J|A|C|S
ARTICLES
Published on Web 12/28/2002

Combined Use of NMR Relaxation Measurements and Hydrodynamic Calculations To Study Protein Association. Evidence for Tetramers of Low Molecular Weight Protein Tyrosine Phosphatase in Solution

Pau Bernadó,[†] Tomas Åkerud,[§,||] José García de la Torre,[⊥] Mikael Akke,[§] and Miquel Pons*[,†,‡]

Contribution from the Departament de Química Orgànica, Universitat de Barcelona, Martí i Franquès, 1-11, 08028, Barcelona, Spain, Laboratori de RMN de Biomolècules, Parc Científic de Barcelona, Josep Samitier, 1-5, 08028,Barcelona, Spain, Department of Biophysical Chemistry, Lund University, Box 124, SE-22100 Lund, Sweden, Structural Chemistry, Biovitrum, Stockholm, Sweden, and Departamento de Química Física, Facultad de Química, Universidad de Murcia, 30071, Múrcia, Spain

Received July 24, 2002; E-mail: mpons@qo.ub.es

Abstract: We describe a novel method for determining weak association constants of oligomeric protein complexes formed transiently under equilibrium conditions. This type of equilibrium process is recognized as being biologically important, but generally hard to study. Heteronuclear spin relaxation rates measured at multiple protein concentrations are analyzed using relaxation rates predicted from hydrodynamic calculations, yielding equilibrium constants and structural characterization of the protein complexes. The method was used to study the oligomerization equilibrium of bovine low molecular weight protein tyrosine phosphatase. X-ray structures of monomeric and dimeric forms of the protein have been reported previously. Using longitudinal and transverse ^{15}N relaxation rates measured at four different protein concentrations, we detected the monomer, dimer, and a previously unknown tetramer and measured the dissociation constants of the equilibria involving these species. A comparison of experimental and predicted relaxation rates for individual backbone amide ^{15}N spins enabled delineation of the tetramerization interface. The results suggest a novel concept for substrate modulation of enzymatic activity based on a "supramolecular proenzyme". The fast and reversible switching of the "supramolecular proenzyme" would have obvious advantages for the regulation of enzymes involved in cell signaling pathways.

Introduction

Self-association leading to discrete oligomers is essential for the activity and regulation of many biological processes. For example, HIV protease is a dimer;[1] p53 acts as a tetramer,[2,3] which can be described as a dimer of dimers. Signal transduction and activator of transcription (STAT) proteins form dimers when activated.[4] Different tyrosine kinases are activated by self-association, and some tyrosine phosphatases are inhibited by the formation of dimers.[5] Protein complexes involved in regulating activity may in some cases have only moderate association constants, so as to ensure efficient dynamic control.

Experimental characterization of self-association equilibria is a difficult task. Most of the experimental observables are concentration-weighted averages of all the species present at equilibrium. Thus, reliable data can only be obtained under experimental conditions, where the concentration of the aggregates is large enough.

Hydrodynamic properties are potentially very sensitive to self-association. However, experimental observables, such as sedimentation or translational diffusion coefficients, provide only a single numerical value that is a complex function of molecular size and shape. Therefore, the number of parameters that can be extracted by fitting a model to the experimental data is limited, and no details about the local structure of the aggregate are available.

NMR has the potential to provide atomic scale information about the oligomers. Under fast exchange conditions, NMR parameters are an average of the contributions from the different species present, weighted by their relative populations. Concentration-dependent chemical shift perturbations can identify residues located at interfaces. However, chemical shift changes usually affect only a small number of nuclei, and the chemical shift difference between the different oligomeric forms cannot be easily predicted in general. Therefore, as in the case of hydrodynamic observables, only simple models can be used to

[†] Universitat de Barcelona.
[‡] Parc Científic de Barcelona.
[§] Lund University.
[||] Biovitrum.
[⊥] Universidad de Murcia.

(1) McPhee, F.; Good, A. C.; Kuntz, I. D.; Craik, C. S. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11477−11481.
(2) McLure, K. G.; Lee, P. W. K. *EMBO J.* **1998**, *17*, 3342−3350.
(3) Chene, P. *Oncogene* **2001**, *20*, 2611−2617.
(4) Leonard, W. J.; O'Shea, J. J. *Annu. Rev. Immunol.* **1998**, *16*, 293−322.
(5) Weiss, A.; Schlessinger, J. *Cell* **1998**, *94*, 277−280.

10.1021/ja027836h CCC: $25.00

extract quantitative information from the data. The local environment of individual nuclei, for example, the proximity of aromatic groups, can also bias a qualitative interpretation of the residues that are most perturbed by self-association.

If the lifetimes of the aggregates are significantly longer than their rotational diffusion correlation times, relaxation rates are also averaged.[6,7] ^{15}N spin relaxation is mainly caused by the modulation of dipolar and chemical shift anisotropy interactions due to molecular tumbling, which in turn is strongly dependent on the molecular volume and shape. As a consequence, longitudinal (R_1) and transverse (R_2) relaxation rates are extremely sensitive to intermolecular association processes. Even small amounts of high molecular weight species will cause a significant relative change in the average relaxation rates. In addition, in contrast to chemical shift perturbations that affect only a subset of nuclei in the molecular interface, changes in relaxation rates will affect all nuclei throughout the protein.

Quantitative use of relaxation rate measurements to characterize intermolecular association has been hindered by the need to know the relaxation rates of the individual species involved.[8−10] However, recent improvements in the application of state-of-the-art hydrodynamic modeling have opened the way to reliable and parameter-free predictions of the relaxation rates of individual nuclei from a known three-dimensional structure.[11,12]

HydroNMR[13] is a public domain program for hydrodynamic calculations based on a shell representation of the molecule that is constructed from the atomic coordinates.[14] Initially, each nonhydrogen atom is replaced by a spherical element of radius a, which we call the atomic element radius, to yield a primary hydrodynamic model. The shell model is then obtained from the primary hydrodynamic model by representing its surface by a set of tangent beads of radius σ. Hydrodynamic properties are calculated for decreasing values of σ and extrapolated to $\sigma = 0$. This procedure avoids the bead overlap problems present in the primary hydrodynamic model that can cause severe problems in the description of the hydrodynamic interactions and in the volume correction. The choice of the proper atomic element radius has been discussed recently, and a consensus value of 3.3 Å has been obtained.[15] This value correctly reproduces different hydrodynamic properties in globular proteins and nucleic acids[16] and in particular provides accurate predictions of the relaxation rates of individual nuclei in proteins of arbitrary shape.[15] Because the value $a = 3.3$ Å is close to the sum of the van der Waals radius of the atoms of the biomolecule plus the radius of a surface probe corresponding to water, the resulting shell matches closely the solvent-accessible surface of the molecule.[17]

Transduction of many extracellular signals involves regulation of the level of tyrosine phosphorylation of target proteins. This is the result of a balance between the competing activities of protein tyrosine kinases (PTKs) and protein tyrosine phosphatases (PTPs) that phosphorylate and dephosphorylate selected tyrosine residues.[18,19] PTPs are divided into membrane bound receptor protein tyrosine phosphatases (RPTPs) and cytoplasmic PTPs.[20,21] Cytoplasmic PTPs include the low molecular weight PTPs, found to be ubiquitous in mammalian cells.[22−24] All PTPs share a signature sequence CX_5−R(S/T) that forms a phosphate binding loop and includes residues essential for the enzyme's activity.[20,25] The S19A mutant of bovine PTP (BPTP), a low molecular weight PTP, was found to crystallize as a dimer in which Tyr131 and Tyr132 from each monomer are inserted into the active site of the other.[26] Tabernero et al. suggested that the dimer found in the crystal could represent a transient state between the dephosphorylation of the tyrosines and the release of the unphosphorylated monomer.[26] The same authors also suggested that the dimer may represent a product inhibition equilibrium state.

Equilibrium ultracentrifugation of both the S19A mutant and wild-type BPTPs at pH 7.3 in Tris buffer has been interpreted in terms of a monomer−dimer equilibrium with an apparent dissociation constant of 1−5 mM.[26] Recent ^{15}N spin relaxation studies of wild-type BPTP have confirmed the presence of a dimer in solution and further suggested that higher-order oligomers may be significantly populated at higher concentrations.[27] In this article, we present a new method, based on a combination of hydrodynamic calculations and standard R_1 and R_2 spin relaxation measurements at different concentrations, that provides, for the first time, evidence for the formation of a tetramer of BPTP in solution. The tetramer is more stable than the previously described dimer, suggesting a regulation mechanism by which the presence of a high-affinity substrate could sharply increase the phosphatase activity. A comparison of experimental and calculated R_2/R_1 values at the higher concentration highlights a cluster of residues at the surface of the dimer, which we tentatively identify as the tetramerization interface. This suggests that the tetramer is formed by the interaction of two perpendicular dimers, leading to a compact tetrahedral arrangement of monomers.

Results

Longitudinal and transverse relaxation rates of BPTP were measured at four different concentrations: 0.17 ± 0.02, 0.34 ± 0.03, 0.66 ± 0.07, and 1.24 ± 0.07 mM. To compare experimental and calculated data, we used the R_2/R_1 values because these are less sensitive to local order parameters and local variations in chemical shift anisotropy than the individual relaxation rates.[28]

(6) Marshall, A. G. *J. Chem. Phys*. **1970**, *52*, 2527−2534.
(7) Wennerström, H. *Mol. Phys*. **1972**, *24*, 69−80.
(8) Fushman, D.; Cahill, S.; Cowburn, D. *J. Mol. Biol*. **1997**, *266*, 173−194.
(9) Pfuhl, M.; Chen, H. A.; Kristensen, S. M.; Driscoll, P. C. *J. Biomol. NMR* **1999**, *14*, 307−320.
(10) Mercier, P.; Spyracopoulos, L.; Sykes, B. D. *Biochemistry* **2001**, *40*, 10063−10077.
(11) Carrasco, B.; García de la Torre, J. *Biophys. J*. **1999**, *76*, 3044−3057.
(12) García de la Torre, J.; Huertas, M. L.; Carrasco, B. *Biophys. J*. **2000**, *78*, 719−730.
(13) García de la Torre, J.; Huertas, M. L.; Carrasco, B. *J. Magn. Reson*. **2000**, *147*, 138−146.
(14) Filson, D. P.; Bloomfield, V. A. *Biochemistry* **1967**, *6*, 1650−1658.
(15) Bernadó, P.; García de la Torre, J.; Pons, M. *J. Biomol. NMR* **2002**, *23*, 139−150.
(16) Fernandes, M. X.; Ortega, A.; López-Martínez, M. C.; García de la Torre, J. *Nucleic Acid Res.* **2002**, *30*, 1782−1788.
(17) García de la Torre, J. *Biophys. Chem*. **2001**, *93*, 159−170.
(18) Charbonneau, H.; Tonks, N. K. *Annu. Rev. Cell Biol*. **1992**, *8*, 463−493.
(19) Hunter, T. *Cell* **1995**, *80*, 225−236.
(20) Neel, B.G.; Tonks, N.K. *Curr. Opin. Cell Biol*. **1997**, *9*, 193−204.
(21) Barford, D.; Das, A. K.; Egloff, M.-P. *Annu. Rev. Biophys. Biomol. Struct*. **1998**, *27*, 133−164.
(22) Rehkop, D. M.; Van Etten, R. L. *Hoppe−Seylers's Z. Physiol. Chem.* **1975**, *356*, 1775−1782.
(23) Chernoff, J.; Li, H. C. *Arch. Biochem. Biophys*. **1985**, *240*, 135−145.
(24) Waheed, A.; Laidler, P. M.; Wo, Y. Y. P.; Van Etten, R. L. *Biochemistry* **1988**, *27*, 4265−4273.
(25) Li, L.; Dixon, J. E. *Semin. Immunol*. **2000**, *12*, 75−84.
(26) Tabernero, L.; Evans, B. N.; Tishmack, P. A.; Van Etten, R. L.; Stauffacher, C. V. *Biochemistry* **1999**, *38*, 11651−11658.
(27) Åkerud, T.; Thulin, E.; Van Etten, R. L.; Akke, M. *J. Mol. Biol*. **2002**, *322*, 137−152.

Experimental R_2/R_1 ratios could not be fitted with $a = 3.3$ Å for any of the concentrations using neither the crystal structure of the monomer nor that of the dimer. For example, the average R_2/R_1 ratio at 0.17 mM is 7.78. Calculated values for the monomer and dimer X-ray structures are 6.63 and 33.70, respectively. Most of the individual relaxation rates could be fitted using the monomer structure and increasingly large values of the atomic element radius, a, as the concentration of the sample increased. Data obtained at 0.17 mM, the lowest concentration studied, already required a value of $a = 3.85$ Å. Previous observations from our group suggested that values of the atomic element radius significantly higher than 3.3 Å are diagnostic for aggregation processes.[15] This observation is in agreement with results from equilibrium ultracentrifugation studies[26] and chemical shift data[27] that have been previously interpreted as evidence for an equilibrium between the BPTP monomer and its dimer in solution.

To derive quantitative information about the self-association behavior of BPTP, we have fitted several equilibrium models to the experimental R_2/R_1 ratios of individual NH nuclei measured at different concentrations. The R_2/R_1 ratios of nuclei in the monomer and the dimer forms of BPTP were calculated from the X-ray structures using HydroNMR with $a = 3.3$ Å. Therefore, the number of adjustable parameters in any of the models could be kept much smaller than the number of experimental points.

HydroNMR assumes that the protein can be represented as a rigid body and that relaxation arises only from the modulation of dipolar couplings and chemical shift anisotropy by global tumbling. Therefore, residues affected by contributions from chemical exchange, related or not to the oligomerization process, and fast local motions have to be identified and removed from the data set (see Experimental Section). The filtered data set thus obtained included a total of 440 R_2/R_1 values from 112 residues.

Monomer−Dimer Model. Relaxation data were fitted initially to a model involving monomer−dimer equilibrium characterized by a single parameter, the dissociation constant of the dimer (K_D), using the calculated relaxation rates of both species derived from crystallographic data.

Assuming fast exchange on the chemical shift time scale and a lifetime for the complex that is long on the time scale of overall rotational diffusion (but short on the time scale of relaxation), the average relaxation rates for individual nitrogen nuclei are given by[8]

$$R_1^{av} = p_M R_1^M + (1 - p_M) R_1^D$$

$$R_2^{av} = p_M R_2^M + (1 - p_M) R_2^D + R_2^{ex}$$

where p_M is the population of protein in the monomeric state, R_r^M and R_r^D ($r = 1,2$) are the relaxation rates in the monomer and dimer, and R_2^{ex} is the contribution to relaxation from the exchange process. R_2^{ex} affects only those residues that have different chemical shifts in the two species and depends also on the populations of the two species and the exchange rate.

Figure 1A−E presents the optimized fitting of the monomer−dimer model to relaxation data. The estimated K_D value is 1.50 ± 0.06 mM, which is in agreement with published equilibrium ultracentrifugation data (1−5 mM)[26] and also with the results of the fitting of the concentration dependence of the chemical shift observed for residues in the dimer interaction surface (1.5 ± 0.1 mM).[27] However, Figure 1B−E clearly reveals a systematic deviation between calculated and experimental data at different concentrations, pointing to the existence of additional association processes leading to higher order oligomers. This deviation was not possible to discern from the chemical shift data alone.

Monomer−Dimer−Tetramer Model. To account for the systematic deviations observed using the monomer−dimer model, we decided to fit the data to a monomer−dimer−tetramer equilibrium model

$$D \rightleftarrows 2\,M \qquad K_D = [M]^2/[D]$$

$$T \rightleftarrows 2\,D \qquad K_T = [D]^2/[T]$$

In contrast to the monomer−dimer model, there was no structural model available for the hypothetical tetramer. Instead, we introduced R_1^T and R_2^T, the longitudinal and transverse relaxation rates of nuclei in the tetramer, as adjustable parameters. We assumed an isotropic model for the higher order aggregate, that is, identical relaxation rates for each nucleus. This is justified, as the average deviation observed between experimental data and the predictions from the monomer−dimer model are much higher than the relative deviations of individual nuclei and, therefore, corrections introduced by considering an anisotropic model would most likely fall below the experimental noise level.

Figure 1F−J shows that relaxation data at the four concentrations are fitted correctly using the monomer−dimer−tetramer model with the following parameters: $K_D = 3.40 \pm 0.2$ mM, $K_T = 0.23 \pm 0.04$ mM, $R_1^T = 0.76 \pm 0.13$ s^{-1}, and $R_2^T = 34.4 \pm 2.1$ s^{-1}. The reduced χ^2 dropped from 5.17 for the monomer−dimer model to 0.99 for the model including the tetramer; the probability that this improvement in the fit is achieved by chance is 10^{-15}, indicating that the latter model correctly explains the experimental data down to the noise level.

Floating Stoichiometry Model. To test alternative stoichiometries for the higher order aggregate, data were fitted to a floating stoichiometry model. In this model, monomer and dimer are in equilibrium with a single species with relaxation rates R_1^X and R_2^X and unspecified stoichiometry, for which the populations at the four different concentrations studied (p_i^X, $i = 1-4$) are taken as adjustable parameters. Assuming an equilibrium involving only three species, monomer, dimer, and an n-mer, we can calculate the stoichiometry, n, from the concentration dependence of the population of the n-mer (p_i^X) obtained by fitting the relaxation data.

Repeated minimization from randomly chosen starting points was used to generate 1000 sets of parameters. All physically reasonable minima show a relatively narrow distribution: $K_D = 4.42 \pm 0.89$ mM, $p_1^X = 0.007 \pm 0.006$, $p_2^X = 0.034 \pm 0.013$, $p_3^X = 0.088 \pm 0.030$, $p_4^X = 0.184 \pm 0.062$, $R_1^X = 0.74 \pm 0.47$ s^{-1}, and $R_2^X = 51.6 \pm 17.3$ s^{-1}. The range of mathematically acceptable ratios of the relaxation parameters is very large ($R_2^X/R_1^X = 401 \pm 3079$). However, the distributions of the population-related parameters and the relaxation rates are

(28) Tjandra, N.; Feller, S. E.; Pastor, R. W.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562−12566.

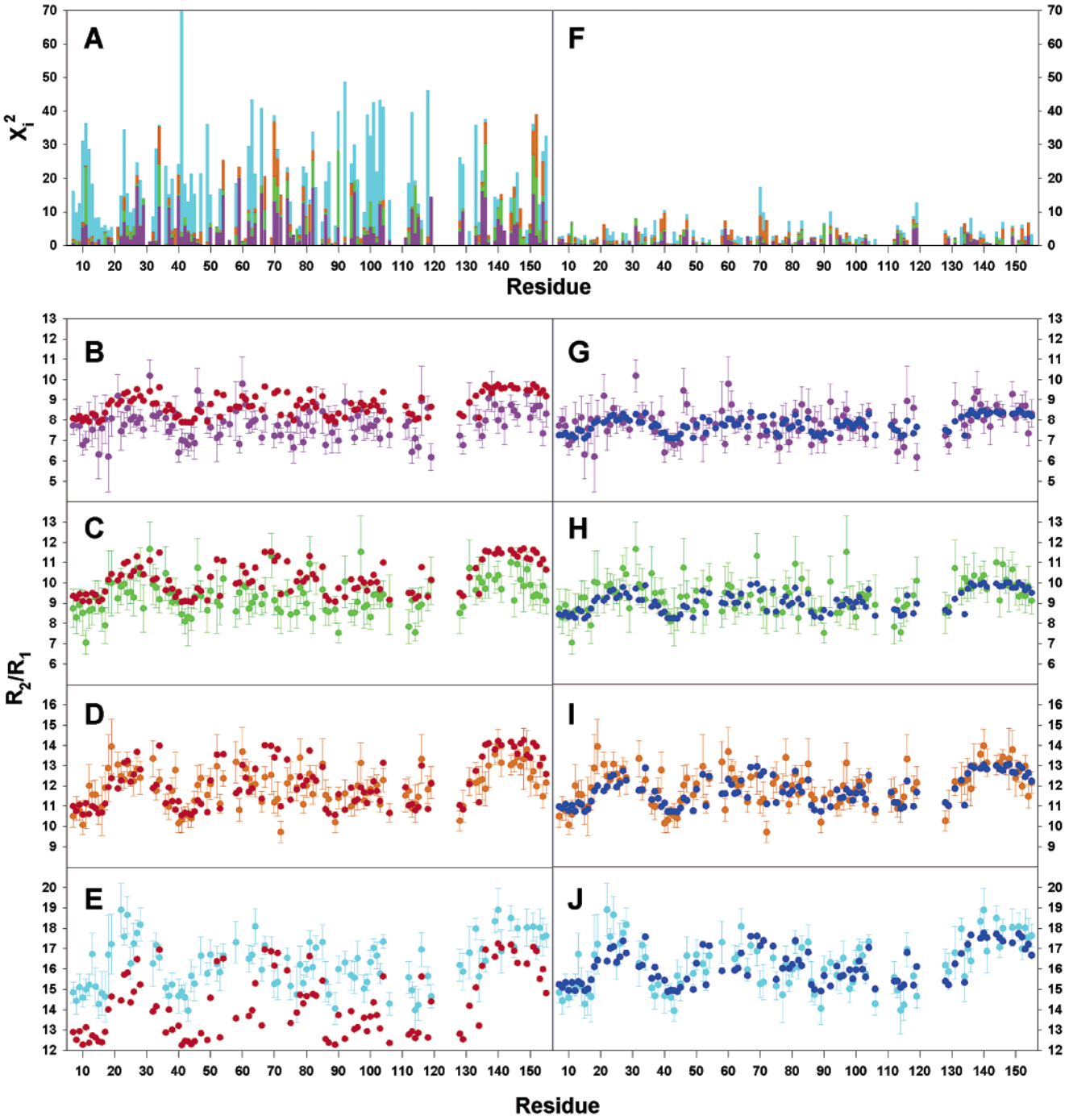

Figure 1. Experimental and calculated values of R_2/R_1 for low molecular weight protein tyrosine phosphatase. Experimental R_2/R_1 values at different concentrations are shown with their error bars (panel, concentration (mM), color code): B and G, 0.17 ± 0.02, magenta; C and H, 0.34 ± 0.03, green; D and I, 0.66 ± 0.07, orange; E and J, 1.24 ± 0.07, blue. Calculated R_2/R_1values are for the monomer−dimer model (B−E, red) and the monomer−dimer−tetramer model (G−J, dark blue). Residuals for each NH are shown in panels A and F, respectively. Contributions to the residuals from experimental data at different concentrations follow the same color code as in panels B−E.

strongly correlated, and the distribution of R_2^X/R_1^X values for certain stoichiometries is much better defined.

Figure 2 shows the correlation between n and R_2^X/R_1^X and between n and K_D in 1000 sets of parameters that minimize the residuals of calculated versus observed R_2/R_1 values. Points with χ^2 lower than that obtained with the monomer−dimer−tetramer model are highlighted in green. This is a weak restriction, considering the larger number of degrees of freedom, but it provides a clear delimitation of acceptable regions.

Values of n between 2.5 and 5 provide statistically acceptable fits. However, the best-fit values of n and R_2^X/R_1^X are found to be strongly anticorrelated, so that small values of n require very large values of R_2^X/R_1^X. Physically acceptable values of R_2^X/R_1^X for each stoichiometry can be extrapolated from the properties of the monomer and the dimer using hydrodynamic theory, as explained in the Experimental Section.[29,30] The R_2^X/R_1^X value calculated for a compact tetramer is 47.5 (corresponding to a τ_c of 25.2 ns). This is in good agreement with

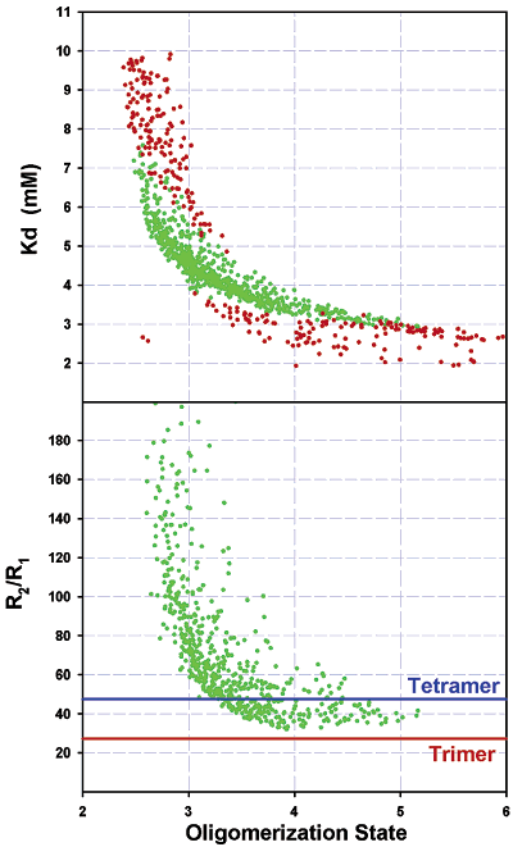

Figure 2. Correlation between parameters (K_D, the dimer dissociation constant, top panel; R_2^X/R_1^X, the relaxation rates of the *n*-mer, bottom panel; and the stoichiometry *n* of the *n*-mer) in the floating stoichiometry model obtained by fitting the experimental R_2/R_1 values, starting from 1000 random initial values. The oligomerization state of the *n*-mer, *n*, was derived analytically from the parameters p_3^X and p_4^X that describe the populations of this species at the two higher concentrations. Points in green have values of χ^2 lower than 430. The values of R_2^X/R_1^X expected for a trimer and a tetramer, as explained in the Results section, are indicated as horizontal lines in the bottom panel.

the R_2^X/R_1^X ratio obtained by averaging all points with *n* of 4 ± 0.3 ($R_2^X/R_1^X = 46 \pm 12$) and also with the value obtained by fitting the data to the monomer−dimer−tetramer model ($R_2^T/R_1^T = 45 \pm 10$). In contrast, the predicted correlation time of the trimer is 18.9 ns, which would give $R_2^X/R_1^X = 27.2$, but the average of the minima with $n = 3 \pm 0.3$ is 230 ± 647 and no points have R_2^X/R_1^X values lower than 30. Thus, only the tetramer structure provides mutually consistent values of *n* and R_2^X/R_1^X.

Figure 3 shows the populations of monomer, dimer, and tetramer as a function of the total concentration of BPTP, as calculated using the dissociation constants derived from the monomer−dimer−tetramer model. Individual points show the average populations, with error bars, of the *n*-mer in the model with floating stoichiometry with the values of R_1^X and R_2^X fixed to those derived from the tetramer model. The nearly perfect agreement confirms that the tetramer stoichiometry is implicit in the concentration dependence of the relaxation data and can be extracted using the proper estimation of the correlation time for the aggregate.

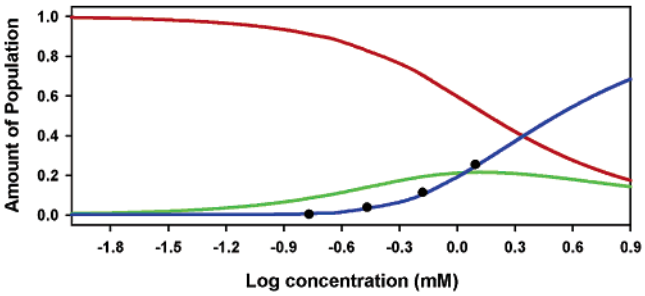

Figure 3. Populations of monomer (red), dimer (green), and tetramer (blue) at different total concentrations, according to the monomer−dimer−tetramer model. Discrete points are the average populations of *n*-mer derived from points fitted using the floating stoichiometry model with the parameters R_2^X and R_1^X fixed to the values of the tetramer. The spread of the distribution is smaller than the width of the symbol.

Insights into the Structure of the Tetramer. The global analysis of relaxation data presented so far provides evidence for the existence of a tetramer form and an estimation of its dissociation constant. However, it does not include any assumption about its structure beyond isotropy. This assumption implies that the tetramer adds a constant R_2/R_1 contribution to all residues. However, five residues show substantial deviations between experimental and calculated R_2/R_1 values only at the highest concentration: residues Ile21, Trp39, Ala59, Ser61, and His72. Val60 is not observed in the 1.24 mM sample due to overlap with Arg150. Residues Thr70 and Ala71 show smaller deviations (Figure 4A). Residues 21, 60, and 61 show values of R_2/R_1 larger than those predicted, while residues 59, 70, 71, and 72 show smaller values. Larger R_2/R_1 values may be the result of exchange broadening at the interface and decreased amplitudes of internal fluctuations. Lower R_2/R_1 values may reflect increased amplitudes of internal fluctuations. The latter effect is perhaps unexpected but not unprecedented; this has been observed previously at interfaces, where the entropic cost of association is compensated by increased internal mobility.[31]

Figure 4B shows the location of these residues on the crystal structure of the dimer. With the exception of Trp39, the outliers in the higher concentration experiment correspond to residues that are clustered together close to the surface in the dimer structure. Nonspecific aggregation is not expected to single out a specific region of the protein. There is no indication that these residues could be subjected to systematic effects arising from peculiar chemical shifts, N−H bond vector orientations, or errors in relaxation rate measurement affecting only these residues. The observation of a cluster of outliers at the highest concentration, where the population of the tetramer would be highest, lends structural credibility to the tetramer model. The plane containing the centers of the symmetry related clusters and the C_2 symmetry axis of the dimer forms an ~45° angle with that formed by the symmetry axis and the major axis of the dimer. If the two dimers approach with their major axes perpendicular to each other, as suggested by its dumbbell shape, the clustered residues of the two dimers would interact. Thus, these residues probably correspond to the tetramerization interface, such that the tetramer structure is a compact pseudo-tetrahedron.

A model constructed by applying tetrahedral symmetry to the dimer structure shows that the clusters of outliers from two

(29) García Bernal, J. M.; García de la Torre, J. *Biopolymers* **1981**, *20*, 129−139.
(30) Carrasco, B.; García de la Torre, J. *J. Chem. Phys.* **1999**, *111*, 4817−4826.
(31) Zídek, L.; Novotny, M. V.; Stone, M. J. *Nat. Struct. Biol.* **1999**, *6*, 1118−1121.

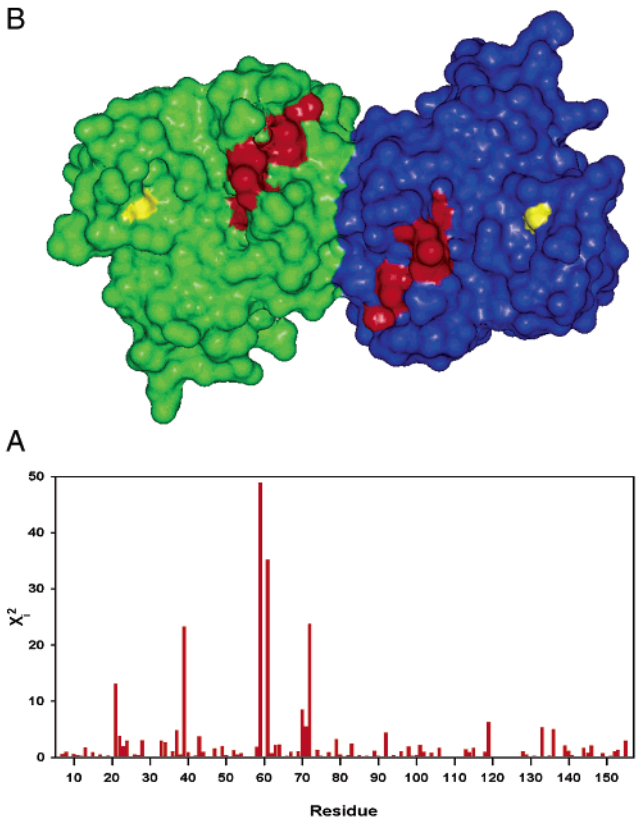

Figure 4. (A) R_2/R_1values of residues 21, 39, 59, 61, 70, 71, and 72 show considerable deviations from the calculated values only at 1.24 mM. (B) Surface representation of the crystal structure of dimeric BPTP with residues 21, 59, 60, 61, 70, 71 and 72 highlighted in red and residue 39 highlighted in yellow. The clustering of residues suggests a tetramerization interface. Note the dumbbell structure of the dimer and the 45° inclination of the line joining the clusters with respect to the dumbbell axis. The clusters from two dimers packed with their main axes perpendicular would be facing each other.

dimers are indeed in contact and the tetramer is highly isotropic. Hydrodynamic calculations predict a correlation time that is slightly overestimated, probably reflecting the fact that the model has suboptimal side chain packing and does not consider plausible distortions, such as minor bending of the dimer axis that would lead to reduced correlation times.

Discussion

NMR Methods to Study Weak Self-Association. The concentration dependence of different parameters that are readily available through NMR experiments can be used to study self-association equilibria. These include chemical shifts, translational diffusion coefficients (obtained from pulsed field gradient echoes), and heteronuclear relaxation rates.

A comparison of the relative merits of the different methods has to take into account the information content of each of the experiments and the precision of the measurements relative to the variation of the parameter induced by a given change in the populations of the different species.

Translational diffusion coefficients are sensitive to molecular size and shape. Modified NMR pulsed field gradient echo experiments that allow measurements of diffusion coefficients of large molecules are available and have been suggested as a suitable method to detect aggregation phenomena under experimental NMR conditions.[32] Translational diffusion coefficients can be calculated from three-dimensional structures using HydroNMR. However, translational diffusion is a global property that does not provide any information on the aggregates at the atomic level and is relatively insensitive to small changes in the concentration of the different species.

Self-association affects usually the chemical shift of only a small number of nuclei. This information can identify residues located at the interface or otherwise affected by the formation of the aggregate(s). In simple systems, the stoichiometry of the different species can be deduced by fitting the concentration dependence of chemical shifts to different models. However, in the case of BPTP, the small number of experimental points and their precision did not provide statistical justification for the use of a model more complex than the simplest monomer−dimer equilibrium, such as the monomer−dimer−tetramer equilibrium.[27]

Relaxation rates are sensitive to the global rotational diffusion properties of the molecule but also to the structural details of the species involved because the relaxation rates of individual ^{15}N nuclei depend on the orientation of the N−H bond vector with respect to the axes of the rotational diffusion tensor. Thus, different ^{15}N nuclei provide nonredundant information that, in the BPTP case, allowed statistically significant differentiation between different models and provided structural details of a previously unknown tetramer.

Figure 5 shows the calculated concentration dependence of different parameters available from NMR experiments using the dissociation constants for the BPTP dimer and tetramer derived in this work. As a comparison, the calculated dependence assuming the best-fit monomer−dimer equilibrium is also included.

Figure 5A shows calculated R_2/R_1 values for residue I41 and C148. The N−H bond of I41 is oriented perpendicularly to the main axis of the rotational diffusion tensor of the dimer, which is the most anisotropic species present. In contrast, the C148 NH bond is collinear with the main axis. When the total concentration of BPTP is increased from 0.1 to 1 mM, R_2/R_1 increases from 6.72 to 13.26 for I41 and from 7.86 to 16.01 for C148, using the full monomer−dimer−tetramer model. Using the simpler monomer−dimer model, with the best fit dissociation constant for the dimer of 1.50 ± 0.06 mM, the predicted R_2/R_1 values for I41 and C148 increase from 7.29 to 11.66 and from 8.70 to 16.23, respectively. Thus, self-association phenomena induce changes that are more than 1 order of magnitude above the experimental uncertainty of R_2/R_1, which is ∼0.74.[33]The differences in the calculated R_2/R_1 values for the two models are also significantly larger than the experimental error. Anisotropy effects increase the discrimination power of relaxation measurements, because information from N−H bond vectors with different orientations is nonredundant. For example, while predicted R_2/R_1 values for I41 in the two models are lower than the experimental error at 0.66 mM, those from C148 are significantly different. Conversely, the two models provide significantly different values for I41 at 1.24 mM but not for C148.

Chemical shift plots (Figure 5B) are normalized, assuming a maximum chemical shift difference between species of 1 ppm.

(32) Altieri, A. S.; Hinton, D. P.; Byrd, R. A. *J. Am. Chem. Soc.* **1995**, *117*, 7566−7567.
(33) Experimental average errors of R_2/R_1 ratios used for fittings were 0.67, 0.85, 0.70, and 0.73, representing 8.54%, 9.10%, 5.85%, and 4.50% of experimental R_2/R_1 values.

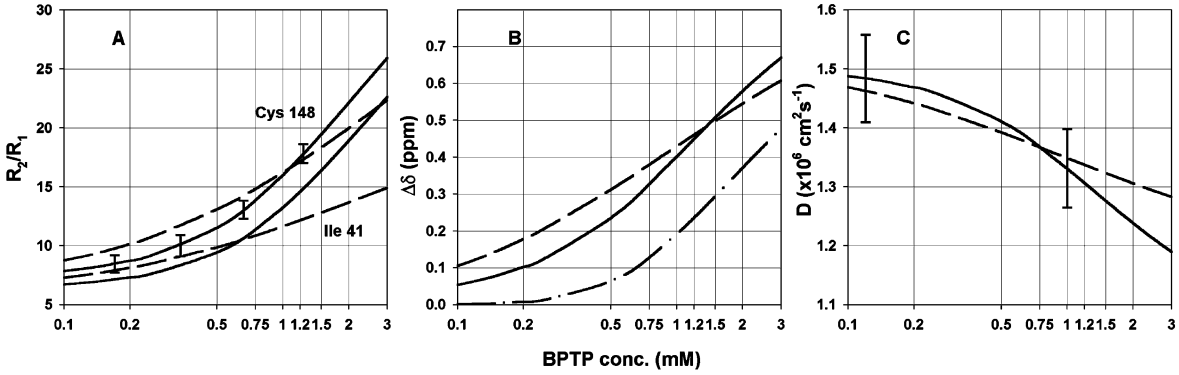

Figure 5. Calculated concentration dependence of different NMR observables for the BPTP system, assuming the monomer−dimer−tetramer model (continuous lines) and the simpler monomer−dimer model (dashed lines). (A) R_2/R_1 for I41 and C148. Uncertainties are indicated at four different concentrations for C148. These are average experimental errors considering all used residues and are also applicable to I41. (B) Average chemical shift, assuming a chemical shift difference of 1 ppm between monomer and dimer sites but no chemical shift difference between dimer and tetramer. The dot−dash curve has been calculated, assuming a monomer−dimer−tetramer model with a 1 ppm difference between dimer and tetramer sites but no difference between monomer and dimer. (C) Translational diffusion coefficient. Error bars represent a relative uncertainty of 5%.

Separate plots represent the situations of residues in the dimerization and the tetramerization interfaces, respectively. In the former case, the chemical shift in the dimer and tetramer are assumed to be the same. In the second case, the monomer and dimer are considered indistinguishable. In both cases, the presence of the "hidden" species manifests itself only through deviations from the pure sigmoidal curve expected from a two-site equilibrium. If the chemical shift difference between nuclei in the tetramer and the other species were smaller than 1 ppm, the expected changes would be scaled accordingly. If the difference between the two models falls below the experimental uncertainty, usually set by the digital resolution of 2D experiments, the detection of the tetramer would be impossible, as it is the case for BPTP.

Figure 5C shows the translational diffusion coefficients calculated using the two self-association models. If an experimental uncertainty of ±5% is assumed, the two models cannot be differentiated, even by changing the concentration from 0.1 to 3 mM.

Possible Biological Significance of BPTP Tetramerization. While there has been considerable progress in understanding the regulation of PTKs, the regulation mechanisms of PTPs are less well-known. It is generally agreed that the intrinsic activity of PTPs is higher than that of PTKs and that PTPs show lower substrate specificity. Therefore, tight regulation of their activity is essential. Regulation is thought to occur by changing the location of these enzymes in the cell and by intra- and intermolecular noncovalent interactions.[34,35]

As explained previously, a BPTP mutant was found to crystallize as a dimer in which Tyr131 and Tyr132 from each monomer are inserted into the active site of the other.[26] Interestingly, the best substrates for two related human low molecular weight PTPs (HCPTPA and HCPTPB) contain phosphorylated tandem tyrosine residues.[36] Pairs of adjacent tyrosines are also found in insulin receptor tyrosine kinase and fibroblast growth factor tyrosine kinase, both of which are reported to interact with low molecular weight PTPs.[37] Tyrosine tandems are also found in numerous kinases that are putative targets for low molecular weight PTPs.[26] In light of this substrate selectivity, the reports of low molecular weight PTPs being activated by phosphorylation at Tyr131 and Tyr132 suggest that phosphorylated BPTP may itself be a substrate,[38−40] thus providing a mechanism for self-regulation of BPTP activity.

The possible biological significance of tetramers in the activity or regulation of BPTP remains speculative. Both dimer and tetramer are expected to be inactive forms, with Tyr131 and Tyr132 blocking the active site. However, a high-affinity substrate will be able to compete with a BPTP molecule for complex formation, thereby effectively dissociating the oligomers and inducing activity. The monomer−dimer−tetramer equilibrium is very cooperative (Figure 3), and this would give rise to a sharp increase in enzyme activity in the presence of a suitable substrate, thereby providing a mechanism for selectivity and substrate induction of the phosphatase activity. On the other hand, phosphorylation at Tyr131 and Tyr132, which is known to activate BPTP, is also expected to destabilize both the dimer and the tetramer. In the absence of a proper substrate, phosphorylated BPTP would become its own substrate, leading to dephosphorylated BPTP, and allowing the regeneration of the inactive tetrameric structure. In fact, the dimer structure, with a bound phosphate, could be interpreted as a stable postreaction complex. This process would switch off the phosphatase activity.

The concentration of BPTP at which the population of tetramer is significant in dilute solution is relatively high. However, tetramer formation could be significantly enhanced because of excluded volume effects in the crowded cell interior.[41] The dissociation constant for a protein tetramer has been estimated to decrease by a factor of 10^3-10^5 in the crowded interior of *E. coli*.[42] Using the tetramerization constant measured here in dilute solution, one could thus estimate a micromolar or lower dissociation constant under physiological conditions. However, this will need to be verified experimentally.

Since the tetramer and dimer forms are inactive, but can be converted into the active monomeric form, one could argue that

(34) Weiss, A.; Schlessinger, J. *Cell* **1998**, *94*, 277−280.
(35) Majetiand, R.; Weiss, A. *Chem. Rev.* **2001**, *101*, 2441−2448.
(36) Schroff, A. D. Ph.D. Thesis, Purdue University, West Lafayette, IN, 1997.
(37) Chiarugi, P.; Cirri, P.; Marra, F.; Raugei, G.; Camici, G.; Manao, G.; Ramponi, G. *Biochem. Biophys. Res. Commun.* **1997**, *238*, 676−682.
(38) Tailor, P.; Gilman, J.; Williams, S.; Couture, C.; Mustelin, T. *J. Biol. Chem.* **1997**, *272*, 5371−5374.
(39) Rigacci, S.; Degl'Innocenti, D.; Bucciantini, M.; Cirri, P.; Berti, A.; Ramponi, G. *J. Biol. Chem.* **1996**, *271*, 1278−1281.
(40) Chiarugi, P.; Taddei, M. L.; Cirri, P.; Talini, D.; Buricchi, F.; Camici, G.; Manao, G.; Raugei, G.; Ramponi, G. *J. Biol. Chem.* **2000**, *275*, 37619−37627.
(41) Ellis, R. J. *Curr. Opin. Struct. Biol.* **2001**, *11*, 114−119.
(42) Zimmerman, S. B.; Trach, S. O. *J. Mol. Biol.* **1991**, *222*, 599−620.

the tetramer is a "supramolecular proenzyme". In contrast to classical proenzymes, the inactive form can be spontaneously reconstituted when the enzyme runs out of its target substrate, with obvious advantages for an enzyme involved in regulating cell-signaling pathways.

Concluding Remarks

Weak protein−protein interactions that are fast on the NMR chemical shift time scale are extremely difficult to detect by conventional means. The oligomeric species may have low populations, even at biologically relevant concentrations, and most of their spectroscopic properties are very similar to those of the monomer. Traditional hydrodynamic methods, such as equilibrium ultracentrifugation or translational diffusion measurements, do not usually afford enough resolution to differentiate between alternative aggregation models. However, the much faster relaxation of the larger species allows their easy detection by standard relaxation methods. Relaxation measurements at different concentrations, combined with hydrodynamic calculations of the contributions from the different oligomers, provide a quantitative estimate of the populations of the different species. From these, one can derive the dissociation constants and stoichiometries. In favorable cases, the accuracy of the predictions is enough to identify the location of residues involved in the interface of the higher order aggregates. Thus, both thermodynamic and structural information can be derived from relaxation data.

In the present study, the experimental test case comprised ^{15}N relaxation rates obtained for a protein undergoing self-association. However, the method is equally well suited for analysis of other nuclei whose relaxation is dominated by well-defined coupling tensors, for example, ^{13}C, and for other types of macromolecular complexes, such as hetero-oligomeric proteins, or complexes between proteins and nucleic acids.

Experimental Section

Sample preparation and ^{15}N relaxation measurements have been described elsewhere.[27] Buffer conditions were 200 mM potassium phosphate, pH 5.0, 10 mM dithiothreitol, and 3 mM sodium azide. The NMR experiments were run at 310 K on a Varian Inova 600 MHz spectrometer.

Relaxation rates for the monomer were calculated using HydroNMR and the PDB coordinates 1PNT.[43] Dimer coordinates were taken from the crystal structure of the S19A mutant (1C0E).[26] A consensus atomic element radius of $a = 3.3$ Å was used in all calculations. Calculated rotational diffusion parameters are the following: monomer, $\tau_c = 7.52$ ns, $D_{\parallel}/D_{\perp} = 0.87$; dimer, $\tau_c = 17.53$ ns, $D_{\parallel}/D_{\perp} = 1.80$. Calculated translational diffusion coefficients for the monomer and dimer were 1.51×10^{-6} cm^2 s^{-1} and 1.14×10^{-6} cm^2 s^{-1}, respectively. HydroNMR is available at the web site http://leonardo.fcu.um.es/macromol.

Estimates of the correlation time for higher order aggregates were obtained from the calculated hydrodynamic properties of the dimer. The correlation time of a compact aggregate formed by merging the volume occupied by n monomers with correlation time $\tau_c^{M'}$ into a single sphere is $n \times \tau_c^{M}$. Changes in the hydration shell will occur at the protein−protein interface upon oligomerization. In addition, side chain complementarity at the interface will reduce the effective volume of the oligomer. To partially account for these effects, the hydrodynamic properties of a partially solvent-depleted monomer were estimated from those calculated for the dimer using the formula $\tau_c^{M'} = \tau_c^{D}/2.78$, derived from the ratio of the harmonic correlation times of a sphere and its dimer.[29,30] A hydrodynamic radius of 33.1 Å and a translational diffusion coefficient of 0.99×10^{-6} cm^2 s^{-1} were calculated for the tetramer from its correlation time using with the well-established Einstein relationships

$$\tau_c = \frac{8\pi\eta_0 r^3}{6kT}$$

$$D = \frac{kT}{6\pi\eta_0 r}$$

where r is the hydrodynamic radius, η_0 is the water viscosity at $T = 310$ K, and k is the Boltzmann constant.

Optimization of parameters was done by minimizing an error function defined as

$$\chi^2 = \sum_i \sum_j \frac{[(R_2/R_1)_{i,j}^{exp} - (R_2/R_1)_{i,j}^{calc}]^2}{E(R_2/R_1)_{i,j}^{exp^2}}$$

where the i and j indices run over different NH groups and different concentrations, respectively, and $E(R_2/R_1)$ is the experimental error in the R_2/R_1 ratio for residue i measured at concentration j.

Using an iterative procedure, we removed residues 5, 84, 105, 107, 108, 109, 110, and 157, which had R_2/R_1 values significantly lower than the calculated ones, indicating fast local motions. All of these residues had heteronuclear NOE values ≤ 0.6. Residues 35, 48, 120, 125, 126, and 132, with higher than predicted R_2/R_1 values, were also removed. All these residues were independently identified as having a non-negligible exchange contribution by comparing experimental cross-correlation and autorelaxation rates.[27] In addition, residues 1−4, which do not appear in the dimer structure, and residues 6, 30, 51, 156, which are clear outliers in the R_2/R_1 predictions, were also eliminated from the data sets at all concentrations. Residues 69, 147, and 148 were eliminated only from the lowest concentration data, and residues 21, 39, 59, 61, and 72, from the highest concentration data set.

Optimizations were performed using Mathematica 3.0 (Wolfram Research). Initial optimizations were performed starting from a random set of initial values for each parameter to avoid local minima. In the monomer−dimer−tetramer model, the complex functional dependency required that the sets of parameters (K_T, R_1^T, and R_2^T), (K_D, K_T), and (K_D, R_1^T and R_2^T) were successively fixed, while the rest was optimized. The process was repeated until convergence. In the floating stoichiometry model, the seven parameters could be adjusted simultaneously. Error propagation was carried out using a Monte Carlo method with repeated minimizations, starting from simulated data points normally distributed around calculated R_1 and R_2 values according to the experimental uncertainty.

Acknowledgment. Prof. R. L. Van Etten kindly provided the BPTP clone. Financial support from the Spanish Ministerio de Ciencia y Tecnología (BIO2001-3115 and BQU2000-0229), the Generalitat de Catalunya (to M.P.), the Swedish Research Council, and the Swedish Foundation for Strategic Research (to M.A.) is gratefully acknowledged. P.B. and T.Å. were recipients of predoctoral grants from the Spanish Ministerio de Educación y Cultura and the Structural Biology Network of the Swedish Foundation for Strategic Research, respectively.

JA027836H

(43) Zhang, M.; Van Etten, R. L.; Stauffacher, C. V. *Biochemistry* **1994**, *33*, 11097−11105.