J|A|C|S
ARTICLES
Published on Web 05/16/2006

Protein−Ligand NOE Matching: A High-Throughput Method for Binding Pose Evaluation That Does Not Require Protein NMR Resonance Assignments

Keith L. Constantine,* Malcolm E. Davis, William J. Metzler, Luciano Mueller, and Brian L. Claus

Contribution from the Bristol Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, New Jersey 08543

Received January 17, 2006; E-mail: keith.constantine@bms.com

Abstract: Given the three-dimensional (3D) structure of a protein, the binding pose of a ligand can be determined using distance restraints derived from assigned intra-ligand and protein−ligand nuclear Overhauser effects (NOEs). A primary limitation of this approach is the need for resonance assignments of the ligand-bound protein. We have developed an approach that utilizes data from 3D ^{13}C-edited, ^{13}C/^{15}N-filtered HSQC−NOESY spectra for evaluating ligand binding poses *without* requiring protein NMR resonance assignments. Only the ^{1}H NMR assignments of the bound ligand are essential. Trial ligand binding poses are generated by any suitable method (e.g., computational docking). For each trial binding pose, the 3D ^{13}C-edited, ^{13}C/^{15}N-filtered HSQC−NOESY spectrum is predicted, and the predicted and observed patterns of protein−ligand NOEs are matched and scored using a fast, deterministic bipartite graph matching algorithm. The best scoring (lowest "cost") poses are identified. Our method can incorporate any explicit restraints or protein assignment data that are available, and many extensions of the basic procedure are feasible. Only a single sample is required, and the method can be applied to both slowly and rapidly exchanging ligands. The method was applied to three test cases: one complex involving muscle fatty acid-binding protein (mFABP) and two complexes involving the leukocyte function-associated antigen 1 (LFA-1) I-domain. Without using experimental protein NMR assignments, the method identified the known binding poses with good accuracy. The addition of experimental protein NMR assignments improves the results. Our "NOE matching" approach is expected to be widely applicable; i.e., it does not appear to depend on a fortuitous distribution of binding pocket residues.

Introduction

Nuclear magnetic resonance (NMR) spectroscopy is an important tool in the drug discovery process, contributing to both lead identification and lead optimization.[1−6] For lead optimization, NMR provides alternatives to X-ray crystallography for obtaining structural information on protein−ligand complexes. Advances in hardware, experimental approaches,[7−10] and data analysis methods[10−12] have increased the throughput and extended the applicability of NMR for protein structure determination. Nevertheless, even when highly optimized,[13] a full NMR-based structure determination of a protein−ligand complex requires a significant commitment of time and resources.

In many cases relevant to lead optimization, one or more experimental structures of the target protein are available. Alternatively, the target protein structure can often be approximated reasonably well by modeling approaches.[14] Given a suitable structure of the target protein, the problem of determining the structure of a protein−ligand complex reduces to one of determining the binding *pose* (i.e., the location, orientation, and internal conformation) of a bound compound of interest, possibly accounting for any protein conformational changes that occur upon binding. To be of optimal value, binding poses must be determined in a time frame that supports iterative cycles of structure-based ligand design.

Many NMR-based approaches have been proposed and developed for rapidly deriving information on binding poses; these include methods that do not require protein resonance assignments. Transferred nuclear Overhauser effect (NOE)

(1) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. *Science* **1996**, *274*, 1531−1534.
(2) Coles, M.; Heller, M.; Kessler, H. *Drug Discovery Today* **2003**, *8*, 803−810.
(3) Pellecchia, M.; Sem, D. S.; Wüthrich, K. *Nat. Rev. Drug Discovery* **2002**, *11*, 211−219.
(4) Stockman, B. J.; Dalvit, C. *Prog. NMR Spectrosc.* **2002**, *41*, 187−231.
(5) Roberts, G. C. K. *Drug Discovery Today* **2000**, *5*, 230−240.
(6) Homans, S. W. *Angew. Chem., Int. Ed.* **2004**, *43*, 290−300.
(7) Ferentz, A. E.; Wagner, G. *Q. Rev. Biophys.* **2000**, *33*, 29−65.
(8) Riek, R.; Pervushin, K.; Wüthrich, K. *Trends Biochem. Sci.* **2000**, *25*, 462−468.
(9) Tugarinov, V.; Muhandiram, R.; Ayed, A.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 10025−10035.
(10) Clore, G. M.; Schwieters, C. D. *Curr. Opin. Struct. Biol.* **2002**, *12*, 146−153.
(11) Güntert, P. *Prog. NMR Spectrosc.* **2003**, *43*, 105−125.
(12) Gronwald, W.; Kalbitzer, H. R. *Prog. NMR Spectrosc.* **2004**, *44*, 33−96.
(13) Medek, A.; Olejniczak, E. T.; Meadows, R. P.; Fesik, S. W. *J. Biomol. NMR* **2000**, *18*, 229−238.
(14) Hillisch, A.; Pineda, L. F.; Hilgenfeld, R. *Drug Discovery Today* **2004**, *9*, 659−669.

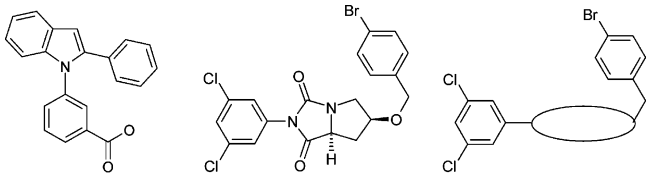
10.1021/ja060356w CCC: $33.50

experiments[15−18] have long been applied to determine the internal conformations of weakly bound ligands. Long-range paramagnetic distance restraints from site-directed spin labeling have been explored computationally as a possible approach.[19] A method based on saturation transfer difference[20] (STD) and residue type-specific labeling has been described (termed "SOS−NMR").[21] Another recently described method for ligand pose evaluation is based on observed and predicted binding-induced chemical shift changes for ligand ^{1}H resonances;[22] the chemical shifts are predicted using quantum mechanics in this procedure.

For proteins with NMR assignments, ligand binding sites can be localized by ligand-induced line-broadening and/or chemical shift changes of protein resonances.[1−6,23] Information on bound ligand locations and orientations can be obtained by examining chemical shift changes induced by closely related ligand analogues,[24] or by empirically estimating protein chemical shift changes expected for trial orientations of ligand aromatic rings.[25] Methods that use a limited number of protein NMR assignments have also been developed. SOS−NMR and chemical shift perturbation mapping have been combined in an approach that selects or rejects binding poses on the basis of van der Waals and restraint energies.[26] Residue type-specific isotopic labeling schemes combined with inferential NMR assignment procedures have been used to obtain binding pose information on ligands in complex with large proteins.[27] A protocol that uses NOEs involving assigned backbone amide protons combined with docking/annealing calculations has been developed to validate ligand binding poses.[28]

With extensive protein ^{1}H, ^{13}C, and ^{15}N resonance assignments of binding site residues available, well-defined ligand binding poses can be determined using isotope-filtered NMR methods[29] to derive NOE distance restraints. A general strategy uses a sample containing a uniformly ^{13}C/^{15}N-labeled protein in complex with an unlabeled ligand. Bound or averaged ligand ^{1}H resonances are assigned with two-dimensional (2D) F_1,F_2-^{13}C/^{15}N-filtered TOCSY, COSY, and/or NOESY experiments.[30] Intra-ligand distance restraints are obtained from the 2D F_1,F_2-^{13}C/^{15}N-filtered NOESY data. Intermolecular protein−ligand distance restraints are derived from a three-dimensional (3D) ^{13}C-edited, ^{13}C/^{15}N-filtered HSQC−NOESY spectrum (hereafter referred to as a 3D X-filtered NOESY). This spectrum contains exclusively NOE peaks between ligand ^{1}H resonances (along the F_3 dimension) and protein ^{1}H^{13}C group resonances (along the F_1 (^{1}H) and F_2 (^{13}C) dimensions).

Chart 1. Compounds Used for the NOE Matching Tests



1 **2** **3**

A main limitation of the isotope-filtered NMR approach is the need to assign resonances for protein residues lining the binding pocket for each protein/ligand complex. (A second major limitation — that of decreasing spectral sensitivity with increasing protein size — may be addressed by modification of the approach described herein.) We have hypothesized that, even *without* protein assignments, the *pattern* of peaks observed in a 3D X-filtered NOESY spectrum contains sufficient information to define the ligand binding pose in most cases, or can at least be used to rule out the vast majority of possible poses. This hypothesis is based on the observation of characteristic chemical shift ranges for ^{1}H^{13}C groups in amino acid residues,[31] and on the heterogeneous composition and spatial distribution of amino acids that line binding pockets.[21]

Herein we describe in detail a method ("NOE matching") for "scoring" trial binding poses based on matching the observed (experimental) pattern of protein−ligand NOEs to predicted (theoretical) patterns of protein−ligand NOEs. The matching process is accomplished with a deterministic algorithm that runs in polynomial time, making it suitable for scoring a very large number of trial poses. The method was tested on muscle fatty acid-binding protein (mFABP) and the leukocyte function-associated antigen 1 I-domain (LFA-1) in complex with small organic compounds (Chart 1). We show that the approach can yield accurate binding pose information even when the "predicted" protein chemical shifts are simply set to mean values derived from the BioMagResBank (BMRB; diamagnetic protein statistics).[31] The results are shown to improve with more accurate protein chemical shifts. Approximate NOE intensity binning is used as well.

Three test cases were used to develop and evaluate protein−ligand NOE matching. The first test case involves an indole-based compound **1** (Chart 1) bound to mFABP, the second involves the LFA-1 I-domain (henceforth simply referred to as LFA-1) in complex with the hydantoin-based inhibitor **2** (Chart 1; compound **5** in ref 32), and the third test case involves LFA-1 in complex with compound **3** (Chart 1), which is similar to compound **2**. Compound **3** contains a proprietary core, which is simply represented by a ellipse in Chart 1.

Advantages, limitations, and possible extensions of our method are discussed, and complementarities with other methods are considered. The overall goal is to develop a widely applicable, robust, and extendable framework for evaluating and defining ligand binding poses that bypasses the requirement for protein NMR assignments and explicit conformational restraints, but which can utilize any such data that are available. Our

(15) Clore, G. M.; Gronenborn, A. M. *J. Magn. Reson.* **1982**, *48*, 402−418.
(16) Clore, G. M.; Gronenborn, A. M. *J. Magn. Reson.* **1983**, *53*, 423−442.
(17) Campbell, A. P.; Sykes, B. D. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 99−122.
(18) Ni, F.; Scheraga, H. A. *Acc. Chem. Res.* **1994**, *27*, 257−264.
(19) Constantine, K. L. *Biophys. J.* **2001**, *81*, 1275−1284.
(20) Mayer, M.; Meyer, B. *Angew. Chem., Int. Ed.* **1999**, *38*, 1784−1788.
(21) Hajduk, P. J.; Mack, J. C.; Olejniczak, E. T.; Park, C.; Dandliker, P. J.; Beutel, B. A. *J. Am Chem.* Soc. **2004**, *126*, 2390−2398.
(22) Wang, B.; Raha, K.; Merz, K. M., Jr. *J. Am. Chem. Soc.* **2004**, *126*, 11430−11432.
(23) Farmer, B. T., II; Constantine, K. L.; Goldfarb, V.; Friedrichs, M. S.; Wittekind, M.; Yanchunas, J., Jr.; Robertson, J. G.; Mueller, L. *Nat. Struct. Biol.* **1996**, *3*, 995−997.
(24) Medek, A.; Hajduk, P. J.; Mack, J.; Fesik, S. W. *J. Am. Chem. Soc.* **2000**, *122*, 1241−1242.
(25) McCoy, M. A.; Wyss, D. F. *J. Biomol. NMR* **2000**, *18*, 189−198.
(26) Schieborr, U.; Vogtherr, M.; Elshorst, B.; Betz, M.; Grimme, S.; Pescatore, B.; Langer, T.; Saxena, K.; Schwalbe, H. *ChemBioChem* **2005**, *6*, 1891−1898.
(27) Pellecchia, M.; Meininger, D.; Dong, Q.; Chang, E.; Jack, R.; Sem, D. S. *J. Biomol. NMR* **2002**, *22*, 165−173.
(28) Bertini, I.; Fragai, M.; Giachetti, A.; Luchinat, C.; Maletta, M.; Parigi, G.; Yeo, K. J. *J. Med. Chem.* **2005**, *48*, 7544−7559.
(29) Breeze, A. L. *Prog. NMR Spectrosc.* **2000**, *36*, 323−372.
(30) Petros, A. M.; Kawai, M.; Luly, J. R.; Fesik, S. W. *FEBS Lett.* **1992**, *308*, 309−314.
(31) Seavy, B.; Farr, E.; Westler, W.; Markley, J. *J. Biomol. NMR* **1991**, *1*, 217−236.
(32) Potin, D.; et al. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1161−1164.

method is a specialized "top-down" approach wherein "...the main aim is not a completely correct spectral assignment but a correct three-dimensional structure...".[12]

Methods

Target and Trial Poses. Several methods were used to determine target poses and generate trial binding poses. In the case of compound **1** (Chart 1) bound to mFABP, the target protein structure for DOCK and target ligand pose were selected from a high-resolution ensemble of NMR structures (details on this NMR structure ensemble are included in the Supporting Information). Trial binding poses for the mFABP/**1** complex were generated by rigid and flexible ligand docking with the program DOCK.[33] (For all DOCK calculations, generic site spheres were generated with no chemical knowledge to enhance the diversity of the generated poses.) The original NMR ensemble was also used as a source of trial poses. In addition, a low-resolution NMR ensemble was generated by simulated annealing[34] with only 10 arbitrarily picked protein−ligand NOE distance restraints; these structures were used as trial poses as well.

For the LFA-1/**2** test case, the X-ray crystallographic structure of this complex[35] served as the target pose. Trial poses for the LFA-1/**2** complex were generated by both flexible and rigid ligand docking using DOCK.[33] Rather than docking compound **2** into the protein coordinates derived from the LFA-1/**2** X-ray structure, the protein coordinates used for docking compound **2** were derived from a publicly available X-ray structure of LFA-1 in complex with lovastatin[36] (PDB code 1CQP) as a starting point. The protein coordinates were moderately diversified by selecting residues within 3.5 Å of lovastatin, removing lovastatin, and then performing conformational sampling of the selected residue side chains using Prime (Schrodinger, Inc.). This yielded 10 protein coordinate sets for ligand docking. The protein coordinates were then held fixed for docking. Compound **2** was held internally rigid for some docking runs and allowed internal flexibility in other docking runs.

A well-defined NMR ensemble (Supporting Information) of the LFA-1/**3** complex was derived by using protein−ligand NOE restraints to place the compound into the publicly available X-ray structure of LFA-1[36] (1CQP) by simulated annealing.[34] A single member of the ensemble was selected as the target pose. Trial poses of LFA-1/**3** were generated by removing compound **3** from the binding site in the target coordinate set and then using the DOCK program[33] to generate alternate poses. Two separate DOCK runs were performed, with **3** being held internally rigid in the first run and flexibly docked in the second run.

Preparation of Experimental 3D X-Filtered NOESY Peak Lists. Using a modified version of the FELIX program (Hare Research, Inc.; M. S. Friedrichs, unpublished), peaks in the 3D X-filtered NOESY spectra were picked interactively, and files containing information on the peak intensities, chemical shifts, and peak assignments were written. (Only the ligand ^{1}H chemical shift assignments are absolutely required for NOE matching). The experimental intensities were classified as very strong, strong, medium, or weak. For reason described below, the intensity classes were assigned "integer intensity" values of 4, 3, 2, or 1, respectively. To enhance the digital resolution in the 3D X-filtered NOESY spectra, they were recorded using 25.0 ppm sweep widths in the F_2 (^{13}C) dimensions, resulting in greater than 1-fold peak aliasing[37] in some cases. Heuristic rules were used to determine the actual ("unaliased") peak ^{13}C chemical shifts. For our test cases, these rules were found to be reliable on the basis of the known protein assignments. (Alternatively, spectra can be recorded with a wider ^{13}C sweep width (e.g., 60.0 ppm); this results in at most 1-fold peak aliasing and allows for straightforward determination of the peak ^{13}C chemical shifts.) For some tests, an idealized synthetic "experimental" spectrum was derived from the target pose by predicting intensities on the basis of the inter-proton distances observed in this pose, and by randomly assigning chemical shifts to the protein ^{1}H^{13}C groups.

Prediction of 3D X-Filtered NOESY Spectra. We are currently using a very simple and fast procedure for predicting the 3D X-filtered NOESY spectrum for a given binding pose. The predicted spectra are based on user-defined distance cutoffs. Effective distances[38,39] between ligand ^{1}H groups and protein ^{13}C-attached ^{1}H groups were computed for each pose, neglecting the effects of fast internal methyl rotations.[39] For all test cases, upper bound cutoffs of 2.5, 3.0, 4.0, and 5.0 Å were used to predict very strong, strong, medium, and weak NOEs, respectively. As with the experimental NOES, the predicted intensities were assigned integer intensity values of 4 (very strong), 3 (strong), 2 (medium), or 1 (weak). Using the above-mentioned cutoffs, for all test systems a greater number of peaks were predicted for the target poses than were experimentally observed. The number of predicted peaks is determined by the cutoffs; for reasons discussed below, we generally want the number of predicted peaks to be greater than or equal to the number of observed peaks (see Spectrum Matching and Pose Scoring section below) for plausible poses. The experimentally determined ligand ^{1}H resonance assignments were used for the predicted spectra. While chemical shift predictions for the protein ^{1}H^{13}C groups could potentially be obtained by a variety of approaches,[40−44] we based most protein chemical shift predictions simply on the mean residue/atom chemical shifts for diamagnetic proteins available from the BMRB. The BMRB-derived shifts can be overwritten with any actual experimental (or more accurately predicted) chemical shifts that are available. Experimentally determined protein ^{1}H and ^{13}C resonance assignments were used for the predicted spectra in some tests.

Spectrum Matching and Pose Scoring. The predicted peaks are associated with specific protein ^{1}H^{13}C groups, whereas the observed peaks, in general, are not. The first step in matching the observed and predicted 3D X-filtered NOESY spectra is to identify protein ^{1}H^{13}C groups in the experimental data set. This is accomplished by grouping the observed peaks using the observed ^{1}H (F_1) and ^{13}C (F_2) chemical shift positions, as illustrated below. This procedure reduces the problem of matching peaks to peaks to one of matching ^{1}H^{13}C groups to ^{1}H^{13}C groups. For a given pose, we obtain the optimal self-consistent matching between the patterns of observed and predicted peaks. Due to the combinatorial complexity ($N!$), the search for the optimal matching cannot be done exhaustively. Fortunately, the matching problem described above can be cast as an equally partitioned bipartite graph weighted matching problem,[45] which can be solved deterministically in polynomial ($O(N^3)$) time.

An equally partitioned bipartite graph is a graph whose nodes are partitioned into two subsets, each containing N nodes. A completely connected bipartite graph is shown in Figure 1A, wherein each node k in one subset is connected by an edge to each node q in the other subset. There are no edges between nodes in the same subset. Each edge is associated with an edge cost $C(k,q)$; the edge costs define the $N \times N$ cost matrix. A matching of an equally partitioned bipartite graph is a

(33) Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. *J. Comput.-Aided Mol. Des.* **2000**, *15*, 411−428.
(34) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27−38.
(35) Sheriff, S. Unpublished results.
(36) Kallen, J.; Welzenbach, K.; Ramage, P.; Geyl, D.; Kriwacki, R.; Legge, G.; Cottens, S.; Weitz-Schmidt, G.; Hommel, U. *J. Mol. Biol.* **1999**, *292*, 1−9.
(37) Cavanaugh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press: New York, 1996; pp 235−236.
(38) Constantine, K. L.; Friedrichs, M. S.; Detlefson, D.; Nishio, M.; Tsunakawa, M.; Furumai, T.; Ohkuma, H.; Oki, T.; Hill, S.; Bruccoleri, R. E.; Lin, P.-F.; Mueller, L. *J. Biomol. NMR* **1995**, *5*, 271−286.
(39) Fletcher, C. M.; Jones, D. N. M.; Diamond, R.; Neuhaus, D. *J. Biomol. NMR* **1996**, *8*, 292−310.
(40) Osapay, K.; Case, D. A. *J. Am. Chem. Soc.* **1991**, *113*, 9436−9444.
(41) Sitkoff, D.; Case, D. A. *J. Am. Chem. Soc.* **1997**, *119*, 12262−12273.
(42) Iwadate, M.; Asakura, T.; Williamson, M. P. *J. Biomol. NMR* **1999**, *13*, 199−211.
(43) Xu, X. P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321−333.
(44) Wang, B.; Brothers, E. N.; Vaart, A. v. d.; Merz, K. M., Jr. *J. Chem. Phys.* **2004**, *120*, 11392−11400.
(45) Papadimitriou, C. H.; Steiglitz, K. *Combinatorial Optimization: Algorithms and Complexity*; Dover Publications: Mineola, NY, 1982; pp 247−255.

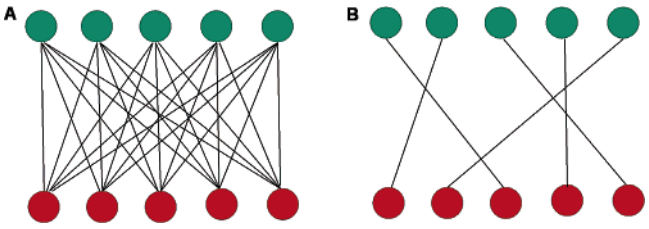


Figure 1. Bipartite graphs with $N = 5$. Node subsets are distinguished by color. (A) A completely connected graph. (B) A completely matched graph.

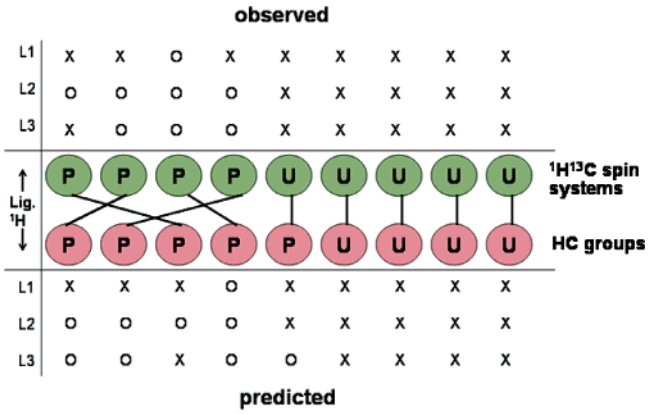


Figure 2. Equally partitioned bipartite graph representing a hypothetical instance of the 3D X-filtered NOESY bipartite graph-weighted matching problem, with $N = 9$. See the main text for additional details.

subset of edges with the property that no two edges share the same node. A complete matching is a matching with N edges (Figure 1B). The combinatorial optimization algorithm[45] finds an optimal complete matching; i.e., a permutation that minimizes the total (summed) cost of the complete matching. Bipartite graph matching has been applied previously to the protein NMR assignment problem.[46−48] Related nondeterministic approaches have also been applied to automated protein NMR assignment[49] and protein fold recognition/refinement.[50]

We now describe how pose scoring based on 3D X-filtered NOESY data is cast as an equally partitioned bipartite graph weighted matching problem. To assist in this description, a hypothetical example is depicted in Figure 2. In this example, the ligand contains three resolved, assigned ^{1}H groups (L1, L2, and L3) that each give rise to one or more observed NOEs. (Cases of ligand ^{1}H equivalence and accidental degeneracy are handled by simply placing the equivalent/degenerate protons into the same group. For example, the *ortho*- and *para*-protons on the *meta*-dichloro ring of compound **3** (Chart 1) are all assigned to 7.10 ppm in the bound state, and all three of these protons are placed into a single group.) In Figure 2, observed or predicted peaks are represented by "O" symbols, whereas missing peaks (not observed, or not predicted) are represented by "X" symbols. Green nodes containing a "P" represent experimental protein ^{1}H^{13}C groups that give rise to one or more observed NOE peaks. Red nodes containing a "P" represent protein HC groups that are predicted to give rise to one or more NOEs on the basis of the protein−ligand effective distances derived from the given pose.

In Figure 2, there are four green (observed) P nodes, and there are five red (predicted) P nodes. As noted above, the matching algorithm used[45] requires an equally partitioned graph. Also, we require the option of mapping any or all of the P nodes to "unassigned" if no other suitable match is found. To facilitate these requirements, "unassigned" nodes (identified by "U" in Figure 2) are added to both node subsets. In the example of Figure 2, there are eight experimentally observed peaks, and nine peaks are predicted by the given pose. One possible optimal complete matching, linking experimentally identified ^{1}H^{13}C groups with predicted HC groups, is also shown in Figure 2.

The process of identifying protein ^{1}H^{13}C groups in an experimental 3D X-filtered NOESY spectrum is illustrated by Figure 3, which shows data for the LFA-1/**3** complex. The peaks labeled with the blue asterisk all have nearly identical F_1 and F_2 chemical shifts of ∼0.02 and ∼70.6 ppm (20.6 ppm unaliased), respectively. Grouping of these peaks on the basis of their F_1 and F_2 chemical shifts identifies a ^{1}H^{13}C group in the experimental spectrum and defines a node on the observed side of the bipartite graph. (This group is the upfield δ-methyl of Leu302.)

In designing a function $C(k,q)$ to define the edge costs, one must account for experimental peaks that are not predicted and for predicted peaks that are not observed. We give more weight to observed peaks than to predicted peaks, and we give more weight to more intense peaks. This approach reflects experimental factors that can attenuate NOE intensities, and possible protein resonance chemical shift overlaps, that can reduce the number of observed peaks. We allow for significant uncertainty when matching intensities. Uncertainties arise due to the effects of spin diffusion, dynamics, and relaxation on the NOE intensities. Also, obtaining a suitable reference distance[51] for scaling intermolecular NOEs is problematic.

The elements of the asymmetric $N \times N$ cost matrix are given by

$$C(k,q) = \sum_i M_i(k,q); \quad i = 1, N_L \tag{1}$$

where N_L is the number of resolved, assigned ligand ^{1}H groups (e.g., $N_L = 3$ in Figure 2). Referring to Figure 2, the matching cost M between an experimental peak and a predicted peak is defined by the following expressions:

$$M_i(\mathrm{X,X}) = 0 \qquad \text{(no experimental peak, no predicted peak)} \tag{2}$$

$$M_i(\mathrm{O,X}) = K_1(\mathrm{IE}_i)^2 \qquad \text{(experimental peak present, no predicted peak)} \tag{3}$$

$$M_i(\mathrm{X,O}) = K_2(\mathrm{IP}_i)^2 \qquad \text{(no experimental peak, predicted peak present)} \tag{4}$$

M_i(O,O): (experimental peak present, predicted peak present)

If $\mathrm{IE}_i > \mathrm{IP}_i$

$$M_i(\mathrm{O,O}) = K_{\mathrm{H}}(f(\mathrm{H})/\sigma\mathrm{H})^2 + K_{\mathrm{C}}(f(\mathrm{C})/\sigma\mathrm{C})^2 + K_3(f'(\mathrm{I}))^2 \tag{5}$$

Else

$$M_i(\mathrm{O,O}) = K_{\mathrm{H}}(f(\mathrm{H})/\sigma\mathrm{H})^2 + K_{\mathrm{C}}(f(\mathrm{C})/\sigma\mathrm{C})^2 + K_4(f'(\mathrm{I}))^2$$

End If

IP_i and IE_i are the integer intensities of predicted and experimental peaks i, respectively. The σH and σC values are ^{1}H and ^{13}C chemical shift uncertainties. These can be set to user-defined values, or they can be set to some multiple of the relevant standard deviation, such as that obtained from the BMRB ($\mathrm{BMRB}_{\mathrm{SD}}$). The K's are adjustable parameters. Defaults values are $K_{\mathrm{H}} = 1$, $K_{\mathrm{C}} = 1$, $K_1 = 12$, $K_2 = 6$, $K_3 = 3$, and $K_4 = 1$. The intensity terms $f'(\mathrm{I})$ are implemented using two ("tight" and "loose") functional forms: $|\mathrm{IP}_i - \mathrm{IE}_i|$ or $\mathrm{argmax}(0,(|\mathrm{IP}_i - \mathrm{IE}_i| - 1))$. In the latter case ("loose" function), $f'(I)$ is non-zero only if the integer

(46) Xu, Y.; Xu, D.; Kim, D.; Olman, V.; Razumovskaya, J.; Jiang, T. *Comput. Sci. Eng.* **2002**, *4*, 50−60.
(47) Hus, J.-C.; Prompers, J. J.; Brüschweiler, R. *J. Magn. Reson.* **2002**, *157*, 119−123.
(48) Langemeade, C.; Donald, B. R. *J. Biomol. NMR* **2004**, *29*, 111−138.
(49) Bartels, C.; Güentert, P.; Billeter, M.; Wüthrich, K. *J. Comput. Chem.* **1997**, *18*, 139−149.
(50) Meiler, J.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 15404−15409.
(51) Neuhaus, D.; Williamson, M. P. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*; VCH Publishers: New York, 1989; p 109.

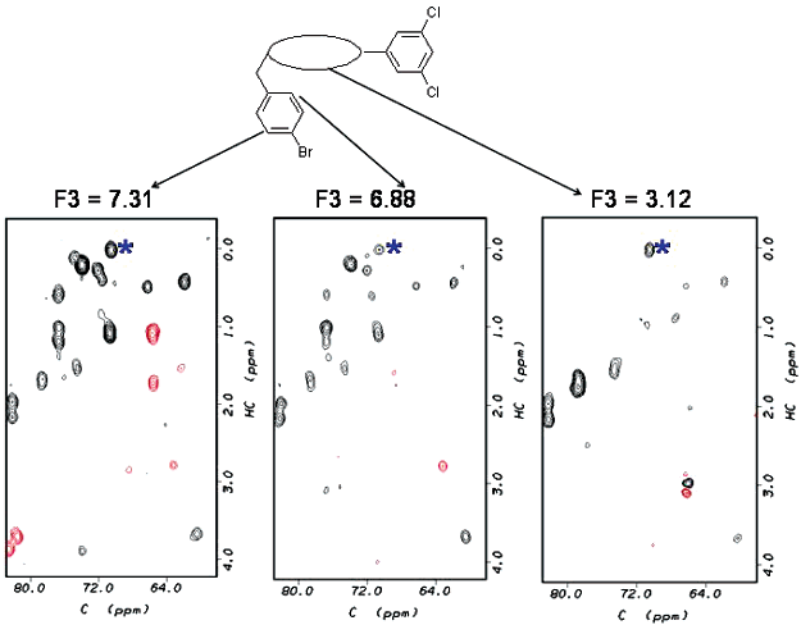


Figure 3. Portions of F_3 (ligand ^{1}H) planes from the 3D X-filtered NOESY (τ_m = 100 ms) of LFA-1/**3**. Planes at the ^{1}H chemical shifts of *meta*- (7.31 ppm) and *ortho*-protons (6.88 ppm) of the *para*-bromo ring, and of a proton group (3.12 ppm) from the proprietary core ring, are shown. Positive peaks (black contours) are aliased an even number of times, and negative peaks (red contours) are aliased an odd number of times, in F_2. Peaks labeled by the blue asterisk all arise from the same protein ^{1}H^{13}C group (example case).

intensity difference magnitude is 2 or 3. The ^{1}H chemical shift term f(H) is implemented as

$$\begin{aligned} f(\mathrm{H}) &= (\mathrm{H_E} - \mathrm{UH_P}) && \text{for } \mathrm{H_E} > \mathrm{UH_P} \\ f(\mathrm{H}) &= 0 && \text{for } \mathrm{UH_P} \geq \mathrm{H_E} \geq \mathrm{LH_P} \\ f(\mathrm{H}) &= (\mathrm{LH_P} - \mathrm{H_E}) && \text{for } \mathrm{H_E} < \mathrm{LH_P} \end{aligned} \quad (6)$$

where H_E is the experimental ^{1}H chemical shift, UH_P is the upper bound on the predicted chemical shift ($UH_P = H_P$ (predicted ^{1}H chemical shift) + σH), and LH_P is the lower bound on the predicted ^{1}H chemical shift ($LH_P = H_P - \sigma$H). The ^{13}C chemical shift term is implemented in an analogous fashion. The total cost of a given pose corresponds to an optimal solution of the complete matching problem, which is a permutation ϵ of {1, 2, ..., N} that minimizes:

$$\mathrm{COST_{pose}} = \sum_j C(j,\epsilon(j)); \quad j = 1, N \quad (7)$$

Results

Target Poses. NMR assignments and details of the structure calculations for the NMR ensembles of the mFABP/**1** and LFA-1/**3** complexes are given in the Supporting Information. In the case of mFABP/**1**, a full NMR-based structure determination of the protein−ligand complex was performed, as described in the Supporting Information. The target pose was selected from this well-defined ensemble (Figure S1A, Supporting Information). After superposition over the protein backbone atoms, the average root-mean-square deviation (RMSD) to the mean coordinates for all ligand heavy atoms is 0.42 Å. The accuracy of this binding pose is supported by comparison to an X-ray structure[52] of a related protein, adipocyte lipid-binding protein (aLBP), in complex with **1** (Figure S1B, Supporting Information). The sequences of aLBP and mFABP are 65% identical, and they have similar binding pockets.[53] The protein backbones superimpose with a RMSD of 1.19 Å. The main differences between the poses of **1** are due to protein conformational differences within the flexible ligand-entry "portal" regions of the two proteins.[53,54]

For LFA-1 with **2** bound, the X-ray structure of the complex[35] (Figure S2A, Supporting Information) served as the target binding pose. In the case of LFA-1 in complex with **3**, a well-defined binding pose ensemble (Figure S2B, Supporting Information) was determined using an available X-ray structure of LFA-1 (PDB entry 1CQP) and protein−ligand NOE restraints, as described in the Supporting Information. After superposition over the protein backbone atoms of residues 129−306, the average RMSD to the mean coordinates for all ligand heavy atoms is 0.19 Å. The target binding pose (Figure S2B, Supporting Information) was selected from this ensemble. An X-ray structure of LFA-1 in complex with a compound that is similar to compounds **2** and **3** has been described.[55]

Trial Poses. For the trial poses, our main goal was to obtain a wide sampling, in terms of RMSDs to the target poses, within the known binding pockets. Both proteins contain only one suitable pocket for high-affinity binding to organic compounds in the relevant size range, so alternate binding sites were not considered in the generation of trial poses. For mFABP/**1**, trial poses were derived both from XPLOR-based[56] simulated annealing[34] and with DOCK.[33] The remaining 20 (non-target) structures from the original well-resolved NMR ensemble were retained as trial poses; the minimum and maximum RMSDs to the target pose are 0.20 and 0.78 Å, respectively, for this set. (Unless stated otherwise, RMSDs correspond to ligand heavy-

(52) Jacobson, B. L. Unpublished results.

(53) Constantine, K. L.; Friedrichs, M. S.; Wittekind, M.; Jamil, H.; Chu, C.-H.; Parker, R. A.; Goldfarb, V.; Mueller, L.; Farmer, B. T., II. *Biochemistry* **1998**, *37*, 7965−7980.
(54) Hodsdon, M. E.; Cistola, D. P. *Biochemistry* **1997**, *36*, 2278−2290.
(55) Last-Barney, K.; Davidson, W.; Cardozo, M.; Frye, L. L.; Grygon, C. A.; Hopkins, J. L.; Jeanfavre, D. D.; Pav, S.; Qian, C.; Stevenson, J. M.; Tong, L.; Zindell, R.; Kelly, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 5643−5650.
(56) Brünger, A. T. *X-PLOR Version 3.1 Manual*; Yale University Press: New Haven, CT 1992.

Table 1. NOE Matching for mFABP/**1** with Ideal Synthetic Data

SDM[a]	$f'(I)$[b]	r[c]	NC_{targ}[d]	NW_{targ}[e]	C_{max}[f]	$RMSD_{max}$[g]	NC_{max}[h]	NW_{max}[i]
0.05[j]	**T**	**0.988**	**73**	**0**	**12335.8**	**6.66**	**19**	**23**
0.05	L	0.988	73	0	12233.4	6.66	19	22
0.25	T	0.973	73	0	8750.7	4.22	14	31
0.25	L	0.975	73	0	8496.0	4.22	15	31
0.50	T	0.951	71	2	7000.2	4.14	19	34
0.50	L	0.954	71	2	6700.5	4.14	19	33
0.75	T	0.931	71	2	5839.7	4.14	10	46
0.75	L	0.936	67	6	5562.7	4.14	11	44
1.00	T	0.914	71	2	5073.8	4.23	6	49
1.00	L	0.961	67	6	4812.7	4.23	6	49
1.25	T	0.906	71	2	4597.0	4.23	7	50
1.25	L	0.911	67	6	4344.0	4.23	7	50
1.50	T	0.900	67	6	4388.7	5.53	8	50
1.50	L	0.905	63	10	4110.9	5.53	10	48
2.00	T	0.889	64	9	4108.6	5.53	6	53
2.00	L	0.892	58	15	3829.0	5.53	8	51
5.00	T	0.825	59	14	3451.3	5.23	5	56
5.00	L	0.828	52	21	3158.7	5.23	5	57

[a] Factor used to multiply the chemical shift standard deviations. [b] T, tight $f'(I)$; L, loose $f'(I)$. [c] Correlation coefficient between $COST_{pose}$ and the RMSD to the target pose. [d] Number of correctly assigned $^1H^{13}C$ groups in the target pose. [e] Number of incorrectly assigned $^1H^{13}C$ groups in the target pose. [f] Maximum $COST_{pose}$. [g] RMSD (Å) between the pose with the maximum $COST_{pose}$ and the target pose. [h] Number of correctly assigned $^1H^{13}C$ groups in the pose with the maximum $COST_{pose}$. [i] Number of incorrectly assigned $^1H^{13}C$ groups in the pose with the maximum $COST_{pose}$. [j] Data set (bold) used to produce Figure 4.

atom RMSDs after superposition of ordered protein backbone atoms; i.e., these are RMSDs in the reference frame of the protein.) By re-annealing the original 21 NMR structures with only 10 arbitrarily picked restraints, an additional 21 trial poses were generated, with minimum and maximum RMSDs to the target pose of 0.73 and 2.10 Å, respectively. DOCK was used to generate trial poses by docking the ligand into the protein coordinates of the target pose using two protocols: (1) with the internal conformation of the ligand fixed to that observed for the trial pose and (2) with full conformational flexibility for the ligand. From the DOCK runs, 20 rigidly docked and 400 flexibly docked trial poses were selected, with RMSDs to the trial pose ranging from 0.28 to 6.66 Å. In total, 461 trial poses were used for the mFABP/**1** tests.

In the case of LFA-1/**2**, trial poses were generated using DOCK.[33] The ligand was docked into the 10 sets of protein coordinates derived from PDB entry 1CQP. For each of the 10 protein coordinate sets, 50 trial poses generated by rigid ligand docking and 100 trial poses generated by flexible ligand docking were obtained. A total of 1500 trial poses were used for the LFA-1/**2** tests. The minimum and maximum RMSDs to the target pose are 0.22 and 8.52 Å, respectively.

For LFA-1/**3**, trial poses were generated using DOCK[33] in a manner analogous to that used for LFA-1/**2**. However, in this case, the ligand was docked into a single protein coordinate set — that of the target pose. Fifty trial poses generated by rigid ligand docking and 300 trial poses generated by flexible ligand docking were generated. In total, 350 trial poses were used for the LFA-1/**3** tests. The minimum and maximum RMSDs to the target pose are 0.18 and 7.63 Å, respectively.

NOE Matching for mFABP/1: Ideal Data. Our initial tests of the algorithm were designed to ensure that it behaves as expected with idealized data. This was accomplished by generating a synthetic "experimental" 3D X-filtered NOESY spectrum based on (1) intermolecular $^1H-^1H$ distances observed in the mFABP/**1** target pose and (2) generating a complete set of synthetic protein resonance assignments by randomly choosing chemical shift values for each ^{13}C and ^1H group in the protein. The synthetic chemical shifts were selected from a uniform distribution ($\pm 2BMRB_{SD}$) for each atom type. Unless noted otherwise, an effective distance upper bound cutoff of 5.0 Å was used for generating all of the predicted spectra discussed in this article. For the 462 mFABP/**1** poses (target and trial poses), the minimum, maximum, and average number of predicted peaks are 148, 198, and 171.4, respectively. For the target pose, 179 peaks are predicted, distributed among 73 $^1H^{13}C$ groups.

In this idealized case, the target pose always yields $COST_{pose} = 0$ for all parameter values, since all of the $^1H^{13}C$ groups derived from the synthetic experimental spectrum have matches within the chemical shift and intensity tolerances in the predicted spectrum for this pose. In addition, we expect all (or nearly all) of the matches for the target pose to correspond to the correct assignment, given sufficiently small σH and σC values. The predicted spectra for the trial poses are generally not expected to yield $COST_{pose} = 0$, since the set of $^1H^{13}C$ groups involved in predicted NOEs will generally differ among poses, and some of the $^1H^{13}C$ groups common to both trial and target poses have different predicted intensities.

Table 1 reports the results obtained by varying the intensity matching function $f'(I)$ ("tight" and "loose") and by varying the σH and σC (uncertainty) values, for idealized mFABP/**1** data. The K parameters (eqs 3−5) were fixed at their default values for all of the tests reported in this article. As mentioned, the uncertainties are defined as the relevant standard deviation multiplied by a standard deviation multiplier (SDM). For these tests, the ^1H and ^{13}C standard deviations were set to the BMRB values, and SDM was varied between 0.05 and 5.00. As expected, the best correlations between $COST_{pose}$ and RMSD values in this case are obtained using the small uncertainties. $COST_{pose}$ is always 0 for the target pose, and all $^1H^{13}C$ groups are correctly assigned for SDM values of 0.05 and 0.25. The correlation between $COST_{pose}$ and RMSD degrades, and fewer $^1H^{13}C$ groups are correctly assigned for the target pose, as the chemical shift uncertainties are increased.

Figure 4 shows a plot of $COST_{pose}$ versus RMSD for SDM = 0.05 and "tight" intensity scoring. While there is no a priori reason to expect a strictly linear correlation between $COST_{pose}$

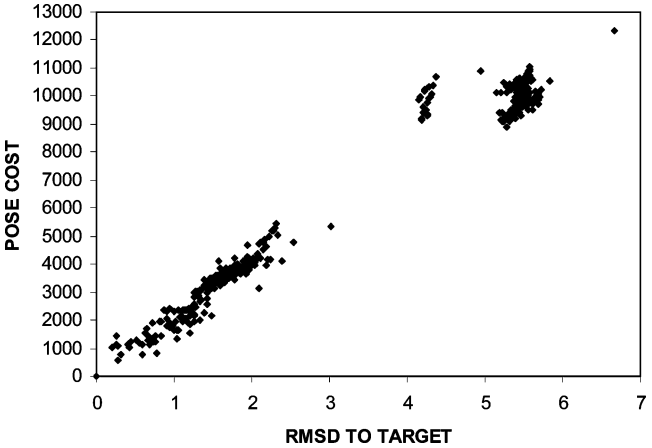


Figure 4. $COST_{pose}$ versus the RMSD (Å) to the target pose for mFABP/**1** with synthetic (ideal) experimental and predicted 3D X-filtered NOESY data, using the data set corresponding to the bold row in Table 1. The symbol for the target pose is at RMSD = 0, $COST_{pose}$ = 0.

and RMSD, a very good correlation is observed for RMSD values out to ~3 Å. This result demonstrates that, with ideal (i.e., completely accurate) data, the method as implemented can discriminate among binding poses with high resolution. The linear correlation breaks down for larger RMSD values. This result was also anticipated since, in general, a correlation between $COST_{pose}$ and RMSD should not be expected among poses that are far from correct.

NOE Matching for mFABP/1: Real Data and Experimental Protein NMR Assignments. For the next series of tests, we used the real experimental 3D X-filtered NOESY data set. After interactive analysis, the experimental 3D X-filtered NOESY spectrum of the mFABP/**1** complex contained 140 peaks, of which 126 have been assigned by interactive analysis. The peaks were grouped into 54 protein $^1H^{13}C$ groups, of which 48 have been assigned. Predicted chemical shifts were set to the known NMR assignments for the assigned groups; otherwise, the average ^1H and ^{13}C chemical shift for that group was taken from the BMRB (diamagnetic protein statistics) and used as the predicted chemical shift. Thus, for these tests, we have highly accurate "predicted" chemical shifts for most protein $^1H^{13}C$ groups. For unassigned prochiral protons/groups, the BMRB predicted shift was arbitrarily selected from the prochiral pair. (The average BMRB chemical shifts are very similar for all prochiral pairs.)

For these tests, the ^1H and ^{13}C standard deviations were set to 0.04 and 0.4 ppm, respectively, for $^1H^{13}C$ groups that are experimentally assigned; otherwise, $BMRB_{SD}$ values were used. Table S1 (Supporting Information) summarizes the results obtained by varying SDM and $f'(I)$ for this case. The maximum value for the correlation coefficient between $COST_{pose}$ and RMSD (r = 0.977) is obtained for SDM = 0.50, $f'(I)$ "tight". While a finer-grained sampling of SDM values may yield slightly higher correlation coefficients, the results presented indicate that an SDM value of 0.50 is near optimal for this test case. Figure 5A shows a plot of $COST_{pose}$ versus RMSD for the SDM = 0.50, $f'(I)$ "tight", test case. A good correlation is observed for RMSD values of ~3 Å or less. Figure 5B compares the target pose and the pose with the minimum $COST_{pose}$; the RMSD between these two poses is 0.95 Å (Table S1).

NOE Matching for mFABP/1: Real Data and Predicted Protein Chemical Shifts Set to BMRB Averages. For these

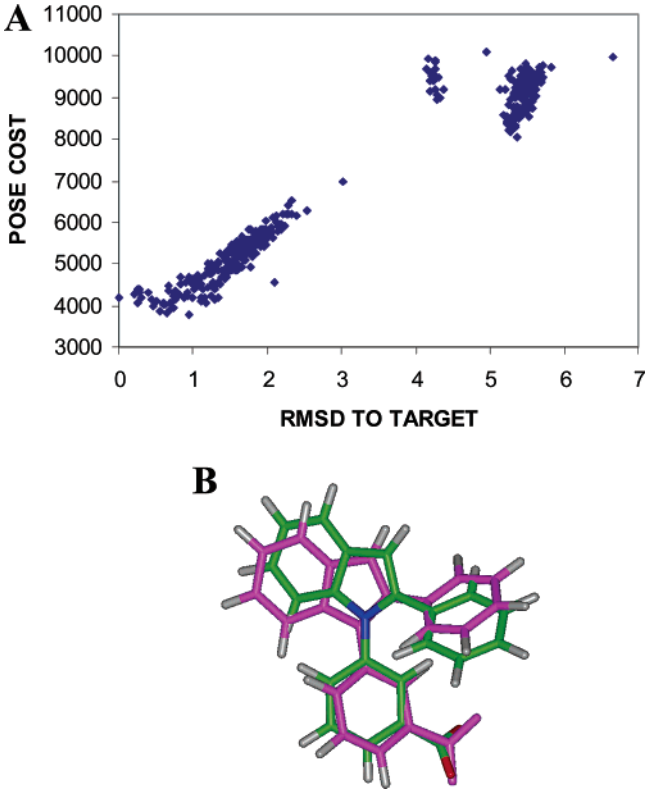


Figure 5. (A) $COST_{pose}$ versus the RMSD (Å) to the target pose for mFABP/**1** with real experimental 3D X-filtered NOESY data and experimental protein NMR assignments for most predicted chemical shifts (see main text), using the data set corresponding to the bold row (largest correlation coefficient) in Table S1. The symbol for the target pose is at RMSD = 0. (B) Superposition of target pose and the minimum cost pose (non-hydrogen atoms colored magenta) from (A).

tests, the real experimental data for mFABP/**1** were used (described above). For all $^1H^{13}C$ groups, predicted chemical shifts were set to the average ^1H and ^{13}C chemical shifts for that group from the BMRB, and $BMRB_{SD}$ values were used for the chemical shift standard deviations. For these tests, we have relatively low accuracy and precision for the predicted chemical shifts.

Table 2 summarizes the results obtained by varying SDM and $f'(I)$ for this case. The maximum value for the correlation coefficient between $COST_{pose}$ and RMSD (r = 0.912) is obtained for SDM = 0.50; as with tests using the experimental protein assignments, SDM = 0.50 appears to be a near-optimal value for this case. Unlike the preceding cases, there are significant degradations of the correlation coefficients between $COST_{pose}$ and RMSD at both low and high values of SDM. At SDM = 0.05, a large majority of the $^1H^{13}C$ groups that are assigned have incorrect assignments. As SDM increases, more $^1H^{13}C$ groups get assigned, since the P-to-P edge costs decrease, while the P-to-U edge costs remain constant (Figure 2; eqs 2−6). At the near-optimal value of SDM = 0.50, a majority of the assigned groups have correct assignments for poses with relatively low $COST_{pose}$ values. At SDM = 5.00, a majority of the assigned groups have incorrect assignments for all poses. In these cases, the intensity terms play a more dominant role; i.e., $^1H^{13}C$ groups may be assigned chemical shift values that are well outside of their expected range. Figure 6A shows a plot of $COST_{pose}$ versus RMSD for the SDM = 0.50, $f'(I)$ "tight", test case. In this case, even with protein chemical shift

Table 2. NOE Matching for mFABP/**1** with Real Data and BMRB-Derived Shifts

SDM[a]	f'(I)[b]	r[c]	C_{max}[d]	C_{min}[e]	RMSD$_{min}$[f]	NC$_{min}$[g]	NW$_{min}$[h]	NC$_{targ}$[i]	NW$_{targ}$[j]	R_{targ}[k]
0.05	T	0.511	14543.0	13288.1	1.65	4	32	5	30	172/462
0.05	L	0.499	14502.0	13243.1	1.65	4	33	5	30	173/462
0.25	T	0.857	9778.3	5865.3	0.95	26	19	28	16	13/462
0.25	L	0.854	9609.5	5661.3	0.95	26	19	28	16	15/462
0.50[l]	**T**	**0.912**	**7163.0**	**3259.2**	**0.75**	**30**	**20**	**35**	**14**	**14/462**
0.50	L	0.911	6943.3	3066.5	0.75	28	22	31	18	18/462
0.75	T	0.892	5891.2	2546.2	0.75	29	21	30	20	15/462
0.75	L	0.888	5655.2	2355.2	0.75	27	23	25	25	15/462
1.00	T	0.875	5210.9	2322.7	0.75	29	21	28	22	20/462
1.00	L	0.868	4990.9	2131.7	0.75	27	23	24	27	22/462
1.25	T	0.861	4811.5	2228.2	0.75	29	22	24	27	25/462
1.25	L	0.851	4577.7	2039.8	0.75	27	24	23	28	29/462
1.50	T	0.849	4460.2	2176.1	0.75	28	23	24	26	29/462
1.50	L	0.837	4265.3	1986.2	0.75	26	25	23	28	32/462
2.00	T	0.832	4155.2	2132.0	0.75	27	25	23	28	34/462
2.00	L	0.818	3962.5	1938.7	0.75	25	28	21	30	39/462
5.00	T	0.774	3510.9	1992.3	1.77	12	42	24	29	74/462
5.00	L	0.735	3312.9	1797.1	1.77	14	41	21	32	93/462

[a] Factor used to multiply the chemical shift standard deviations. [b] T, tight f'(I); L, loose f'(I). [c] Correlation coefficient between COST$_{pose}$ and the RMSD to the target pose. [d] Maximum COST$_{pose}$. [e] Minimum COST$_{pose}$. [f] RMSD (Å) between the pose with the minimum COST$_{pose}$ and the target pose. [g] Number of correctly assigned ^{1}H^{13}C groups in the pose with the minimum COST$_{pose}$. [h] Number of incorrectly assigned ^{1}H^{13}C groups in the pose with the minimum COST$_{pose}$. [i] Number of correctly assigned ^{1}H^{13}C groups in the target pose. [j] Number of incorrectly assigned ^{1}H^{13}C groups in the target pose. [k] Target pose rank in terms of COST$_{pose}$. [l] Data set (bold) used to produce Figure 6.

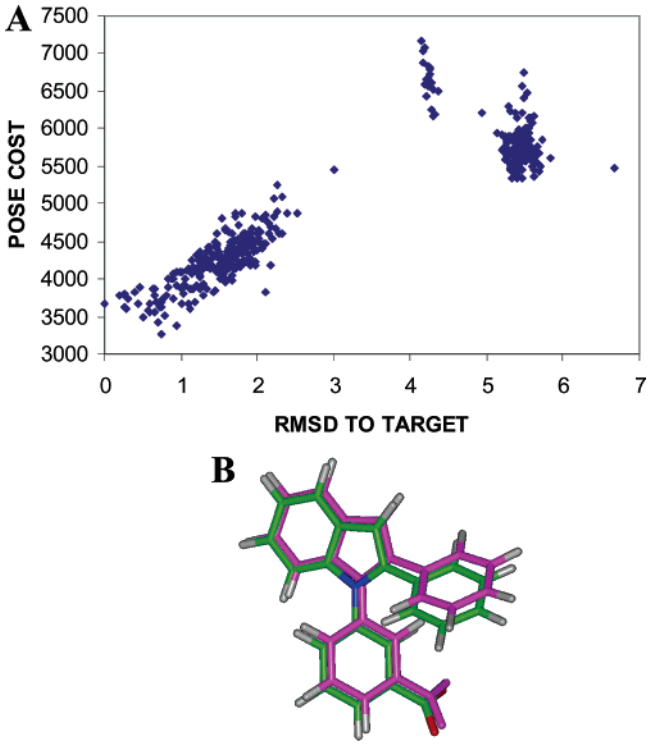


Figure 6. (A) COST$_{pose}$ versus the RMSD (Å) to the target pose for mFABP/**1** with real experimental 3D X-filtered NOESY data and predicted protein chemical shifts set to the corresponding BMRB average values, using the data set corresponding to the bold row in Table 2. (B) Superposition of target pose and the minimum cost pose (non-hydrogen atoms colored magenta) from (A).

predictions of low accuracy and low precision, a good correlation is observed for RMSD values of ~3 Å or less. Figure 6B compares the target pose and the pose with the minimum COST$_{pose}$ (RMSD = 0.75 Å; Table 2).

To test the expected performance of NOE matching on data with less sensitivity, lower intensity peaks were systematically deleted (in increments of 10%) from both the experimental and predicted 3D X-filtered NOESY peak lists. NOE matching was then applied to the resulting peak lists using SDM = 0.50, f'(I) "tight". The results of these tests are summarized in Table S2 (Supporting Information). The results indicate that, in the case of mFABP/1 with real experimental data, NOE matching results degrade gradually and can identify the correct pose when the minimum observable NOE intensity corresponds to a distance greater than or equal to ~3.5 Å.

NOE Matching for LFA-1/2: Real Data and Predicted Protein Chemical Shifts Set to BMRB Averages. This test case is more challenging than the mFABP/**1** system, for several reasons. Fewer peaks (69) are present in the 3D X-filtered NOESY spectrum of LFA-1/**2**; these have been associated with 51 experimental ^{1}H^{13}C groups. Compound **2** has greater internal flexibility than compound **1** (see Chart 1). Also, no protein−ligand NOEs have been detected for any of the methine or methylene protons of compound **2**; as a result, there is no direct information on the placement of the compound **2** core (Chart 1). With many protein−ligand complexes, we have observed that ligand methine and methylene protons often do not yield intermolecular NOEs due to intrinsically broad lines and weak signal intensities. Therefore, the methine and methylene protons of compound **2** were excluded entirely from the NOE matching calculations. Also, we chose to create trial poses for this system by docking into protein structures that are different from the target pose (see Methods section). The X-ray structure[35] of LFA-1/**2** was determined shortly after the NMR studies were initiated; experimental protein NMR assignments were therefore not completed and verified.

Real NMR data and BMRB-derived predicted protein chemical shifts and standard deviations were used for the NOE matching tests. Using a 5.0 Å upper-bound distance cutoff, the minimum, maximum, and average numbers of predicted peaks are 33, 87, and 64, respectively, over the trial and target poses. For the target pose, 87 peaks are predicted, distributed among 62 ^{1}H^{13}C groups.

The NOE matching results for this system are summarized in Table S3 (Supporting Information). The best correlation coefficient between COST$_{pose}$ and RMSD (r = 0.858) is obtained for SDM = 0.25, f'(I) = "tight". Figure 7A shows a plot of COST$_{pose}$ versus RMSD, and Figure 7B compares the

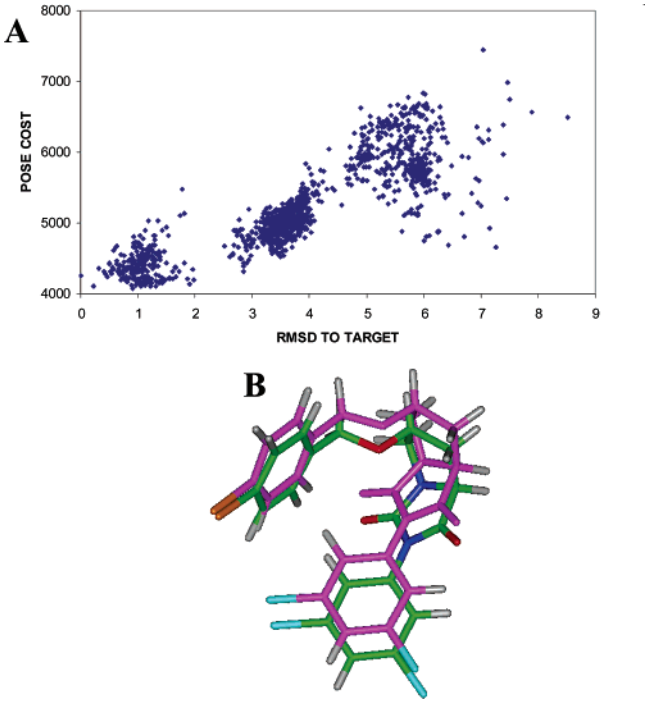


Figure 7. (A) $COST_{pose}$ versus the RMSD (Å) to the target pose for LFA-1/**2** with real experimental 3D X-filtered NOESY data and predicted protein chemical shifts set to the corresponding BMRB average values, using the data set corresponding to the bold row (largest correlation coefficient) in Table S3. (B) Superposition of the target pose and the minimum cost pose (carbon, nitrogen, and oxygen atoms colored magenta) from (A).

target pose and the pose with the minimum $COST_{pose}$ (RMSD = 0.91 Å; Table S3, Supporting Information). In four cases (Table S3, Supporting Information), the target pose is also the one that has the minimum $COST_{pose}$ value. The target pose is well-ranked in all cases except SDM = 0.05, and the pose with the minimum $COST_{pose}$ value is similar to the target pose in all cases except SDM = 0.05 and SDM = 5.00 (Table S3).

While NOE matching was successful in this case, the overall results with LFA-1/**2** are not as good as those obtained with mFABP/**1**. The correlation between $COST_{pose}$ and RMSD is not as high, especially for poses similar to the target pose (compare Figure 7A with Figure 6A). For LFA-1/**2**, some poses that are quite different from the target pose are relatively well-ranked (see Figure 7A); such poses may be problematic in the absence of a known target pose. These results point to the need for adequate conformational sampling and for methods for evaluating the results of NOE matching that are independent of a known target pose (see Discussion section).

NOE Matching for LFA-1/3: Real Data and Predicted Protein Chemical Shifts Set to BMRB Averages. The final test system (LFA-1/**3**) utilizes an analogue of **2** with better properties for NOE matching. The exact chemical structure of **3** (Chart 1) cannot be revealed at this time. Compound **3** is a more favorable case than compound **2** since (1) it is less flexible and (2) the core moiety of compound **3** contains two methyl groups that give rise to protein−ligand NOE interactions. Also in this case, we used the protein coordinates of the target pose for generating trial poses, rather than using alternate protein structures.

A total of 74 peaks were picked in the experimental 3D X-filtered NOESY spectrum of LFA-1 in complex with **3**, and 71 of these peaks were subsequently assigned by interactive

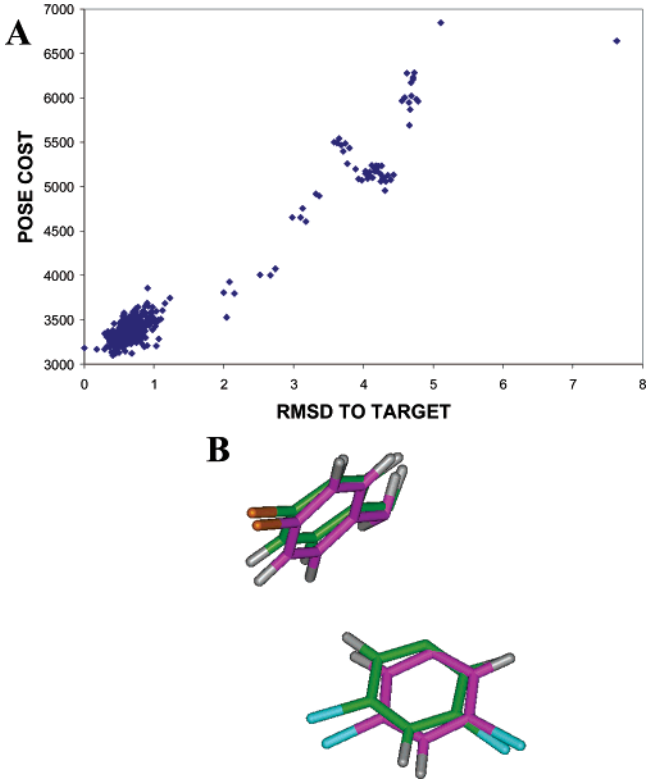


Figure 8. (A) $COST_{pose}$ versus the RMSD (Å) to the target pose for LFA-1/**3** with real experimental 3D X-filtered NOESY data and predicted protein chemical shifts set to the corresponding BMRB average values, using the data set corresponding to the bold row (largest correlation coefficient) in Table S4. (B) Superposition of the target pose and the minimum cost pose (carbon, nitrogen, and oxygen atoms colored magenta) from (A), showing only the non-proprietary moieties.

analysis. The experimental peaks were grouped into 44 protein $^{1}H^{13}C$ groups; 41 of these have been assigned interactively. Using a 5.0 Å upper-bound distance cutoff, the minimum, maximum, and average numbers of predicted peaks are 92, 139, and 125, respectively, over the trial and target poses. For the target pose, 122 peaks are predicted, distributed among 68 $^{1}H^{13}C$ groups.

Table S4 (Supporting Information) summarizes the results obtained using the real experimental data and BMRB-derived protein chemical shift predictions and standard deviations. Overall, excellent results are obtained for this system. For SDM values between 0.25 and 1.00 (inclusive), the correlation coefficients between $COST_{pose}$ and the RMSD values are all >0.900, with the best correlation (r = 0.973) obtained with the near-optimal parameters SDM = 0.25, f'(I) = "tight". With these parameters, the target pose is ranked 11/351. Figure 8A shows a plot of $COST_{pose}$ versus RMSD using the near-optimal parameters. Figure 8B shows a comparison of the target pose and the pose with the lowest $COST_{pose}$ value (RMSD = 0.41 Å). For this case, NOE matching yielded a high-resolution identification of the binding pose without utilizing any experimental protein NMR assignments.

Discussion

Assessment of Protein−Ligand NOE Matching. The test cases that were used in this study were chosen on the basis of the availability of (1) good-quality 3D X-filtered NOESY spectra and (2) well-defined target poses. Additionally, in two of the

three test cases, most of the peaks in the 3D X-filtered NOESY spectra were previously assigned by traditional methods, allowing an evaluation of the ability of NOE matching to reproduce known assignments. The specific compositions and distributions of residue types in the binding pockets were not used as a basis for choosing the test cases. We expect, in general, that small-molecule binding sites in proteins are sufficiently heterogeneous[21] to permit informative pose scoring by NOE matching. While it is not possible to prove this claim on the basis of three test cases, our results point to wide applicability of the NOE matching approach. It may fail for simple ligands that yield only a few NOEs, or for those that populate multiple binding modes.

Protein conformational changes are important considerations for any ligand docking/scoring protocol. While not attempting to address this issue in any general way here, we stress that adequate conformational sampling is a prerequisite for NOE matching. LFA-1/**2** is an illustrative example of this. For LFA-1/**2**, it was necessary to moderately vary the initial protein structure (1CQP). Specifically, Leu 302 in the flexible C-terminal α-helix of LFA-1[57] had to be moved in order to sample the correct binding pose of compound **2**. In the absence of a known binding pose, methods for evaluating the adequacy of the conformational sampling will be required.

Heterogeneous residue distributions in binding pockets[21] are the basis for molecular recognition and, as noted above, are a basic assumption of the NOE matching approach. For mFABP/**1**, the experimentally assigned intermolecular NOEs involve 22 different residues and 12 different residue types, and for LFA-1/**3** they involve 17 different residues and 7 different residue types. Due to the multiple occurrences of some residue types in the binding pockets, and overlapping chemical shift ranges for most of the atom types present, in general one could not assign the protein $^{1}H^{13}C$ groups observed in the 3D X-filtered NOESY data using just chemical shift information. Methyl groups are considered as an illustrative example. The distributions of methyl-containing residues for the mFABP/**1** and LFA-1/**2** pockets are shown in Figures S3 and S4 (Supporting Information), respectively. Methyl groups with overlapping chemical shift ranges are diversely distributed throughout both binding pockets. These situations preclude assigning these groups on the basis of simple inspections of chemical shifts.

These considerations indicate that the successful pose characterizations obtained for all test cases are due to the information inherent in the *overall patterns* of NOEs observed. In addition, for the mFABP/**1** and LFA-1/**3** test cases, significant numbers of correct assignments were obtained for the target and low-cost poses, even when using the BMRB-derived chemical shift predictions (Tables 2 and S2−S4). This has implications for extending the method.

An indication that the method should be robust is the insensitivity to the exact choice of parameter values used. The *K* parameters (eqs 3−5) were not varied; they were fixed at values aimed at achieving two main goals: (1) giving more weight to observed peaks than to predicted peaks and (2) allowing P nodes to be matched to U nodes (Figure 2) if no suitable match is found in terms of chemical shifts and intensities. Since P-to-U matches contain intensity terms only (eqs 3 and 4), these terms need to be weighted more heavily than the intensity terms in P-to-P matches (eq 5), which contain both chemical shift and intensity matching terms.

The σH and σC values and the functional form of the intensity matching term $f'(I)$ in eq 5 were varied. Major differences were not typically observed between results obtained with "tight" versus those obtained with "loose" intensity matching; however, the "tight" matching function performed slightly better overall in terms of yielding correct assignments for the $^{1}H^{13}C$ groups. The σH and σC values determine the degree of chemical shift mismatch above which P nodes will be matched to U nodes. In the case of the ideal (completely accurate) synthetic "observed" and "predicted" data, increasing the σH and σC parameter values results in increasingly degraded performance (Table 1). As expected for this case, the target pose always obtains a $COST_{pose}$ of 0, since observed P nodes always match to predicted P nodes that contain the same peak pattern, shifts, and intensities. However, the correlation between $COST_{pose}$ and RMSD to the target pose decreases, and the number of correct assignments decreases for the target pose and low-cost poses, with increasing σH and σC. These results are expected, as the edge weights for alternate P-to-P matches decrease for the trial poses as the σH and σC values are increased.

When most protein resonance assignments are known experimentally, varying the σH and σC values from 0.05 to 5.0 had little effect on the results (Table S1). Overall, better results are obtained when known assignments are used relative to simply using the BMRB values for the predicted shifts (compare Tables S1 and 2). These results, along with the results obtained for the synthetic data sets, indicate that NOE matching is more robust and effective when the protein chemical shifts are known, or when they can be predicted more accurately (vide infra).

The most stringent tests of NOE matching are those that were performed using the real experimental 3D X-filtered NOESY data sets, with the predicted chemical shifts set to average values derived from the BMRB (Tables 2 and S2−S4). For our test cases, the near-optimal SDM values are 0.25 or 0.50 (case-dependent). Values of σH and σC that are too small produced unsatisfactory results, due to the inability to correctly match many of the observed shifts to the BMRB mean shifts. Large values of σH and σC also produced unsatisfactory results, since the ability to discriminate between possible assignments by chemical shift matching degrades; i.e., too many P-to-P matches are associated with low edge costs. At very large σH and σC values, the $COST_{pose}$ values asymptotically approach non-zero values (data not shown). A non-zero cost remains associated with each pose due to intensity mismatches.

The effects of deleting low-intensity subsets of peaks were tested (Table S2, Supporting Information). A gradual degradation in the performance NOE matching was observed. The target pose ranked well, and the pose with the lowest $COST_{pose}$ value was similar to target pose when as many as 80% of the observed and predicted peaks were deleted. These results indicate that NOE matching can succeed with lower sensitivity data. When 90% of the peaks were deleted (leaving 14 experimental and an average of 18 predicted peaks), NOE matching performed poorly (Table S2). With this limited number of peaks, the correct pose was not identified using BMRB-derived predicted chemical shifts. We note that ligand binding poses can be validated using a relatively small number of *assigned* protein−ligand NOEs;[28]

(57) Legge, G. B.; Kriwacki, R. W.; Chung, J.; Hommel, U.; Ramage, P.; Case, D. A.; Dyson, H. J.; Wright, P. E. *J. Mol. Biol.* **2000**, *295*, 1251−1264.

likewise, we expect NOE matching to require fewer peaks when some experimental assignments are available.

For the tests described in this article, the performance of the NOE matching procedure was judged primarily by comparisons to known target poses and (for two of the three systems) known protein NMR resonance assignments. As an initial recommendation for future applications, the SDM parameter may be optimized by sampling values between 0.10 and 1.00 and computing the correlation between $COST_{pose}$ and RMSD to a target pose. The $COST_{pose}$ versus RMSD plots should be examined for outliers, nonlinear relationships, and lack of correlation at high RMSD values. Procedures for optimizing and evaluating NOE matching in cases of unknown target poses with unknown protein resonance assignments are being explored; one possible approach is to use one or more low-cost poses in place of the target pose when computing the correlation between $COST_{pose}$ and RMSD values. Clustering methods[58,59] may be applied to the trial poses. Such an analysis will allow the identification of representative cluster members, which can then be subjected to more computationally demanding evaluations. Also, in de novo pose determinations using NOE matching, data on closely related ligand analogues can provide valuable information; e.g., the LFA-1/**3** results could be used to rule out the problematic (e.g., relatively low cost, high RMSD) poses obtained for LFA-1/**2**. Other possible approaches for discriminating among poses are discussed in the next section (Possible Extensions and Complementarities).

Protein−ligand NOE matching has advantages and limitations. Pose characterization assumes that suitable coordinate sets for the target protein are available. For NOE matching, two additional requirements must be fulfilled: (1) a sufficient number protein−ligand NOEs must be experimentally observed, and (2) poses that are similar to the true binding pose must be sampled in order to be recognized. For the latter, docking methods[60] that thoroughly sample pose space, including the protein conformation, can be used. Extensive conformational sampling requires an approach that can rank many thousands of trial poses; NOE matching meets this requirement.

NOE matching has a number of important features. It utilizes, from the outset, all of the available protein−ligand NOEs (assigned or unassigned) arising from all $^{1}H^{13}C$ groups in a uniformly $^{13}C/^{15}N$-labeled protein. Predicted protein chemical shifts can be overwritten with any available protein chemical shift assignments, affording a direct way of incorporating such information. Similarly, any assigned intra-ligand and protein−ligand NOEs can be used as explicit restraints to direct the sampling of poses.[28] Protein−ligand NOEs can be observed under both fast and slow exchange conditions. Therefore, protein−ligand NOE matching is applicable to most exchange regimes, the exception being when severe exchange-broadening cannot be eliminated. By defining nodes in terms of HC groups instead of individual peaks, a degree of self-consistency is automatically imposed on matching; i.e., all of the experimental NOEs known to arise from the same experimental $^{1}H^{13}C$ group are matched, as a set, to one particular HC group in the given pose. Direct peak-to-peak matching does not impose this restriction.

Possible Extensions and Complementarities. Our overall goal is to build a widely applicable framework that facilitates the rapid evaluation of ligand binding poses. Ideally, this framework should support alternative strategies suitable to the cases at hand, and it should ultimately facilitate the study of large protein−ligand complexes. As presented in this article, protein−ligand NOE matching is primarily a *top-down* strategy[12] that uses minimal experimental data and that does not require sequence-specific protein resonance assignments. It is also primarily a scoring or filtering strategy,[61] as opposed to a restrained or directed search strategy based on explicit restraints. In this section, we comment on possible extensions to NOE matching and on how this framework can facilitate pose evaluations, including traditional bottom-up strategies and strategies based on alternate isotopic labeling schemes.

Intermolecular NOE data are typically acquired early in the process of "bottom-up" pose determination, since there is no point in continuing if these are not observed. NOE matching can be performed before protein assignments are obtained. The bottom-up process can be continued in parallel. As protein assignments are obtained, they can be used to assign some of the observed intermolecular NOEs, and hence to restrict the search space[28,62] and provide accurate predicted chemical shifts. (While not utilized here, intermolecular NOEs involving backbone amide groups could also be used.[28]) Thus, while at a given time only the assigned subset of protein resonances can be used to derive unambiguous restraints, all of the peaks in the 3D X-filtered NOESY are used to evaluate and identify the most consistent pose(s).

Resonance assignments, especially side-chain assignments, can be difficult or impossible to obtain for larger proteins. Also, the sensitivity of the 3D X-filtered NOESY experiment with uniformly $^{13}C/^{15}N$-labeled samples degrades significantly with larger proteins. For larger proteins, the observation and identification of protein−ligand NOEs may be accomplished by different approaches. Stabilizing agents[63,64] may allow spectra to be recorded at higher temperatures, which can enhance sensitivity. Selective (non-uniform) isotopic labeling strategies can be used to increase both spectral sensitivity and the information content of NOE peaks. NOE matching is being modified to use information from 2D NOESY and/or 3D X-filtered NOESY data sets acquired using multiple, non-uniformly labeled protein samples.

Once a smaller set of the most consistent poses are identified, additional approaches become feasible. Prediction of the absolute chemical shifts for each pose could be used to rescore selected poses. These poses could be filtered using ligand proton chemical shift changes predicted by quantum-mechanical methods.[22] If some protein assignments are available, the consistency of binding poses with observed protein atom chemical shift changes could be evaluated as well.[25] A small set of consistent poses could be subject to more thorough analysis using more

(58) Han, J.; Kamber, M. *Data Mining: Concepts and Techniques*; Morgan Kaufmann: New York, 2001; pp 335−393.
(59) Hyvönen, M. T.; Hiltunen, Y.; El-Deredy, W.; Ojala, T.; Vaara, J.; Kovanen, P. T.; Ala-Korpela, M. *J. Am. Chem. Soc.* **2001**, *123*, 810−816.
(60) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.
(61) Dobridumov, A.; Gronenborn, A. M. *Proteins: Struct., Funct. Genet.* **2003**, *52*, 18−32.
(62) Lugovskoy, A. A.; Degterev, A. I.; Fahmy, A. F.; Zhou, P.; Gross, J. D.; Yuan, J.; Wagner, G. *J. Am. Chem. Soc.* **2002**, *124*, 1234−1240.
(63) Matthews, S. J.; Leatherbarrow, R. J. *J. Biomol. NMR* **1993**, *3*, 597−600.
(64) Lane, A. N.; Sengodagounder, A. *J. Magn. Reson.* **2005**, *173*, 339−343.

accurate and complete force fields and computationally intensive conformational sampling techniques,[60] with the resulting poses being evaluated on the basis of both NOE matching and the theoretical binding energies. In addition to providing $COST_{pose}$ values, NOE matching provides possible assignments for many of the experimental $^{1}H^{13}C$ groups (and hence possible NOE peak assignments) for each pose. By associating likelihoods with the possible assignments, explicit restraints could be derived from those assignments with high likelihoods. These restraints could then be used to limit the search space in a subsequent round of trial pose generation. By repeating this process, iterative refinement strategies[28,50] are feasible. Finally, several aspects of the NOE matching process may be recast in terms of Bayesian probabilities, as recently demonstrated for NOE peak identification[65] and NMR-based protein structure determination.[66]

Concluding Remarks

The studies described herein lay the groundwork for a widely applicable, general framework for "top-down"[12] determinations of ligand binding poses using protein−ligand NOE data. The NOE matching approach is able to use all of the available NOE data for pose evaluation, without the need for the time- and resource-consuming "bottom-up" process of establishing protein NMR resonance assignments by traditional approaches. Recent advances in experimental NMR methods[7−10,67] have made it possible to obtain sequence-specific resonance assignments for large proteins, and advances in automated NMR data analysis methods[12] have enhanced the throughput of the sequential assignment/structure determination process. Nevertheless, in the context of lead optimization, the ability to rapidly provide information on ligand binding poses remains a key challenge for biomolecular NMR. The methodology described and demonstrated in this article represents a novel, promising approach aimed at addressing this crucial issue.

Acknowledgment. We thank Valentina Goldfarb, Murali Dhar, Bennett Farmer, Wesley Cosand, Steven Sheriff, Bruce Jacobson, Deborah Loughney, and Patricia McDonnell for supporting aspects of this work, and Mark Friedrichs and Robert Bruccoleri for carefully reading the manuscript.

Note Added after ASAP Publication. After this paper was published ASAP on May 16, 2006, the structure for compound **2** was corrected in Chart 1 and in the Supporting Information, and a name was added to the Acknowledgment. The corrected version was published on May 31, 2006.

Supporting Information Available: Implementation of the NOE matching procedure and code execution timings; summaries of NMR structure determinations and target pose selections; supplementary figures; tables of additional NOE matching test results; complete ref 32; ^{1}H, ^{13}C, and ^{15}N protein NMR resonance assignments for mFABP/**1** and LFA-1/**3**; protein-bound ^{1}H NMR chemical shifts for all three ligands; coordinate files (ligands only) for the target poses of **1** and **2** with the atom naming conventions used herein; and lists of experimentally observed 3D X-filtered NOESY peaks for all three complexes. This material is available free of charge via the Internet at http://pubs.acs.org.

JA060356W

(65) Grishaev, A.; Llinás, M. *J. Biomol. NMR* **2004**, *28*, 1−10.
(66) Rieping, W.; Habeck, M.; Nilges, M. *Science* **2005**, *309*, 303−306.
(67) Kay, L. E. *J. Magn. Reson.* **2005**, *173*, 193−207.