Purification and structural characterization of the CD11b/CD18 integrin α subunit I domain reveals a folded conformation in solution

Michael B. Fairbanks, John R. Pollock, Mark D. Prairie, Terrence A. Scahill, Lubo Baczynskyj, Robert L. Heinrikson, Brian J. Stockman*

Upjohn Laboratories, The Upjohn Company, 301 Henrietta Street, Kalamazoo, MI 49007, USA

Received 22 May 1995

Abstract **The α subunits of the leukocyte CD11/CD18 integrins contain a ~200 amino acid 'inserted' or I domain. The I domain of the cell-surface Mac-1 (CD11b/CD18) integrin has been shown to be the major recognition site for several adhesion ligands, including iC3b, fibrinogen, factor X, and ICAM-1. The I domain from the Mac-1 α subunit has been expressed in *Escherichia coli* as a soluble GST-fusion protein containing a factor X^a sensitive cleavage site. Analytical characterization of the purified I domain reveals that it is obtained in very high quality at high yields. CD and NMR spectra indicate that I domain adopts a predominantly folded structure in solution, independent of the remainder of the α subunit. Addition of Ca^{2+} and Mg^{2+} did not significantly perturb the structural conformation.**

Key words: Adhesion biology; Circular dichroism; Fusion protein; I domain; Integrin; Protein NMR

1. Introduction

Cell adhesion molecules play a role in a wide variety of cellular processes, including development, metastasis, hemostasis, and inflammation [1]. Understanding how these molecules function at the molecular level in these processes may have therapeutic value [2]. Leukocyte integrins, or β_2-integrins constitute a subfamily of these adhesion molecules which are distributed exclusively on the surface of leukocytes. The leukocyte integrins, LFA-1 ($\alpha_L\beta_2$, CD11a/CD18), Mac-1 ($\alpha_M\beta_2$, CR3, CD11b/CD18), p150,95 ($\alpha_X\beta_2$, CD11c/CD18) and CD11d/CD18 are closely related cell surface receptors that mediate neutrophil attachment to the endothelium [3–6]. All four glycoproteins are $\alpha\beta$ heterodimers consisting of a common β subunit (CD18, M_r 95,000) non-covalently associated with a unique but structurally homologous α subunit (CD11a, M_r 175,000; CD11b, M_r 160,000; CD11c, M_r 155,000; CD11d, M_r 150,000) [7–12]. At least 14 distinct α and 8 distinct β integrin subunits have been identified [4], but only the α subunits of VLA-1 ($\alpha_1\beta_1$) [13], and VLA-2 ($\alpha_2\beta_1$) [14], and the four leukocyte integrins contain a domain of ca. 200 amino acids termed the I (inserted, interactive) domain [15,16]. The I domain shows sequence similarities to domains of von Willebrand factor, cartilage matrix protein, and complement factor B [9,10,17].

Of the four known members of the β_2-integrins, Mac-1 has been shown to be a receptor for a variety of proteins associated with cell surfaces and the extracellular matrix. Several lines of evidence suggest that the I domain plays an integral role in Mac-1 function. Diamond et al. [18] showed that a number of monoclonal antibodies (mAbs) that block Mac-1 dependent adhesion map to the I domain. At about the same time, Michishita et al. [19] found that point mutations in the I domain region of recombinant Mac-1 abolished binding to its iC3b ligand. A recombinant form of the I domain of Mac-1 was shown to bind to functionally blocking mAbs, as well as to fibrinogen and soluble ICAM-1, other known ligands of Mac-1 [20]. Finally, the I domain has been identified as the site on Mac-1 for binding of a hookworm-derived inhibitor of neutrophil adhesion called NIF (21).

Understanding the molecular basis for the I domain/ligand interactions is of critical importance to the design of small molecules that may alter these interactions. The present report describes the purification of unlabeled and ^{15}N-enriched I domain from a glutathione *S*-transferase fusion construct expressed in *Escherichia coli*, and its subsequent structural characterization by circular dichroism (CD) and NMR spectroscopy to show that it adopts a predominantly folded structure in solution. It should, therefore, be feasible to obtain structural insights regarding I domain recognition without the necessity to determine the structure of the entire 250 kDa $\alpha\beta$ heterodimer.

2. Materials and methods

2.1. Expression and purification of GST/I domain fusion proteins

The Mac-1 α subunit is a transmembrane protein, with the I domain inserted toward the N-terminus of the protein. The I domain was recombinantly expressed in *E. coli* as a fusion protein to glutathione *S*-transferase (GST) as shown in Fig. 1. The pGEX-3X vector was used to construct a plasmid containing GST followed by a factor X^a sensitive site and finally residues Ser^{133}–Glu^{337} of CD11b which correspond to the I domain [22]. Cells expressing the fusion protein were harvested from 3.0–4.5 liters of fermentation broth and lysed by sonication. The cell lysate was clarified by centrifugation at 40,000 × *g* followed by filtration through a 0.45 μm membrane. The filtrate was then loaded directly onto a glutathione Sepharose (Pharmacia) affinity column (50 ml; 1.6 × 26 cm) pre-equilibrated with PBS/0.2% β-octyl glucoside, pH 7.4. The column was loaded at 1.0 ml/min and protein detection was monitored at 280 nm. The column was washed with 4 column volumes (CV) of equilibration buffer and step-eluted with 2 CV of 50 mM Tris buffer, pH 8.0, containing 20 mM reduced glutathione. Eluent fractions were subjected to 10–20% SDS-PAGE under non-reducing conditions.

*Corresponding author. Fax: (1) (616) 385-7522.

Abbreviations: CD, circular dichroism; CHAPS, 3-[(3-cholamidopropyl)-dimethyl-ammonio]-1-propane-sulfonate; GARP, globally optimized alternating phase rectangular pulse; GST, glutathione *S*-transferase; HSQC, heteronuclear single-quantum correlation; ICAM, intercellular adhesion molecule; LFA, lymphocyte function-associated antigen; NMR, nuclear magnetic resonance; PAGE, polyacrylamide gel electrophoresis; PBS, phosphate buffered saline; SDS, sodium dodecyl sulfate; Tris, tris(hydroxymethyl)aminomethane; VLA, very late antigen.

SSDI 0014-5793(95)00747-4

Fractions containing GST/I domain were pooled and stored at 4°C. The purity of the fusion protein thus obtained is approximately 95%, with the remainder of the protein appearing to be free-GST. Liberation of the I domain was achieved enzymatically by directly adding factor X^a (Boehringer-Mannheim; sequencing grade) to a final concentration of 0.5% (by weight) and allowing the digestion to proceed for 16–24 h at 23°C. I domain was then purified by directly loading the digestion mixture onto an S-Sepharose Fast Flow cation exchange column (50 ml; 1.6 × 26 cm) pre-equilibrated with 50 mM sodium phosphate buffer, pH 6.5. The column was run at 1.0 ml/min and the effluent monitored at 280 nm (Fig. 2). The column was washed with 4 CV of equilibration buffer, followed by a linear gradient of increasing NaCl from 0 to 0.1 M developed over 50 min. I domain was then eluted with a 2 CV linear salt gradient run from 0.1–1.0 M NaCl in the presence of 50 mM sodium phosphate, pH 6.5. Column fractions were collected and analyzed by 10–20% SDS-PAGE; I domain was observed as a single band at approximately 25 kDa (Fig. 3). Purified I domain was concentrated to 20 mg/ml using an Amicon stirred-cell ultrafiltration module containing a YM05 membrane. The same purification procedure was used for metabolically labeled [^{15}N]I domain.

2.2. Electrospray ionization mass spectrometry

All electrospray ionization mass spectra (ESIMS) were recorded on the Vestec 201A mass spectrometer. Aliquots (0.05–0.5 μg) of a solution of the I domain in 50% acetonitrile/50% water/0.1%TFA were injected via a loop injector into the ion source. The mass spectrometer was scanned from m/z 500 to 2000 at 2 s/scan. The data were acquired with the Teknivent Vector 2 data system. Ten scans were averaged and transferred to the Harris 800 computer for further processing. The average molecular weights (av. MW) were determined using the centroid and deconvolution programs.

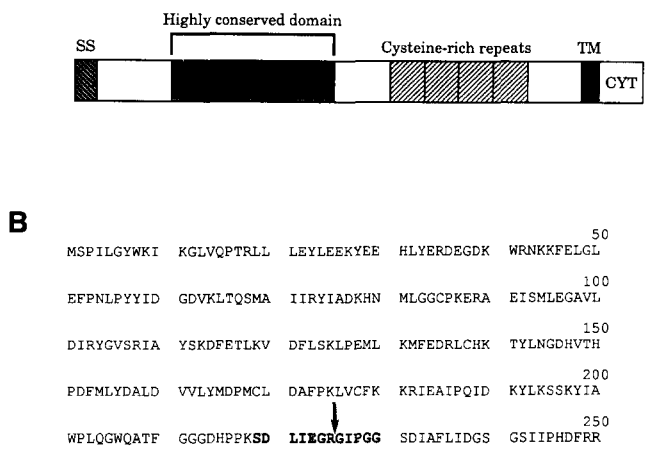

Fig. 1. A schematic diagram of the leukocyte integrin Mac-1 β_2 α-subunit (CD11b) with the amino acid sequence of the GST/I domain fusion protein. (A) The I domain resides in the N-terminal region of all leukocyte integrins. The signal sequence, transmembrane, and cytoplasmic domains are denoted as SS, TM, and CYT, respectively. (B) The GST/I domain fusion construct consists of 435 amino acids. The N-terminal 218 amino acids comprise GST, followed by a 12 amino acid linker region (bold) designed to allow easy access of factor X^a, and then the 205 residue I domain (Ser^{133}–Glu337). The ⇓ denotes the specific cleavage site for factor X^a, leaving the I domain with a 5 amino acid N-terminal extension.

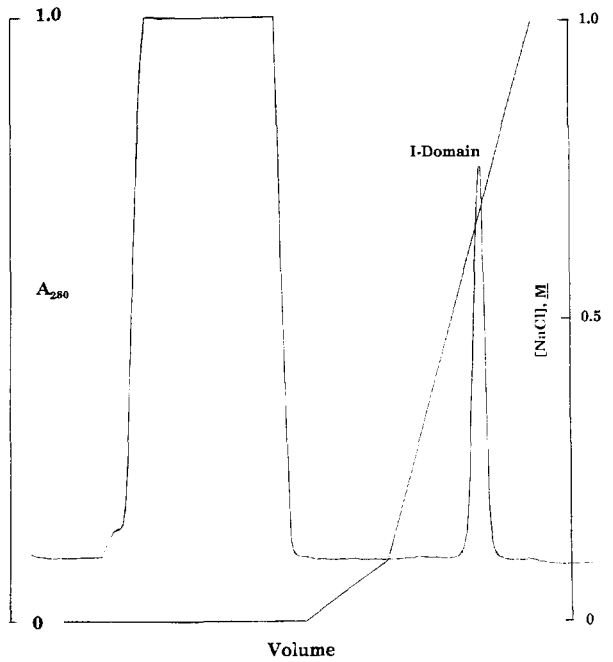

Fig. 2. Purification of I domain by ion exchange chromatography. Factor X^a digested GST/I domain was loaded directly onto a column of S-Sepharose Fast Flow. Chromatographic conditions were optimized to allow for complete isolation of the I domain. The relatively low absorbance of the I domain (as compared to the liberated GST in the column flowthrough) is due primarily to the absence of tryptophan residues.

2.3. NMR spectroscopy

Samples for NMR analysis contained 0.8 mM I domain, 150 mM NaCl and 50 mM phosphate buffer at pH 6.2. The total volume was 440 μl (400 μl ^1H_2O and 40 μl ^2H_2O). NMR spectra were recorded at 27°C on a Bruker AMX-600 spectrometer. In all experiments, continuous wave low-power saturation was used during the relaxation delay to attenuate the intensity of the ^1H_2O resonance. For experiments on [^{15}N]I domain, GARP decoupling [23] was used during the acquisition period to decouple ^{15}N. Proton chemical shifts were referenced to the ^1H_2O signal at 4.76 ppm. Nitrogen chemical shifts were referenced to external 2.9 M $^{15}NH_4Cl$ in 1 M HCl at 24.93 ppm relative to liquid ammonia. Data were processed on a Silicon Graphics Iris Crimson workstation using the software package FELIX from Hare Research Inc.

The two-dimensional ^1H–^{15}N HSQC spectrum [24] was acquired using ^1H and ^{15}N sweep widths of 9,090 Hz and 1,785 Hz, respectively. For each of 128 complex t_1 values, 128 transients were recorded. Quadrature in t_1 was accomplished using the method of States et al. [25].

2.4. Circular dichroism measurements

The CD spectrum of I domain was measured between 20 and 22°C on a Jasco Model J-720 CD spectropolarimeter from 260 to 190 nm in a 0.086 mm cell. Spectra are an average of 16 scans with baseline subtraction. The spectrophotometer was calibrated at 290 nm with d-10-camphorsulfonic acid [26]. The concentration of the protein was 42 μM, 8 mM NaCl and 3 mM phosphate buffer at pH 6.5, in the absence or presence of 10 mM $CaCl_2$ or 10 mM $MgCl_2$. Addition of calcium to the protein formed a turbid solution which was clarified by centrifugation.

Molar intensities were computed from the given concentration and mean residue molecular weight of 113.2 g/mol. The secondary structure was calculated using the method of Compton and Johnson [27] and their data base of 16 proteins. This calculation determines the percentage of the five secondary structures: α-helix, anti-parallel β-sheet, parallel β-sheet, β-turn, and other (random). An acceptable analysis contains no large negative values and a total secondary structure percent-

age of approximately 100%. The 'variable selection method' of Manavalan and Johnson [28] was also applied, which allows for a decrease in the number of variables by removal of proteins with CD contributions absent in the protein being analyzed, until an acceptable percentage is obtained.

3. Results and discussion

The *E. coli* expression system with plasmid pG-3x-CD11b provides reasonable levels of GST/I domain fusion protein. Purification of the GST/I domain construct over immobilized GSH proved to be very straightforward; good yields were observed with very high purities. In fact, the only contaminant observed upon SDS-PAGE analysis was a low level (<5%) of free-GST, which would be expected to co-purify with the fusion protein. N-Terminal sequence analysis of the fusion protein pool yielded a single sequence corresponding to the N-terminus of GST. The absence of secondary sequences was good evidence that there was little or no proteolytic degradation of the fusion protein occurring during the initial purification step.

Liberation of the I domain from its fusion partner was achieved enzymatically via incubation with factor X^a. Factor X^a is a serine protease which, by design of the construct, should specifically cleave the fusion protein at the Arg-Gly bond leading into N-terminal extension of the I domain (Fig. 1). Factor X^a was added directly to the purified fusion protein to a final concentration of 0.5% (by weight), and incubation was carried out at room temperature for 16–24 h. SDS-PAGE provided a useful technique for monitoring the completeness of digestion (Fig. 3, lane 3,4).

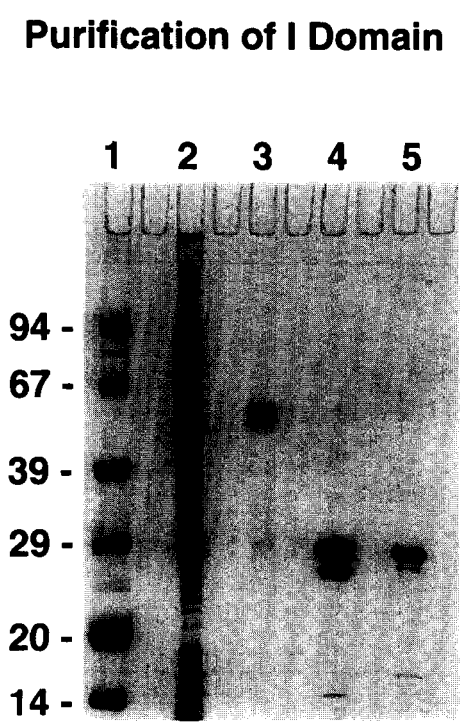

Fig. 3. Coomassie blue stain of a 10–20% SDS-PAGE gel monitoring the purification of recombinant CD11b I domain. Lane 1 = molecular weight markers × 10^3; lane 2 = the *E. coli* lysate containing the GST/I domain; lane 3 = purified GST/I domain; lane 4 = Factor X^a digestion of GST/I domain; lane 5 = purified I domain (S-Sepharose pool).

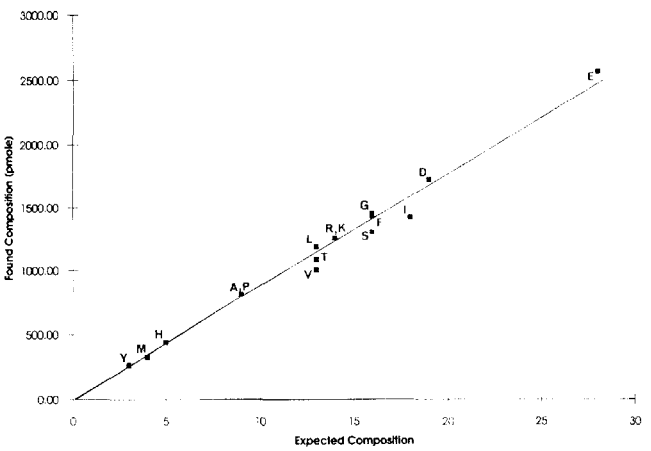

Fig. 4. Amino acid compositional analysis of purified I domain. I domain preparations were hydrolyzed with 6 N HCl for 22 h in vacuo at 110°C prior to analysis. Individual amino acid results (ordinate) were plotted against the theoretical composition of I domain (abscissa). Least squares calculation yielded a correlation of 0.994.

Due to the charge disparity between GST (pI 6.5), factor X^a (pI 7.5), and I domain (pI 9.0), ion exchange chromatography was selected as the final purification step. The digestion mixture was loaded directly onto a column of S-Sepharose Fast Flow under conditions which would allow the GST to flow through unretarded. The elution profile is shown in Fig. 2, where the I domain elutes as a single, homogeneous peak. In fact, the elution of the I domain at around 150 mM NaCl in the gradient results in its recovery in a physiologic-type buffer. The relatively low absorbance at 280 nm of the I domain can be explained by its lack of tryptophan residues ($\varepsilon_{280} = 0.235$ $l \cdot mol^{-1} \cdot cm^{-1}$). Further processing of the purified I domain was limited to conventional dialysis or ultrafiltration to concentrate the protein to desired levels. No differences in expression levels or purification were seen with unlabeled versus ^{15}N-labeled I domain. Typically, cells harvested from 4 liters of fermentation generated 40 mg of final I domain product.

Analytical characterization of the purified I domain revealed that the protein was consistently of very high purity and integrity. N-Terminal sequence analysis confirmed the presence of a single species exactly matching the primary structure of the I domain. The absence of secondary sequences was also evidence that the I domain was not significantly proteolyzed during its purification. This evidence is supported by the single species seen on SDS-PAGE at the expected molecular weight (Fig. 3, lane 4), as well as the data generated by amino acid compositional analysis which reveals a very high correlation to the theoretical, full-length I domain ($r^2 = 0.99$; Fig. 4). Electrospray ionization mass spectrometry was used to further define the I domain in terms of its protein chemistry. Average molecular weights of 23,769 and 24,036 daltons were obtained for unlabeled and ^{15}N-labeled I domain, respectfully, which are within one mass unit of the theoretical masses. Combined, these results show that the I domain is intact at the N- and C-termini, and is essentially free of post-translational modifications (acetylation, deamidation) which could alter the structural and/or functional integrity of this protein.

The conformational properties of purified Mac-1 α subunit I domain were studied by both CD and NMR spectroscopy.

Fig. 5 shows the one-dimensional ^1H NMR spectrum of the I domain. The general resonance dispersion observed, is indicative of folded structure. The isolated methyl resonances at high field (denoted by asterisks) are ring-current shifted more than a ppm from their normal position, indicating that they are situated near an aromatic ring system. This type of interaction is not observed in denatured or random coil proteins, and is thus very indicative of a folded protein. In fact, high-field resonances such as these are excellent markers of protein integrity. If the I domain were to denature with time, temperature, variation in solvent or any other condition, a change in the positions of these resonances would be observed. Other distinct features of the one-dimensional spectrum include the observation of several $^1H^{\alpha}$ resonances with β-sheet-like chemical shifts between 5.0 and 6.0 ppm, and the spread of $^1H^N$ resonances from 8.0 to 10.0 ppm.

Addition of $CaCl_2$ to the I domain solution did not produce obvious changes in the resulting NMR spectra, suggesting the sequestered I domain may not bind calcium or the cation binding-site is occupied. In addition, NMR spectra in the presence of 15 mM CHAPS detergent were identical to those without detergent, evidence that the I domain does not aggregate at concentrations used for NMR studies.

The CD spectra of the Mac-1 α subunit I domain in the absence and presence of Ca^{2+} and Mg^{2+} is shown in Fig. 6. Scans were terminated at 190 nm due to the high salt concentration. The addition of divalent cations yielded no significant conformational change. The secondary structure calculations suggest approximately 45% of the protein is α-helix, 7% β-sheet, 23% β-turn, and 25% other or random in composition.

Extraction of detailed structural information using NMR spectroscopy of proteins this size requires isotopic ^{13}C and/or ^{15}N enrichment [29]. Isotopic labeling increases the likelihood of assigning the majority of the resonances and concomitantly increases the number of structural constraints that can be determined. Fig. 7 displays the two-dimensional ^1H–^{15}N HSQC spectrum of the [^{15}N]I domain. Many well-dispersed correlations

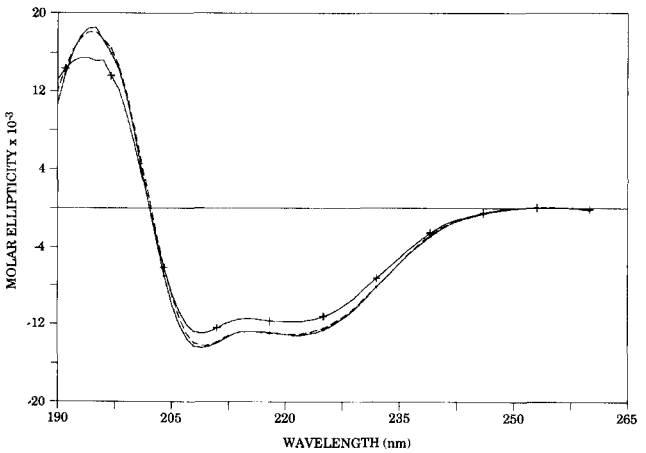

Fig. 6. The CD spectrum of 42 μM CD11b I domain in the absence (solid line) and presence of 10 mM $CaCl_2$ (solid line/cross) or 10 mM $MgCl_2$ (dash line), 8 mM NaCl, 3 mM phosphate buffer, pH 6.5.

are observed, which is again indicative of a folded protein. It is apparent that the correlations documented in Fig. 7 manifest a range of intensities and/or line-widths. This signifies differences in mobility across the structure of the I domain, consistent with the view that some regions exist in a more stable folded structure than others.

In summary, the CD results and preliminary NMR studies show the Mac-1 α subunit I domain acquires a folded conformation in solution. Both CD and NMR evidence suggests that the majority of the I domain secondary structure is α-helical in nature. Furthermore, the NMR spectra suggest that I domain exits as a monomer and does not aggregate in solution. Although previous reports have shown the presence of a novel divalent cation-binding site within the I domain that is required for metal-dependent ligand binding [19,22,30], gross conformational changes were not observed upon addition of Ca^{2+} or Mg^{2+} to I domain, possibly signifying the site is occupied.

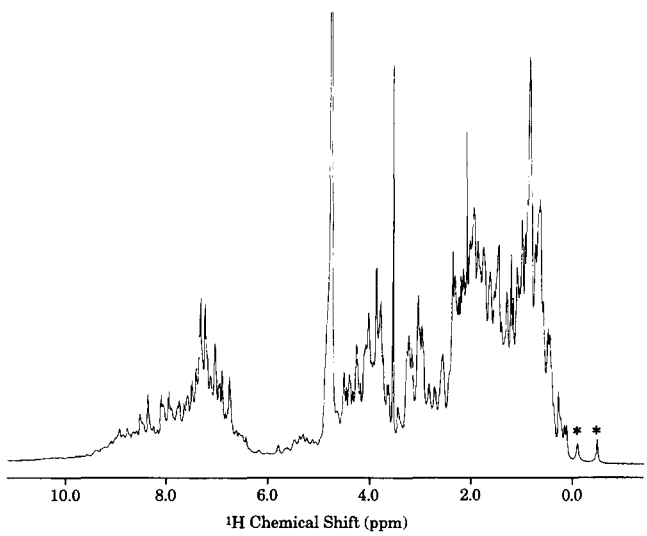

Fig. 5. The 600 MHz ^1H NMR spectrum of 0.8 mM CD11b I domain in 90% ^1H_2O/10% ^2H_2O, 150 mM NaCl, 50 mM phosphate buffer, pH 6.2. The large out-of-phase resonance at 4.76 ppm arises from the water protons. Peaks designated with an asterisk are ring-current shifted from their normal position.

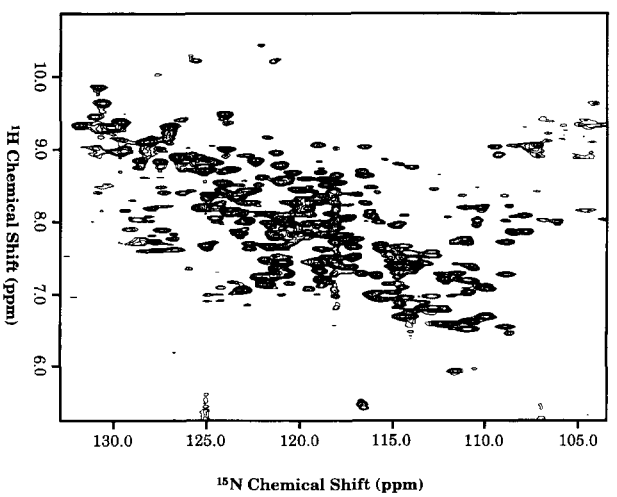

Fig. 7. Region of the 14.2 T ^1H–^{15}N HSQC spectrum of uniformly ^{15}N-enriched CD11b I domain in 90% ^1H_2O/10% ^2H_2O, 150 mM NaCl, 50 mM phosphate buffer, pH 6.2. Observed correlations arise from the backbone amide groups.

Recently it was shown that recombinant Mac-1 I domain contains an iC3b ligand binding site that is divalent cation dependent and temperature independent [31]. Sequencing of short strands of peptides from the I domain has determined that residues 330–343 (Fig. 1) bound iC3b independent of the cation, with iC3b most likely binding to the less well conserved N- and C-terminus hydrophilic regions of the peptide. It was suggested that the I domain assumes a functionally active conformation outside the Mac-1 α subunit [31]. Collectively, our results from CD and NMR support the idea that I domain retains its functionality as an isolated subunit [13,14,18,20]. While this manuscript was in preparation, the crystal structure of the Mac-1 I domain (termed the A domain, and slightly different in amino acid sequence from the one described herein) was reported [32]. It will be of interest now to compare the solution structures of the I domain and I domain/ligand complexes with the structure determined crystallographically.

Acknowledgments: We thank Kimberly A. Curry and Che-Shen C. Tomich for expressing GST/I domain in *E. coli* and providing cell extracts containing unlabeled and isotopically-enriched I domain. We also thank Donald C. Anderson for advice and encouragement.

References

[1] Loftus, J.C., Smith, J.W. and Ginsberg, M.H. (1994) J. Biol. Chem. 269, 25235–25238.
[2] Springer, T.A. (1990) Nature 346, 425–434.
[3] Springer, T.A. (1990) Nature 346, 425–434.
[4] Arnaout, M.A. (1990) Blood 75, 1037–1050.
[5] Arnaout, M.A. (1990) Blood 75, 1037–1050.
[6] Bilsland, C.A., Diamond, M.S. and Springer, T.A. (1994) J. Immunol. 152, 4582–4589.
[7] Corbi, A.L., Kishimoto, T.K., Miller, L.J. and Springer, T.A. (1988) J. Biol. Chem. 263, 12403–12411.
[8] Arnaout, M.A., Gupta, S.K., Pierce, M.W. and Tenen, D.G. (1988) J. Cell. Biol. 106, 2153–2158.
[9] Pytela, R. (1988) EMBO J. 7, 1371–1378.
[10] Larson, R.S., Corbi, A.L., Berman, L. and Springer, T.A. (1989) J. Cell. Biol. 108, 703–712.
[11] Corbi, A.L., Garcia-Aguilar, J. and Springer, T.A. (1990) J. Cell. Biol. 265, 2782–2788.
[12] Sanchez-Madrid, F., Nagy, J.A., Robbins, E., Simon, P. and Springer, T.A. (1983) J. Exp. Med. 158, 1785–1803.
[13] Kern, A., Briesewitz, R., Bank, I. and Marcantonio, E.E. (1994) J. Biol. Chem. 269, 22811–22816.
[14] Kamata, T., Puzon, W. and Takada, Y. (1994) J. Biol. Chem. 269, 9659–9663.
[15] Hemler, M.E. (1990) Annu. Rev. Immunol. 8, 365–400.
[16] Shaw, S.K., Cepak, K.L., Murphy, E.A., Russell, G.J., Brenner, M.B. and Parker, C.M. (1994) J. Biol. Chem. 269, 6016–6025.
[17] Colombatti, A. and Bonaldo, P. (1991) Blood 77, 2305–2315.
[18] Diamond, M.S., Garcia-Aguilar, J., Bickford, J.K., Corbi, A.L. and Springer, T.A. (1993) J. Cell. Biol. 120, 1031–1043.
[19] Michishita, M., Videm, V. and Arnaout, M.A. (1993) Cell 72, 857–867.
[20] Zhou, L., Lee, D.H.S., Plescia, J., Lau, C.Y. and Altieri, D.C. (1994) J. Biol. Chem. 269, 17075–17079.
[21] Muchowski, P.J., Zhang, L., Chang, E.R., Soule, H.R., Plow, E.F., and Moyle, M. (1994) J. Biol. Chem., 269, 26419–26423.
[22] Altieri, D.C. (1991) J. Immunol. 147, 1891–1898.
[23] Shaka, A.J., Barker, P.B. and Freeman, R. (1985) J. Magn. Reson. 64, 547–552.
[24] Bodenhausen, G. and Ruben, D.L. (1980) Chem. Phys. Lett. 69, 185–188.
[25] States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) J. Magn. Reson. 48, 286–292.
[26] Krueger, W.C. and Pschigoda, L.M. (1971) Anal. Chem. 43, 675–677.
[27] Compton, L.A. and Johnson Jr., W.C. (1986) Anal. Biochem. 155, 155–167.
[28] Manavalan, P. and Johnson Jr., W.C. (1987) Anal. Biochem. 167, 76–85.
[29] Clore, G.M. and Gronenborn, A.M. (1991) Science 252, 1390–1399.
[30] Smith, J.W. and Cheresh, D.A. (1991) J. Biol. Chem. 266, 11429–11432.
[31] Ueda, T., Rieu, P., Brayer, J. and Arnaout, M.A. (1994) Proc. Natl. Acad. Sci. USA 91, 10680–10684.
[32] Lee, J.-O., Rieu, P., Arnaout, M.A. and Liddington, R. (1995) Cell, 80, 631–638.