Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 11673–11677, December 1992
Biochemistry

Secondary structure of Src homology 2 domain of c-Abl by heteronuclear NMR spectroscopy in solution

(oncogene/c-*abl*)

MICHAEL OVERDUIN, BRUCE MAYER, CARLOS B. RIOS, DAVID BALTIMORE, AND DAVID COWBURN*

Laboratories of The Rockefeller University, 1230 York Avenue, New York, NY 10021

Contributed by David Baltimore, July 14, 1992

ABSTRACT **The Src homology 2 (SH2) domain is a recognition motif thought to mediate the association of the cytoplasmic proteins involved in signal transduction by binding to phosphotyrosyl-containing sequences in proteins. Assignments of nearly all ^1H and ^{15}N resonances of the SH2 domain from the c-Abl protein-tyrosine kinase have been obtained from homonuclear and heteronuclear NMR experiments. The secondary structure has been elucidated from the pattern of nuclear Overhauser effects, from vicinal coupling constants, and from observation of slowly exchanging amino hydrogens. The secondary structure contains two α-helices and eight β-strands, six of which are arranged in two contiguous, antiparallel β-sheets. Residues believed to be involved in phosphotyrosyl ligand binding are on a face of one β-sheet. The alignment of homologous sequences on the basis of secondary structure suggests a conserved global fold in a family of SH2 domains.**

The Src homology 2 (SH2) domain is a modular protein of ≈100 amino acids that binds with high affinity to tyrosine-phosphorylated proteins (refs. 1–5; reviewed in ref. 6). Interactions between SH2 domains and their ligands are thought to mediate the formation of complexes involved in intracellular signal transduction. In the best characterized examples, activated tyrosine-kinase growth factor receptors autophosphorylate tyrosine residues, providing high-affinity binding sites for SH2-containing proteins such as phospholipase C-γ (PLC-γ), GTPase-activating protein (GAP), and the p85 subunit of phosphatidylinositol 3-kinase (reviewed in refs. 6 and 7). The importance of these interactions has been demonstrated by mutating specific sites on the platelet-derived growth factor receptor, which eliminates both p85 binding and mitogenic activity (8).

SH2 domains are found in a wide variety of proteins (6), including the nonreceptor class of tyrosine kinases. In most cases, mutation or deletion of the SH2 domain of tyrosine kinase oncogenes decreases their transforming activity (9–13). In the SH2 of Abl, a series of point mutations were constructed in the FLVRESES sequence. Mutants that were impaired in their *in vitro* binding to tyrosyl-phosphorylated proteins were concomitantly impaired in transforming activity, strongly suggesting that a functional SH2 is required for full transforming activity (4, 13). A number of highly conserved sequences are found in SH2 domains (6), consistent with a common folded structure among SH2 domains. Moreover, residues implicated in ligand binding by mutagenic studies are also highly conserved (3, 14).

In this paper, the secondary structure of the c-Abl SH2 domain is derived from nuclear Overhauser effect (NOE), vicinal coupling, and amide exchange data obtained by a variety of homonuclear and heteronuclear NMR methods. In light of the mutagenic and sequence homology data, the secondary structure reveals part of a putative ligand binding site and conserved structural features.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Sample Preparation. Recombinant c-Abl SH2 domain was obtained by expression and cleavage of a glutathione *S*-transferase fusion product (4, 13). The sequence of the 109-residue protein is shown in Fig. 3. The first 3 and last 5 residues are derived from fusion sequence. Uniform ^{15}N labeling was achieved by growing the cells in labeled Celtone medium (Martek) or in M9 minimal medium with 18 mM $^{15}NH_4Cl$. A sample with [^{15}N]leucine labeling was obtained by using M9 medium without NH_4Cl but with each amino acid at 100 μg/ml, including L-[^{15}N]leucine (Cambridge Isotope Laboratories, Cambridge, MA).

Solutions used for NMR contained up to 4 mM c-*abl* SH2 product in phosphate-buffered saline at pH 7.2 (uncorrected for isotope effects), 50 μM sodium azide, and 10% ^2H_2O in the ^1H_2O samples. The protein has a stable conformation over a wide pH and temperature range as monitored both by CD and by one-dimensional NMR. ^2H_2O samples were prepared by lyophilizing three times and dissolving the residue in 99.996% ^2H_2O. The internal ^1H chemical shift reference used was sodium 2,2-dimethyl-2-silapentane-5-sulfonate, and indirect referencing was used for ^{15}N (15).

NMR Spectroscopy. NMR experiments were run on a General Electric Omega 500 spectrometer. Quadrature detection was achieved by the States time-proportional phase incrementation (TPPI) method (16). The water signal was suppressed by using selective on-resonance irradiation during a relaxation delay of 0.6–1.5 s and, in NOE spectroscopy (NOESY) experiments, also during the mixing period. A 60-ms SCUBA pulse train (17) was employed in NOESY and double-quantum-filtered, two-dimensional correlated spectroscopy (DQF-COSY) experiments to allow recovery of saturated H^α resonances. Experiments were run at 25°C with sweep widths of 7017 and 2000 Hz for ^1H and ^{15}N, respectively, unless indicated otherwise.

Homonuclear experiments were run in both H_2O and ^2H_2O with standard pulse sequences and phase cycling. A range of t_1 increments from 200 to 512, each consisting of 2048 complex points, were typically acquired, with 96–256 scans per increment. In ^2H_2O DQF-COSY spectra 4096 points were acquired. Z-filtered homonuclear Hartmann–Hahn spectroscopy (HOHAHA) experiments used an MLEV-17 sequence for spin lock with mixing times of 35, 65, and 85 ms in ^2H_2O and between 30 and 90 ms in ^1H_2O. NOESY spectra were recorded

Abbreviations: SH2, Src homology domain 2; NOE, nuclear Overhauser effect; PLC, phospholipase C; GAP, GTPase-activating protein; COSY, two-dimensional correlated spectroscopy; NOESY, NOE spectroscopy; HMQC, heteronuclear multiple-quantum coherence spectroscopy; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; DQF, double-quantum-filtered.

*To whom reprint requests should be addressed.

in ^2H_2O with mixing times (τ_m) of 40, 80, 120, 180, and 240 ms and in ^1H_2O with τ_m of 40, 80, 100, 120, and 180 ms.

Heteronuclear experiments consisted of heteronuclear multiple-quantum coherence spectroscopy (HMQC), HMQC-*J*, HSMQC, ^{15}N-filtered HOHAHA and NOESY, two- and three-dimensional HMQC-NOESY, and three-dimensional HMQC-HOHAHA experiments (see refs. 18–20). The τ_m in HMQC-NOESY experiments was 120 ms, and in three-dimensional HMQC-HOHAHA the spin lock was 30 ms. The three-dimensional spectra were recorded with a 34 × 2 × 64 × 2 × 2048 complex matrix, with 32 scans per increment. Estimates of the $^3J(H^\alpha–H')$ coupling constants were made from HMQC-*J* spectra processed by using a 50° shifted sine bell in the acquisition dimension (ω_2) and a Gaussian equivalent to a linewidth of −4 Hz in the evolution dimension (ω_1). Coupling constants were derived from the peak separation at maximum height by appropriate corrections (19, 20). The amide hydrogen exchange rates were measured by following the intensity of cross peaks in HMQC experiments initiated at 5, 11, 22, 42, 83, 123 min, and 12 hr after exchanging a fully protonated, lyophilized sample with 99.996% ^2H_2O.

All data were processed with RUNMR software (21). Signal processing typically consisted of a Gaussian apodization in ω_2 and a cosine or cosine-squared apodization in combination with an inverted exponential function ($1 - ce^{-t}$) in ω_1. A single zero-filling and polynomial baseline correction was typically used in ω_1, whereas linear baseline correction was used in ω_2 where necessary. NOESY cross-peak intensity was estimated from the volume and maximum height relative to the height of cross peaks between hydrogens separated by a fixed distance.

RESULTS AND DISCUSSION

A combination of two- and three-dimensional NMR experiments was used to identify and sequentially assign nearly all of the ^1H and ^{15}N resonances of 98 of the total 109 amino acid residues. For 94 of these residues, all the backbone ^1H resonances and >96% of side-chain ^1H resonances residues could be assigned. The sequential assignment of the four additional residues (Gly^{28}, Gly^{43}, Asp^{65}, and Ser^{72}), all of which occur in loop regions, is tentative due to their partial backbone assignment. Line broadening due to chemical exchange precluded the identification of NOEs from the amide resonances of unassigned residues, all of which are located in the N- and C-terminal segments (Gly^1 to Ser^5, Gly^{105}, and Asp^{109}) and putative loop regions (Ser^{26}-Ser^{27}, Ser^{40}, and Ser^{64}).

Spin System Assignment. The c-Abl SH2 sequence contains 109 amino acid residues of the following types: 10 Gly, 6 Ala, 5 Thr, 17 Ser, 3 Asp, 5 Asn, 2 Phe, 8 His, 1 Trp, 7 Tyr, 7 Glu, 1 Gln, 6 Val, 5 Ile, 10 Leu, 3 Lys, 9 Arg, and 4 Pro. Candidate spin systems for all amino acid side chains could be identified in HOHAHA and DQF-COSY spectra in ^2H_2O by standard methods (22). Difficulties were encountered with the discrimination of the 17 Ser side-chain spin systems due to the crowded portion of the spectrum in which their H^α and H^β nuclei resonate. Also, overlap of some of the aromatic H^δ resonances necessitated the confirmation of their attached spin systems by subsequent sequential assignment and by NOEs from ring protons to spatially proximal residues. Several of the Leu, Lys, and Arg side-chain spin systems could not be readily differentiated either because of resonance overlap or because of limited transfer efficiency along the side chains in HOHAHA experiments. Nine of the 10 expected leucyl amide protons were identified from HMQC spectra of [^{15}N]Leu protein. The remaining, overlapped Leu^6 was discriminated through $^{15}N–H^\alpha$ correlations in HMQC-*J* spectra.

Spin systems were completed by connecting the amide hydrogens to their H^α protons in the DQF-COSY and HMQC-*J* spectra and to all side-chain protons in HOHAHA spectra, all recorded in ^1H_2O. Only 69 cross peaks between amide and H^α protons could be detected in DQF-COSY spectra, the Gly and Ser cross peaks being particularly weak. Greater signal intensity was observed in the HOHAHA

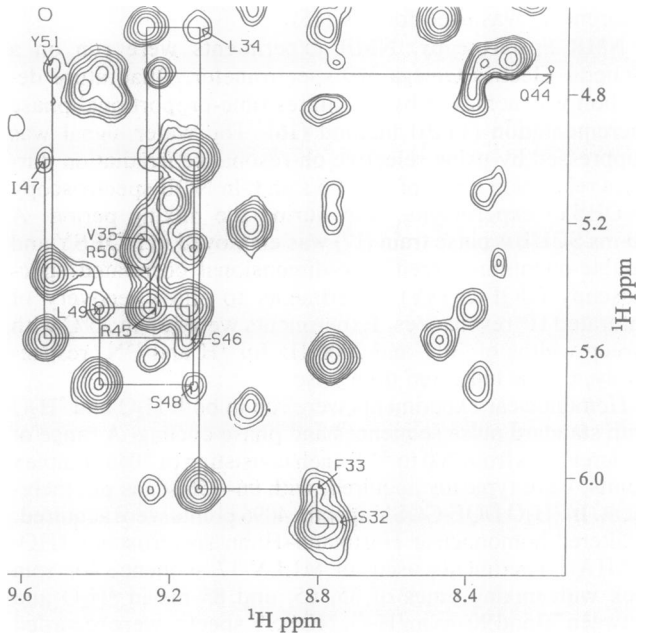

FIG. 1. Fingerprint region of a NOESY (τ_m = 80 ms) spectrum of 4 mM c-Abl SH2 domain in ^1H_2O buffer at 30°C. Sequential $d_{\alpha N}(i, i + 1)$ connectivities for residues 32–35 and 44–51 are shown with intraresidue cross peaks labeled. Overlapping connectivities for Val^{35} and Arg^{36} were resolved in three-dimensional HMQC-NOESY spectra (data not shown). The Gln^{44} H^α resonance intensity is diminished due to overlap with the water resonance.

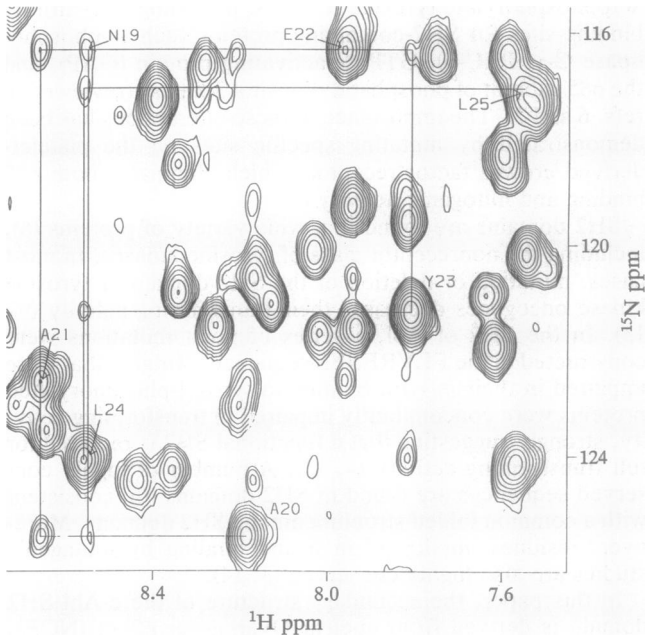

FIG. 2. Portion of two-dimensional HMQC-NOESY (τ_m = 120 ms) spectrum of 2.5 mM uniformly ^{15}N-labeled c-Abl SH2 domain in ^1H_2O buffer. Sequential $d_{NN}(i, i + 1)$ NOEs indicated for the residues 19–25 connect the intraresidue amide cross peaks that are labeled. The ^{15}N resonances of Asn^{19} and Gln^{23} overlap, and the intraresidue amide cross peaks of Tyr^{23} and Leu^{25} are not fully resolved from those of Arg^{54} and Ser^{10}, respectively.

spectra, although amide correlations of several spin systems with H^{α} protons near the water resonance could be observed only at 30°C. Line broadening due to chemical exchange of the amides at neutral pH precluded the identification of amide protons of several residues in the amino and carboxyl termini and the loop regions. From HMQC-*J* spectra 52 $^{15}N'$–H^{α} correlations were observed that confirmed many already established assignments and also provided a number of new connections to amides. Finally, in the three-dimensional HMQC-HOHAHA spectra the connectivities from overlapped amide protons were resolved into separate planes based on dispersed ^{15}N chemical shifts.

Sequential Resonance Assignments. Sequence-specific assignments were obtained from NOEs to neighboring residues by standard procedures (22). Two unique residues, Trp^{11} and Gln^{44}, and numerous unique dipeptide fragments served as seeds for the sequential assignment procedure. The majority of strong sequential NOEs were $d_{\alpha N}(i, i + 1)$, such as those depicted in Fig. 1 connecting β-strand residues. Other residues were sequentially assigned by $d_{NN}(i, i + 1)$ NOEs as illustrated for a helical segment in Fig. 2. Breaks in the sequential assignment occurred occasionally where weak amide proton intensity prevented the observation of sequential NOEs. Ambiguities resulting from overlapping H^{α} chemical shifts were resolved on the basis of the compatibility of the spin system to which the H^{α} belonged, on the presence of other $(i, i + 1)$ NOEs or by medium-range NOEs.

The amide protons of 25 residues exhibited slow exchange rates, which indicates that they are either buried or involved in hydrogen bonds. All of these residues are located in regions of defined secondary structure, with the slowest exchange rates being found in the central residues of each strand and the second helix.

Secondary Structure. Elements of secondary structure were defined by several NMR criteria as summarized in Fig. 3. Eight β-strands were predicted by strong $d_{\alpha N}(i, i + 1)$ NOEs and large $^3J(H'–H^{\alpha})$ values. Helical regions were identified by coincidence of strong $d_{NN}(i, i + 1)$ NOEs, by NOEs to hydrogens three and four residues distant, and by smaller $^3J(H'–H^{\alpha})$ values.

Two antiparallel β-sheets were assembled from six β-strands by using long-range NOE information as depicted in Fig. 4. Hydrogen bonds are illustrated where the amide hydrogen exchanges slowly and where there is at least one adjacent intrastrand NOE. A total of 12 hydrogen bonds stabilize the interaction of the three strands—βI(32–38), βII(44–51), and βIII(54–60)—of the first sheet. A second, smaller β-sheet, assembled from the three strands βIV(61–63), βV(67–70), and βVI(74–77), is apparently less stable due to its smaller number of intrastrand hydrogen bonds. The slowest exchange and highest density of NOEs are in the interaction of βI, βII, and βIII. A clear irregularity in the extended conformation between the βIII and βIV strands is evident from the pattern of sequential NOEs and the small

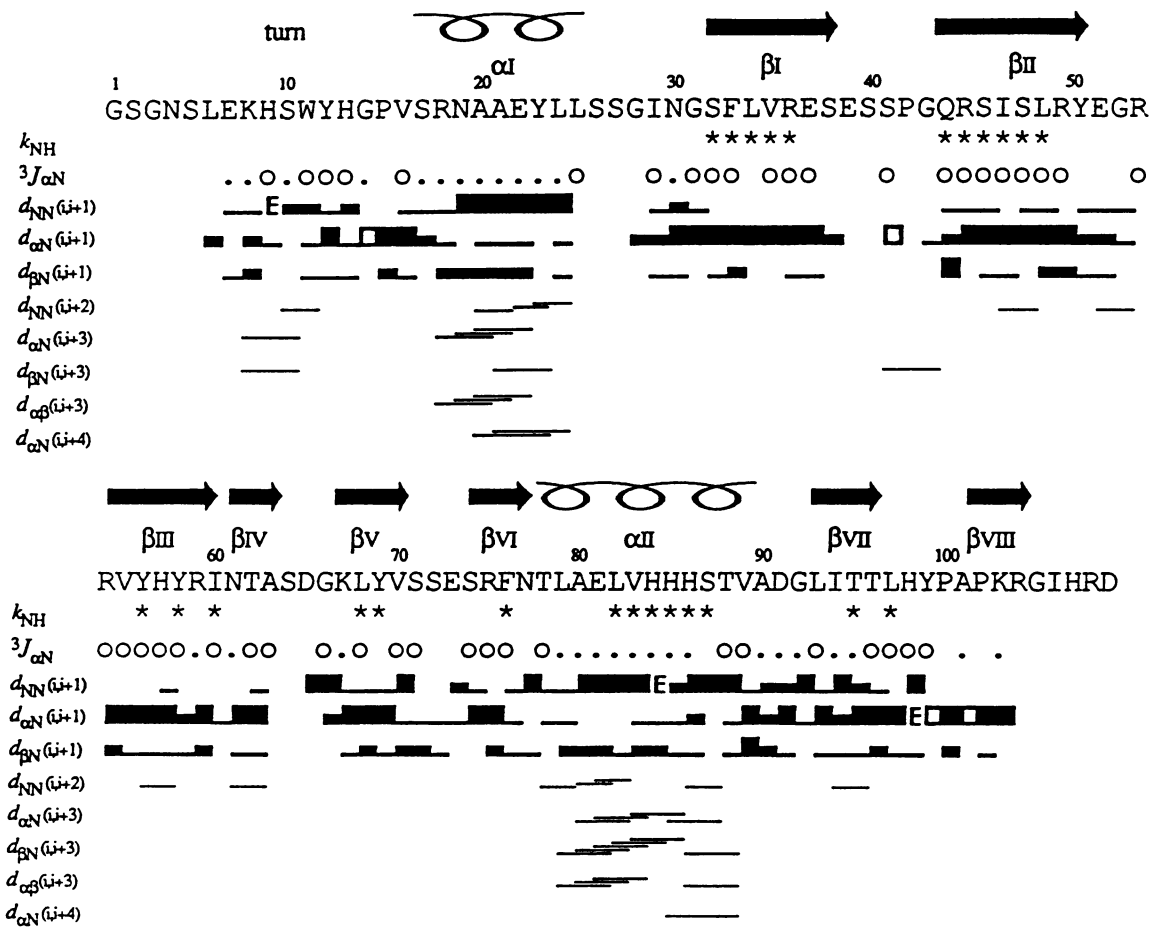

FIG. 3. Amino acid sequence and summary of amide proton exchange rates (k_{NH}), vicinal coupling constants between H^{α} and H' ($^3J_{\alpha N}$), and short- and medium-range NOEs. Below the sequence, stars indicate slow amide proton exchange rates. For coupling constants, a dot indicates an observed coupling < 6 Hz, and a circle indicates an observed coupling ≥ 7 Hz. For $(i, i + 1)$ NOEs between adjacent residues, thin, medium, and thick bars represent weak, moderate, or strong intensity. Open squares indicate $d_{\alpha\delta}(i, i + 1)$ NOEs for proline residues. The letter E represents NOEs that might be expected but were not observed, due to chemical shift overlap between the partners or with the water resonance.

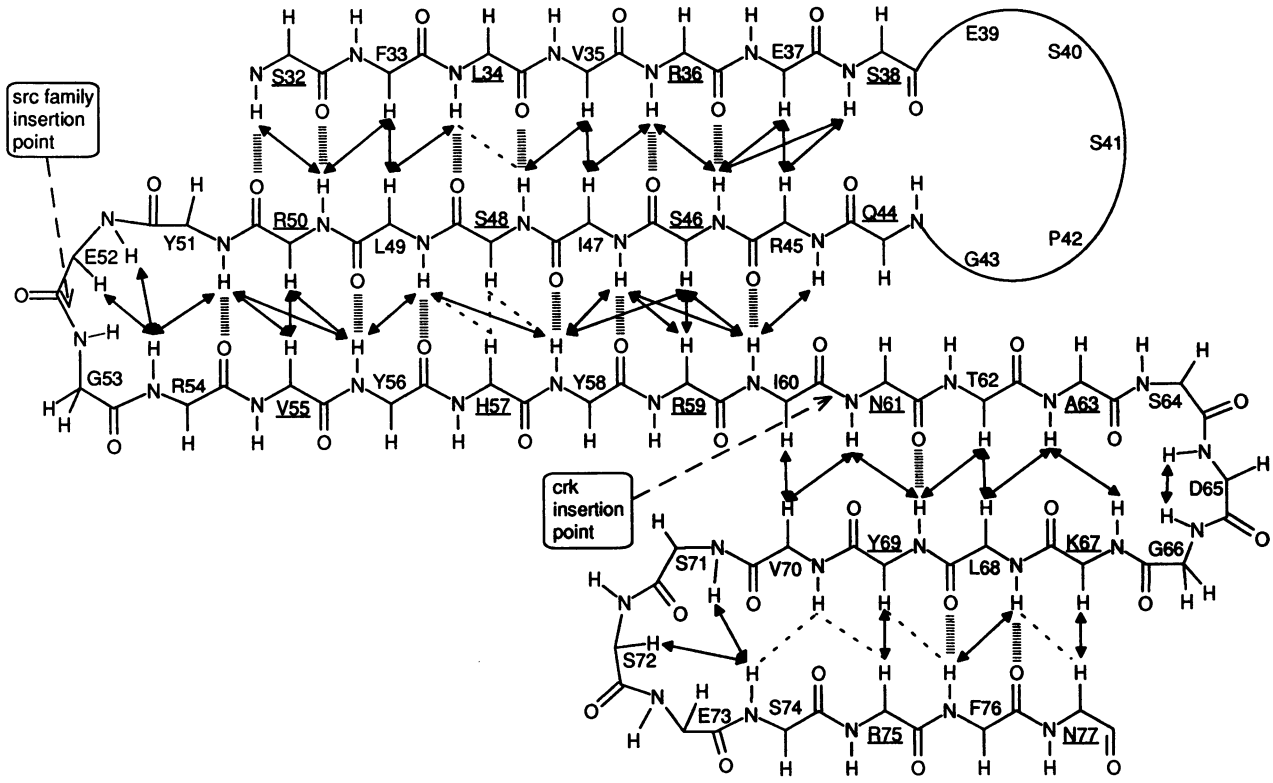

FIG. 4. Schematic diagram of the antiparallel β-sheet region of the c-Abl SH2 domain. Solid arrows indicate observed long-range NOEs. Dashed lines represent NOEs that were expected but not unequivocally identified, due to resonance overlap or line broadening due to chemical exchange. Hydrogen bonds from slowly exchanging amide protons that are supported by at least one interstrand NOE are indicated by hashed lines. The side chains of underlined residues project up from the page and contribute to the more polar face of the β-sheets.

$^3J(H'-H^\alpha)$ of Arg[59] and Asn[61] (see Fig. 3). All the β-strands within each of the two β-sheets are connected by +1 turns—that is, without crossovers. Strands βVII and βVIII could not be positioned unambiguously at this level of analysis.

The β-sheet region shows one face that is predominantly hydrophilic (residues underlined in Fig. 4, with side chains projecting out of the page), whereas the other face is hydrophobic. Central strands of each sheet (i.e., βII and βV) in particular display a pattern of alternating polar and apolar residues. Exceptions to this amphiphilicity are Leu[34], Arg[45], Arg[54], and Val[55]. The βI strand encompasses part of the FLVRESES motif. The five ring hydrogens of Phe[33] are dispersed due to slow rotation and show many long-range NOEs, as does Val[35], suggesting that these residues project into the hydrophobic core. The intervening residues, Leu[34] and Arg[36], on the other hand, display weak correlations between H^α and distal side-chain protons. The solvent-exposed position suggested by these latter two weakly coupled spin systems is consistent with their inferred role in ligand binding (13).

Two α-helices are found for the sequences Ser[17]–Leu[25] and Leu[78]–Val[89] and are indicated as αI and αII (Fig. 3), respectively. A network of hydrogen bonds implied by six slowly exchanging amides stabilizes αII. In addition, the slowly exchanging NH^δ of His[86] indicates that the polar face of αII is buried. A turn between residues Lys[8] and Tyr[12] is suggested by several medium-range NOEs.

Alignment of Structural Elements with Other SH2 Sequences. A group of SH2 domains (see ref. 6) are aligned with the Abl sequence on the basis of conserved secondary structure in Fig. 5. Particular emphasis has been placed on the preservation of glycine and proline residues at turns (23) and to the retention of the amphiphilic character of β-strands and α-helices. Deletions and insertions are placed in loop and turn regions.

Two classes of junction between βII and βIII are evident in Fig. 5. Most SH2 domains are similar to the Abl sequence, with a tight turn formed at Glu[52]-Gly[53]. In contrast, SH2 domains from the Src family have a hydrophilic insertion while maintaining a carboxyl-terminal Gly. The larger size of this turn in Src suggests the name "Src family insertion point." The point of juncture between the two β-sheets is conserved except in the cases of Crk and the amino-terminal SH2 domains of PLC-γ1 and PLC-γ2. Greater sequence variability is found in the second β-sheet, especially in βIV, although the amphiphilicity and positions of turns are generally conserved.

The amphiphilic nature of the α-helices is also maintained. The α-helix spanning Ser[17]–Leu[25] in Abl contains conserved leucine and alanine residues. Likewise, a hydrophobic side of αII consisting of residues Leu[79], Leu[82], Val[83], His[85], and His[86] in Abl is conserved. A common tertiary structure among the SH2 domain family is supported by the high degree of conservation of the composition, length, and spacing of the secondary structure elements.

Previous mutational studies have established the significance of the FLVRESES site for phosphotyrosyl binding (13). The disruption of binding by conservative mutations of the Arg[36], Ser[38], and Ser[40] residues of this motif can now be seen to be consistent with their apparently exposed positions by the βI strand. The proximity of Arg[59] and the conserved His[57] (3) on the same face of the large β-sheet suggests their involvement in ligand binding. Further details of the phosphotyrosyl binding site and of control of peptide specificity may be expected to result from additional structural studies of complexes between SH2 domains and their peptide ligands. The determination of the secondary structure of the c-Abl SH2 domain, with the demonstration of its conservation, provides a basis for further mutagenic studies. Studies

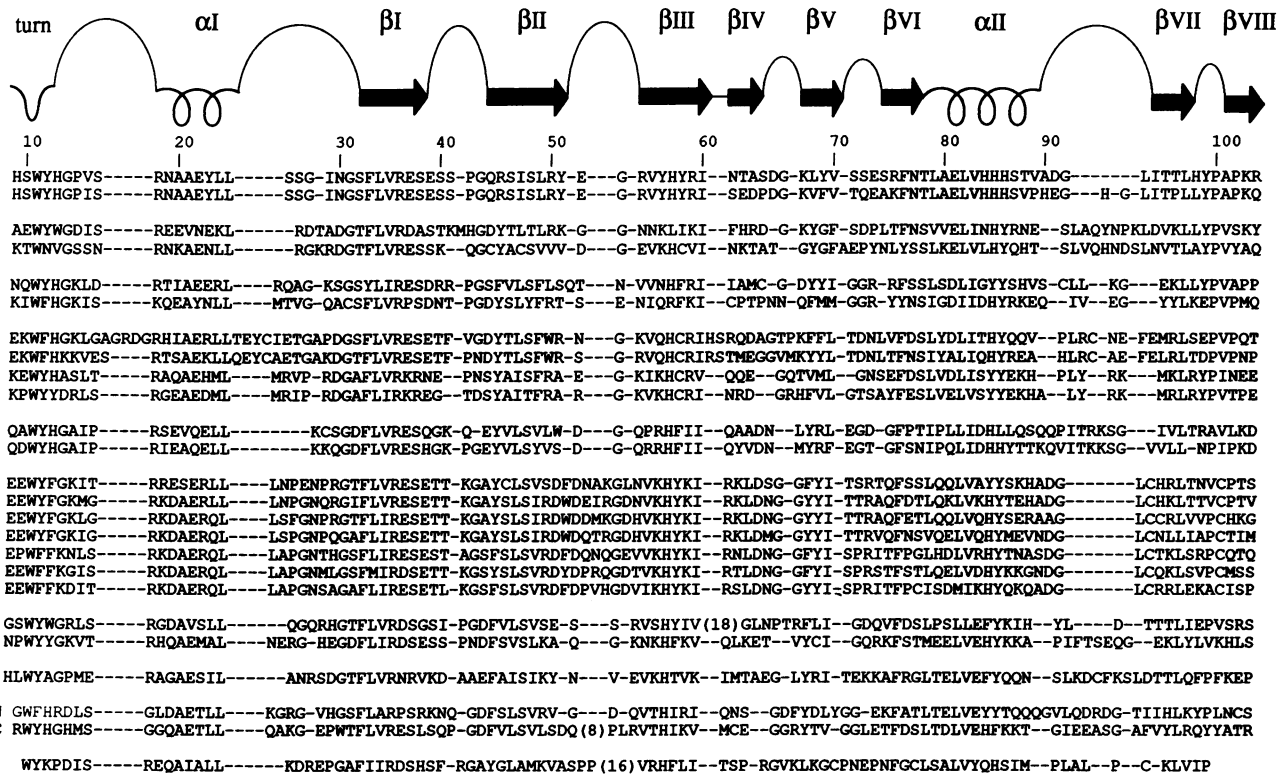

FIG. 5. Alignment of SH2 sequences based on derived secondary structure. Sequences were from previous tabulations (6) or, for protein-tyrosine phosphatase (PTP), from ref. 24. Amino acid residues are numbered relative to the c-abl SH2 domain sequence. The N and C suffixes indicate amino- and carboxyl-terminal SH2 domains of those proteins containing two domains. The prefix D denotes fruit fly sequences. Segments were removed from Crk, tensin, and PTP1C-N sequences for clarity.

on the tertiary structure should further illuminate the function of this domain.

Note Added in Proof. The tertiary structure of Abl SH2 has been determined (25).

We are grateful to Prof. John Kuriyan, Dr. Nalin Pant, and Dr. Yuying Gosser for discussion. This work was supported by grants (D.B., CA51462; D.C., DK20357 and GM47021) and fellowships (B.M., CA0887501; C.B.R., GM14313) from the National Institutes of Health. NMR resources were purchased with grants from the National Institutes of Health, the Keck Foundation, and the National Science Foundation.

1. Anderson, D., Koch, C. A., Grey, L., Ellis, C., Moran, M. F. & Pawson, T. (1990) *Science* **250**, 979–982.
2. Margolis, B., Li, N., Koch, A., Mohammadi, D. R., Hurwitz, A., Zilberstein, A., Ullrich, T., Pawson, T. & Schlessinger, J. (1990) *EMBO J.* **9**, 4375–4380.
3. Matsuda, M., Mayer, B. J. & Hanafusa, H. (1991) *Mol. Cell. Biol.* **11**, 1607–1613.
4. Mayer, B. J., Jackson, P. K. & Baltimore, D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 627–631.
5. Moran, M. F., Koch, C. A., Anderson, D., Ellis, C., England, L., Martin, G. S. & Pawson, T. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8622–8626.
6. Koch, C. A., Anderson, D., Morgan, M. F., Ellis, C. & Pawson, T. (1991) *Science* **252**, 668–674.
7. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R. & Soltoff, S. (1991) *Cell* **64**, 281–302.
8. Fantl, W. J., Escobedo, J. A., Martin, G. A., Turck, C. W., Rosario, M., McCormick, F. & Williams, L. T. (1992) *Cell* **69**, 413–423.
9. De Clue, J. E. & Martin, G. S. (1989) *J. Virol.* **63**, 542–554.
10. Hirai, H. & Varmus, H. E. (1990) *Mol. Cell. Biol.* **10**, 1307–1318.
11. Koch, C. A., Moran, M., Sadowski, I. & Pawson, T. (1989) *Mol. Cell. Biol.* **9**, 4131–4140.
12. Wang, H. C. R. & Parsons, J. T. (1989) *J. Virol.* **63**, 291–302.
13. Mayer, B. J., Jackson, P. K., van Etten, R. A. & Baltimore, D. (1992) *Mol. Cell. Biol.* **12**, 609–618.
14. Kashishian, A., Kazlauskas, A. & Cooper, J. A. (1992) *EMBO J.* **11**, 1373–1382.
15. Live, D. H., Davis, D. G., Agosta, W. C. & Cowburn, D. (1984) *J. Am. Chem. Soc.* **106**, 1939–1941.
16. Marion, D., Ikura, M., Tschudin, R. & Bax, A. (1989) *J. Magn. Reson.* **85**, 393–399.
17. Brown, S. C., Weber, P. L. & Mueller, L. (1988) *J. Magn. Reson.* **77**, 166–169.
18. Gronenborn, A. M., Bax, A., Wingfield, P. T. & Clore, G. M. (1989) *FEBS Lett.* **243**, 93–99.
19. Kay, L. E. & Bax, A. (1990) *J. Magn. Reson.* **86**, 110–126.
20. Forman-Kay, J. D., Gronenborn, A. M., Kay, L. E., Wingfield, P. T. & Clore, G. M. (1990) *Biochemistry* **29**, 1566–1572.
21. Cowburn, D., Glushka, J., DiGennaro, F. & Rios, C. B. (1991) in *Computational Aspects of the Study of Biological Macromolecules by NMR*, ed. Hoch, J. C. (Plenum, New York).
22. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
23. Chothia, C. & Lesk, A. M. (1987) *J. Mol. Biol.* **196**, 901–917.
24. Shen, S.-H., Bastien, L., Posner, B. I. & Chrétien, P. (1991) *Nature (London)* **352**, 735–739.
25. Overduin, M., Rios, C. B., Mayer, B. J., Baltimore, D. & Cowburn, D. (1992) *Cell* **70**, 697–704.