Ernst - From the fitting, one cannot decide whether there are even more conformations present. So we also performed some MD computer simulations. All the computations which we have done using different MD programs invariably gave us two rigid conformations, two well defined conformations which did not differ much between the different programs. Only the timescale of interconversion was a little different due to the different force fields.

Xu - My second question is when you investigate the multiple conformations using the molecular dynamics approach, is that as good as molecular dynamics if you use Monte Carlo simulations to investigate the multiple conformations?

Ernst - We did some Monte Carlo simulations as well. One obtains similar results.

Yuan Xu - So the two approaches are pretty much the same.

Ernst - Yes.

Xu - Thank you.

STRUCTURAL, DYNAMIC, AND FOLDING STUDIES OF SH2 AND SH3 DOMAINS

Julie D. Forman-Kay,[1] Steven M. Pascal,[1,2] Alex U. Singer,[1]
Toshio Yamazaki,[2] Ouwen Zhang,[1,2] Neil A. Farrow,[2] and Lewis E. Kay[2]

[1]Biochemistry Research Division
Hospital for Sick Children
Toronto, Ontario M5G 1X 8 CANADA
[2]Departments of Medical Genetics, Biochemistry and Chemistry
University of Toronto
Toronto, Ontario M5S 1A8 CANADA

INTRODUCTION

Recent advances in NMR methodology have made it a powerful approach for the study of biomolecular structure and dynamics (Bax and Grzesiek, 1993; Bax, 1994; Farrow et al., 1994a). In addition to NMR being a structural tool, as a solution spectroscopy it is exquisitely sensitive to dynamic processes - not only fast, low amplitude motions which can often be described by analysis of X-ray crystallographic B factors (Ringe and Petsko, 1985), but also slower, larger amplitude motions. These may include conformational exchange on millisecond time-scales or longer between states as dissimilar as folded and unfolded states of proteins and reflecting motions of tens of angstroms. We have exploited this distinguishing capability of NMR spectroscopy to describe the dynamic processes observed during structural studies of two isolated domains of signal transduction proteins, a Src Homology 2 (SH2) domain of phospholipase Cγ in complex with a phosphopeptide from the platelet-derived growth factor receptor (PDGFR) and an isolated Src Homology 3 (SH3) domain from the drosophila protein Drk[1].

SH2 and SH3 domains were first identified as regions of significant sequence similarity in the non-catalytic portion of the src family of protein tyrosine kinases (Pawson et al., 1993). These small, modular, independently folding domains have since been found

[1]The results described in this lecture are presented in greater detail in three manuscripts: S.M. Pascal, T. Yamazaki, A.U. Singer, L.E. Kay, and J.D. Forman-Kay, 1995, Structural and dynamic characterization of the phosphotyrosine binding region of an SH2 domain-phosphopeptide complex by NMR relaxation, proton exchange and chemical shift approaches, in preparation; O. Zhang and J.D. Forman-Kay, 1995, Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer, *Biochemistry*, submitted; N.A. Farrow, O. Zhang, J.D. Forman-Kay and L.E. Kay, 1995, Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer, *Biochemistry* 34:868.

NMR As a Structural Tool for Macromolecules: Current Status and Future Directions
Edited by B.D. Nageswara Rao and Marvin D. Kemple, Plenum Press, New York, 1996

in a number of other proteins involved in signaling, including enzymes and the so-called adapter molecules which function primarily by binding to multiple proteins. SH2 domains, of approximately 100 residues, and SH3 domains, of about 60 residues, function not through catalytic activity but by mediating protein-protein interactions in signal transduction. SH2 domains bind to specific sites of phosphorylated tyrosine on their biological targets, while SH3 domains bind to poly-proline type II helices (Yu et al., 1994).

C-TERMINAL SH2 DOMAIN OF PHOSPHOLIPASE Cγ1

When a growth factor receptor binds to the extracellular region of a receptor protein tyrosine kinase, it induces dimerization of these receptors and subsequent cross-phosphorylation by the kinase domains on exposed tyrosine residues (Ullrich and Schlessinger, 1990). This creates specific phosphotyrosine binding sites for SH2 domains in a number of signaling proteins. In the case of phospholipase Cγ (PLCγ), the Tyr-1021 site in the C-terminal region of the β-platelet-derived growth factor receptor is an important biological target (Valius and Kazlauskas, 1993; Valius et al., 1993). Upon binding of the C-terminal SH2 domain of PLCγ to the phosphorylated Tyr-1021 site, the kinase domain of the receptor can phosphorylate PLCγ, activating its catalytic metabolism of phosphoinositol-bis-phosphate to the second messengers diacylglycerol and inositol-triphosphate (Rhee, 1991). In order to understand at the atomic level the mechanism of the sequence specificity in the biological recognition taking place between the PLCγ SH2 domain and this phosphorylated tyrosine site, we have studied the isolated SH2 domain, labeled with ^{15}N and ^{13}C, in complex with a synthetic 12-residue phosphopeptide derived from the sequence around the Tyr-1021 site of the receptor (Asp-Asn-Asp-pTyr-Ile-Ile-Pro-Leu-Pro-Asp-Pro-Lys; gift of Dr. Steve Shoelson). Binding studies with this peptide have shown that the interaction of PLCγ with peptide has similar affinity to the *in vivo* receptor interaction, with a K_D of approximately 50-100 nM (Piccione et al., 1993).

NMR is particularly useful for the investigation of protein-protein interactions because we can employ isotopic labeling strategies to isolate intramolecular and intermolecular interactions, enabling us to probe the structure of the individual components (the SH2 domain and the peptide) and the SH2-peptide contacts separately (W. Lee et al., 1994). The SH2 domain is isotopically labeled with ^{15}N and ^{13}C and the synthetic peptide contains the natural abundance isotopes ^{14}N and ^{12}C. We utilize three types of NMR experiments to assign the resonances and structurally characterize the system: (1) isotope-edited experiments which probe interactions between protons bound to ^{15}N and ^{13}C for assignment and structure determination of the labeled SH2 domain, (2) isotope-filtered experiments which probe interactions between protons bound to ^{14}N and ^{12}C for the assignment and structure of the peptide and (3) half-filtered experiments which probe interactions between protons bound to either ^{15}N or ^{13}C and protons bound to either ^{12}C or ^{14}N to detect specific interactions between the SH2 domain and peptide. An example of this third class of experiments is the 3D ^{13}C-half-filtered NOESY experiment with proton and ^{13}C resonances of the SH2 domain along the F_1 and F_2 axes, respectively, correlated to proton resonances of the phosphorylated peptide along the F_3 axis. This experiment allowed us to identify over one hundred NOEs between the peptide and the protein, including numerous strong NOEs observed between the H^δ and H^ε resonances of the phosphotyrosine (pTyr) and residues of the SH2 domain.

In addition, NMR allows us to describe the dynamics of the interaction. Tight binding does not always lead to rigid complexes. It is of interest to observe any motions

within this tight binding complex in order to better understand the potential entropic role in maintaining high affinity binding. The dynamic studies described below focus on the motion of residues of the SH2 domain and are of the isotope-edited class of experiments.

Structural Basis of Specificity in Protein Recognition

Utilizing NOE data along with coupling constants and stereospecific assignments we have refined our previously published structures of the PLC-phosphopeptide complex (Pascal et al., 1994) and are now able to define the backbone coordinates to a precision of 0.6 Å overall (residues 11-99) and to 0.3 Å within the binding interface. The electrostatic potential surface of the SH2 domain (Nicholls et al., 1991) displays a very deep strongly positive binding pocket for the pTyr residue created by four arginine residues. The invariant arginine found in all SH2 domains (Koch et al., 1991), defined using the Eck nomenclature for SH2 residues (Eck et al., 1993) as Arg-βB5 (or Arg-37 in our numbering), is located at the bottom of this pocket with three additional arginines, Arg-αA2 (Arg-18), Arg-βB7 (Arg-39), and Arg-βD6 (Arg-59) surrounding it and contributing to the intense positive electrostatic potential. In addition to this deep pocket, a long hydrophobic groove is observed which has significant interactions with the residues C-terminal to the pTyr extending from isoleucine at the +1 position, isoleucine at the +2 position, proline at the +3 position, leucine at the +4 position and proline at position +5. There are also NOEs observed to the backbone of the aspartic acid at the +6 position. The extent of the hydrophobic binding groove for these hydrophobic residues C-terminal to the phosphotyrosine is quite significant and helps to explain the sequence specificity of the binding of this SH2 domain of PLCγ to the Tyr-1021 site of the PDGFR. The root-mean-squared deviation of all non-hydrogen atom coordinates within this hydrophobic binding interface to the mean coordinates is 0.6 Å. Thus, we are able to clearly describe the structural mechanisms of sequence specific recognition for this PLCγ-Tyr-1021 interaction. The structure also demonstrates our current general understanding of SH2 domain-target binding specificity. Not all phosphorylated tyrosine residues bind to all SH2 domains. Binding is strongly dependent on the particular amino acid sequence context of the pTyr, notably the residues C-terminal to it (Songyang et al., 1993, 1994).

SH2 domains, since they are very important molecules in signal transduction, have been the subject of a number of other structural studies (reviewed in Kuriyan and Cowburn, 1993; Yu and Schreiber, 1994). Complexes of the src and lck tyrosine kinase SH2 domains with the high-affinity binding peptide containing the sequence pTyr-Glu-Glu-Ile have been determined in the laboratories of John Kuriyan (Waksman et al., 1993) and Steve Harrison (Eck et al., 1993), respectively. These structures demonstrate a two-pronged plug model of SH2 binding with two deep pockets, one for the pTyr and one for the hydrophobic sidechain of the isoleucine residue at the +3 position C-terminal to the pTyr. Our structure of the PLCγ-PDGFR peptide complex shows a very different mode of binding where all of the residues from +1 to +5 are important for sequence-specific recognition. A complex with a high-affinity binding peptide to the N-terminal SH2 domain of the syp protein tyrosine phosphatase was recently solved in the laboratory of John Kuriyan (C.-H. Lee et al., 1994). This structure also demonstrated interactions of residues at the +1 through +5 positions of the peptide to the SH2 domain.

Dynamic Studies of Arginines

During the course of our structural studies, we found numerous resonances which were broadened or doubled. These observations stimulated our studies of the dynamic processes within this complex. We have specifically focused on the study of arginine dynamics (Pascal et al., 1995) because of the important electrostatic role of the four arginines lining the pTyr binding site in the SH2-peptide interaction. We have interpreted the results using the following principles to guide us: (1) conformational averaging in the intermediate exchange regime on the NMR time-scale leads to broadening of resonances, (2) fast rotation about the Nε-Cζ bond of an arginine leads to degeneracy of the Nη resonances and fast rotation about the Cζ-Nη bond leads to degeneracy of the Hη proton resonances and (3) hydrogen bonding of the Hη or Hε protons causes distinctive chemical shifts. In a 3D H(CCO)NH-TOCSY spectrum (Grzesiek et al., 1992; Logan et al., 1992; Montelione et al., 1992) of the SH2-peptide complex we observe that a number of sidechain resonances of the arginines in the pTyr binding site, especially the δ and γ proton resonances, have extremely broad linewidths. In the ^{15}N-^{1}H HSQC spectrum in the upfield region where Nη and Nε resonate (Figure 1) most of the arginines display degenerate Nη and Hη resonances at the expected chemical shifts for arginines that are interacting only with solvent and many are significantly broadened, indicative of intermediate time-scale rotation about their Nε-Cζ bonds. However, Arg-37 which is located at the base of the pTyr binding pocket has two distinct Nη chemical shifts and four distinct Hη chemical shifts, indicative of significantly slowed rotation about the Nε-Cζ and Cζ-Nη bonds. Two of the Hη resonances are extremely down-field shifted, one on each of the two Nη, which suggests involvement in two phosphate hydrogen bonds.

The Nε and Hε resonances observed in the upfield region of the HSQC spectrum are much better resolved than the Nη and Hη resonances. In order to better understand the dynamics and the pTyr-arginine interactions in the binding site, we exploited the chemical shift dispersion of these resonances and performed ^{15}N T_1, T_2 and heteronuclear ^{15}N-^{1}H NOE relaxation studies of the arginine Nε sites in the SH2 domain. We analyzed this data using the model-free approach of Lipari and Szabo (1982a,b) and found that three of the four arginines in the pTyr binding site, Arg-αA2 (Arg-18), Arg-βB5 (Arg-37) and Arg-βB7 (Arg-39) have order parameters in the range of 0.7 to 0.8, while Arg-βD6 (Arg-59) has an order parameter of 0.3. Utilizing these relaxation results in conjunction with our previous structural data and interpretation of chemical shifts of the guanidino group resonances, we can construct a model for the interactions of the arginines in the pTyr binding site. The highly conserved Arg-37 at the base of the binding site appears to be very rigid and donates two phosphate hydrogen bonds, one from each of the two Nη NH_2 groups. Arg-18 and Arg-39 also are stabilized by interactions within the binding site, probably with a phosphate and potentially with the electrons of the tyrosine ring itself in an aromatic-amino interaction. These interactions may be more transient, with an off-rate on a millisecond time-scale, leading to broadening of the aliphatic resonances, such that a number of different potential hydrogen bonding arrangements could be sampled. Arg-59 appears not to be involved in significant stabilizing hydrogen bonding interactions with the pTyr, but the role of the guanidino group may be for electrostatic attraction of the pTyr to the binding site. This suggestion correlates well with surface plasmon resonance experiments that have been used to investigate SH2-pTyr peptide binding (Felder et al., 1993). These studies show high affinities but also high off-rates which can be explained by compensating high on-rates which approach or exceed the diffusion limit. On-rates of this magnitude may be due to an

electrostatic drawing mechanism which would require the presence of a large positive field in the pTyr binding site, with contributions from a number of positively charged residues.

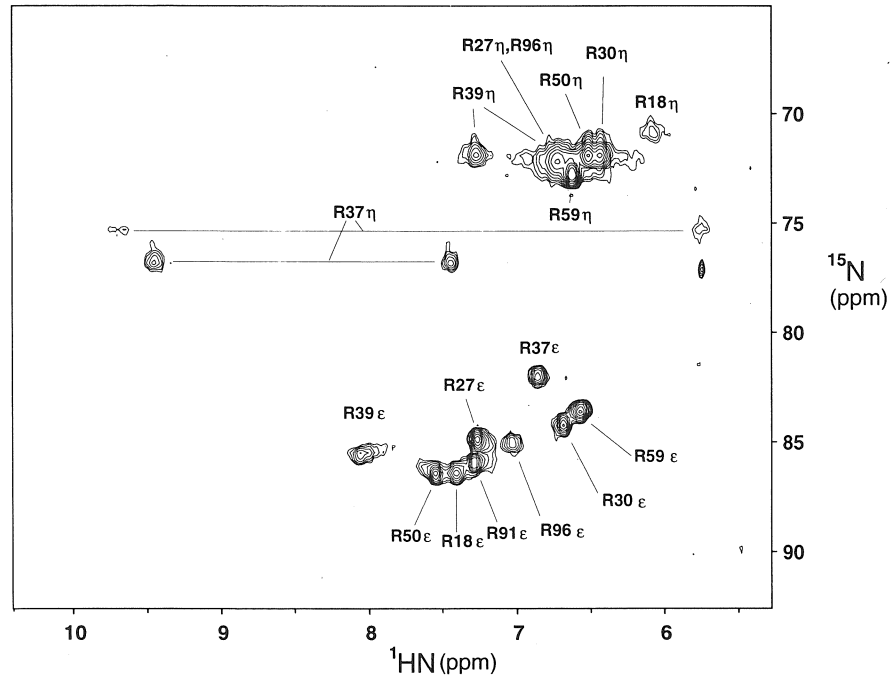

Figure 1. NεHε and NηHη region of the ^{1}H-^{15}N HSQC spectrum of the complex between the C-terminal SH2 domain of PLCγ and a 12-residue phosphopeptide from the Tyr-1021 high-affinity binding site of the PDGFR. The figure is modified from Yamazaki et al. (1995) where assignment methods are described.

The conclusions that we have drawn about the different kinds of motion within the binding site are currently qualitative. We hope to interpret them more quantitatively in order to understand the role of entropy in high-affinity binding. This may be important in the rational design of specific inhibitors of SH2 binding, since drug design requires an understanding of not just the enthalpic interactions derived from structural studies of sequence specific recognition but also entropic components of the free energy of binding.

THE N-TERMINAL SH3 DOMAIN OF DRK

Drk, the Drosophila homologue of *C. elegans* SEM-5 and mammalian GRB2, is composed of three domains, an N-terminal SH3 domain, a central SH2 domain and a C-terminal SH3 domain (Olivier et al., 1993; Clark et al., 1992; Simon et al., 1993). We have expressed the isolated N-terminal SH3 domain of the protein, including the first 59 residues of this molecule up to the conserved tryptophan which is the start of the SH2 sequence. We expected to perform structural studies similar to those described for the PLCγ SH2 domain. However, when we analyzed our first NMR spectrum of the sample in 50 mM phosphate buffer, pH 6, we observed an equilibrium between folded and unfolded states. Since we had discovered a system which enables the study of an unfolded state under the identical

conditions as the folded state, at near physiological conditions, we have exploited it instead for protein folding. NMR is a very useful technique for probing folding since unfolded and partially folded states, which cannot be crystallized, can be characterized in solution (Shortle, 1993; Neri et al., 1992). Analysis of these states to elucidate the interactions which are formed very early in folding can aid our understanding of the protein folding pathway. Characterization of the unfolded state is also important since knowledge of the beginning state, as well as any intermediates, is necessary for truly understanding any thermodynamic transition. The end states, folded proteins, have been very well characterized in many instances, but much less is known about unfolded states. A significant problem for interpretation of any results in this area is the fact that unfolded states are highly sensitive to the conditions of the denaturation (Shortle, 1993). Our equilibrium between folded and unfolded states of the SH3 domain, enabling us to study both states under the identical conditions, is a very unique one.

Structural Characterization of Folded and Unfolded States Existing in Equilibrium

The ^{15}N-^{1}H HSQC spectrum of this 59 residue SH3 domain in 50 mM sodium phosphate, pH 6, displays twice as many peaks as one would expect for a single folded state (Figure 2a). In addition, there are a number of amide proton resonances clustered between 8.0 and 8.5 ppm. In a homonuclear TOCSY we observed many amide-amide correlations, with resonances in one of the dimensions also clustered around 8-8.5 ppm. On the basis of previous structural studies the SH3 domain was known to be composed predominantly of β-strands (reviewed in Kuriyan and Cowburn, 1993; Yu and Schreiber, 1994). Thus, we would not expect artifactual amide-amide NOE effects in our TOCSY, especially since we utilized a clean TOCSY sequence (Griesinger et al., 1988). Therefore, these peaks must be indicative of conformational exchange.

In order to confirm that we were observing slow exchange on the NMR time-scale between folded and unfolded states, we performed titration experiments with agents known to destabilize or stabilize folded structure including pH, temperature, guanidine hydrochloride (Gdn) and Hofmeister series salts (Arakawa and Timasheff, 1985). The sidechain indole of the single tryptophan in this molecule displays two isolated downfield resonances, one for each of the folded and unfolded states. Thus, we could utilize 1D NMR spectra in the titrations and we observed the intensity of the more downfield of the two increase at high temperature, low pH or high Gdn concentration but completely disappear in the presence of 400 mM sodium sulfate. In an ^{15}N-^{1}H HSQC spectrum in 2 M Gdn (Figure 2b), approximately one peak per residue is observed, clustered between 8 and 8.5 in the proton dimension with very good dispersion in the ^{15}N dimension. In 400 mM sodium sulphate (Figure 2c), we observe a typical, well-dispersed spectrum of a β-sheet protein also containing about one peak per residue. If these two spectra are summed, the resulting spectrum is remarkably similar to the original HSQC in 50 mM sodium phosphate, pH 6.

We assigned the equilibrium folded and unfolded states simultaneously in these original conditions using 3D ^{15}N-edited gradient/sensitivity enhanced TOCSY- and NOESY-HSQC experiments (Kay et al., 1992; Zhang et al., 1994). The assignment could be performed relying on the cross-peaks between only folded or only unfolded resonances, as well as on TOCSY or NOESY exchange peaks between the two states. The secondary structure of the folded state of the SH3 domain (Figure 3a) based on $^{3}J_{HN\alpha}$ coupling constants, which are related to phi torsion angles, and patterns of sequential NH_i-NH_{i+1} and $C\alpha H_i$-NH_{i+1} NOES is very similar to that of other SH3 domains that have been solved. The domain consists largely of β-strands and can topologically be described as a β-barrel.

We also performed a titration experiment using a fragment of the biological target of this SH3 domain, the C-terminal tail of SOS, which contains a number of proline rich regions that form poly-proline helices (Olivier et al., 1993). There were significant chemical shifts of resonances of the folded state of the SH3 domain, demonstrating specific binding of the folded state in this equilibrium to its biological target.

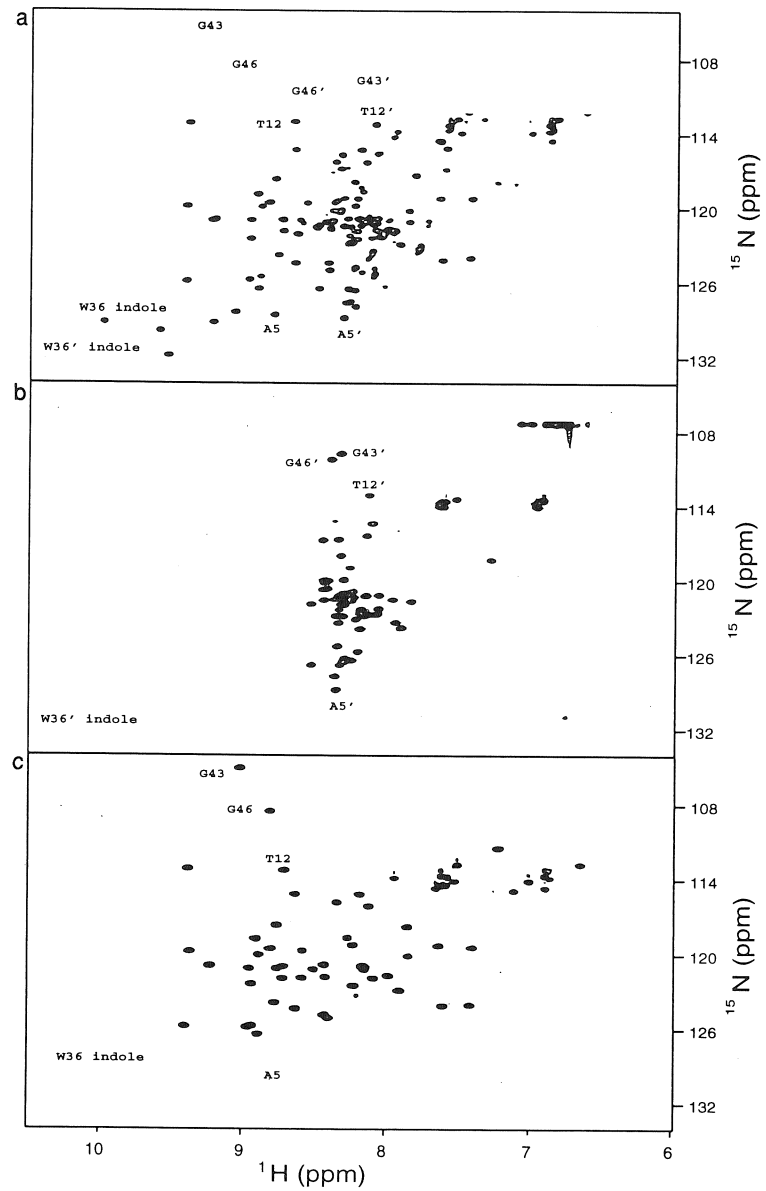

Figure 2. ^{1}H-^{15}N HSQC spectra of the N-terminal SH3 domain of drk in (a) 50 mM sodium phosphate, pH 6, (b) 2M Gdn, 50 mM sodium phosphate, pH 6 and (c) 400 mM sodium sulfate, 50 mM sodium phosphate, pH 6. Figure was modified from Zhang and Forman-Kay (1995).

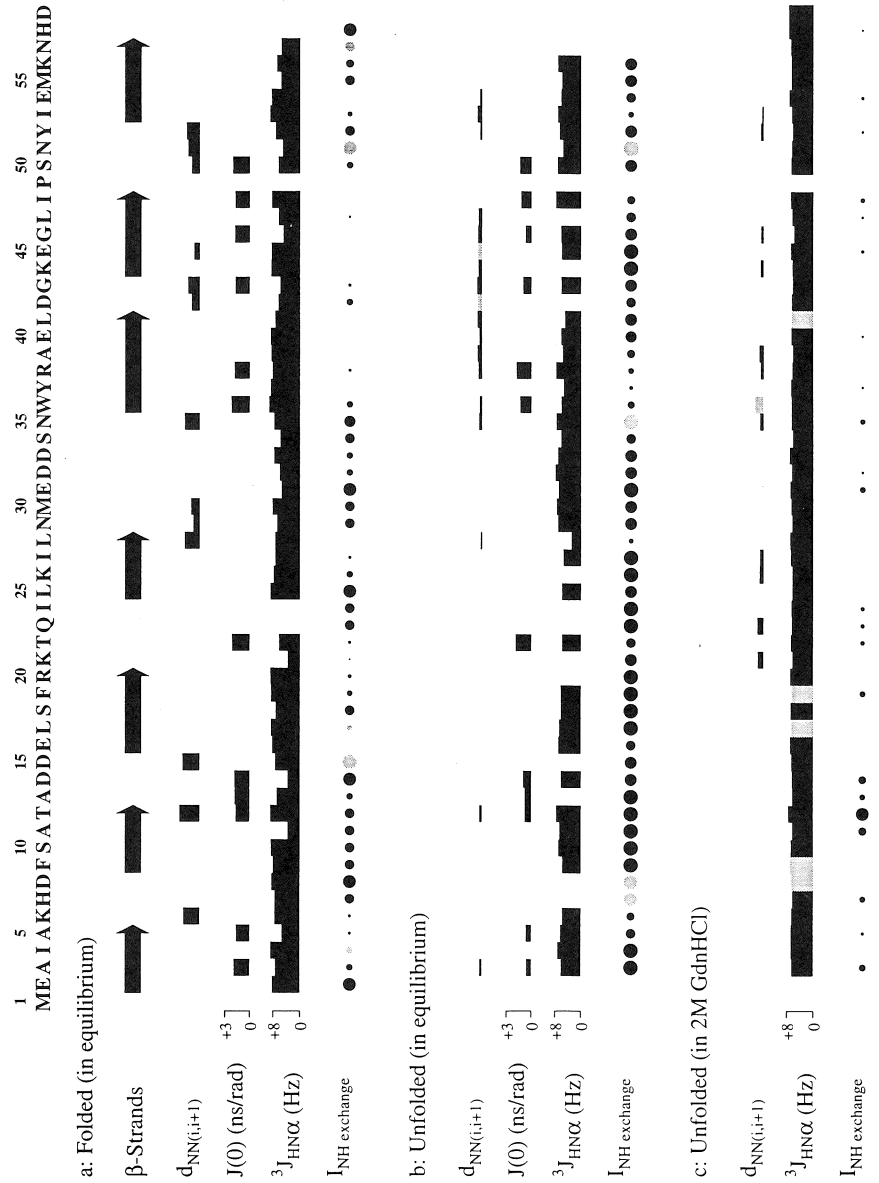

Figure 3. Sequence, secondary structure, sequential NH_i-NH_{i+1} NOEs, J(0) values, $^3J_{HN^\alpha}$ couple constants and intensities of amide-water exchange peaks for the N-terminal SH3 domain of drk (a) in the equilibrium folded state, (b) in the equilibrium unfolded state and (c) in the unfolded state in 2M Gdn.

The unfolded state, however, is very different from the folded states of other SH3 domains (Figure 3b). The changes in CαH proton chemical shift from random coil values (Wishart and Sykes, 1994) are quite small and the $^3J_{HN\alpha}$ coupling constants cluster around 6 and 7 Hz, indicative of conformational averaging, except for Leu-28 which has a very small coupling of less than 4 Hz. Amide proton exchange with solvent water in this state is much greater than in the folded state which is protected by hydrogen bonds within stable β-structure. Of special interest is a stretch of NH_i-NH_{i+1} sequential NOES in the region from Asn-35 to Ile-48, spanning a highly stable β-strand, β-turn, β-strand region in the folded structure and for which an extended stretch of NH_i-NH_{i+1} NOES is not observed in the folded state. While deviation from purely random structure in unfolded proteins is typically described as "residual structure," since these NOEs are not present in the folded state it is not accurately described as such.

Relaxation Studies

It is important to note that many of the NMR observables that were used for interpretation of structure in the unfolded state average in a complex manner over the rapidly exchanging ensemble of multiple unfolded protein conformations. Chemical shift is the easiest to interpret in this regard because it is a simple linear, population weighted average, while $^3J_{HN\alpha}$ coupling constants are averaged as a cosine modulated function of the phi torsion angle and NOEs are either an $\langle r^{-6} \rangle$ or $\langle r^{-3} \rangle$ average (Kessler et al., 1988). In addition, the value of the NOE is not merely a reflection of interproton distances, since it also contains contributions from dynamic behavior. In the macromolecular limit the NOE is directly proportional to the J(0) spectral density term. Thus, the extended stretch of sequential NOEs can not be interpreted as a clear preference of these residues for sampling conformations in the α-region of (ϕ,ψ) conformational space. If the J(0) spectral density values in this region are much higher than those in other regions, these NOEs must be interpreted simply as a consequence of motion. In order to establish that the NOEs are reflective of structural preferences, we performed ^{15}N relaxation experiments to measure the values of J(0) as a function of residue.

Relaxation experiments are typically analyzed by quantitating and fitting the intensity of peaks in T_1, T_2 and ^{15}N-^1H NOE spectra recorded with various delay times. In this particular system with exchange between folded and unfolded states, we have an additional complication. New pulse sequences were developed (Farrow et al., 1994b) to enable the separation of the exchange rate and the relaxation rates. However, in order to fit the data, the intensity of the two diagonal peaks (i.e. the folded and the unfolded peak) and at least one of the two exchange cross-peaks must be unambiguously quantified. This limits the number of resonances that can be characterized because of the poor spectral dispersion and overlap of many peaks in the unfolded state. We were able to analyze 12 backbone amides and the tryptophan sidechain indole (Farrow et al., 1995). The folded SH3 domain behaves as expected for a molecule of its size, with an overall correlation time of 5.5 ns. Interestingly, the unfolded domain also has the same average correlation time, which suggests that this state is rather compact as opposed to an extended random coil. Detailed analysis of the values of the relaxation rates for the unfolded state demonstrate that the residues towards the center of the molecule display very similar behavior to that observed for the folded state, while residues towards the ends of the molecule deviate much more significantly from the behavior of the folded state. The forward and backward exchange rates can also be determined from the fitting procedure. Since the transition is from folded

to unfolded states of the protein, the exchange rates are identical to the rates of unfolding and folding. The rate of folding of this isolated SH3 domain is approximately 0.9 sec^{-1}, which is very similar to a value measured for the rate of folding of interleukin-1β, another all β-sheet protein (Varley et al., 1993). The rate of folding of the SH3 domain can be measured at each of the 12 residues. We observe that the rates at each site are not identical and the differences in the measurements are greater than our experimental error.

Rather than analyzing this relaxation data in terms of the Lipari and Szabo model-free formalism and assuming isotropic tumbling which may not hold in the case of an unfolded protein, we have mapped the spectral density functions in a similar way to that described by Peng and Wagner (1992a,b). We have grouped the three high frequency terms into one, $J(\omega^h)$, and determined $J(0)$, $J(\omega^N)$ and $J(\omega^h)$. The spectral density mapping approach that we employed shows the following relationships for the folded state, $J(0) >> J(\omega^N) >> J(\omega^h)$. The values of $J(0)$ are relatively constant over the protein, providing confidence that NOE intensities in the folded state may be interpreted directly in terms of structural properties. It is noteworthy that similar trends were observed for the unfolded state, with $J(0) > J(\omega^N) > J(\omega^h)$ and $J(0)$ values for some residues approaching their values in the folded state. This suggests that the unfolded state is not well described by a random coil model having non-interacting residues and that the motion in this state may be closer to that of a folded protein.

Addressing our original question regarding the distribution of the $J(0)$ values in the unfolded state, we measured $J(0)$ values for four residues in the region from Asn-35 to Ile-48 (Trp-36, Arg-48, Glu-43 and Gly-46) where NH_i-NH_{i+1} sequential NOEs are observed. These four values vary greatly from about 0.7 to 1.8 ns/rad, from close to the lowest and highest of all measured values. There is also a very high $J(0)$ value of 1.9 ns/rad for residue Thr-22 for which no NH_i-NH_{i+1} sequential NOE is observed. This is evidence supporting a significant role for structural preferences (i.e. interproton distances) giving rise to these NOEs since there appears to be no direct correlation between the presence of the NOEs and the value of $J(0)$. Thus, these NOEs can be interpreted as reflecting preferential sampling of the α-region of (ϕ,ψ) conformational space.

Implications for Protein Folding

We have also analyzed the equilibrium constants for the unfolding/folding transition as a function of residue by performing fully relaxed HSQC experiments and quantitating the size of the folded and unfolded peaks. Again, we noted differences across the protein. Although it is difficult to interpret, the fact that we see differences in exchange rates as a function of residue and differences in equilibrium constants as a function of residue implies that there is not perfect cooperativity in this folding reaction. We have explored this further by measuring the heat capacity change during thermal unfolding of the SH3 domain in high salt and observed what is, at a global level, a two-state cooperative transition (P. Morin and E. Friere, personal communication). Thus, the NMR results may be providing higher resolution data than is often seen using differential scanning calorimetry or from thermal melting curves using CD spectroscopy.

In order to compare the unfolded state that is present in denaturant (2 M Gdn) with the unfolded state that exists in equilibrium with the folded state, we assigned and analyzed the chemical shifts, sequential NOEs, $^3J_{HN\alpha}$ coupling constants and amide exchange with solvent for a 2 M Gdn sample of the N-terminal SH3 domain (Figure 3c). Some sequential NH_i-NH_{i+1} NOEs are present in the region from Asn-35 to Ile-48, but the long contiguous stretch in that region is not preserved. Again very little difference between the CαH proton

chemical shift and random coil values (Wishart and Sykes, 1994) is observed as expected. All of the $^3J_{HN\alpha}$ coupling constants reflect averaging of phi torsion angles and the very low coupling at Leu-28 has disappeared. Very interestingly, the observed amide exchange rates with solvent water, with the exception of a region near Thr-12, is much less than that in the aqueous equilibrium unfolded state. We also assigned and performed similar comparisons of the fully stabilized folded state of the SH3 domain in 400 mM sodium sulfate with the folded state in equilibrium with the unfolded in low ionic strength. Comparison of the backbone chemical shifts shows that these two folded states are virtually identical. However, the unfolded state in equilibrium with the folded is not identical to the unfolded state in Gdn based on chemical shift comparisons. There are significant deviations near Leu-28, the residue which has a coupling constant of less than 4 Hz in low ionic strength and a much larger one in Gdn, reflecting conformational averaging. While both unfolded states show differences in backbone chemical shifts from the random coil values (Wishart and Sykes, 1994; Braun et al., 1994), currently these differences are not easily interpretable. With the recent rapid advances in our understanding of chemical shifts (de Dios et al., 1993a,b), however, it is likely that we will be able to learn many things in the future by such comparisons.

In summary, the unfolded state of the N-terminal SH3 domain in low ionic strength aqueous buffer is not a true random coil state. It appears to be more compact, with non-random structure. It is also significantly different in describable ways from the unfolded state in 2 M Gdn. Of particular interest, this state contains a stretch of residues demonstrating rapidly averaging conformations with a structural preference for the helical region of (ϕ,ψ) space. These residues span a β-strand, β-turn, β-strand region in the folded state, so their structural preference in the unfolded state is decidedly non-native. In addition, we have measured the exchange rates as a function of residue, which are identical to the folding and unfolding rates. The average rate agrees well with previous measurements of the rate of folding of β-sheet proteins but differences in the rates as a function of residue imply deviations from exact cooperativity. We are currently pursuing this issue of microscopic non-cooperativity.

ACKNOWLEDGMENTS

We thank our collaborator Dr. Tony Pawson at the Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto for the original clones of the PLCγ SH2 and Drk SH3 domains and for his support of these studies and Dr. Steve Shoelson at the Joslin Diabetes Center and Department of Medicine, Brigham and Womens Hospital and Harvard Medical School who provided the phosphorylated peptide from the Tyr-1021 site of the PDGFR.

REFERENCES

Arakawa, T., and Timasheff, S.N., 1985, Theory of protein solubility, *Methods Enzymol.* 114:49.

Bax, A., 1994, Multidimensional nuclear magnetic resonance methods for protein studies, *Curr. Opin. Str. Biol.* 4:738.

Bax, A., and Grzesiek, S., 1993, Methodological Advances in Protein NMR, *Acc. Chem. Res.* 26:131.

Braun, D., Wider, G., and Wüthrich, K., 1994, Sequence-corrected ^{15}N "random coil" chemical shifts, *J. Am. Chem. Soc.* 116:8466.

Clark, S.G., Stern, M.J., and Horvitz, H.R., 1992, C. elegans cell-signalling gene sem-5 encodes a protein with SH2 and SH3 domains, *Nature* 356:340.

de Dios, A.C., Pearson, J.B., and Oldfield, E., 1993a, Secondary and tertiary structural effects on protein NMR chemical shifts: an ab initio approach, *Science* 260:1491.

de Dios, A.C., Pearson, J.B., and Oldfield, E., 1993b, Chemical shifts in proteins - an ab initio study of carbon-13 nuclear magnetic resonance chemical shielding in glycine, alanine and valine residues, *J. Am. Chem. Soc.* 115:9768.

Eck, M.J., Shoelson, S.E., and Harrison, S.C., 1993, Recognition of a high-affinity phosphotyrosyl peptide by the src homology-2 domain of p56^{lck}, *Nature* 362:87.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E., 1994a, Backbone dynamics of a free and phosphopeptide-complexed src homology 2 domain studied by ^{15}N NMR relaxation, *Biochem.* 33:5984.

Farrow, N.A., Zhang, O., Forman-Kay, J.D., and Kay, L.E., 1994b, A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal deacy and chemical exchange rates of systems in slow equilibrium, *J. Biomol. NMR* 4:727.

Farrow, N.A., Zhang, O., Forman-Kay, J.D., and Kay, L.E., 1995, Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer, *Biochem.* 34:868.

Felder, S., Zhou, M., Hu, P., Ureña, J., Ullrich, A., Chaudhuri, M., White, M., Shoelson, S.E., and Schlessinger, J., 1993, SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange, *Mol. Cell. Biol.* 13:1449.

Griesinger, C., Otting, G., Wüthrich, K., and Ernst, R.R., 1988, Clean TOCSY for proton spin system identification in macromolecules, *J. Am. Chem. Soc.* 110:7870.

Grzesiek, S., Anglister, J., and Bax, A., 1992, Correlation of backbone amide and aliphatic side-chain resonances in ^{13}C/^{15}N enriched proteins by isotropic mixing of ^{13}C magnetization, *J. Magn. Reson. Series B* 101:114.

Kay, L.E., Keifer, P., and Saarinen, T., 1992, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc.* 114:10663.

Kessler, H., Griesinger, C., Lautz, J., Muller, A., van Gunsteren, W.F., and Berendsen, H.J.C., 1988, Conformational dynamics detected by nuclear magnetic resonance NOE values and J coupling constants, *J. Am. Chem. Soc.* 110:3393.

Koch, C.A., Anderson, D., Moran, M.F., Ellis, C., and Pawson, T., 1991, SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins, *Science* 252:668.

Kuriyan, J., and Cowburn, D., 1993, Structures of SH2 and SH3 domains, *Curr. Opin. Str. Biol.* 3:828.

Lee, C-H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S.E., and Kuriyan, J., 1994, Crystal structures of peptide complexes of the amino-terminal SH2 domain of the Syp tyrosine phosphatase, *Structure* 2:423.

Lee, W., Revington, M., Arrowsmith, C.H., and Kay, L.E., 1994, A pulsed-field gradient isotope-filtered 3D ^{13}C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes, *FEBS Lett.* 350:87.

Lipari, G., and Szabo, A., 1982a, Model-free approach to the interpretation of nuclear magnetic relaxation in macromolecules: 1. Theory and range of validity, *J. Am. Chem. Soc.* 104:4546.

Lipari, G., and Szabo, A., 1982b, Model-free approach to the interpretation of nuclear magnetic relaxation in macromolecules: 2. Analysis of experimental results, *J. Am. Chem. Soc.* 104:4559.

Logan, T.M., Olejniczak, E.T., Xu, R., and Fesik, S.W., 1992, Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments, *FEBS Letts.*, 314:413.

Montelione, G.T., Lyons, B.A., Emerson, S.D., and Jashiro, M., 1992, An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins, *J. Am. Chem. Soc.* 114:10974.

Neri, D., Billeter, M., Wider, G., and Wüthrich, K., 1992, NMR determination of residual structure in a urea-denatured protein, the 434-repressor, *Science* 257:1559.

Nicholls, A., Sharp, K.A., and Honig, B., 1991, Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons, *Protein Struct. Funct. Genet.* 11:281.

Olivier, J.P., Raabe, T., Henkemeyer, M., Dickerson, B., Mbamalu, G., Margolis, B., Schlessinger, J., Hafen, E., and Pawson, T., 1993, A Drosophila SH2-SH3 adaptor protein implicated in coupling the Sevenless tyrosine kinase to an activator of Ras guanine nucleotide exchange - Sos, *Cell* 73:179.

Pascal, S.M., Singer, A.U., Gish, G., Yamazaki, T., Shoelson, S.E., Pawson, T., Kay, L.E., and Forman-Kay, J.D., 1994, NMR Structure of an SH2 Domain of Phospholipase C-γ1 Complexed with a High-Affinity Binding Peptide, *Cell* 77:461.

Pascal, S.M., Yamazaki, T., Singer, A.U., Kay, L.E., and Forman-Kay, J.D., 1995, Structural and dynamic characterization of the phosphotyrosine binding region of an SH2 domain-phosphopeptide complex by NMR relaxation, proton exchange and chemical shift approaches, in preparation.

Pawson, T., Olivier, P., Rozakis-Adcock, M., McGlade, J., and Henkemeyer, M., 1993, Proteins with SH2 and SH3 domains couple receptor tyrosine kinases to intracellular signaling pathways, *Phil. Trans. R. Soc. Lond. B* 340:279.

Peng, J.W., and Wagner, G., 1992a, Mapping of spectral density functions using heteronuclear NMR relaxation measurements, *J. Magn. Reson.* 98:308.

Peng, J.W., and Wagner, G., 1992b, Mapping of the spectral densities of N-H bond motions in Eglin c using heteronuclear relaxation experiments, *Biochem.* 31:8571.

Piccione, E., Case, R.D., Domchek, S.M., Hu, P., Chaudhuri, M., Backer, J.M., Schlessinger, J., and Shoelson, S.E., 1993, Phosphatidylinositol 3-kinase p85 SH2 domain specificity defined by direct phosphopeptide/SH2 domain binding, *Biochem.* 32:3197.

Rhee, S.G.,1991, Inositol phospholipid-specific phospholipase C: interaction of the γ1 isoform with tyrosine kinase, TIBS 16:297.

Ringe, D., and Petsko, G.A., 1985, Mapping protein dynamics by X-ray diffraction, *Prog. Biophys. Molec. Biol.* 45:197235.

Shortle, D., 1993, Denatured states of proteins and their roles in folding and stability, *Curr. Opin. Str. Biol.* 3:66.

Simon, M.A., Dodson, G.S., and Rubin, G.M., 1993, An SH3-SH2-SH3 protein is required for p21^{Ras1} activation and binds to Sevenless and Sos proteins in vitro, *Cell* 73:169.

Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J., Neel, B.G., Birge, R.B., Fajardo, J.E., Chou, M.M., Hanafusa, H., Schaffhausen, B., and Cantley, L.C., 1993, SH2 domains recognize specific phosphopeptide sequences, *Cell* 72:767.

Songyang, Z., Shoelson, S.E., McGlade, J., Olivier, P., Pawson, T., Bustelo, R.X., Barbicid, M., Sabe, H., Hanafusa, H., Yi, T., Ren, R., Baltimore, D., Ratnofsky, S., Feldman, R.A., and Cantley, L.C., 1994, Specific motifs recognized by the SH2 domains of csk, 3BP2, fes/fps, Grb-2, SHPTP1, SHC, Syk and vav, *Mol. Cell. Biol.* 14:2777.

Ullrich, A., and Schlessinger, J., 1990, Signal transduction by receptors with tyrosine kinase activity, *Cell* 61:203.

Valius, M., Bazenet, C., and Kazlauskas, A., 1993, Tyrosines 1021 and 1009 are phosphorylation sites in the carboxy terminus of the platelet-derived growth factor receptor β subunit and are required for binding of phospholipase Cγ and a 64-kilodalton protein, respectively, *Mol. Cell. Biol.* 13:133.

Valius, M., and Kazlauskas, A., 1993, Phospholipase C-γ-1 and phosphatidylinositol 3 kinase are the downstream mediators of the PDGF receptor's mitogenic signal, *Cell* 73:321.

Varley, P., Gronenborn, A.M., Christense, H., Wingfield, P.T., Pain, R.H., and Clore, G.M., 1993, Kinetics of folding of the all-β sheet protein interleukin-1 β, *Science* 260:1110.

Waksman, G., Shoelson, S.E., Pant, N., Cowburn, D., and Kuriyan, J., 1993, Binding of a high affinity phosphotyrosyl peptide to the src SH2 domain: crystal structures of the complexed and peptide-free forms, *Cell* 72:779.

Wishart, D.S., and Sykes, B.D., 1994, Chemical shifts as a tool for structure determination, *Methods Enzymol.* 239:363.

Yamazaki, T., Pascal, S.M., Singer, A.U., Forman-Kay, J.D., and Kay, L.E., 1995, NMR pulse schemes for the sequence-specific assignment of arginine guanidino ^{15}N and ^{1}H chemical shifts in proteins, *J. Am. Chem. Soc.*, in press.

Yu, H., and Schreiber, S.L., 1994, Signalling an interest, *Nature Str. Biol.* 1:417.

Yu, H., Chen, J.K., Feng, S., Dalgarno, D.C., Brauer, A.W., and Schreiber, S.L., 1994, Structural basis for the binding of proline-rich peptides to SH3 domains, *Cell* 76:933.

Zhang, O., and Forman-Kay, J.D., 1995, Structural characterization of folded and unfolded states of an SH3 domain in equilibrium in aqueous buffer, *Biochem.*, submitted.

Zhang, O., Kay, L.E., Olivier, J.P., and Forman-Kay, J.D., 1994, Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradients NMR techniques, *J. Biomol. NMR* 4:845.