium and to inhibit crystal growth, functions that may be important in maintaining calcium homeostasis in the mouth.<sup>5</sup> They are also involved in dental pellicle.<sup>5</sup> Glycosylated PRPs ensure oral lubrication<sup>15</sup> and bind oral bacteria.<sup>16</sup> The only reported function of basic PRP is their ability to bind tannins.

The structural and physicochemical basis of tannin interactions with human salivary proline-rich proteins in relation with astringency is poorly understood. The experiments were achieved in a very simple medium: water added with 100 mM of NaCl and acidified to pH 3.5. These conditions were chosen in connection with wine, on which we were focusing in this study. 100 mM is the mean ionic strength of saliva<sup>17</sup> and of wine.<sup>18</sup> pH in the mouth when drinking wine was checked to be driven by wine pH and therefore to be around 3.5. As the strength of the interaction between flavan-3-ols and proteins increases with the number of galloyl rings,<sup>19,20</sup> a flavan-3-ol monomer, epigallocatechin gallate (EGCG), was chosen as a tannin model. Protein IB-5, a 70 residues basic human salivary proline-rich protein, was studied previously by Charlton et al.<sup>21</sup> in the presence of epicatechin and hydrolysable tannins (compounds containing a central core of glucose or other polyhydric alcohol esterified with gallic acid or hexahydroxydiphenic acid) in 10% of dimethylsulfoxide (DMSO). Using NMR, these authors determined association constants between the protein binding sites and the ligands by a titration method. Here IB-5 was produced by the recombinant method.<sup>22</sup> Its amino acid sequence presents tandem repeats as shown on Figure 1. Its pI was calculated to be 11.2. The primary structure of IB-5 is characterized by low complexity,<sup>23</sup> calculated to be 0.157 for the dipeptides (the higher complexity available being 1.0) and 0.531 for the

```

1      SPPG KPQGPPQQE
14     GNKPQGPPPP
24     G KPQGPPPA
33     GGNPQQPQAPPA
45     G KPQGPPPPPQ
56     GGRPPRPAQGQQPPQ

```

FIGURE 1 IB-5 sequence,<sup>22</sup> aligned to show tandem repeats.

tripeptides. This low complexity is due to a strongly reduced amino acid repertoire. Only nine different amino acids are present in the protein, resulting in 14 PP dipeptide repeats and five PQG tripeptides. As the other salivary proline-rich proteins, IB-5 belongs to the natively unfolded protein family.<sup>24–26</sup> Natively unfolded proteins have several functions that result from a near absence of secondary structure.<sup>24,26–28</sup> In some cases, thanks to their flexible backbone, they can undergo a disorder-to-order transition upon binding.<sup>29,30,31</sup> Our aim here was to evidence a structural change for the protein when flavan-3-ols bind to PRPs.

## MATERIALS AND METHODS

### Chemicals

EGCG was purchased from Sigma-Aldrich (St. Louis, MO). Human salivary proline-rich protein, IB-5, was produced using the yeast *Pichia pastoris* as a host organism and purified as described previously.<sup>22</sup> N-terminal sequencing of the protein showed the presence of two sequences in the protein sample, corresponding to IB-5 and to the same sequence with three additional N-terminal amino acids.<sup>22</sup> MALDI TOF mass spectrometry confirmed the presence in the sample of a protein with a molecular weight of 6923 Da (IB-5) and a second protein with a molecular weight of 7237 Da (IB-5 with three more amino acids, SAR, on its N-terminal end). This protein mixture was used without further purification, calculating protein concentration using a mean molecular weight of 7000 Da.

### Sample Preparation

IB-5 concentration was set at 0.5 mg/mL (71  $\mu$ M) in all experiments. CD spectra were recorded in two aqueous media: HCl pH 3.5, 100 mM of NaCl, and HCl pH 3.5, 5 mM of NaCl. These media will be respectively called 100 mM of NaCl and 5 mM of NaCl in this article. NMR spectra were recorded in H<sub>2</sub>O/D<sub>2</sub>O (90/10, v/v) acidified to pH 3.5 with HCl and with an ionic strength of 100 mM (adjusted with NaCl).

### Circular Dichroism

Far ultraviolet (200–260 nm) circular dichroism spectra were recorded on a Chirascan Circular Dichroism spectrometer (Applied Photophysics, Leatherhead, UK) at room temperature in a circular cell (0.5 mm length path). All spectra were averaged from five scans. The EGCG spectrum was recorded as a control and subtracted from the EGCG/protein mixture spectrum. The protein concentration was fixed at 0.5 mg/mL and the EGCG/IB-5 molar ratio at 6.

### Nuclear Magnetic Resonance

Homonuclear 1D and two-dimensional (2D) proton nuclear magnetic resonance (NMR) experiments were recorded on a Bruker Avance spectrometer, operating at 600 MHz, and equipped with a cryoprobe. Heteronuclear <sup>1</sup>H-<sup>13</sup>C HSQC experiments were recorded on a Bruker Avance spectrometer operating at 500 MHz and

equipped with a cryoprobe. In all experiments, solvent suppression was achieved with an excitation sculpting sequence.<sup>32</sup> All experiments were processed using the NPK software<sup>33</sup> and NMRnotebook (NMRtec S.A.S., Ilkirch-Grafenstaden, France).

Two samples were studied, sample A is a 71  $\mu$ M solution of IB-5 in H<sub>2</sub>O/D<sub>2</sub>O (90/10, v/v), at pH 3.5, 100 mM NaCl, 0.1 mM ethanol (4.6 ppm w/w). Sample B is equivalent to A, with 12 equivalent of EGCG added. All experiments were run at a temperature of 297 K.

A <sup>1</sup>H spectral width of 7183.9 Hz was used for all experiments. TOCSY experiments were performed using 512 increments, 32 scans, for a total acquisition time of 7 h. Three different experiments were acquired with TOCSY contact time values of 45, 60, and 75 ms. The three experiments were merged after processing.

NOESY experiments were acquired using 400 increments (512 for sample B), 64 scans (80 scans for sample B) and a mixing time of 200 ms for a total time of 12 h (19 h for sample B).

ROESY experiments were acquired using 440 increments (512 for sample B), 32 scans (64 scans for sample B), and a mixing time of 200 ms with a mixing power of 8 kHz for a total time of 6.5 h (15 h for sample B).

<sup>1</sup>H-<sup>13</sup>C HSQC experiments were acquired using 256 increments, 512 scans for a total time of 61 h.

The DOSY NMR experiments were measured using the LED (Longitudinal Eddy-current Delay) sequence<sup>34,35</sup> with 40 increments and 128 scans, with a diffusion time ( $\Delta$ ) of 150 ms and bipolar PFG of total duration ( $\delta$ ) of 2.6 ms for a total time of 2.5 h. The gradient waveform was trapezoidal, and the gradient intensity was scanned from 2 to 95% of the nominal 44 G/cm value delivered by the system. The DOSY NMR data were processed by Inverse Laplace Transform (ILT) on 256 points.<sup>36</sup>

Hydrodynamic radii,  $r_H$ , were computed from diffusion coefficients using the approach proposed by Wilkins et al.<sup>37</sup> Ethanol was used as an internal reference for viscosity variations, with its  $r_H$  set to 2.0 Å.

The tentative assignment of IB-5 NMR signals was performed from the TOCSY experiment, with the help of the Rescue software.<sup>38,39</sup>

Chemical shift analysis was performed using either the work from Wishart et al.<sup>40</sup> considering amino acids followed or not by a proline; or from the work of Schwarzsinger et al.<sup>41</sup> In the latter case, chemical shifts for each amino acid were computed using an additive approach, based on the primary sequence analysis. From the predicted values, the mean chemical shift values for each kind of amino acid were also computed, and are presented in Tables I and II.

## RESULTS AND DISCUSSIONS

### Circular Dichroism Spectra

IB-5 CD spectra recorded in two media differing only in their ionic strength are not identical (Figures 2 and 3), indicating that the protein adopts different conformations in these media. Both spectra are consistent with natively unfolded proteins: they present a deep minimum near 200 nm and a molar ellipticity close to zero in the vicinity of 222 nm.<sup>25,42</sup> However, the minimum of the spectrum of IB-5 in 100 mM

NaCl is deeper than that in 5 mM NaCl, which reflects a higher degree of structure for IB-5 in 100 mM NaCl.<sup>42</sup> The deep minimum in the spectrum in 100 mM NaCl resembles that of poly(L-proline) II helices (PPII).<sup>42</sup> Nevertheless, it does not present a positive band around 220–230 nm, as expected for PPII, and this suggests a higher degree of disorder for IB-5 in 100 mM NaCl compared to PPII. It has already been noted that unfolded peptides have significant amounts of PPII helices in the form of short helical regions<sup>43</sup> and that such preformed structural elements could be of importance in partner recognition upon binding.<sup>44</sup> It has also been shown that such proteins may undergo folding upon binding.<sup>29,31,42</sup>

IB-5 structure was then investigated after addition of EGCG to both media (Figures 2 and 3). EGCG asymmetric carbon interacts with circularly polarized light, resulting in a non-null contribution to the CD spectrum. The spectrum of EGCG can therefore be used as a control and subtracted from that of EGCG/protein mixture. In 100 mM NaCl, the CD spectrum of IB-5 changed upon addition of EGCG. A positive band appeared in the 215–230 nm region, replacing the negative shoulder (Figure 2). This indicates that the secondary structure of IB-5 was modified upon binding and corresponds to a higher extent of structuration. On the contrary, in 5 mM NaCl, the CD spectra before and after addition of EGCG were identical (Figure 3). Therefore, an increase in structural content upon binding is only observed when IB-5 shows preformed structural elements prior to binding. As discussed elsewhere,<sup>45</sup> this difference in protein structure due to ionic strength results in a different colloidal behavior: aggregates that are observed in 100 mM NaCl are far more stable along time than those observed in 5 mM NaCl. Protein residual structures are then likely to be of importance in the binding mechanism. Their implication in the binding process was further explored by NMR spectroscopy.

### 1D <sup>1</sup>H Spectra of IB-5 and EGCG/IB-5 Mixtures

NMR experiments were all recorded in 100 mM NaCl, in which IB-5 presents residual structures. Our aim was to investigate the implication of these structures in both the binding mechanism and the folding process.

As reported previously by Charlton et al.,<sup>21</sup> severe signal overlap is observed in the 1D <sup>1</sup>H NMR spectrum of IB-5 (data not shown). Because of the water suppression technique, all the NMR lines between 4 and 5.5 ppm are missing. After addition of EGCG, small spectral modifications are observed (data not shown): notably broadening of resonances in the amide region (8–9 ppm), broadening and upfield shift of some aliphatic signals (1–4 ppm).

**Table I** Proton Signals Corresponding to IB-5 Amino Acids

	Number in IB-5	Name on Spectra	Amide Proton (Peptide Bond)	H $\alpha$	H $\beta$	H $\gamma$	H $\delta$	H $\epsilon$	Amide Proton (Side Chain)
Ala (A)	4 (1 AP)	Aref	8.43	4.30	1.39				
		ArefP	8.16	4.43	1.35				
		Aa	8.39	4.26	1.36				
		Ab	8.34	4.36	1.36				
		Ac	8.67	4.36	1.37				
		Ad	8.29	4.54	1.31				
Gly (G)	13 (4 GP)	Gref	8.37	3.97					
		GrefP	8.22	4.10					
		Ga	8.15	3.63/3.98					
		Gb	8.33	3.84/3.93					
		Gc	8.36	3.85/3.93					
		Gd	8.26	3.87/3.94					
		Ge	8.16	3.90					
		Gf	8.22	3.90					
		Gg	8.33	3.91					
		Gh*	8.41	3.92					
		Gi	8.42	3.93					
Lys (K)	4 (4 KP)	Gj*	8.20	3.98/4.15					
		Kref	8.29	4.32	1.75/1.84	1.44	1.68	3.00	
		KrefP	8.12	4.45	1.74/1.80	1.45	1.69	2.97	
		Ka	8.21	4.58	1.76	1.41	1.67		
		Kb	8.08	4.60	1.77	1.42	1.66	2.97	7.49
Asp (N)	2 (1 NP)	Nref	8.37	4.76	2.75/2.83				
		NrefP	8.32	4.89	2.67/2.82				
		Na	8.26	4.64	2.70/2.75				
		Nb	8.33	4.94	2.64/2.80				
Pro (P)	29 (14 PP)	Pref		4.40	1.94/2.29	2.02	3.63		
		PrefP		4.53	1.91/2.31	2.01	3.60		
		Pa*		4.38	1.89/2.26	1.99	3.79/3.61		
		Pb*		4.66	1.87/2.30	2.00	3.58		
		Pc		4.75	1.86/2.35	2.02			
		Pd		4.87	2.11/2.42	1.92/1.81	3.52		
Glu (Q or E)	14 + 1 (2 QP)	Qref	8.48	4.33	1.99/2.12	2.36	7.52	6.85	
		QrefP	8.32	4.46	1.93/2.09	2.38	7.53	6.88	
		Qa	8.25	4.25	1.91/2.12	2.32			
		Qb*	8.42	4.26	1.94/2.05	2.34			
		Qc	8.40	4.28	1.93/2.07	2.31			
		Qd	8.31	4.29	1.97/2.08	2.32			
		Qe*	8.50	4.30	1.95/2.08	2.38			
		Qf	8.19	4.31	2.04/1.91	2.28			
		Qg	8.42	4.35	1.94/2.05	2.43			
		Qh	8.42	4.58	1.90/2.05	2.36			
		Qi	8.33		1.92/2.06	2.33			
Arg (R)	2 (2 RP)	Rref	8.23	4.34	1.76/1.86	1.63	3.20	8.07	
		RrefP	8.22	4.46	1.81	1.67	3.21	8.07	
		Ra	8.26	4.57	1.79	1.67	3.18	7.14	
		Rb	8.12	4.58	1.80	1.66	3.18	7.14	
Ser (S)	1 (0 SP)	Sref	8.21	4.61	3.89	3.87			
		S		4.40	3.85				

These assignments were obtained from the TOCSY spectra of IB-5. Data in brackets indicate the number of occurrence of the amino acid followed by a proline residue in the primary sequence of IB-5. Xref and XrefP are chemical shifts for random coil residues computed from the mean value of the chemical shift on the primary sequence.<sup>41</sup> XrefP corresponds to the chemical shifts of a residue that is followed by a proline.

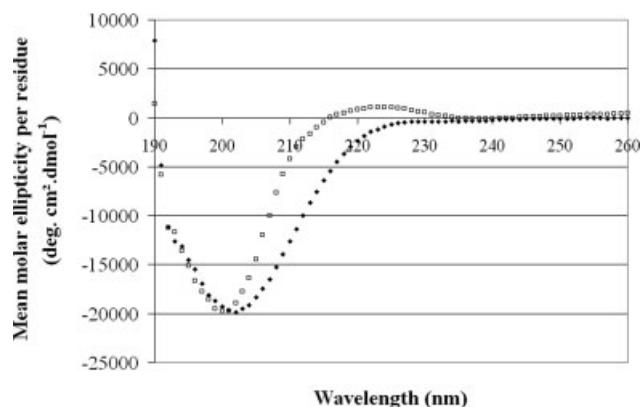
\*Indicates that the intensity of the signal is high, probably meaning that the signal corresponds to more than one of the considered amino acid.

**Table II** Carbon Signals Corresponding to IB-5 Amino Acids, All Other Parameters as in Table I

	Number in IB-5	Name on Spectra	C $\alpha$	C $\beta$	C $\gamma$	C $\delta$	C $\epsilon$
Ala (A)	4 (1 AP)	Aref	53.16	18.98			
		ArefP	50.68	19.30			
		Aa-c	60.73	16.52			
		Ad	nd	15.51			
Gly (G)	13 (4 GP)	Gref	45.35				
		GrefP	43.25				
		G	42.56				
		Gi	41.90				
Lys (K)	4 (4 KP)	Kref	56.98	32.77	24.90	28.92	41.87
		KrefP	54.77	33.20			
		Ka	nd	29.98	21.79	24.07	nd
		Kb	nd	29.98	nd	nd	39.29
Asp (N)	2 (1 NP)	Nref	54.66	40.86	179.22		
		NrefP	51.31	39.1			
		Na-b	nd	36.42			
Pro (P)	29 (14 PP)	Pref	63.33	31.84	27.20	50.34	
		PrefP	61.64	32.20			
		Pa-c	nd	29.41	24.70	47.97	
		Pd	nd	28.20	28.16	47.97	
Glu (Q or E)	14 + 1 (2 QP)	Qref	57.35	30.00	36.06	182.47	
		QrefP	54.16	29.50			
		Q	53.29	27.01	31.29		
		Rref	56.79	30.68	27.21	43.15	160.11
Arg (R)	2 (2 RP)	RrefP	54.39	30.90			
		Ra-b	nd	27.66	26.60	40.75	
Ser (S)	1 (0 SP)	Sref	58.72	63.80			
		S	nd	nd			

## 2D <sup>1</sup>H Spectra

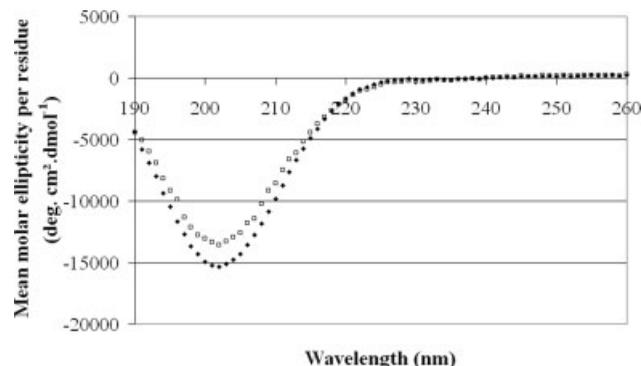
**IB-5 Protein TOCSY Spectrum.** A set of experiments was acquired on this system: TOCSY, NOESY, and ROESY. The NOESY spectra present very little information. Apart from a few contact peaks in the HN-H $\alpha$  region and in the proline H $\gamma$ -H $\delta$  region, no correlation can be observed in this spectrum, particularly in the HN-HN area (data not shown).



**FIGURE 2** Circular dichroism spectra of (◆) IB-5 and of (□) mixture of IB-5 and EGCG after subtraction of EGCG signal (EGCG/protein molar ratio = 6) recorded in NaCl 100 mM, HCl pH 3.5.

Along with the low complexity of the primary sequence, the completely uninformative NOESY experiment prevented the possibility to realize sequential assignment, even partially.

On the contrary, IB-5 TOCSY spectrum presents numerous well-defined correlations. Coupled spin systems were extracted and assigned to amino acids based on the chemical shift values and the correlation patterns. The CRAACK procedure<sup>46</sup> was used to confirm assignments. Chemical shift values were then



**FIGURE 3** Circular dichroism spectra of (◆) IB-5 and of (□) mixture of IB-5 and EGCG after subtraction of EGCG signal (EGCG/protein molar ratio = 6) recorded in NaCl 5 mM, pH 3.5.

compared to random coil values<sup>40</sup> for amino acids followed or not by a proline, and to the prediction made from the primary sequence analysis.<sup>41</sup> The observed chemical shift spreading is larger than the predicted one. It could be thus determined that some residues have chemical shift values that depart from the random coil values. Results are shown in Tables I and II and detailed later.

Some amino acids (alanine, asparagine, glycine, glutamine, and proline) are found to have several sets of chemical shifts, indicating that distinct chemical environments are experienced by these amino acids. More precisely, several glycine spin systems present pairs of H $\alpha$  resonances with non-equivalent chemical shifts; several residues with long chain (Gln, Lys) present CH<sub>2</sub> chemical shifts far from the expected random coil values given by Wishart et al.<sup>40</sup>; more diversity is found in the spectrum than expected from the primary sequence. To illustrate this last point, nine different glutamine signals are observed (Table I), although only five different environments are found in the primary sequence (namely QG, QQ, QP, QA, and QNterm). Chemical shifts observed for the glutamine H $\alpha$  range from 4.25 to 4.58, extending the predicted range of 4.46 to 4.47 for Q followed by a P or of 4.30 to 4.35 predicted for all other Q residues. Similarly, 10 different glycine signals are observed for only six different environments (GP, GK, GN, GG, GR, GQ), and their H $\alpha$  chemical shifts range from 3.63 to 4.15, which has to be compared with a predicted range from 3.95 to 4.10.

Four types of prolines were detected, corresponding to four H $\alpha$  signals (4.38, 4.66, 4.75, and 4.87 ppm). According to Wishart et al.,<sup>40</sup> the chemical shift of the H $\alpha$  of a proline followed by an alanine is 4.42 ppm, while it is 4.73 ppm when followed by another proline. Alternatively, the Schwarzingger et al. approach<sup>41</sup> predicts proline H $\alpha$  chemical shifts between 4.51 and 4.56 when followed by another proline, and from 4.38 to 4.42 otherwise.

Residues preceding a proline in the sequence are expected to present a downfield shift to the H $\alpha$  chemical shift of about 0.1 ppm. However, this effect alone cannot explain the chemical shift spreading that is experimentally observed. It can thus be inferred that some partial secondary structures are likely to be present in IB-5. Moreover, many observed shifts from random coil position are further downfield compared to corrected random coil values, in particular for H $\alpha$  chemical shifts of Ad, Ka, Kb, Nb, Pc, Pd, Qh, Ra, Rb, and Gj (see Table I).

This pattern is suggestive of  $\beta$ -sheet structures. As already stated, IB-5 presents poly(L-proline) II helices (it should be noted that this secondary structure is not restricted to prolines). According to Avbelj et al.,<sup>47</sup> the electrostatic field distribution induced by PPII helices is comparable to the one found around  $\beta$ -sheets, and in consequence, the chemical shift index (CSI) pattern of PPII helices should be close to that of  $\beta$ -sheet.

For the residues considered here, the observed chemical shift values can be explained either as the cumulated effect of a following proline and a further polypeptidic chain structuration, or as the only effect of a structuration, the former explanation being the most likely. In both cases, the pattern observed here can be assigned to a PPII partial structuration, in agreement with the CD observation. However this is probably not the only structuration present, as upfield shifts are also observed, for instance for glycines or glutamines.

A few impurities are also observed in the spectra. These impurities are in a low concentration and could be assigned to the presence of some *cis*-proline conformers. However, no evidence can enforce this hypothesis. On the other hand, these impurities do not display any particular pattern upon EGCG binding, and were not further studied.

**IB-5/EGCG Mixture TOCSY Spectrum.** Comparison of the TOCSY spectra recorded before and after addition of EGCG allows analysis of signal changes. No intensity modification could be observed and strong signal changes probably correspond to protons of residues involved in binding.

Most signals are shifted upfield (0.01–0.06 ppm). This observation cannot be attributed to a bulk effect or a solvent effect, as all ethanol and impurity signals remain completely unchanged. The upfield shift upon binding may be due to the shielding cone of the aromatic rings of EGCG. Indeed, face-to-face stacking is described in the literature for PRP/flavan-3-ol interactions,<sup>10,48–50</sup> which may result in the location of the interacting residues in the shielding cone of EGCG. Most signals are broadened after binding. Charlton et al.<sup>21</sup> reported similar observations on a IB-5/pentagalloyl-glucose mixture. This may be due to protein oligomerization in the presence of EGCG because protons of residues implied in different types of aggregates are in different environments, leading to broader signals.

The signals of backbone amide protons are affected more than most other signals (Figures 4 and 5). These protons are known to be highly sensitive to the environment, in particular to the presence and nature of hydrogen bonds with the solvent or the carboxyl functions of either the peptide bonds or the amino acids side chains. The addition of EGCG is likely to modify this network as it provides new hydrogen bond acceptors (e.g. phenol hydroxyl, phenol carboxyl functions).

No modification (Figure 4) was detected for the side chain protons of residues: asparagine Na, alanine Aa, and glutamine Qe.

Weak modifications (Figure 4) were observed on the side chain signals of arginines, asparagines Nb, alanines Ab, Ac, and prolines Pa.

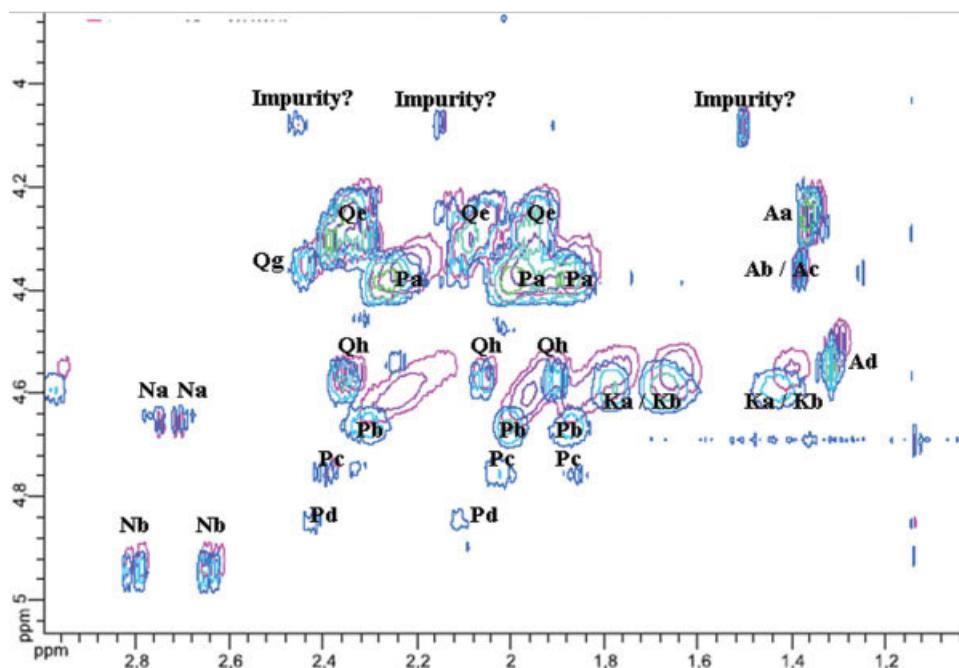


FIGURE 4 Comparison of IB-5 TOCSY spectra recorded in H<sub>2</sub>O/D<sub>2</sub>O (90/10, v/v), HCl pH 3.5, NaCl 100 mM, ethanol 0.1 mM before (in blue) and after (in pink) addition of EGCG (EGCG/protein molar ratio = 12): contour plot of the cross peaks region between H $\alpha$  and H $\beta$ , H $\delta$  or H $\gamma$ .

The most striking modifications were observed for the side chain signals of lysines, alanines Ad, glycines Gj (not shown), and prolines Pb, Pc, Pd (Figure 4). The signals of Pc and Pd were weak on the initial IB-5 spectrum. The interaction with EGCG causes the broadening of these signals such that they totally disappear after addition of EGCG. All signals

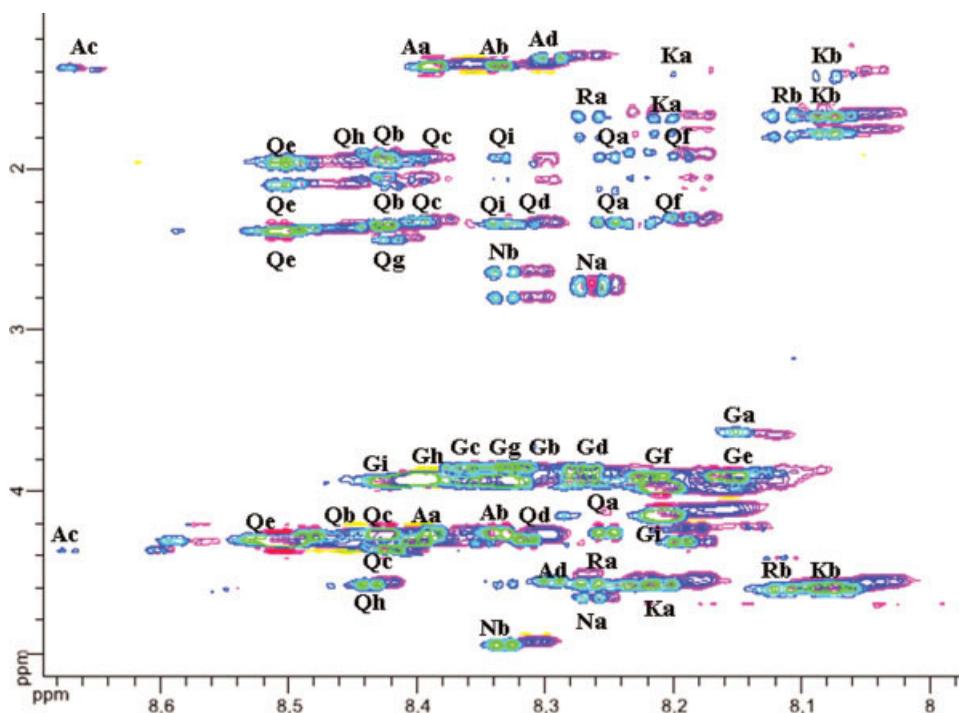


FIGURE 5 Comparison of IB-5 TOCSY spectra recorded in H<sub>2</sub>O/D<sub>2</sub>O (90/10, v/v), HCl pH 3.5, NaCl 100 mM, ethanol 0.1 mM before (in blue) and after (in pink) addition of EGCG (EGCG/protein molar ratio = 12): contour plot of the cross peaks region between HN and H $\alpha$ , H $\beta$ , H $\delta$  or H $\gamma$ .

listed earlier are shifted upfield and probably experience an aromatic shielding effect.

The residues experiencing the largest modification upon binding are also those which were mentioned earlier as probably being engaged in a PPII secondary structure, while being followed by a proline.

The preferential implication of amino acids preceding proline in the primary sequence of proline-rich peptides has been previously reported for a 19 amino acid peptide interacting with a flavan-3-ol monomer or hydrolysable tannins.<sup>10</sup> In this work, the hypothesis of a conformational change of the peptide upon binding was discarded after analysis of the values of the coupling constants. The chemical shift change for these amino acids was attributed to the proximity of the polyphenol aromatic rings. Although proline residues followed by proline were found to be preferred interaction sites, the authors proposed a favorable geometry of these residues that are described as being in a  $\beta$ -sheet conformation. Although these observations are consistent with ours, the PRP used here is larger than that used in the previous study and a structural change upon binding was observed by circular dichroism. Other authors such as Fuxreiter et al.<sup>44</sup> have hypothesized that preformed structural elements in natively unfolded proteins are anchorage points for the ligands during the binding process. Our results illustrate the importance of such residual structures in the binding process: these structural elements form preferred interaction sites and seem to determine the occurrence of a folding process upon interaction as shown by circular dichroism.

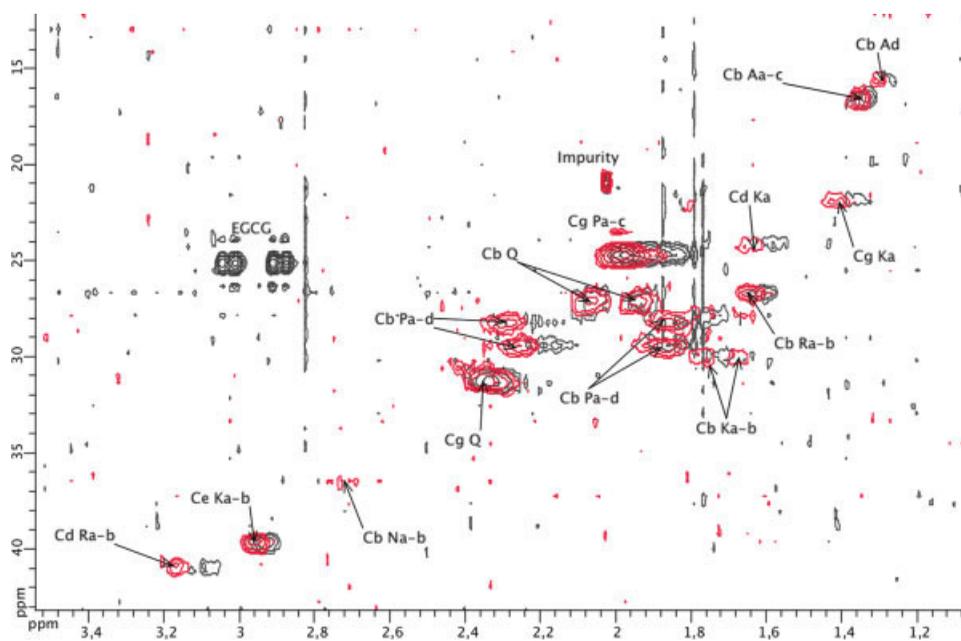
### HSQC Spectra

An additional set of  $^1\text{H}$ - $^{13}\text{C}$  HSQC experiments was recorded on the same samples as described earlier to evaluate the influence of EGCG/protein interaction on  $^{13}\text{C}$  chemical shifts (Figure 6). Because of the strong NMR signal of the solvent, most of the  $\text{H}\alpha/\text{C}\alpha$  correlations were wiped off the spectra. Thus, no comparison with random coil type chemical shifts has been obtained based on these spectra.

Comparing the HSQC spectrum of IB-5 upon addition of EGCG mixture to that obtained without EGCG (Figure 6) shows that  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts of the residues are affected in a similar manner by the presence of EGCG. However, because the chemical shift spreading is much larger in carbon spectra than in proton ones, the interaction effect is less visible on this nucleus. This feature could be assigned to the ring effect of the aromatic systems in EGCG.<sup>51</sup>

### DOSY Spectra

DOSY experiments were performed to monitor the formation of protein aggregates after addition of EGCG. In these experiments, the presence of a small amount of ethanol allowed apparent hydrodynamic radii to be determined for protein and the protein aggregates, while having no effect on the interaction. The hydrodynamic radii,  $r_{\text{H}}$ , are determined considering that  $r_{\text{H}}$  is proportional to the reciprocal of the diffusion coefficient, and using ethanol as an internal reference.<sup>37</sup> Diffusion coefficients and hydrodynamic radii are given with a 5% uncertainty.



**FIGURE 6** Superposition of 2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC NMR spectra of IB-5 before (blue lines) and after (red/pink lines) addition of EGCG recorded in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (90/10, v/v), HCl pH 3.5, NaCl 100 mM, ethanol 0.1 mM.

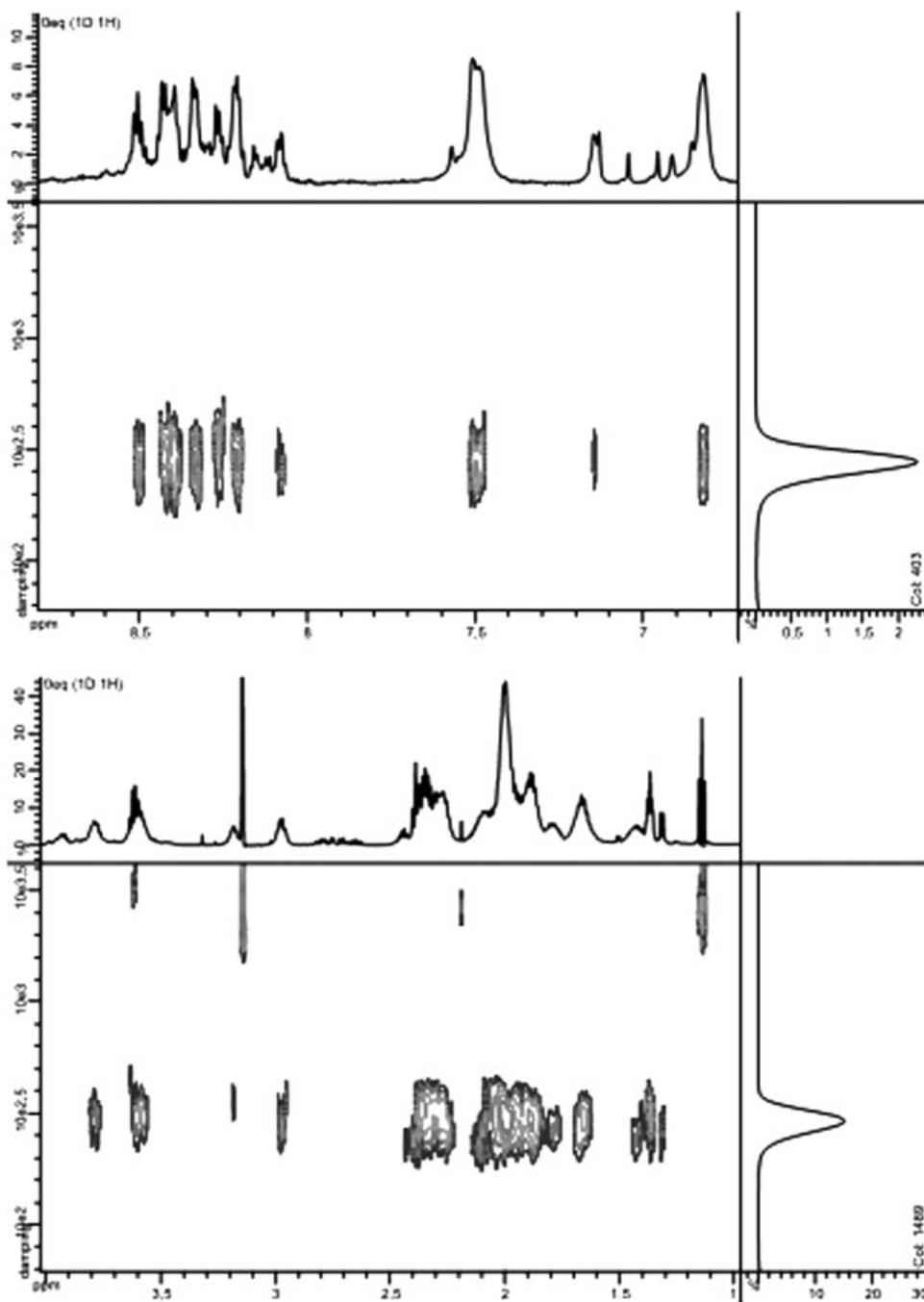
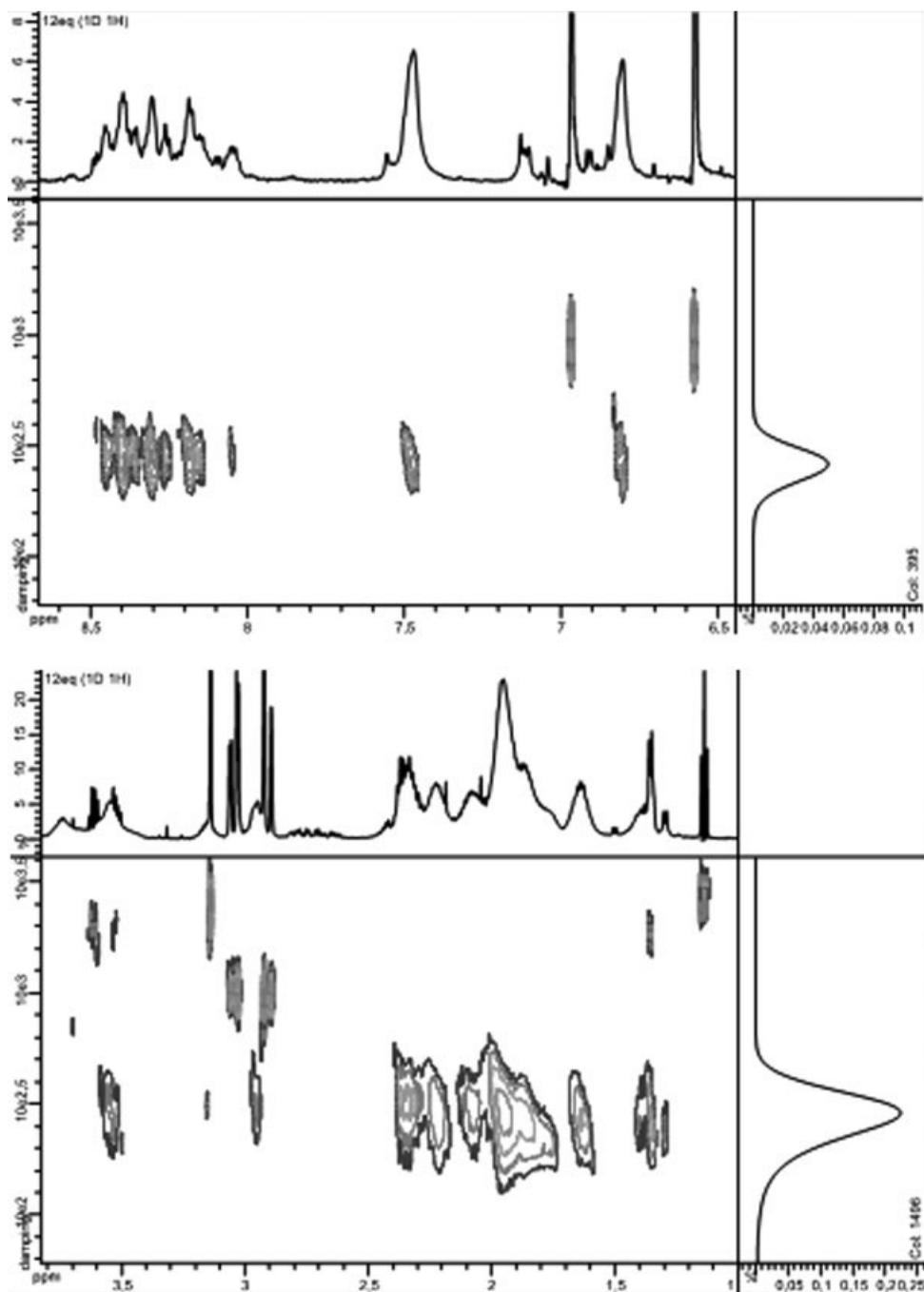


FIGURE 7  $^1\text{H}$  DOSY NMR spectrum of “pure” IB-5 recorded in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (90/10, v/v), HCl pH 3.5, NaCl 100 mM, ethanol 0.1 mM. Amide region (6.9 to 8.4 ppm) (top). Aliphatic region (1.2 to 4 ppm) (bottom).

**IB-5 Spectrum.** In the DOSY spectrum of IB-5, the protein diffusion coefficient appears centered around  $300 \mu\text{m}^2/\text{s}$ , but presents some polydispersity (Figure 7). From this value, the mean hydrodynamic radius of the protein can be estimated to be  $16.5 \text{ \AA}$ , in good agreement with previous DLS measurements.<sup>52</sup> Minor signals are observed at  $230\text{--}250 \mu\text{m}^2/\text{s}$  ( $r_{\text{H}} = 20\text{--}22 \text{ \AA}$ ) for protons at 1.44, 2.07, 2.42,

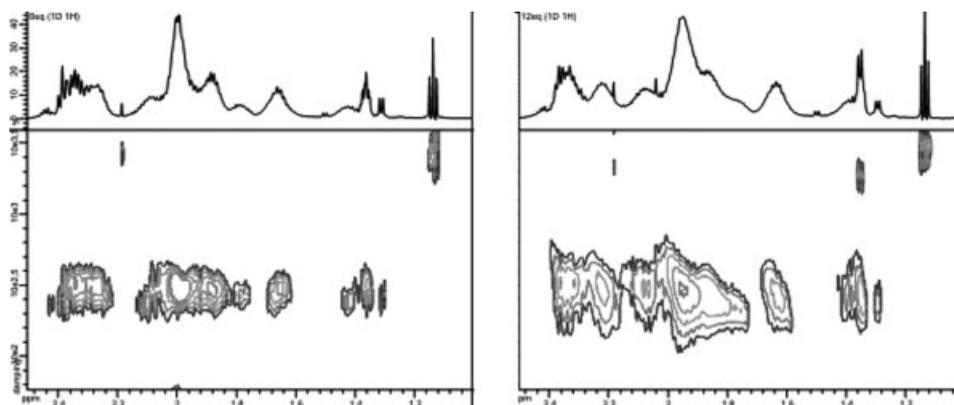
and 8.50 ppm. Other minor signals are observed at  $340\text{--}380 \mu\text{m}^2/\text{s}$  ( $r_{\text{H}} = 13\text{--}15 \text{ \AA}$ ) for protons at 2.30, 3.20, 8.26, and 8.40 ppm. This signal dispersion is attributed to the existence of numerous conformational forms for the protein in a slow equilibrium. On this spectrum, ethanol diffusion appears well defined, with a diffusion coefficient of  $2500 \mu\text{m}^2/\text{s}$ .



**FIGURE 8**  $^1\text{H}$  DOSY NMR spectrum of IB-5 after addition of EGCG recorded in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (90/10, v/v), HCl pH 3.5, NaCl 100 mM, ethanol 0.1 mM (EGCG/IB-5 molar ratio = 12). Amide region (6.9 to 8.4 ppm) (top). Aliphatic region (1.2 to 4 ppm) (bottom).

**IB-5/EGCG Spectrum.** A DOSY spectrum was measured after addition of 12 equivalents of EGCG to the IB-5 solution (Figure 8). In this spectrum, ethanol diffusion coefficient is at  $2610 \mu\text{m}^2/\text{s}$  a value only slightly higher than in the absence of EGCG, showing that the solution viscosity is perturbed slightly. EGCG is observed at  $975 \mu\text{m}^2/\text{s}$  ( $r_{\text{H}} = 5.35 \text{ \AA}$ ).

The protein diffusion coefficient appears centered around  $305 \mu\text{m}^2/\text{s}$ , with considerable polydispersity. From this value, the mean hydrodynamic radius of the protein can be estimated to be  $17 \text{ \AA}$ , a value marginally larger than that obtained in the absence of EGCG and within the margin of error of the measurement. Many signals are observed with



**FIGURE 9** Comparison of the 1.9 ppm signal on DOSY NMR spectra recorded in H<sub>2</sub>O/D<sub>2</sub>O (90/10, v/v), HCl pH 3.5, NaCl 100 mM, ethanol 0.1 mM (a) before and (b) after addition of EGCG (EGCG/IB-5 molar ratio = 12).

diffusion coefficients distinct from the mean value. A large signal centered at 1.77 ppm spreads from 260  $\mu\text{m}^2/\text{s}$  down to 190  $\mu\text{m}^2/\text{s}$  ( $r_{\text{H}} = 20\text{--}27 \text{ \AA}$ ), while signals in the same diffusion range appear at 8.35, 7.47, 6.80, 3.70, 3.50, 2.16, 1.35, and 1.30 ppm. On the other hand, signals with diffusion coefficient from 415 to 335  $\mu\text{m}^2/\text{s}$  ( $r_{\text{H}} = 1315.5 \text{ \AA}$ ) are observed for protons around 2.00–2.20, 3.55, 3.75, 8.20, and 8.45 ppm.

A comparison of the two experiments shows that the monomer is observed in both cases, but, while free IB-5 exhibits some polydispersity, this is increased considerably by interaction with EGCG.

Along the diffusion axis, the addition of EGCG induces a broadening of the diffusion signal, as observed at 1.9 ppm (Figure 9) and signals are extended towards larger aggregates. Moreover, diffusion coefficients seem to be correlated with chemical shifts (Figure 9), the largest forms being associated with largest upfield shifts. This indicates that several forms of the protein are present after binding, with different chemical environments for some protons and broadening due to the heterogeneity of the sample, as observed on the 1D spectrum. The largest modifications upon interaction correspond to the structured prolines referred to as “Pb” in the assignment ta-

**Table III** Molecular Hydrodynamic Radii that Were Determined by DOSY for Protein IB-5 by Comparison to Ethanol Before and After Adding 12 EGCG/Protein

	IB-5 ( $\text{\AA}$ )	IB-5 + 12 EGCG/Protein ( $\text{\AA}$ )
IB-5 main signal	16.5	17
Dispersion	13–22	13–27
EGCG		5.35

Values are determined as  $\pm 5\%$ .

ble. Table III summarizes the hydrodynamic radii that were determined by DOSY for protein IB-5.

The DOSY experiments revealed the existence of several types of aggregates after binding of EGCG to IB-5 with different conformations or degrees of aggregation. This result is in agreement with Charlton et al.’s work<sup>50</sup> that showed by molecular simulation based on ROESY data that aggregates of a seven amino acid peptide and EGCG could adopt several conformations, some being more stable than others. Exchange between the various aggregates can be excluded, as an exchange faster than the diffusion measurement time (here 150 ms) would lead to the observation of a mean diffusion coefficient rather than a polydisperse system as it is observed here.

The NMR experiments presented here illustrate the role of residual structures of the natively unfolded protein IB-5 in binding to EGCG. This observation is in agreement with the hypothesis of Fuxreiter et al.<sup>44</sup> who proposed that preformed structural elements on natively unfolded proteins could be anchorage points for ligands. It is also worth emphasizing that an increase in structural content is observed (CD data) in the presence of these preformed structural elements prior to binding (100 mM NaCl) but not in the absence of these residual structures (5 mM NaCl). The DOSY experiments revealed the existence of several types of aggregates after EGCG binding to IB-5, resulting in an increased polydispersity although an important amount of quickly diffusing species remained in solution. NMR was shown in this work to be a powerful tool for investigating these complex systems with large unfolded proteins. Further, experiments are under way to better identify the interacting amino acids and the effects of residual structures on the binding mechanism.

We would like to thank Nathalie Declerck and Yannick Bessin (CBS Montpellier) for their help for CD spectroscopy, Marie José Vallier, Francis Canon, Thérèse Marlin, and H  l  ne Boze (INRA Montpellier) for their technical support in protein production and purification; Bernard Cabane (ESPCI Paris), Aude Vernhet, and C  line Poncet-Legrand (INRA Montpellier) for their help in the colloidal study of the protein; and Andrew Atkinson for careful reading of the manuscript.

## REFERENCES

- Hagerman, A. E.; Robbins, C. T. *J Chem Ecol* 1987, 13, 1243–1259.
- Clausen, T. P.; Reichardt, P. B.; Bryant, J. P.; Provenza, F. In *Plant Polyphenols*; Hemingway, R. W.; Lacks, P. E., Eds.; Plenum Press: New York, 1992; pp 639–651.
- Bennick, A. *Crit Rev Oral Biol Med* 2002, 13, 184–196.
- Mehansho, H.; Butler, L. G.; Carlson, D. M. *Annu Rev Nutr* 1987, 7, 423–440.
- Bennick, A. *Mol Cell Biochem* 1982, 45, 83–99.
- Gawel, R. *Aust J Grape Wine Res* 1998, 4, 74–95.
- Haslam, E.; Williamson, M. P.; Baxter, N. J.; Charlton, A. J. In *Phytochemicals in Human Health Protection*; Romeo, J. T., Ed.; Kluwer Academic/Plenum Publishers: New York, 1999; pp 289–318.
- Beeley, J. A. *Oral Dis* 2001, 7, 69–70.
- Butler, L. G.; Mole, S. *JIEP*; St. Catharines: Canada, 1988; pp 111–114.
- Baxter, N. J.; Lilley, T. H.; Haslam, E.; Williamson, M. P. *Biochemistry* 1997, 36, 5566–5577.
- Charlton, A. J.; Baxter, N. J.; Haslam, E.; Williamson, M. In *Polyphenols in Food*; Amado, R.; Andersson, H.; Bardocz, S.; Serra, F., Eds.; Office for Official Publications of the European Communities: Luxembourg, 1998; pp 179–185.
- Edgar, W. M. *Br Dent J* 1992, 172, 305–312.
- Azen, E. A.; Amberger, E.; Fisher, S.; Prakobphol, A.; Niece, R. L. *Am J Hum Genet* 1996, 58, 143–153.
- Chan, M.; Bennick, A. *Eur J Biochem* 2001, 268, 3423–3431.
- Hatton, M. N.; Loomis, R. E.; Levine, M. J.; Tabak, L. A. *Biochem J* 1985, 230, 817–820.
- Douglas, C. W. I. *Adv Dent Res* 1994, 8, 254–262.
- Gal, J. Y.; Fovet, Y.; Adib-Yadzi, M. *Talanta* 2001, 53, 1103–1115.
- Flanzy, C. *Oenologie–Fondements Scientifiques et Technologiques*; Lavoisier: Paris, 1998.
- Poncet-Legrand, C.; Edelman, A.; Putaux, J.-L.; Cartalade, D.; Sarni-Manchado, P.; Vernhet, A. *Food Hydrocolloids* 2005, 20, 687–697.
- Bacon, J. R.; Rhodes, M. J. C. *J Agric Food Chem* 1998, 46, 5083–5088.
- Charlton, A. J.; Baxter, N. J.; Lilley, T. H.; Haslam, E.; McDonald, C. J.; Williamson, M. P. *FEBS Lett* 1996, 382, 289–292.
- Pascal, C.; Bigey, F.; Ratomahenina, R.; Boze, H.; Moulin, G.; Sarni-Manchado, P. *Protein Expr Purif* 2006, 47, 524–532.
- Nandi, T.; Dash, D.; Ghai, R.; Rao, C.; Kannan, K.; Brahmachari, S. K.; Ramakrishnan, C.; Ramachandran, S. *J Biomol Struct Dyn* 2003, 20, 657–668.
- Tompa, P. *Trends Biochem Sci* 2002, 27, 527–533.
- Uversky, V. N. *Protein Sci* 2002, 11, 739–756.
- Tompa, P. *J Mol Struct Theochem* 2003, 666, 361–371.
- Uversky, V. N. *Eur J Biochem* 2002, 269, 2–12.
- Tompa, P. *FEBS Lett* 2005, 579, 3346–3354.
- Dyson, H. J.; Wright, P. E. *Curr Opin Struct Biol* 2002, 12, 54–60.
- Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W. *J Mol Graph Model* 2001, 19, 26–59.
- Fink, A. L. *Curr Opin Struct Biol* 2005, 15, 35–41.
- Hwang, T. L.; Shaka, A. J. *J Magn Reson Ser A* 1995, 112, 275–279.
- Tramesel, D.; Catherinot, V.; Delsuc, M.-A. *J Magn Reson* 2007, 188, 56–67.
- Gibbs, S.; Johnson, C. *J Magn Reson* 1991, 93, 395–402.
- Balayssac, S.; Delsuc, M.; Gilard, V.; Prigent, Y.; Malet-Martino, M. *J Magn Reson* 2009, 196, 78–83.
- Delsuc, M. A.; Malliavin, T. E. *Anal Chem* 1998, 70, 2146–2148.
- Wilkins, D. K.; Grimshaw, S. B.; Receveur, V.; Dobson, C. M.; Jones, J. A.; Smith, L. J. *Biochemistry* 1999, 38, 16424–16431.
- Links-Pons, J. L.; Delsuc, M. A. *J Biomol NMR* 1999, 15, 15–26.
- Marin, A.; Malliavin, T. E.; Nicolas, P.; Delsuc, M. A. *J Biomol NMR* 2004, 30, 47–60.
- Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. *J Biomol NMR* 1995, 5, 67–81.
- Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Chung, J.; Wright, P. E.; Dyson, H. J. *J Am Chem Soc* 2001, 123, 2970–2978.
- Receveur-Brechot, V.; Bourhis, J. M.; Uversky, V. N.; Canard, B.; Longhi, S. *Proteins Struct Funct Bioinform* 2006, 62, 24–45.
- Kentsis, A.; Mezei, M.; Osman, R. *Proteins Struct Funct Bioinform* 2005, 61, 769–776.
- Fuxreiter, M.; Simon, I.; Friedrich, P.; Tompa, P. *J Mol Biol* 2004, 338, 1015–1026.
- Pascal, C.; Poncet-Legrand, C.; Sarni-Manchado, P.; Cheynier, V.; Vernhet, A. In *Macromolecules and Secondary Metabolites of Grapevine and Wines*; Jeandet, P.; Cl  ment, C.; Conreux, A., Eds.; Tec&Doc: Paris, 2007; pp 229–236.
- Benod, C.; Delsuc, M. A.; Pons, J. L. *J Chem Inform Model* 2006, 46, 1517–1522.
- Avbelj, F.; Kocjan, D.; Baldwin, R. L. *Proc Natl Acad Sci USA* 2004, 101, 17394–17397.
- Murray, N. J.; Williamson, M. P.; Lilley, T. H.; Haslam, E. *Eur J Biochem* 1994, 219, 923–935.
- Charlton, A.; Baxter, N. J.; Khan, M. L.; Moir, A. J. G.; Haslam, E.; Davis, A. P.; Williamson, M. P. *J Agric Food Chem* 2002, 50, 1593–1601.
- Charlton, A. J.; Haslam, E.; Williamson, M. P. *J Am Chem Soc* 2002, 124, 9899–9905.
- duVernet, R.; Boekelheide, V. *Proc Natl Acad Sci USA* 1974, 71, 2961–2964.
- Pascal, C.; Poncet-Legrand, C.; Imbert, A.; Gautier, C.; Sarni-Manchado, P.; Cheynier, V.; Vernhet, A. *J Agric Food Chem* 2007, 55, 4895–4901.

*Reviewing Editor: David Wemmer*