Journal of Biomolecular NMR, **22:** 211–223, 2002.
KLUWER/ESCOM

Solution structure and dynamics of melanoma inhibitory activity protein

Julie C. Lougheed[a], Peter J. Domaille[b] and Tracy M. Handel[a,*]
[a]*Department of Molecular and Cell Biology, 229 Stanley Hall, University of California, Berkeley, CA 94720, U.S.A.;* [b]*Dupont Pharmaceuticals Company, Experimental Station E353/56B, Wilmington, DE 19880-0353, U.S.A.*

Received 9 October 2001; Accepted 5 December 2001

Key words: ambiguous restraints, automated NOE assignments, CD-RAP, crystal structure, MIA, NMR structure, SH3 subdomain

Abstract

Melanoma inhibitory activity (MIA) is a small secreted protein that is implicated in cartilage cell maintenance and melanoma metastasis. It is representative of a recently discovered family of proteins that contain a Src Homologous 3 (SH3) subdomain. While SH3 domains are normally found in intracellular proteins and mediate protein-protein interactions via recognition of polyproline helices, MIA is single-domain extracellular protein, and it probably binds to a different class of ligands.

Here we report the assignments, solution structure, and dynamics of human MIA determined by heteronuclear NMR methods. The structures were calculated in a semi-automated manner without manual assignment of NOE crosspeaks, and have a backbone rmsd of 0.38 Å over the ordered regions of the protein. The structure consists of an SH3-like subdomain with N- and C-terminal extensions of approximately 20 amino acids each that together form a novel fold. The rmsd between the solution structure and our recently reported crystal structure is 0.86 Å over the ordered regions of the backbone, and the main differences are localized to the most dynamic regions of the protein. The similarity between the NMR and crystal structures supports the use of automated NOE assignments and ambiguous restraints to accelerate the calculation of NMR structures.

Introduction

Melanoma inhibitory activity (MIA) is a small secreted protein that is normally produced by cartilage. MIA expression begins at chondrogenesis and continues in mature cartilage tissue, suggesting a fundamental role in cartilage development and maintenance (Bosserhoff et al., 1997b). Since its expression is inhibited by retinoic acid, a repressor of cartilage cell phenotype, MIA is also known as cartilage-derived retinoic acid-sensitive protein (CD-RAP) (Dietz and Sandell, 1996).

MIA is also expressed by malignant melanoma, but generally not by normal melanocytes (Golob et al., 2000). Historically, MIA was identified as a factor secreted from melanoma cells that inhibited their growth *in vitro* (Blesch et al., 1994). However, subsequent studies indicated that it may actually promote melanoma metastasis. One study demonstrated a correlation between increased plasma levels of MIA and a more advanced metastatic disease state in humans (Bosserhoff et al., 1997a). Increased expression of the protein in melanoma cells has also been shown to enhance their metastatic potential when injected into hamsters (Guba et al., 2000).

MIA has been shown to inhibit cell binding to the extracellular matrix, suggesting a mechanism by which it may promote metastasis (Bosserhoff et al., 1998). Since the extracellular matrix affects cell growth, the anti-adhesive properties of MIA may be the source of the observed growth inhibition. MIA has been reported to bind the extracellular matrix protein fibronectin, leading to the hypothesis that direct competition or occlusion of integrin binding sites on fibronectin is the source of the anti-adhesive activity (Stoll et al., 2001).

*To whom correspondence should be addressed: E-mail: handel@paradise1.berkeley.edu

Currently there are two known paralogs of MIA, with a third one predicted from human genomic sequences (Lougheed et al., 2001; Pan, 2000; Robertson et al., 2000). For the mature secreted human forms, the percent identity with MIA ranges from 33% to 45%. All homologs contain a subdomain with low sequence homology to SH3 domains; for example, MIA and its closest SH3 homologue, Vav-3, are 35% identical. SH3 domains are small (∼60 residue) proteins that mediate protein-protein interactions in signal transduction cascades and membrane-cytoskeleton structures through recognition of left-handed polyproline type II (PPII) helices (Musacchio et al., 1994). In contrast, MIA is a 107 residue single-domain extracellular protein (Lougheed et al., 2001). Outside the SH3 homology region, it has short N- and C-terminal extensions of approximately twenty amino acids each. Both the SH3 subdomain and the N- and C-terminal extensions are conserved among the different paralogs.

We recently solved the crystal structure of MIA and confirmed that its subdomain is strikingly similar to canonical SH3 domains (Lougheed et al., 2001). Based on amino acid conservation of MIA homologues mapped onto the structure, we identified two possible ligand interactions sites on opposite sides of the protein. The larger of the two patches is located in a region that corresponds to the polyproline helix binding site on canonical SH3 domains. However, the identity of many of the conserved residues differs from those in canonical SH3 domains, resulting in a much flatter molecular surface that most likely does not recognize polyproline helices.

Here we report the solution structure, which was solved without manual assignment of NOE cross-peaks. The NMR structure was determined independent of the crystal structure, is of high quality, and agrees remarkably well with the crystal structure. The regions that deviate most from the crystal structure are dynamically disordered as indicated by backbone dynamics experiments, and we discuss a possible binding sequence that occurs in one of these regions. A solution structure of MIA at pH 7, solved by a more traditional manual assignment approach, has also been reported by Stoll et. al. (Stoll et al., 2001). However, as the coordinates (1HJD) are currently on hold, comparison of the two sets of NMR structures is not possible.

Materials and methods

Protein expression and purification

A synthetic gene of human MIA was constructed, cloned into the pET3a vector (Novagen), and transformed into BL21(DE3) *E. coli* cells that contained a separate vector (pACYC) for constitutive expression of the lac repressor. The presence of the lac repressor increased the amount of MIA produced, presumably by reducing leaky expression of the protein. Uniformly labeled ^{15}N and $^{15}N/^{13}C$ labeled protein was prepared by growing the cells in MOPS minimal media (Neidhardt et al., 1974) containing ^{13}C-glucose (2 g l^{-1}) and/or ^{15}N-ammonium sulfate (1 g l^{-1}), 100 μg ml^{-1} of ampicillin, and 20 μg ml^{-1} kanamycin. Cells were grown at 37 °C to an OD_{600} of 0.8, at which time protein expression was induced for 4 hours by the addition of 1.0 mM isopropyl β-D thiogalactopyranoside. The protein was purified and refolded from inclusion bodies as previously described (Lougheed et al., 2001). NMR samples (0.8–0.9 mM) were prepared as follows: MIA was concentrated in centricon-10 concentrators (Amicon) and then equilibrated in the sample buffer either by repeated centricon spins or by running the sample through a Quicksep spin column (IsoLab Inc) packed with G-25 resin (Pharmacia) and pre-equilibrated in sample buffer. Sample buffer consisted of 90% H_2O/10% D_2O or 100% D_2O containing 20 mM sodium acetate-d3 pH 4.3, 0.02% sodium azide, and in some cases, 2.0 mM EDTA-d12.

NMR spectroscopy

NMR experiments were carried out on a 600 MHz Bruker DMX spectrometer at 25 °C. Backbone and C_β assignments were obtained from 3D CBCANNH (Grzesiek and Bax, 1992b) and CBCA(CO)NNH (Grzesiek and Bax, 1992a) experiments. Sidechain assignments were made primarily using 4D HCC(CO)NNH (Clowes et al., 1993) and 3D HCCH-TOCSY (Bax et al., 1990) experiments. Aromatic sidechain assignments were derived from a number of experiments: a 2D aromatic version of the constant time ^{13}C-HSQC (Vuister and Bax, 1992) and a relayed version of this experiment, 2D CBHD and CBHE experiments (Yamazaki et al., 1993), a 3D ^{13}C separated NOESY-HSQC centered on the aromatic carbons (D_2O sample) and collected at a shorter mixing time (70 ms) to favor intraresidue NOEs, and a 3D ^1H-TOCSY-relayed ct-[^{13}C, ^1H]-HMQC (Zerbe et al.,

1996) (D_2O sample). Distance restraints were derived from a 3D ^{15}N separated NOESY-HSQC (Talluri and Wagner, 1996), two 3D ^{13}C separated NOESY-HSQC spectra (Zuiderweg et al., 1990) (D_2O sample) with the ^{13}C transmitter centered on either the aliphatic or the aromatic carbons, a 4D ^{13}C-^{13}C separated HMQC-NOESY-HSQC (Clore et al., 1991) (D_2O sample), and a 4D ^{13}C-^{15}N separated HMQC-NOESY-HSQC (Kay et al., 1990). Mixing times ranged from 70 to 150 ms. The number and distribution of NOE derived distance restraints is described in the results section.

A total of thirty-two backbone ϕ angle restraints and twenty-three χ restraints were obtained from quantitative J correlation experiments (Bax et al., 1994). ϕ angle restraints were obtained from the $^3JH_NH_\alpha$ coupling constants determined from a 3D H_NH_α experiment (Kuboniwa et al., 1994). For stereospecific assignment of the β methylene protons and determination of χ_1, the 3D HACAHB (D_2O) (Grzesiek et al., 1995) and 3D HNHB (Archer et al., 1991) were used to get quantitative estimates of $^3J\alpha\beta$ and $^3JH_NH_\beta$, respectively. Stereospecific assignment of valine methyls and χ_1 torsion angle restraints for Ile, Val, and Thr residues were obtained from measurements of $^3JC_\gamma N$ (Vuister et al., 1993) and $^3JC_\gamma C'$ (Grzesiek et al., 1993). Additionally, ^3JCC measurements from a long-range $^{13}C^{13}C$ correlation experiment (Bax et al., 1992) provided one Ile χ_2 restraint.

Hydrogen bond restraints (2 restraints per amide) were derived from 47 slowly exchanging amides, identified from 2D ^1H-^{15}N HSQC experiments immediately following exchange of the protein into D_2O using a Quick-sep spin column (see above). Hydrogen bond acceptors were identified from partially refined NMR structures and corroborated by the crystal structure. In a more automated approach, one could automatically identify potential acceptors within a given radius of hydrogen bond donors, and treat them as ambiguous restraints (Linge et al., 2001). Table 1 summarizes the restraints used in the structure calculations.

Data processing and analysis

All NMR spectra were processed using the program AZARA (Wayne Boucher, unpublished results). The directly-detected dimensions were processed with a conventional Fourier transform. Maximum entropy was generally used to reconstruct the indirect dimensions. Peak picking and sequential assignments were done within the program ANSIG 3.3 (Kraulis et al., 1994). Structure calculations were done in X-PLOR 3.851 (Brunger, 1992) with ARIA (Ambiguous Restraint for Iterative Assignment) extensions (Nilges et al., 1997) as described in the results and discussion.

Dynamics

Backbone dynamics information was derived from the $^1H/^{15}N$ heteronuclear NOE and the ^{15}N T1 and T2 relaxation rates. Peak heights were determined by fitting the crosspeaks to two-dimensional gaussians with the program Priism (Chen et al., 1996). For the heteronuclear NOE, uncertainties in crosspeak heights were taken as the baseline noise levels and used to estimate the uncertainties in the NOE by standard propagation of error. The T1 and T2 values were determined by recording 10 spectra with relaxation delay times ranging from 10.2 ms to 867 ms for T1 and 8 to 160 ms for T2. Crosspeak intensities were fit to a single decaying exponential and the uncertainties were taken as the error of the fit.

The coordinates, assignments, and experimental restraints have been deposited with the Protein Data Bank (Accession Code 1K0X) and BioMagResBank (Accession Code 5220).

Results and discussion

Disulfide bond pattern

Although it contains a subdomain with sequence homology to intracellular SH3 domains, MIA and its homologues are extracellular proteins containing disulfides. Since the disulfide connectivity was not known when the NMR studies were initiated, we analyzed the disulfide pattern by mass spectrometry of proteolyzed fragments. These experiments demonstrated connectivity between Cys13-Cys18 and Cys36-Cys107. The same pattern became evident from the close proximity of the appropriate sulfurs after a few iterations of the structure calculations done without explicit disulfide bonds. Once the fold had been established, covalent restraints between the sulfurs were introduced and did not produce any NOE violations. Thus these restraints did not influence the convergence to the correct structures or the automated assignment of NOE crosspeaks.

Solution conditions

Systematic analysis of ^{15}N-HSQC spectra and heteronuclear T2's indicated that MIA was most soluble

Table 1. **Experimental restraints and structural statistics**

Number of experimental restraints	**First round**	**Final round**
NOE distance restraints[a]		
Ambiguous	4929	905
Unambiguous	358	2300
H Bond[b]	94 (47 H-bonds)	
Backbone ϕ angles from coupling constants	32	
Sidechain χ angles from coupling constants	23	
Average level of ambiguity [c]		
3D ^{13}C (aliphatic) separated NOESY-HSQC	87	
3D ^{13}C (aromatic) separated NOESY-HSQC	150	
3D ^{15}N separated NOESY-HSQC	42	
4D ^{13}C-^{13}C separated HMQC-NOESY-HMQC	22	
4D ^{13}C-^{15}N separated HMQC-NOESY-HSQC	70	
Average ambiguity for combined data sets	53	3
R.m.s. deviations from experimental data	**Final round**	
Average distance restraint violation	0.018 ± 0.001 Å	
Average dihedral angle restraint violation	$0.34 \pm 0.07°$	
R.m.s. deviations from ideal stereochemistry		
Bonds	0.0048 ± 0.0001 Å	
Angles	$0.55 \pm 0.01°$	
Impropers	$0.34 \pm 0.01°$	
Final Energy (kcal mol^{-1})		
Lennard–Jones[d]	-767 ± 7	
Ramachandran statistics[e]		
Residues in most favored regions	81.5%	
Residues in additionally allowed regions	16.4%	
Residues in generously allowed regions	1.0%	
Residues in disallowed regions	1.1%	
Coordinate precision (residues 7–66, 78–87, 96–101)		
Backbone	0.38 ± 0.07 Å	
Heavy atoms	0.87 ± 0.06 Å	

[a]The difference in the total number of NOE restraints at the start and at the end of the calculation is due to the elimination of duplicates during the course of the structure calculations.
[b]For each of the 47 identifiable H-bonds, two restraints were applied: $1.7 \leq D_{H-O} \leq 2.2$ Å and $2.7 \leq D_{N-O} \leq 3.2$ Å.
[c]The average level of ambiguity is the average number of possible assignment pairs for each NOE crosspeak in the initial round of structure calculations. In some sense, these numbers are overestimated in that for degenerate protons belonging to the same heavy atom, each proton is counted as a possible assignment. For example, a restraint between two methyl protons has an average level of ambiguity of nine.
[d]The Lennard–Jones van der Waals energies were not included in the target function during simulated annealing.
[e]Ramachandran statistics were calculated with PROCHECK.

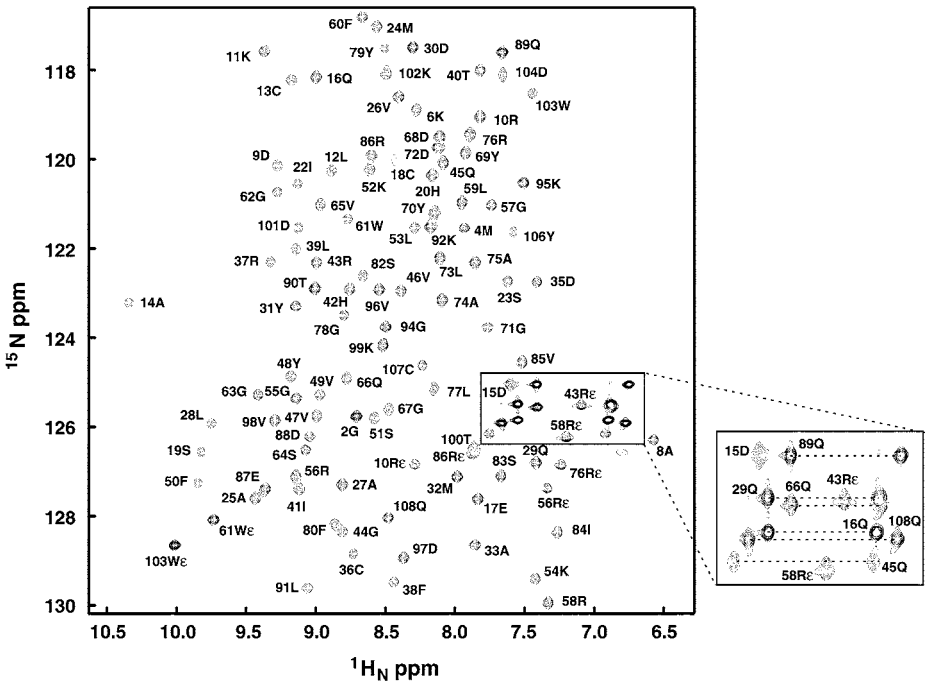

Figure 1. ^{15}N-HSQC of MIA in 20 mM sodium acetate-d3, pH 4.3. Many crosspeaks are aliased in the ^{15}N dimension due to the optimized spectral width of 842 Hz, however they are not indicated for simplicity. The crosspeak for Phe 50 and Leu 7 are not shown since Phe 50 was below the contour level of the figure, and the crosspeak for Leu 7 appeared upfield at a ^{1}H ppm of 4.0. Asn and Gln sidechain NH_2 crosspeaks are expanded in the box.

at pH 4.3 and low salt concentration. Analytical ultracentrifugation showed that MIA is monomeric at pH 4.3, at least up to a concentration of 33 μM. However, the average T2 was shorter than expected for a protein of this size, suggesting a tendency to aggregate. Using the optimized buffer conditions, the concentration of protein was systematically varied and the best linewidth and signal to noise in^{15}N-HSQC spectra was obtained at approximately 0.8–0.9 mM.

Chemical shift assignments

Figure 1 shows the ^{15}N-HSQC of MIA at pH 4.3, the conditions under which the NMR experiments were carried out. The spectra are well dispersed, all amide crosspeaks are visible, and complete backbone assignments were made in a straightforward manner.

The non-aromatic sidechain resonances were assigned primarily from the 3D HCCH TOCSY and 4D HCC(CO)NNH. In cases where crosspeaks were missing from these spectra, assignments were obtained from the 4D ^{13}C-^{15}N separated NOESY, the 3D ^{13}C separated NOESY, or the 4D ^{13}C-^{13}C separated-NOESY. The 4D ^{13}C-^{15}N separated NOESY was particularly helpful in cases where crosspeaks were missing from the 4D HCC(CO)NNH because it enabled assignment of sidechain proton/carbon pairs via NOE crosspeaks to assigned amide protons. In total, all non-aromatic sidechain resonances were assigned except for the following: Lys 11 (CD, HD1, HD2, CE, HE1, HE2), Gln 29 (CG, HG1, HG2), and Arg 37 (NE/HE). For these unassigned nuclei, the random coil chemical shifts (Wishart et al., 1995) were used but a large chemical shift tolerance was applied (0.5 ppm for ^{1}H and 2.0 ppm for ^{13}C) for matching to the NOE crosspeaks (see below).

MIA has a large number of aromatic residues (six tyrosines, five phenylalanines, two tryptophans, and two histidines) which were completely assigned, although not without difficulty. The initial assignments made for Phe 105 (CE/HE and CZ/HZ) and Phe 50 CZ/HZ were not firm due to a combination of overlap and missing through bond correlations. Additionally, the data did not distinguish between Phe 50 CD/HD and CE/HE. We were able to tolerate these inconclusive assignments by applying a large chemical shift uncertainty to the estimated sequential assignments. At the end of structure calculations, NOE restraints supported the assignments for Phe 105

CE/HE, whereas very few restraints included Phe 105 CZ/HZ, and these remained ambiguous. The inconclusive assignments for Phe 50 were tolerated but not resolved by the end of the structure calculations, presumably because of the spatial proximity of the HD and HE protons to each other.

Restraints and structure calculations

Sequential assignments were made using the program ANSIG3.3 (Kraulis et al., 1994). NOE spectra were semi-automatically peak picked using the scalar coupling assignments as a mask to minimize picking artifacts and to avoid crosspeaks associated with unassigned residues, but no explicit filtering of the peak list was done. Furthermore, with the exception of a few short range NOE crosspeaks, which were used to augment the sequential assignments (described above), no manual NOE assignments were made. Instead, the NOE crosspeaks were output from ANSIG and converted to ambiguous restraint lists in external programs by matching the peak positions to the sequential assignments within a user-defined tolerance, the choice of which can be critical (see below).

Structures were calculated with restrained molecular dynamics-simulated annealing procedures in X-PLOR and restraint tables were modified at each iteration using ARIA manipulations (Nilges et al., 1997). Initial calculations included dihedral angle restraints and all contributors to all NOE crosspeaks, the vast majority of which were ambiguous. Using ARIA, restraints were gradually discarded in subsequent iterations if they contributed little to the crosspeak intensity based on a percent contribution cutoff. In addition, duplicate restraints were removed such that only the shortest restraint was kept.

Once the global fold had been established, non-stereoassigned isopropyl and methylene groups were explicitly swapped by setting the force constants for improper angles to zero (Folmer et al., 1997) and subjected to a Metropolis acceptance criterion (Raine and Smith, unpublished) based on the NOE energy and a temperature factor optimized for convergence. In the last few iterations, χ angle and hydrogen bond restraints were included. If the circular variance (Laskowski et al., 1993) of a given χ was close to 0 in the top twenty structures, the χ was included as a loose ($\pm 30°$) angular restraint in subsequent calculations (12 additional restraints were added in this manner). Final structures were refined using the PARALLHDG5.1 force field which contains non-bonded

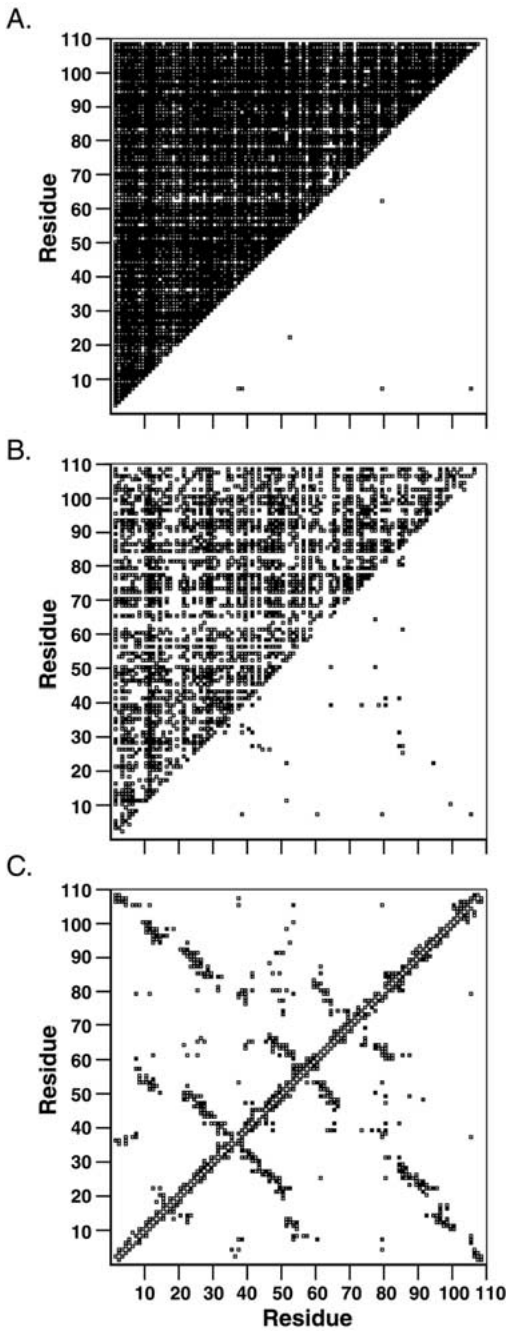

Figure 2. Contact plots of the restraints between residues from (A) the 3D ^{13}C separated NOESY and (B) the 4D ^{13}C-^{13}C separated NOESY from the first round of structure calculations, and (C) from all the NOESY data combined in the final iteration. The number of ambiguous and unambiguous restraints are shown above and below the diagonal, respectively. The following coloring scheme is used to indicate the number of assignment possibilites: white, 1–4; light gray, 5–8; dark gray, 8–12; and black, >12. Because of the resolution and coloring scheme, it is not possible to visually distinguish the number of unambiguous (2300) and ambiguous (905) restraints in (C).

parameters from the PROLSQ refinement program and a dihedral potential (Linge and Nilges, 1999).

Some additional details about the calculations are particularly noteworthy. The tolerance of the match between the scalar coupling assignments and the NOE crosspeak positions was set larger than the expected statistical variation in chemical shift to account for non-uniform differences between MIA samples. Earlier calculations that used restraint lists generated with tighter tolerances failed to improve beyond a backbone rmsd of ~0.8 Å and showed variable NOE violations. We hypothesized that some correct NOE assignments had been excluded in the initial restraint lists and therefore increased the tolerances to 0.6 ppm for ^{13}C and 0.1 ppm for ^{1}H, and restarted the calculations. Table 1 shows the number of ambiguous and unambiguous assignments at the beginning and end of the calculation, as well as the average level of ambiguity. The decrease in ambiguity by the end of the calculation is also indicated by the contact plots of the restraints from the first round of structure calculations (Figures 2A and 2B) compared to the contact plot for all restraints used in the final round (Figure 2C). The initial ambiguity was high; in the case of the 3D ^{13}C separated NOESY, for example, the average number of contributors to each of the crosspeaks was 87 at the beginning of the calculation (Table 1, Figure 2B). Nevertheless, 25 out of 30 structures converged to within a backbone rmsd of 2.0 Å in the initial iteration. By the end of the calculation, the average ambiguity was approximately 3 for all NOE datasets combined (Table 1), and the final NMR structures were well-defined (see below). Figure 3 shows the per residue distribution of unambiguous long range contacts after the final iteration, and the legend summarizes the total number of unambiguous intra-residue, short, medium, and long range restraints.

Another important detail about the calculations is that the initial iteration was extremely time consuming (~12.5 h per structure on a MIPS R10000 250 MHz processor) because of the large number of possible contributors from each NOE crosspeak. As restraints that contributed little to crosspeak intensities were discarded, the number and extent of ambiguous restraints were reduced (Table 1, Figure 2) along with the calculation time (90 min for each structure). Significantly, by using just the 4D data, the time required for the initial structure calculations could be reduced by about a factor of 6, and the ensemble had the same rmsd as the ensemble generated with all the data. This is not surprising, given that the 4D data sets had significantly more unique long range NOEs and fewer ambiguous NOEs than the 3D data sets (Figure 2).

Description of the structure

A stereoview of the NMR ensemble of the 20 lowest energy structures of MIA is shown in Figure 4 next to a MOLSCRIPT (Kraulis, 1991) ribbon of the secondary structure elements. The structure is well determined and had no NOE or torsion angle violations greater than 0.3 Å and 3°, respectively (Table 1). The backbone rmsd of the ensemble is 0.32 Å for secondary structural elements (43 residue pairs) and 0.38 Å for the well-ordered regions (76 residue pairs). Only the N- and C-termini and the loop between β4 and β5 have deviations of 1 Å or greater from the mean coordinates (see below). A stereochemical analysis of the ensemble using PROCHECK (Laskowski et al., 1993) shows that 81.5% of the residues fall in the most favored region of the Ramachandran plot, 16.4% are in the additionally allowed regions, 1.0% are in the generously allowed region, and 1.1% are in the disallowed region. Residues in the less favored regions generally correspond to those with higher mobility; when analyzing only the ordered regions of the ensemble, 89.3% are in the most favored region, 9.9% are in the additionally allowed regions, 0.1% are in the generously allowed region, and 0.7% are in the disallowed region.

The quality of the solution structures is underscored by the fact that the crystal structure could be solved by molecular replacement using the NMR structures as search models (data not shown) and verifies that our automated procedure was successful. Importantly, this automation strategy is not limited to proteins with extensive beta sheet content like MIA; it has also been successfully used to calculate the structure of a highly helical protein with significant chemical shift degeneracy and fewer long range restraints (Luh et al., 1997). Practical details of our implementation of structure calculations with ambiguous restraints will be described for MIA and other proteins from our laboratories in a future publication (in preparation).

Because the structure of MIA has been described previously, only a brief analysis of the structure will be presented here. The overall structure consists of seven β-strands, two short 3_{10} helices, and several loops (Figures 4 and 5). The SH3 subdomain contains five β-strands (β2-β6) which form two three-stranded β-sheets that pack at right angles to each other and are highly curved as in beta-barrels. Between β2 and β3 is

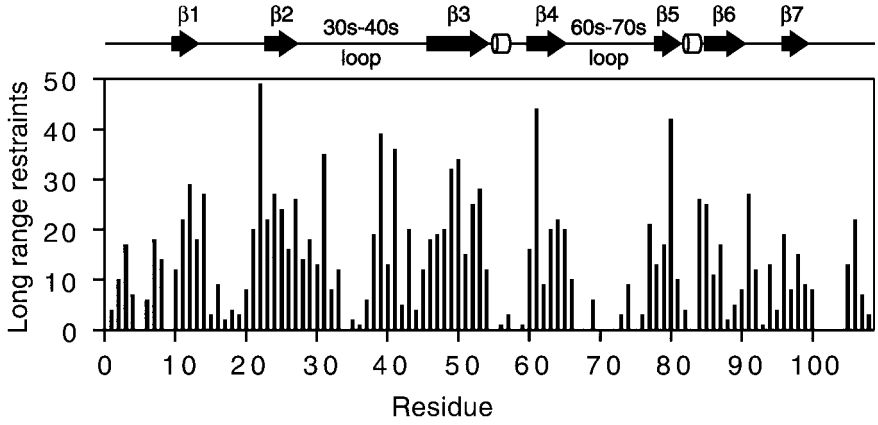

Figure 3. The number of unambiguous long range restraints as a function of residue at the end of the calculation. The total number of unambiguous restraints are distributed as follows: [i − j] = 0 (965), [i − j] = 1 (472), [i − j] = 2–4 (170), [i − j] > 4 (693). The secondary structure of MIA is shown above the figure.

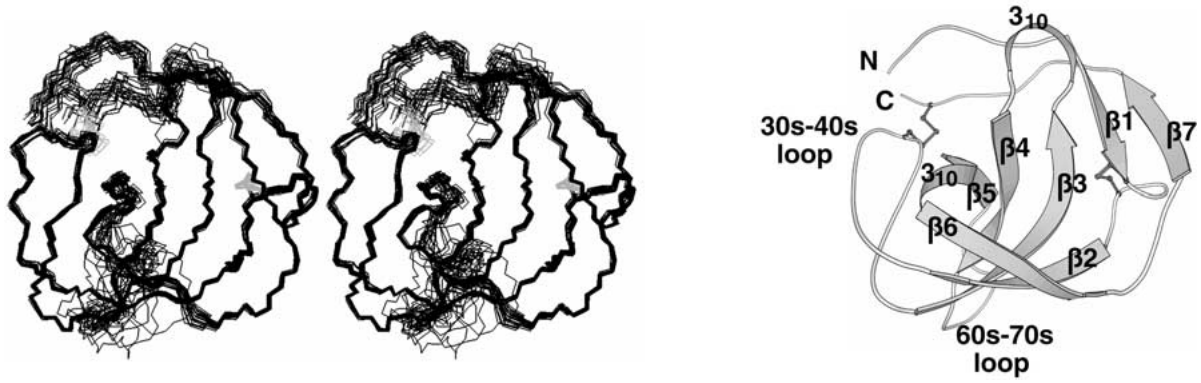

Figure 4. The solution structure of MIA. A stereoview of the NMR ensemble of the 20 lowest energy structures of MIA is shown next to a MOLSCRIPT ribbon of the secondary structure elements in the same orientation. The disulfide bonds are shown in gray in the ensemble and as balls and sticks in the MOLSCRIPT representation. The backbone rmsd of the ensemble is 0.32 Å for secondary structural elements (43 residue pairs) and 0.38 Å (76 residue pairs) for the well-ordered regions. The N- and C-termini and the 60s–70 s loop between β4 and β5 sample multiple conformations in the ensemble. The lowest energy NMR structure is shown as the MOLSCRIPT representation.

an 18 residue structured 30s-40s loop that corresponds to the RT loop of SH3 domains. Like canonical SH3 domains, this loop has an irregular antiparallel hairpin-like structure (Musacchio et al., 1994). Between β3 and β4 are three residues of a 3_{10} helix. Between β4 and β5 is a large 12 residue 60s-70s loop that samples multiple conformations, particularly between residues 67 and 76. This loop corresponds to the distal loop in SH3 domains, which is typically a tight β-turn distant from the PPII helix-binding site (Larson and Davidson, 2000).

The extensions to the MIA SH3 subdomain build on one end of the SH3 barrel (Figure 5). Two beta strands (β1 and β7) add to one of the sheets of the SH3 domain with the following connectivity (β7-β1-β3-β4-β5). The extensions also contain three of the four cysteines that form the two disulfides. The C-terminus is tethered to the 30s-40s loop by the disulfide between the penultimate residue, Cys 107, and Cys 36, which is located at the apex of the RT loop. The other disulfide (Cys 13-Cys 18) forms a small loop that directly follows β1. The eight residues preceding β1 lack regular secondary structure and have higher rms deviations from the mean, but overall, there is a tendency for the N-terminus to lie near the C-terminus. The C-terminal nine residues that follow β7 also sample multiple conformations. However, the mobility of the C-terminus is restricted by the Cys 36-Cys 107 disulfide.

Despite these additional structural elements, the MIA SH3 subdomain is strikingly similar to that of canonical SH3 domains. The rmsd between the MIA SH3 subdomain and the Sem-5 SH3 domain (acces-

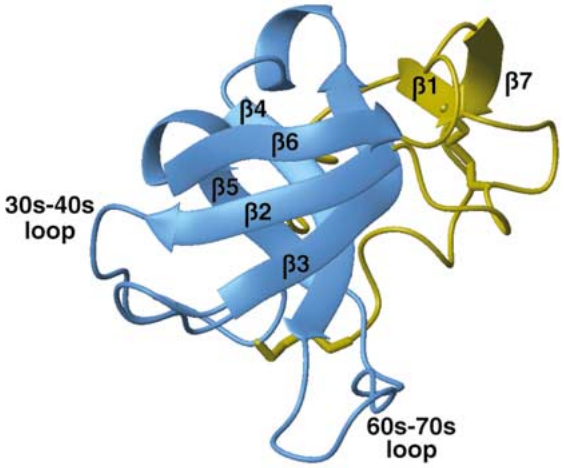

Figure 5. MOLMOL (Koradi et al., 1996) representation of the lowest energy structure of MIA in a different orientation than in Figure 3. Strands β2-β6 make up the SH3 subdomain, which is colored in blue, while the extensions to the subdomain and the disulfide bonds are colored in gold.

sion code 1SEM) is 0.71 Å using the lowest energy NMR structure and 37 of the 56 possible α-carbon pairs in the superposition. The main structural differences between canonical SH3 domains and the MIA SH3 subdomain are found in the length of the beta strands and in some of the loops. For a more thorough discussion of the structural similarities and differences, see (Lougheed et al., 2001).

Backbone dynamics

^{15}N Relaxation measurements were made in order to determine the flexibility of the backbone and to assess whether regions of the structure with higher rmsd from the mean are truly more dynamic (Kay et al., 1989). Relaxation data were obtained for every assigned backbone amide except for Leu 7, Cys 18, and Phe 105, which had low crosspeak intensity. Only a qualitative discussion of the backbone dynamics follows since a full treatment is complicated by the tendency of MIA to aggregate, which was indicated by lower than expected values of T2.

The heteronuclear ^{1}H-^{15}N-NOE and the ^{15}N T1/T2 are plotted against the amino acid sequence in Figures 6A and 6B. Significant backbone flexibility on the nanosecond to picosecond time scale is observed in the N-terminus, the 60s-70s loop, and for the residues C-terminal to β7 (heteronuclear NOEs < 0.65), where the N-terminus and the 60s-70s loop appear to be the most flexible. The higher rmsd of the ensemble from the mean in these regions (Figure 6C) is therefore due to true flexibility. Although significant flexibility in the N-terminus is suggested by the heteronuclear NOEs of ∼0.4 to 0.6, these values are not so low as to indicate complete disorder. This is consistent with the observed tendency of the N-terminus to lie near the C-terminus in the ensemble. Some backbone flexibility is also indicated for the last three residues of β6 where a beta bulge is located at Asp 88.

Conformational exchange processes on the microsecond to millisecond time scale are indicated by an elevated T1/T2 value for Ser 19, many residues in the 60s-70s loop, and the residues C-terminal to β7 (Figure 6B). For the 60s-70s loop and the residues C-terminal to β7, this is mirrored by the greater rmsd of the ensemble from the mean (Figure 6C).

Comparison of the solution and crystal structure

The solution and crystal structure of MIA are remarkably similar, attesting to the success of the automated approach. Figure 6D shows a plot of the difference in backbone rmsd as a function of residue and Figure 7A shows the superposition of the crystal structure and the lowest energy NMR structure. The rmsd between the crystal and NMR structure is 0.86 Å for the backbone atoms when the ordered regions are superimposed (76 residue pairs).

The position and length of the secondary structure elements in the NMR and crystal structures, as defined by Kabsch and Sander (1983), are virtually identical, and consistent among members of the NMR ensemble. The largest variation in the ensemble was seen in the n-Src loop, which did not have the characteristic hydrogen bonding pattern for a 3_{10} helix in every structure. The conformations of the sidechains are also well determined in the NMR structure and compare well with those in the crystal structure. Figure 7B shows the superposition of an arbitrarily chosen subset of the core residues.

The largest differences between the crystal and solution structure are found at the N-terminus and in the 60s-70s loop, which are the most dynamic regions of the protein, as discussed above. Not surprisingly, the 60s-70s loop is defined by fewer restraints than other regions of the protein (Figure 3); nevertheless it does show some conformational preference due to mutually supporting restraints including 21 unambiguous long range and 10 unambiguous medium range restraints. The first and last few residues of the loop have backbone conformations that match the crystal structure fairly well. However, between residues 68 and 73 the

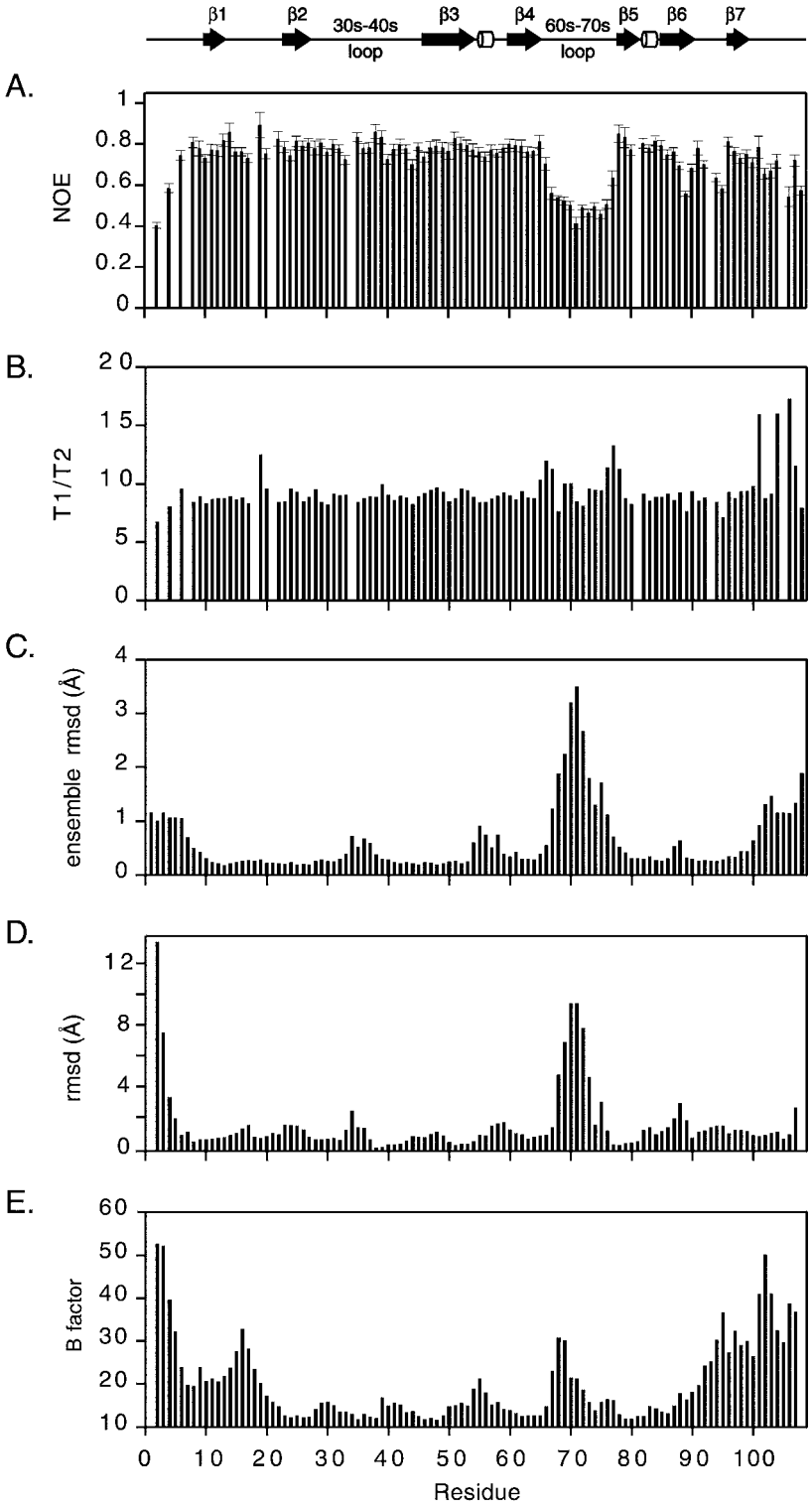

Figure 6. Sequence plots of (A) the heteronuclear ^{1}H-^{15}N-NOE and (B) the ^{15}N T1/T2, (C) the backbone rmsd of the ensemble from the mean, (D) the backbone rmsd between the crystal structure and the lowest energy NMR structure, and (E) the average backbone B factor as a function of the residue. The highest flexibility is seen at the N-terminus and in the 60s–70s loop. The crystal structure had two molecules (A and B) in the asymmetric unit. Molecule A was used for the plots, however, no qualitative difference was seen when molecule B was used. The secondary structure of the protein is shown above the figure.

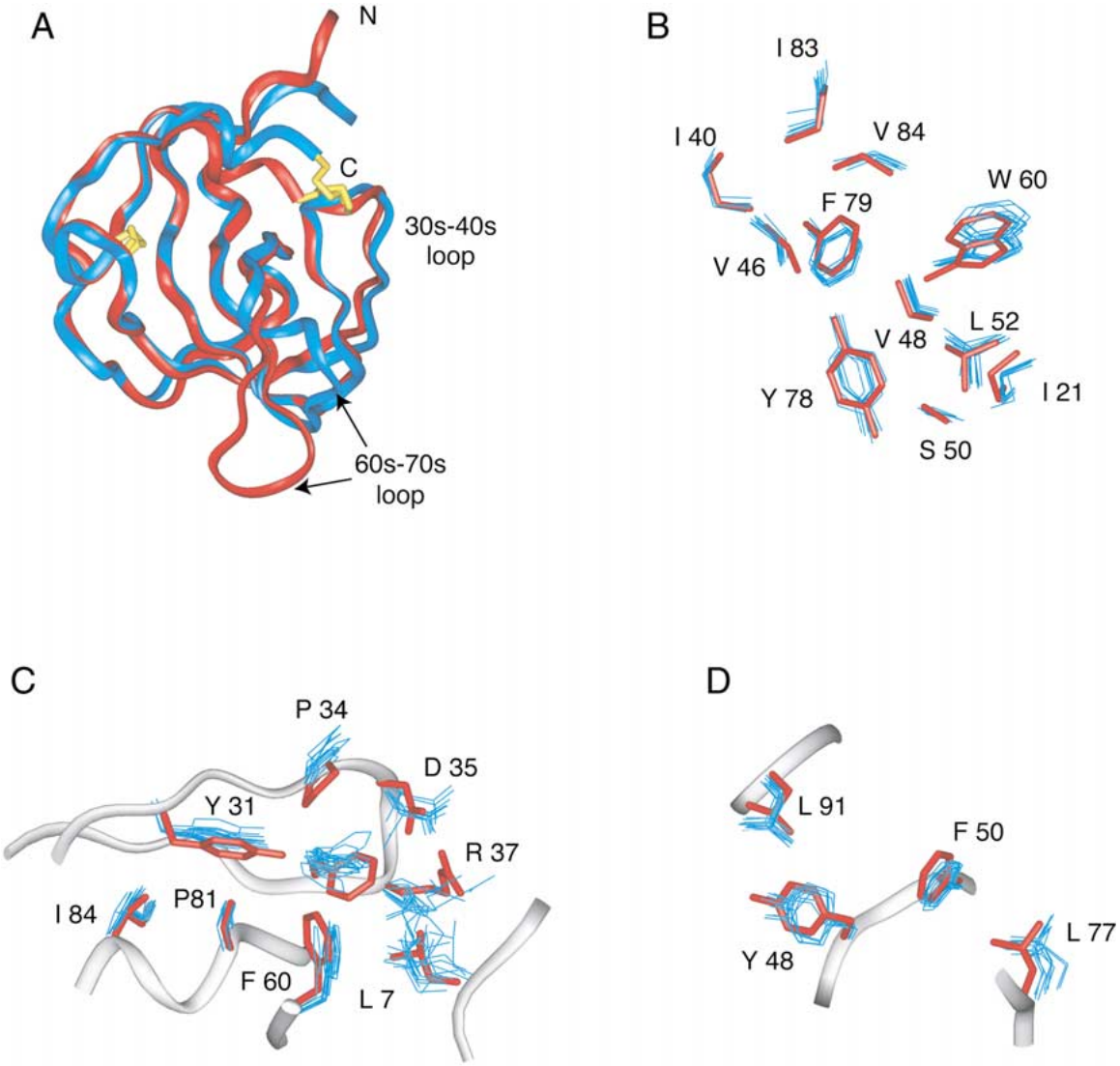

Figure 7. Comparison of the NMR structure (blue) with the crystal structure (red). (A) The superposition of the crystal structure with the lowest energy NMR structure. The rmsd between the two structures is 0.87 Å for the backbone atoms when the ordered regions are superimposed (76 residue pairs). The largest differences are seen at the N-terminus and in the 60s–70s loop. (B) An arbitrarily selected set of core residues is shown. The 10 lowest energy NMR structures (blue) are superimposed with the crystal structure (red). (C and D) The putative ligand interaction sites displayed with the 10 lowest energy NMR structures (blue) superimposed on the crystal structure (red). The crystal structure backbone is shown. (C) The larger of the two conserved patches located in the region that corresponds to the polyproline helix-binding site in conventional SH3 domains. (D) The smaller of the two conserved patches.

deviations between the NMR and crystal structure are on the order of 5–10 Å. In the crystal structure, the loop is in a more extended conformation relative to the NMR ensemble due to a crystal contact involving Tyr70 at the loop apex. The presence of this conformation cannot be excluded in the solution structure, but there was no evidence for it in the data, as we failed to detect any violations exceeding 0.3 Å. Differences between the N-termini involve the first two residues, which protrude outward from the molecule in the crystal structure while they lie closer to the C-terminus in the solution structure.

In general, regions with higher B factors correlate with the more dynamic regions of MIA in solution (the N-terminus, the C-terminus, and the 60s-70s loop, see Figure 6E). However, the B factors for β1 and β7 are slightly higher than might be expected from the NMR dynamics and ensemble rms deviations. These

sheets had clear electron density in the original map and were unambiguously traced. The higher B factors most likely indicate some disorder that cannot be modeled by a static structure.

Possible ligand interaction sites

The biological mechanism and ligand(s) of MIA are not currently known. MIA has been demonstrated to inhibit cell binding to the extracellular matrix *in vitro* and to promote metastasis *in vivo*. These anti-adhesive properties may arise by integrin-mediated effects, since integrins are transmembrane receptors that tether cells to the extracellular matrix and function in cell motility, growth, and signal transduction (Varner and Cheresh, 1996). Identification of ligand binding sites on MIA might provide insight into the origin of the anti-adhesive properties. Previously, two potential ligand binding sites were identified based on the largest areas of amino acid conservation mapped onto the crystal structure (Lougheed et al., 2001). Here we compare these two sites in the solution and crystal structures, and describe a third sequence that is a potential integrin binding site.

The two putative ligand interaction sites that were identified previously appear to be relatively rigid. The backbone conformations for these two sites are similar in the crystal and NMR structures, and the backbone dynamics data demonstrate that the sites are well-ordered. Most sidechains of residues that make up these sites are well-determined in the ensemble of NMR structures and have similar conformations in the crystal structure (Figures 7C and 7D). One of these sites is located in a region that corresponds to the PPII helix binding site in canonical SH3 domains. However, our previous analysis of the crystal structure suggested that this site is not able to bind PPII helices. The close similarity of this region of the protein between the NMR and crystal structures lends further support to this conclusion.

Another possible ligand interaction site occurs in the 60s-70s loop. The amino acids in this loop resemble a sequence in laminin, YGYYGDALR , that has been implicated in laminin recognition of the $\alpha 2\beta 1$ integrin (Underwood et al., 1995). As shown below, the sequence ^{69}YYGD^{72} , is strictly conserved among the MIA orthologs.

MIA_human	^{66}QGDYYGDLAARL^{77}
MIA_mouse	^{66}QGDYYGDGAARL
MIA_bovine	^{66}QGDYYGDLAAHL
MIA_rat	^{66}QGDYYGDLAARL
Laminin	^{66}YG - YYGD - ALR -

It also occurs at the apex of the loop, similar to integrin binding sites, which are frequently found in β-turn structures or more flexible loops (Ruoslahti, 1996). Additionally, the integrin $\alpha 2\beta 1$ is expressed by many cell types (Zutter and Santoro, 1990), including both chondrocytes (Durr et al., 1993) and malignant melanocytes (Albelda et al., 1990), which are the two main cell types that express MIA. Therefore, a mechanism for MIA's anti-adhesive properties might involve an interaction with integrin receptors. A binding activity could interfere directly with integrin cell attachment sites or possibly initiate cell signaling. An alternative mechanism could involve interaction of MIA with the extracellular matrix protein fibronectin and masking of fibronectin's integrin binding sites (Stoll et al., 2001). However, fibronectin binds a wide range of macromolecules (Ruoslahti, 1988), and the significance of its interaction with MIA *in vivo* remains to be determined. Since the observed anti-adhesive activity occurs at nanomolar concentrations of MIA (Bosserhoff et al., 1998), receptor-mediated signal transduction may also be involved.

Concluding remarks

MIA is a member of a recently identified family of proteins that contain a subdomain with low sequence homology to SH3 domains and is the first secreted protein demonstrated to have an SH3 subdomain. The solution structure of MIA was solved in a semi-automated manner without manual assignment of its NOE crosspeaks. Since we previously solved the crystal structure of MIA, we were able to test the validity of our automated approach in this case of high levels of ambiguity. The solution structure is virtually indistinguishable from the crystal structure, except in the most dynamic regions of the protein, demonstrating that accurate protein structures can be reliably calculated from ambiguous assignments. The greatest variation between the crystal and NMR structure is found in a loop, which resembles a sequence implicated in integrin binding.

Acknowledgements

We thank Michael Nilges for providing initial ARIA scripts, Brian Smith, Andrew Raine, Wayne Boucher, and Ernest Laue for providing essential software, and Andy Byrd, Amanda Altieri, and Steven Sparks for contact map display software. This work was supported by grants awarded to T.M.H. from the National Institutes of Health Grant AI 37110 and the UCOP CLC Program.

References

Albelda, S.M., Mette, S.A., Elder, D.E., Stewart, R., Damjanovich, L., Herlyn, M. and Buck, C.A. (1990) *Cancer Res.*, **50**, 6757–6764.

Archer, S.J., Ikura, M., Torchia, D.A. and Bax, A. (1991) *J. Magn. Reson.*, **95**, 636–641.

Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) *J. Magn. Reson.*, **88**, 425–431.

Bax, A., Max, D. and Zax, D. (1992) *J. Am. Chem. Soc.*, **114**, 6923–6925.

Bax, A., Vuister, G.W., Grzesiek, S., Delaglio, F., Wang, A.C., Tschudin, R. and Zhu, G. (1994) *Meth. Enzymol.*, **239**, 79–105.

Blesch, A., Bosserhoff, A.K., Apfel, R., Behl, C., Hessdoerfer, B., Schmitt, A., Jachimczak, P., Lottspeich, F., Buettner, R. and Bogdahn, U. (1994) *Cancer Res.*, **54**, 5695–5701.

Bosserhoff, A.K., Golob, M., Buettner, R., Landthaler, M. and Hein, R. (1998) *Hautarzt*, **49**, 762–769.

Bosserhoff, A.K., Kaufmann, M., Kaluza, B., Bartke, I., Zirngibl, H., Hein, R., Stolz, W. and Buettner, R. (1997a) *Cancer Res.*, **57**, 3149–3153.

Bosserhoff, A.K., Kondo, S., Moser, M., Dietz, U.H., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Buettner, R. and Sandell, L.J. (1997b) *Dev. Dynam.*, **208**, 516–525.

Brunger, A.T. (1992) *X-PLOR Manual, Version 3.0*, Yale University Press, New Haven, CT.

Chen, H., Hughes, D.D., Chan, T.A., Sedat, J.W. and Agard, D.A. (1996) *J. Struct. Biol.*, **116**, 56–60.

Clore, G.M., Kay, L.E., Bax, A. and Gronenborn, A.M. (1991) *Biochemistry*, **30**, 12–18.

Clowes, R.T., Boucher, W., Hardman, C.H., Domaille, P.J. and Laue, E.D. (1993) *J. Biomol. NMR*, **3**, 349–354.

Dietz, U.H. and Sandell, L.J. (1996) *J. Biol. Chem.*, **271**, 3311–3316.

Durr, J., Goodman, S., Potocnik, A., von der Mark, H. and von der Mark, K. (1993) *Exp. Cell Res.*, **207**, 235–244.

Folmer, R.H., Hilbers, C.W., Konings, R.N. and Nilges, M. (1997) *J. Biomol. NMR*, **9**, 245–258.

Golob, M., Buettner, R. and Bosserhoff, A.K. (2000) *J. Invest. Dermatol.*, **115**, 42–47.

Grzesiek, S. and Bax, A. (1992a) *J. Am. Chem. Soc.*, **114**, 6291–6293.

Grzesiek, S. and Bax, A. (1992b) *J. Magn. Reson.*, **99**, 201–207.

Grzesiek, S., Kuboniwa, H., Hinck, A.P. and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 5312–5315.

Grzesiek, S., Vuister, G.W. and Bax, A. (1993) *J. Biomol. NMR*, **3**, 487–493.

Guba, M., Bosserhoff, A.K., Steinbauer, M., Abels, C., Anthuber, M., Buettner, R. and Jauch, K.W. (2000) *Br. J. Cancer*, **83**, 1216–1222.

Kabsch, W. and Sander, C. (1983) *Biopolymers*, **22**, 2577–2637.

Kay, L.E., Clore, G.M., Bax, A. and Gronenborn, A.M. (1990) *Science*, **249**, 411–414.

Kay, L.E., Torchia, D.A. and Bax, A. (1989) *Biochemistry*, **28**, 8972–8979.

Koradi, R., Billeter, M. and Wuthrich, K. (1996) *J. Mol. Graph.*, **14**, 29–32.

Kraulis, P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946–950.

Kraulis, P.J., Domaille, P.J., Campbell-Burk, S.L., Van Aken, T. and Laue, E.D. (1994) *Biochemistry*, **33**, 3515–3531.

Kuboniwa, H., Grzesiek, S., Delaglio, F. and Bax, A. (1994) *J. Biomol. NMR*, **4**, 871–878.

Larson, S.M. and Davidson, A.R. (2000) *Protein Sci.*, **9**, 2170–2180.

Laskowski, R.A., Macarthur, M.W., Moss, D.S. and Thornton, J.M. (1993) *J. Appl. Crystallogr.*, **26**, 283–291.

Linge, J.P. and Nilges, M. (1999) *J. Biomol. NMR*, **13**, 51-59.

Linge, J.P., O'Donoghue, S.I., Nilges, M. (2001) *Meth. Enzymol.*, **339**, 71–90.

Lougheed, J.C., Holton, J.M., Alber, T., Bazan, J.F. and Handel, T.M. (2001) *Proc. Natl. Acad. Sci. USA*, **98**, 5515–5520.

Luh, F.Y., Archer, S.J., Domaille, P.J., Smith, B.O., Owenm D., Brotherton, D.H., Raine, A.R., Xu, X., Brizuela, L., Brenner, S.L. and Laue, E.D. (1997) *Nature*, **389**, 999–1003.

Musacchio, A., Wilmanns, M. and Saraste, M. (1994) *Prog. Biophys. Mol. Biol.*, **61**, 283–297.

Neidhardt, F.C., Bloch, P.L. and Smith, D.F. (1974) *J. Bacteriol.*, **119**, 736–747.

Nilges, M., Macias, M.J., O'Donoghue, S.I. and Oschkinat, H. (1997) *J. Mol. Biol.*, **269**, 408–422.

Pan, Y. (2000) International Publication Number WO 00/12762, A novel protein related to melanoma-inhibiting protein and uses thereof. U.S. Patent Office.

Robertson, N.G., Heller, S., Lin, J.S., Resendes, B.L., Weremowicz, S., Denis, C.S., Bell, A.M., Hudspeth, A.J. and Morton, C.C. (2000) *Genomics*, **66**, 242–248.

Ruoslahti, E. (1988) *Annu. Rev. Biochem.*, **57**, 375–413.

Ruoslahti, E. (1996) *Annu. Rev. Cell Dev. Biol.*, **12**, 697–715.

Stoll, R., Renner, C., Zweckstetter, M., Bruggert, M., Ambrosius, D., Palme, S., Engh, R.A., Golob, M., Breibach, I., Buettner, R., Voelter, W., Holak, T.A. and Bosserhoff, A.K. (2001) *EMBO J.*, **20**, 340–349.

Talluri, S. and Wagner, G. (1996) *J. Magn. Reson. Series*, **B112**, 200–205.

Underwood, P.A., Bennett, F.A., Kirkpatrick, A., Bean, P.A. and Moss, B.A. (1995) *Biochem. J.*, **309**, 765–771.

Varner, J.A. and Cheresh, D.A. (1996) *Curr. Opin. Cell Biol.*, **8**, 724–730.

Vuister, G.W. and Bax, A. (1992) *J. Magn. Reson.*, **98**, 428–435.

Vuister, G.W., Wang, A.C. and Bax, A. (1993) *J. Am. Chem. Soc.*, **115**, 5334–5335.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) *J. Biomol. NMR*, **5**, 67–81.

Yamazaki, T., Forman-Kay, J.D. and Kay, L.E. (1993) *J. Am. Chem. Soc.*, **115**, 11054–11055.

Zerbe, O., Szyperski, T., Ottiger, M. and Wüthrich, K. (1996) *J. Biomol. NMR*, **7**, 99–106.

Zuiderweg, E.R.P., McIntosh, L.P., Dahlquist, F.W. and Fesik, S.W. (1990) *J. Magn. Reson.*, **86**, 210–216.

Zutter, M.M. and Santoro, S.A. (1990) *Am. J. Pathol.*, **137**, 113–120.