Automatic assignment of methyl-NMR spectra of supramolecular machines using graph theory

Iva Pritisanac[1], Matteo T. Degiacomi[1], T. Reid Alderson[1], Marta G. Carneiro[2], Eiso AB[2], Gregg Siegal[2], Andrew J. Baldwin[1*]

[1]Department of Chemistry, Physical & Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, Oxfordshire, OX1 3QZ, UK

[2]ZoBio BV, BioPartner 2 building, J.H. Oortweg 19, 2333 CH Leiden, The Netherlands

*Correspondence to be addressed: andrew.baldwin@chem.ox.ac.uk

ABSTRACT: Methyl groups are powerful probes for the analysis of structure, dynamics and function of supramolecular assemblies, using both solution- and solid-state NMR. Widespread application of the methodology has been limited due to the challenges associated with assigning spectral resonances to specific locations within a biomolecule. Here, we present Methyl Assignment by Graph MAtching (MAGMA), for the automatic assignment of methyl resonances. A graph matching protocol examines all possibilities for each resonance in order to determine an exact assignment that includes a complete description of any ambiguity. MAGMA gives 100% accuracy in confident assignments when tested against both synthetic data, and 9 cross-validated examples using both solution- and solid-state NMR data. We show that this remarkable accuracy enables a user to distinguish between alternative protein structures. In a drug discovery application on HSP90, we show the method can rapidly and efficiently distinguish between possible ligand binding modes. By providing an exact and robust solution to methyl resonance assignment, MAGMA can facilitate significantly accelerated studies of supramolecular machines using methyl-based NMR spectroscopy.

INTRODUCTION

NMR spectroscopy can simultaneously probe both the structure and dynamics of biomolecules at atomic resolution[2-4], revealing how they function, describing interactions with binding partners and determining the locations of binding sites for therapeutic agents. Solution-state NMR has long been restricted to relatively small proteins up to ~30 kDa in molecular weight. Yet ca. 70% of single polypeptide chains in the human proteome are larger than this threshold, as are the majority of complexes and oligomers in the interactome[2]. Methyl-TROSY based solution-state NMR experiments greatly extend what is possible, enabling detailed structural and dynamical studies of protein complexes up to 1 MDa in molecular weight[7-10,55]. Similar gains are also being realised in solid-state NMR applications[11-12].

The significant bottleneck for application of methyl-based experiments, particularly for studies on larger proteins, is the need to relate the resonances observed in NMR spectra (Figure 1a) to specific atoms within the molecule (Figure 1d). This challenge, known as resonance assignment, is a prerequisite for interpretation of NMR data in functional studies. For smaller proteins, resonances are assigned using experiments that correlate nuclei between the backbone and side-chains[16]. For larger proteins, this is rarely possible and so assignment of methyl resonances is typically achieved by either transferring the assignments obtained from smaller protein fragments to spectra of the intact protein[8, 10], or by monitoring spectral changes after mutation of individually targeted residues[18]. Both approaches are laborious, time con-

suming and expensive, and require that the introduced mutations perturb neither the structure, aggregation propensity or expression levels of a protein. Nevertheless, methyl-based NMR methods are seeing wide uptake in both academia and industry.

In functional and dynamical NMR studies, structural models of the molecule of interest are generally available. In this case, comparing the structure to inter-methyl distance restraints derived from, for example, solution-state nuclear Overhauser enhancement (NOE) experiments can provide a means for resonance assignment[21]. This approach can be further enhanced through the inclusion of paramagnetic relaxation enhancement (PRE) experiments[19]. Two dedicated automatic structure-based methods are available using this approach for methyl group assignment (FLAMEnGO2.0 and MAP-XSII)[23-24]. Both rely on Metropolis Monte Carlo sampling to optimize approximate scoring functions to provide an assignment. These pioneering software approaches can unfortunately provide a user with confident but erroneous assignments, which is undesirable (Figures 3, 5) [23-24].

Here, we introduce Methyl Assignment by Graph MAtching (MAGMA), a novel structure-based automatic assignment approach. MAGMA builds upon the emerging applications of graph theory[29] in chemistry[30], structural biology[31] and biology[32], and uses discrete mathematics to perform an exact search of all possible assignments. The user is provided with an exact set of solutions, containing both confident (single option) and ambiguous (multiple options) assignments when there are multiple ways to explain the ex-
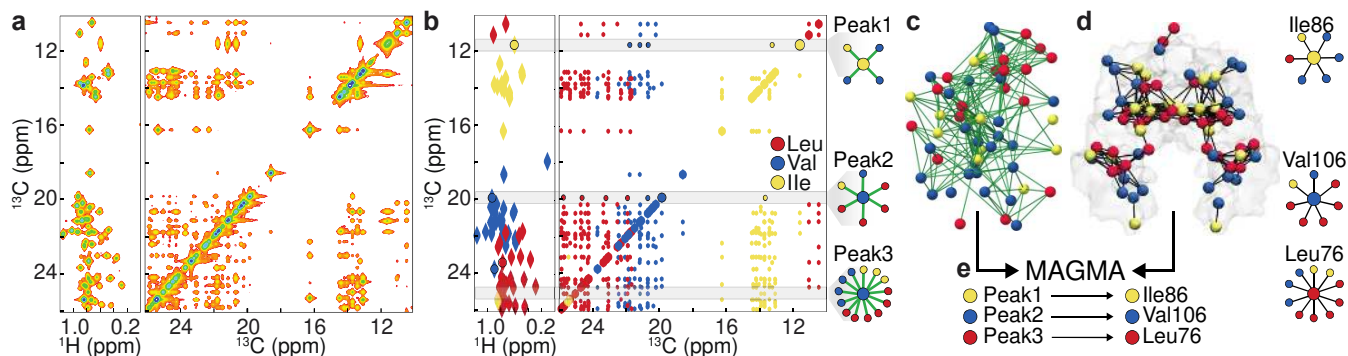
Figure 1 Automated methyl resonance assignment of solution-state NMR spectra by MAGMA. **(a)** Projections of a 3D (H)CCH HMQC NOE solution-state NMR spectrum of the R₂ dimer of ATCase, recorded from a deuterated sample with [¹³CH₃]-labelled Ile-δ, Leu-δ and Val-γ residues. Off-diagonal resonances reveal proximal pairs of methyl groups. These connectivities are used as distance restraints by MAGMA. **(b)** This is interpreted by MAGMA to include information about residue type. Analysis of regions associated with a single resonance (shaded) identifies the local network of connections for each methyl resonance. **(c)** The local networks can be combined to give a 'data graph'. **(d)** Similarly, inter-methyl connections from a protein structure can be obtained at a predefined distance threshold (detailed methods) and represented as a graph. **(e)** MAGMA compares all ways of mapping between the two graphs (figure S1), to give assignments that explain the maximum amount of data possible.

perimental data. We apply MAGMA to both synthetic (Figure 4) and a benchmark of 8 experimental NMR datasets with cross-validated assignments (Figure 4,5), the largest assembled of this type. Inter-methyl distance restraints were derived from either NOE based solution, or dipolar recoupling enhancement through amplitude modulation (DREAM) solid-state NMR experiments. Remarkably, MAGMA confidently assigned up to 93% of resonances, with 100% accuracy overall, on data obtained from complexes of molecular weight as high as 360 kDa.

This accuracy enables two new applications. We show that the algorithm can distinguish between alternative structural forms of 3 proteins (Figure 6). In an independent pharmaceutical application involving the human chaperone HSP90, we show both that MAGMA efficiently provides assignments, and that when combined with ligand/protein NOE data, it can rapidly and efficiently distinguish between possible ligand binding modes (Figure 7). Overall, MAGMA effectively removes the bottleneck that has limited widespread application of methyl-based NMR experiments. These results can facilitate structural studies of high molecular weight molecular machines, and their interactions with ligands and other biomolecules, and we anticipate that our core algorithm will be applicable to a wide set of problems involving structure determination.
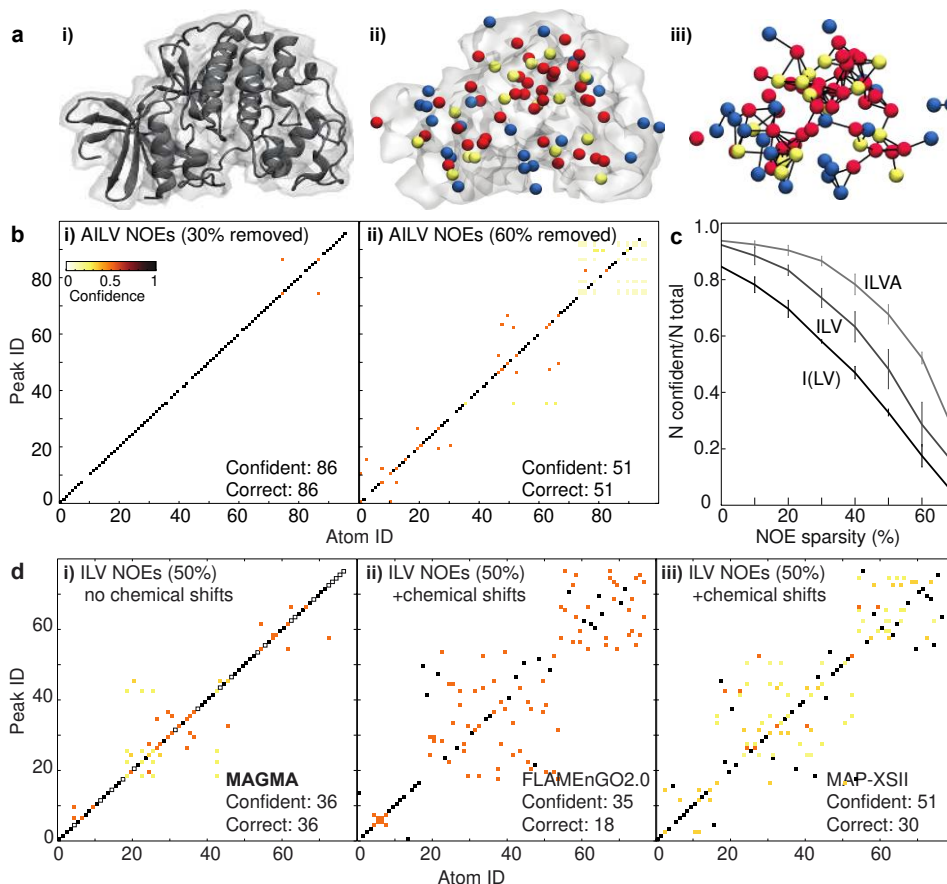
RESULTS

Resonance assignment as a graph-matching problem. When performing a methyl-based NMR study, a deuterated protein sample is prepared with ¹³CH₃ labelled methyl residues, typically isoleucine δ, valine γ₁/₂, and leucine δ₁/₂, although alanine β, threonine γ, isoleucine γ, and methionine ε methyl labeling methods are available[33]. NMR spectra corre-
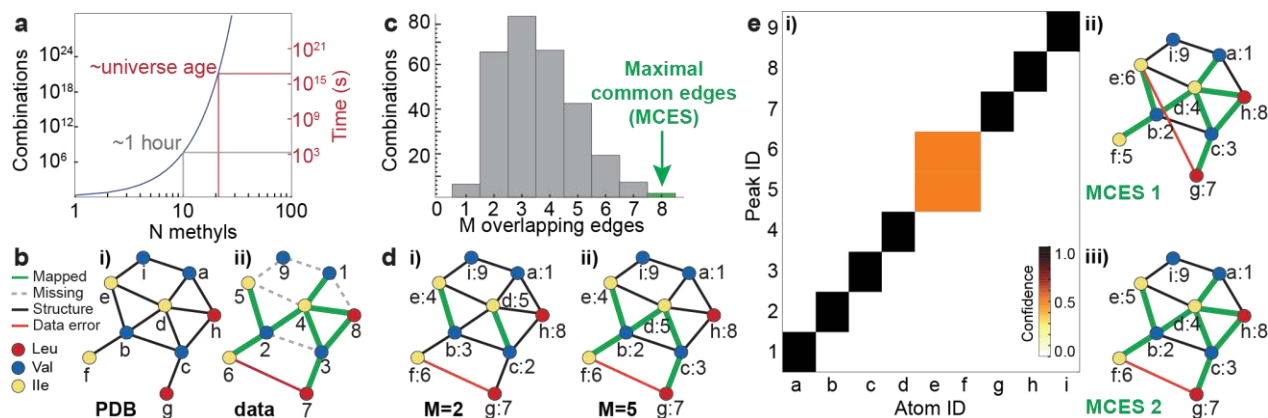
Figure 2. Graph matching for resonance assignment is a combinatorial problem. **(a)** If there are N methyls, the total number of assignment combinations is $N!$ (blue). If scoring each assignment takes 10^{-3} seconds, the assignment of 10 residues would take 1 hour whereas if N>21, the total time exceeds the age of the universe (~10^{17} seconds). **(b)** An illustrative structure (i) and data graphs (ii) containing 9 methyls, with 13 and 9 edges respectively. The data graph has fewer edges than the structure graph (dotted lines) and contains a deliberate error (red line), an edge not present in the structure graph. The data graph can be mapped onto the structure, with common edges indicated (green). **(c)** Each of the 288 ways of mapping the vertices have different numbers of common edges, *M*, indicating the number of 'explained' data points. **(d)** Two poor assignments; *M=2* (i) and *M=5* (ii). **(e)** There are two maximal common edge subgraphs (MCES, ii, iii), where *M=8* that comprise the solution. Pooling the results from these two solutions provides the final result, showing correspondence between methyl resonances (y-axis) and specific atoms (x-axis), (i) where 7 methyl resonances are confidently assigned, and two positions are ambiguous as they are structurally identical (orange). The 'correct' assignment follows the x=y line.

Figure 3. Performance of MAGMA on synthetic data derived from CDK2 (**a**) i) Secondary structure, ii) locations of Ile-δ (yellow), Leu-δ (red) and Val-γ (blue) methyl ^{13}C atoms (spheres) and iii) the structure graph at 8 Å. (**b**) The results of MAGMA on synthetic data, derived from the structure graph at i) 30% and ii) 60% sparsity (supporting information), with Ala, Leu, Val and Ile methyl groups labelled (AILV). Methyl resonances are on the y-axis (Peak ID), and methyl carbon atoms are on the x-axis (Atom ID) where correct assignments lie along the diagonal. Where a result is ambiguous, the correct assignment is always within the set of possibilities. Confident assignments have one option. (**c**) A comparison of the performance of MAGMA on randomly simulated synthetic datasets that vary in both their labelling (ILV and AILV) and sparsity. Mean and standard deviations of 10 independently simulated datasets are shown. (LV) denotes that the L and V residue types are not distinguished. Even with highly sparse data, a significant number of confident assignments can be achieved. Increasing the number of different types of residue, and decreasing the sparsity increase the success of MAGMA. (**d**) Performance of MAGMA (i), FLAMEnGO2.0 (ii) and MAP-XSII (iii) on a representative 50% sparse ILV inter-methyl data graph using CH3shift for chemical shift prediction. MAGMA provides confident assignments with 100% accuracy.

lating the ^{1}H and ^{13}C resonant frequencies of each methyl, which encode structural and dynamical properties, can be rapidly acquired (Figure 1a). The spectral frequencies encode the structural and dynamical properties of the molecule. NMR experiments can then be performed that reveal which methyl groups are close together in space (Figure 1a). These define a 'data graph', where each unique ^{1}H^{13}C correlation is a vertex, and each observed inter-methyl correlation is an edge connecting two vertices (Figures 1b, c, detailed methods). A complementary 'structure graph' can be derived from a structural model in which the vertices are the carbon atoms of the relevant methyl residues, and the edges comprise all connections between vertices within a distance threshold derived by MAGMA (Figure 1d, S3; detailed methods). Assigning methyl resonances requires comparison of the two graphs to find a mapping of the vertices that results in the maximum overlap of their edges (Figure 1e, Supporting information S.1). Physically, this mapping provides the assignments that explain the maximum number of experimentally observed inter-methyl

distance restraints. Resonance assignment is therefore a 'graph-matching' problem[34], where we seek the set of mappings that define the 'maximal common edge subgraphs' (MCES, supporting information). Importantly, as each single mapping in the set explains a maximum number of inter-methyl restraints, we require all such mappings, which together reveal any ambiguity in the assignment.

The MCES between two graphs could be obtained exactly by individually testing all assignment combinations (Figure 2). The total number of assignments to be tested increases with *N!*, where *N* is the number of methyl groups in the protein, and so while a problem involving 10 methyls can be evaluated in less than one hour, a problem containing 21 methyls (a 12 kDa protein of average residue composition) would take longer than the age of the universe (Figure 2a). A more efficient sampling strategy is necessary in order to solve the assignment problem exactly. A further complication is that the data graph will generally contain fewer edges than the structure graph, and which edges are missing will not be

known *a priori*. Effects including spin diffusion, chemical exchange and local disorder can all obscure individual inter-methyl correlations in NMR spectra. Moreover, low signal to noise ratios and overlap of resonances can lead to individual restraints being incorrectly interpreted.

MAGMA utilises exact graph matching methods designed to exclude large regions of the search space that would otherwise lead to suboptimal assignments early in the

search[35-36]. This provides an exact result that is tolerant to the main types of expected uncertainty.

Exact graph matching algorithms efficiently sample all possible methyl assignment solutions. We rely on two exact graph-matching algorithms, *VF2* and *McGregor*[35-36]. In the best case, where a data graph can be entirely embedded in the structure graph, the MCES is 'subgraph iso-morphic' (Supporting information S.1). In such a situation,
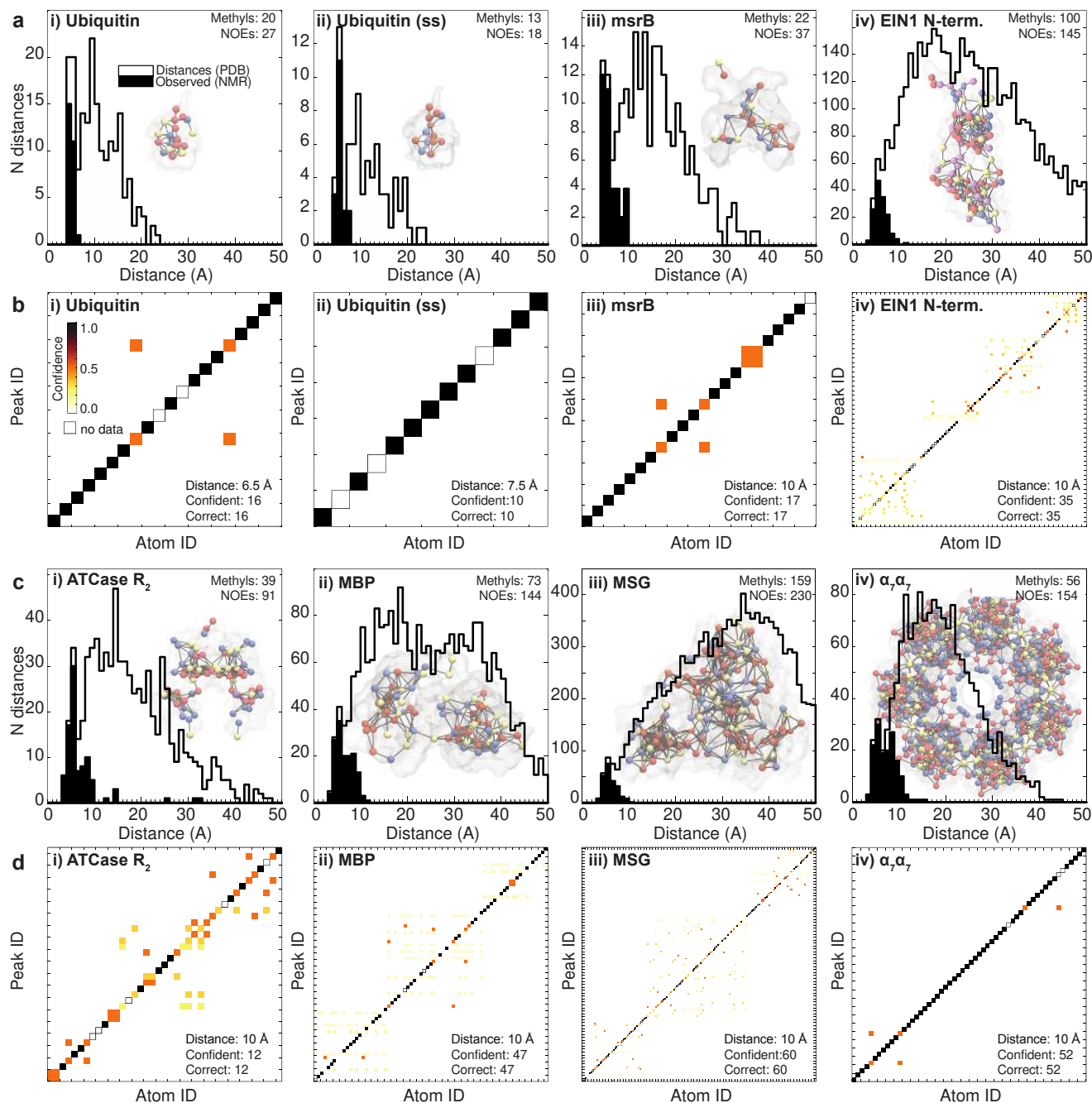
Figure 4. Performance of MAGMA on experimental data (Table 1). (**a/c**) Histograms of inter-carbon distances are calculated from structures (black line), and compared with those experimentally observed (solid black). Structures of each protein are shown approximately to scale, with the Ile, Leu, Ala, and Val groups in red, yellow, purple, and blue, respectively. (**b/d**) Assignment results from MAGMA. Methyl resonances are on the y-axis (Peak ID), and methyl carbon atoms are on the x-axis (Atom ID). Cross-validated assignments lie along the diagonal. The colourbar indicates the confidence in assignments from MAGMA. White boxes are resonances that have no inter-methyl restraints and so cannot be assigned. The total number of confidently assigned methyl groups, where only one assignment option is possible (black) is indicated together with the number of correct confident assignments. The accuracy of confident assignments in this benchmark was 100%.

where the two graphs are highly similar, the efficient *VF2* algorithm can be used[36]. This is not the case when there are erroneous inter-methyl distance restraints, which can arise through incorrect identification of the start and finish resonance of a cross-peak, or through the inadvertent assignment of unfortunate conjunctions of noise in spectra. Here, MAGMA executes an algorithm based on that of McGregor[35] (detailed methods, Figure S1). These algorithms are well suited for sparse, labelled graphs as encountered in this application (supporting information). Our implementation features a novel optimized ordering of the search that occurs both before and during the MCES search that substantially reduces the time required for the calculation, rendering it feasible for large proteins (detailed methods, Figure S4). In brief, two inputs are provided by the user, a set of inter-methyl distance restraints from NMR experiments that define a data graph (Figures 1a, S1, Table S1) and a structure file from which atom positions and Euclidean distances are used to construct a structure graph (S.5, Figures 1d, S1). MAGMA returns all assignment possibilities for each methyl resonance, which are drawn from the set of vertex mappings that define the MCES. Where exactly one methyl resonance is assigned to one residue in the structure, the assignment is defined here as 'confident'. Otherwise, the assignment is 'ambiguous'. For each resonance, all possibilities are given a score equal to the reciprocal of the number of assignment possibilities. Identifying ambiguity inherent in the data is highly desirable as it reflects the maximum information content of the experimental data, identifying regions where the information is sparse or missing, enabling the design of additional experiments if required (Figure S2).

To illustrate our method, a simplified problem involving nine methyl groups is considered, where the data graph contains fewer edges than the structure graph, and a deliberate error (Figure 2). Due to the inclusion of an erroneous restraint, the data graph is not subgraph isomorphic to the structure graph and *VF2* cannot be applied. Of the 288 possible solutions, MAGMA finds the 2 solutions that comprise the MCES early in the search (Figure 2c). This provides 7 confident assignments and 2 that can be interchanged, as they are structurally identical.

Proof-of-principle on simulated data. To demonstrate that MAGMA is well suited for automatic methyl resonance assignment, we performed calculations using simulated inter-methyl distance restraints. These were obtained from the structure graph of the 34 kDa human Cyclin-dependent kinase 2 (CDK2, PDB: 2C6o)[37], with a distance threshold of 8 Å. Several common experimental methyl labelling schemes involving labelling of Ile-δ, Leu-δ, Val-γ and Ala-β residues were compared (Figure 3a). Increasingly sparse data graphs were simulated by randomly removing edges, weighted to enhance the probability of retaining shorter distances (supporting information S.2). In the assignments returned by MAGMA, even when an assignment was ambiguous, the correct solution was always present within the set of solutions for each labeling scheme and sparsity level tested (Figure 3b, d). The confident assignments returned by MAGMA were 100% accurate in all tested cases. Increasing the sparsity of the simulated data graph reduced the number of methyls that could be confidently assigned (Figure 3b-d), as expected given the lower information content of the data. Notably, having a large number of distin-

Table 1. The inter-methyl data defining our benchmark, and the performance of MAGMA.

For each protein the molecular weight, structure, labelling scheme and type of NMR experiment used is specified. HSP90α was analysed independently of the benchmark to cross validate the method. Solution-state NMR experiments obtain restraints via NOE spectroscopy, and the solid-state NMR example obtained restraints via a 4D DREAM experiment. The number of inter-methyl distance restraints and the number of independent subgraphs within the data are indicated. (**i**) Assignable residues are those with inter-methyl restraints. (**ii**) The average number of inter-methyl restraints per assignable methyl group. (**iii**) The distance threshold derived by MAGMA for each case (see text). (**iv**) Data sparsity, here defined as number of observed NOEs divided by the number of total number of possible distances, calculated at the optimal distance threshold, **iii**. (**v, vi**) The number of confident (unambiguous) assignments returned by MAGMA, together with the total assigned fraction. (**vii**) The time to arrive at the final result, including all ambiguity (Figure 4b), took a maximum of 2 days (α7α7). (**viii**) The time for the calculation to run to completion. *See Supp S.

	Mw (kDa)	PDB identifier	Labelling scheme	Methyl groups	Labelled residues	Data type	Inter-methyl restraints	Data subgraphs	Assignable resonances(i)	Mean vertex degree (ii)	Distance threshold (Å) (iii)	Sparsity (iv)	Confident assignments (v)	Assigned fraction (vi)	Time to converge (s) (vii)	Total time (s) (viii)
Ubiquitin	8.6	1ubq[5]	I,L,V	33	20	NOE[13]	25	3	18	2.78	6.5	0.56	16	0.8	2×10^{-2}	2×10^{-2}
Ubiquitin(ss)	8.6	1ubq[5]	L,V	26	13	DREAM[12]	18	1	10	3.6	7.5	0.78	10	0.77	2×10^{-2}	2×10^{-2}
MsrB	16.6	3eoo[14]	I,L,V	39	22	NOE[15]	37	2	21	3.52	10	0.62	17	0.77	4×10^{-3}	4×10^{-3}
EIN1	27	1eza[17]	I,L,V,A	146	100	NOE[19]	145	6	84	3.45	10	0.37	39	0.35	4×10^{3}	5×10^{3}
ATCase R2	30	1d09[20]	I,L,V	66	39	NOE[22]	91	1	34	5.35	10	0.58	12	0.31	7×10^{4}	2×10^{5}
MBP	43.4	1ez9[25]	I,L,V	123	73	NOE[15]	144	4	70	4.11	10	0.56	47	0.64	1×10^{5}	2×10^{5}
MSG	81.4	1y8b[26]	I,L,V	274	159	NOE[26-27]	230	12	141	3.26	10	0.47	60	0.38	7×10^{4}	4×10^{6}
α7α7	360	1yau[28]*	I,L,V	96	56	NOE[8]	154	1	54	5.81	10	0.8	52	0.93	2×10^{5}	-
HSP90α	22.1	1yer	I,L,V	76	47	NOE	71	5	42	3.4	10	0.42	20	0.48	6×10^{1}	6×10^{1}

guishable residue types increased the success of the method. Simulated data with Ile-δ, Leu-δ, Val-γ, and Ala-β methyl groups provided a higher proportion of confident assignments than those from Ile-δ, Leu-δ and Val-γ, and being able to explicitly distinguish Leu from Val residues, an experimentally achievable goal[38], was advantageous (Figure 3c).

We compared the performance of MAGMA to two other available automated methyl assignment approaches, FLAMEnGO2.0[23] and MAP-XSII[24] (Figure 3d, Supporting information S.3). These both require chemical shift data (Figure S5) in addition to inter-methyl distance restraints. These were predicted using either ShiftX2[39] or CH3Shift[40], with each method yielding similar results in the final assignments. The results from a representative data graph simulated at 50% sparsity with Ile-δ, Leu-δ, and Val-γ methyl-labelling allow performance comparison of the three algorithms. The number of correct/incorrect assignments was 36/0 in the case of MAGMA, 18/17 for FLAMEnGO2.0 and 30/21 for MAP-XSII (Figures 3, S4). Similar results were seen for other randomly simulated data graphs. These results demonstrate that MAGMA can provide both accurate and reliable methyl assignments.

Testing of MAGMA on experimental NMR data. MAGMA was tested against a benchmark of experimental data obtained from either 3D or 4D NOE solution-state[8, 15, 22-23, 26], or proton detected 4D DREAM solid-state[12] NMR spectra of methyl-labelled proteins[5, 15, 17, 41-43] (Table 1), whose assignments were previously determined and cross-validated (figure 4, Table 1)[8, 12-13, 15, 22, 26]. The benchmark is comprised of Ubiquitin (Ubq)[5, 12-13], methionine-R- sulfoxide reductase B protein (MsrB)[15], the N-terminal domain of *E. coli* Enzyme I (EIN1)[19], a dimer of regulatory chains of aspartate transcarbamoylase from *E. coli* (ATCase R$_2$)[22, 41], maltose binding protein (MSB)[15], malate synthase G (MSG)[26, 43], and ($\alpha_7\alpha_7$)[8, 42]. This benchmark is, to our knowledge, the largest assembled for testing automatic methyl assignment programs and is freely available for download with MAGMA.

The 8 experimental data sets are diverse in molecular weight, shape, amino acid labeling scheme, and data sparsity (Table 1). Using the known assignments, the distribution of inter-methyl distance restraints observed in the data was compared with the distribution of the total possible inter-methyl distances within the protein structures (Figures 4a, c). In general, the majority of short ranges inter-methyl distances are observed (<8 Å), while relatively few restraints are observed at distances greater than 10 Å.

The method requires a distance threshold to be set for calculation of the structure graph (Table 1). Running the calculation with a distance that is too short can give incorrect solutions, whereas running with a threshold that is too long resulted in both artificially increased ambiguity and long calculation times. We ran our benchmark calculations for a wide range of distance thresholds in order to determine optimal values, allowing us to define the following empirical rule (Figure S3a). MAGMA determines its threshold as the shorter, either 10 Å, or the minimum distance required to explain 100% of the inter-methyl distance restraints (detailed methods). This empirical rule results in all proteins in our benchmark being run at 10 Å apart from ubiquitin, the smallest protein in our dataset, which was run at 6.5 Å and 7.5 Å, solution- and solid-state NMR datasets, respectively. This definition results in 100% accuracy for all tested benchmark
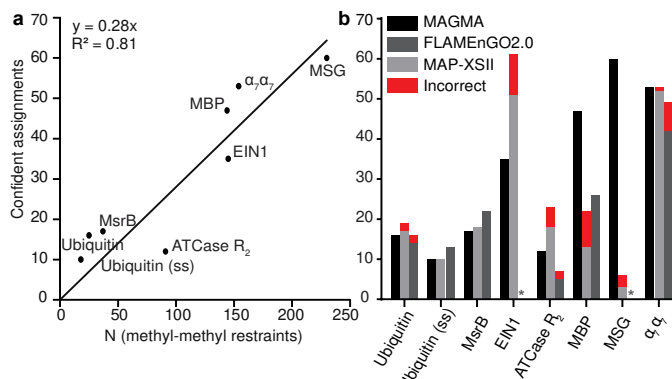
Figure 5. (a) The number of confident assignments from MAGMA in our benchmark versus the total number of restraints in the calculation. On average, 3.6 were required for each confident assignment. **(b)** Comparison of MAGMA (black) to alternatives FLAMEnGO2.0 (dark grey), and MAP-XSII (light grey). The number of incorrect confident assignments are indicated (red). No confident assignments for EIN1 and MSG were reported by FLAMEnGO2.0 (*). MAGMA provides more confident assignments on average, with exceptional performance on large targets (MBP, MSG, $\alpha_7\alpha_7$), with 100% accuracy overall.

cases, and also for the N-domain of HSP90α, which was not a part of the benchmark (Figure 6).

Using this method to automatically select for distance threshold, the number of confident assignments returned by MAGMA (Figure 4b, d; Table 1) was found to vary with the specific case. On average, over the benchmark, 3.5 inter-methyl restraints were needed for each confident assignment (Figure 5a). Notably, one of the larger proteins under study, the $\alpha_7\alpha_7$ double ring of 20S proteasome core particle resulted in 93% confidently assigned methyl groups (Figure 4d iv); Table 1). Zero NOEs were observed from the three unassigned methyl resonances. This exceptional performance on $\alpha_7\alpha_7$ can be attributed to the remarkably high ratio of experimental restraints to total inter-methyl distances within the protein structure model (80%, Table 1). The performance of MAGMA on the benchmark was compared to that of FLAMEnGO2.0 and MAP-XSII (Figures 5b, S4; Supporting information S.3). Overall, MAGMA provided more confident assignments than the alternative programs. As with the simulated data, confident assignments identified by MAGMA were 100% accurate.

Chemical shifts do not help assignment. The resonance frequencies of methyl groups are in principle a useful addition for assignment, and both FLAMEnGO2.0 and MAP-XSII use these directly[23-24]. Methyl chemical shifts have a relatively low range of values in both the proton and carbon dimensions, which are frequently close to the statistical prediction accuracy of prediction algorithms. The results from the benchmark were re-scored using either ^{13}C, ^{1}H or combined ^{13}CH$_3$ chemical shifts, predicted on the basis of protein structure using different protocols (ShiftX2 and CH3Shift)[39-40]. We applied the chemical shifts in a highly conservative fashion. Within a set of ambiguous assignments, a new assignment was accepted if one and only one assignment option was within the prediction accuracy. This procedure resulted in incorrect confident assignments (Supporting information S.4, Figure S5). While chemical shifts are exceedingly useful

for structural and dynamical studies, we conclude that at present, they are insufficiently reliable for assignments using these commonly used programs.

MAGMA can guide site-directed mutagenesis to yield more methyl assignments. In the case where a specific assignment is ambiguous, MAGMA can predict which sites should be selected for site-directed mutagenesis experiments to maximise the number of new expected assignments. Illustrating this using a simulated 50% sparse ILV dataset from CDK2, our protocol predicts three single point mutations of the most ambiguous residues that will provide five new confident assignments (Figure S2). MAGMA allows any known assignments to be specified, which will reduce ambiguity of unknown assignments and can lead to a significant increase in the final number of confident assignments.

MAGMA can identify which protein conformation is predominantly present during the NMR measurement. Multiple structural forms of a protein are commonly available. From our benchmark, the ATCase R_2 dimer exists in 'R' and 'T' conformations (Figure 6a) and MSG adopts a different conformation when bound to certain ligands (Figure 6b). In addition, the N-terminal domain of HSP90α has two conformations, termed here 'open' and 'closed', with the largest differences in the binding pocket residues 104-114 (Figure 6c). The MAGMA assignments for each structure

were compared with assignments obtained independently through triple resonance experiments [6, 22, 27]. For the expected structural form, the results from MAGMA were in perfect agreement with the independent assignments. By contrast, the alternative conformations led to erroneous assignments (Figure 6). If multiple conformations of a protein are known and its structural form is uncertain, the results can be combined (Figure 6 a-c ii)), reducing the number of confident assignments but preventing errors caused by uncertainty in the structure.

Multiple structural forms can be distinguished by the differences in their MAGMA assignments. In each of the described cases, multiple residues exist where confident assignments are in disagreement between the two forms. In such a case, by obtaining an independent assignment through, for example, mutagenesis, it is possible to effectively exclude structural forms. In this way, MAGMA results combined with a small number of additional independent assignments determine ATCase R_2 dimer to be in its 'R' state, MSG to be in its 'free' form, and the HSP90α N-terminal domain to be in the 'closed' conformation (Figure 6). In each case, the structure predicted by MAGMA is supported by the observed $^1H^{13}C$ chemical shifts. The accuracy of MAGMA therefore enables it to be used to distinguish structures.

MAGMA can discriminate between different ligand binding modes to aid drug discovery. Identifying the correct binding mode of ligands when bound to proteins using sparse NMR data is of significant pharmacological interest[6]. HSP90α is a ubiquitous molecular chaperone and an oncology target[44]. Inhibitors targeting the ATPase are presently in various stages of clinical trials[1]. Here, we studied the N-terminal domain of HSP90α bound to a fragment-hit aminotriazine compound[1]. Different binding modes of this compound have been determined previously by crystallography[1] and using NMR[6] (Figure 7).

Protein/ligand NOEs[6] from methyl labelled HSP90α in the presence of the fragment hit were recorded together with protein/protein NOEs. Both sets of distance restraints were analysed by MAGMA for each predicted binding mode (detailed methods). The results were then compared to the known ligand assignments, which are obtained from inspection of its isolated ^1H spectrum (Figure 7aii). This reveals that only one of the structures is consistent with the MAGMA assignment (Figure 7b iii). Moreover, the combined protein and ligand data provided 5 new confident protein assignments (25 in total) in the binding pocket[45]. This result reveals the utility of MAGMA in drug discovery, allowing filtering of ligand conformations by a combi-
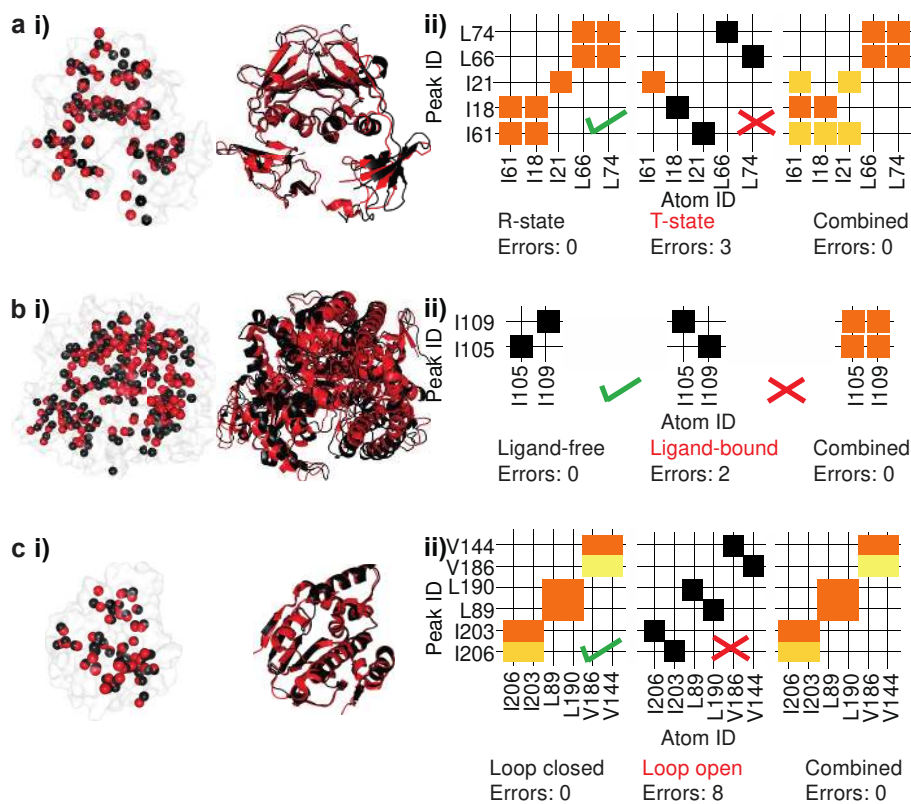

Figure 6. MAGMA results can discriminate between protein conformations. Two classes of structure are available for ATCase R_2 dimer (**a,** R 1D09/T 1TUG), MSG (**b,** bound 1D8C/free 1Y8B), and the N-terminal domain of HSP90α (**c** open 3B24/closed 1YER). Relevant methyl groups (spheres), and backbone orientations in the two structures are shown. Regions where MAGMA reports different assignments are indicated (ii**)**. In a conservative case, these results can be pooled. Notably, sites are revealed with incorrect assignments for one of the two structural forms. Independent data such as a point mutant at one of the conflicting sites is then sufficient to unambiguously distinguish the two possible structures.

nation of protein/protein NOEs, protein/ligand NOEs, and known ligand resonance assignments. This method requires neither independently known protein assignments, nor new point mutants to be generated.

DISCUSSION

Inter-methyl connections can be obtained experimentally using either solution- or solid-state NMR. By relating these inter-methyl connectivity data to a structure, MAGMA can
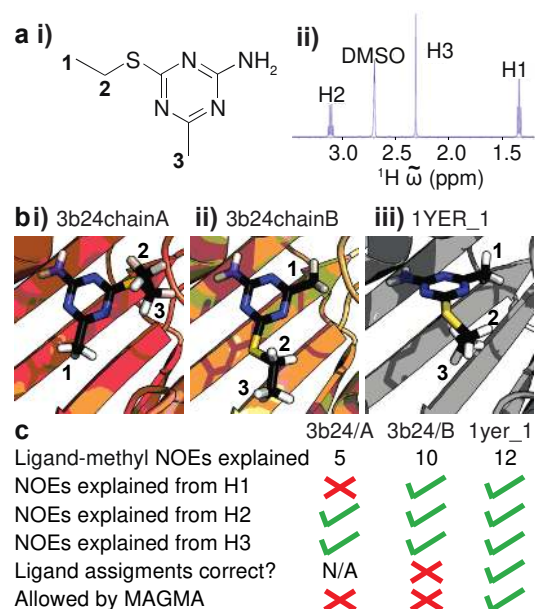
Figure 7. MAGMA can distinguish between ligand binding modes. An aminotriazine based fragment hit compound was bound to the N-terminal domain of HSP90α. **a)** A 1D ¹H NMR spectrum of the compound, allows a straightforward assignment of H1, H2 and H3. **b)** Three structures show the compound in different arrangements (PDB: i) 3B24 chain A[1], ii) 3B24 chain B[1], NMR iii) 1YER_1[6]). MAGMA was run on each, including both protein/protein and protein/ligand NOEs. One structure could be excluded as the magma result explained few ligand/protein NOEs, and none from H1. Of the remaining two similar structures, one resulted in incorrect ligand assignments. 1YER_1 was the only structure consistent with the NMR data, following screening by MAGMA.

accurately and reliably provide an exact set of assignments. Using either simulated data, or our benchmark of experimental data derived from both solution- and solid-state NMR experiments, we achieve 100% accuracy, with as many as 93% confident assignments revealed. The HSP90α assignments were obtained using the protocol developed on the benchmark. These results were independently cross-validated, demonstrating the method reliably predicts assignments beyond the benchmark. MAGMA outperforms alternative approaches in accuracy on our benchmark. Notably, MAGMA's performance improves as the protein target gets larger (MBP, MSG, $\alpha_7\alpha_7$).

Existing methods for automatic resonance assignment both for methyl groups and the backbone aim to minimise a continuous scoring function[19, 23-24]. Our approach is fundamentally different, using discrete mathematics to sample all possible assignment combinations, providing assignments that include all possible ambiguity. We anticipate that our

core graph-matching algorithm will be highly applicable to a range of future data and structural analysis applications. From a computational complexity point of view, our algorithm is NP-complete[34] and in the worst case scenario the execution time will scale with $N!$, where N is the number of methyl groups to be assigned. The design of MAGMA makes such a case unlikely, as demonstrated by the execution times of our benchmark (Figure S4c). The total calculation time ranges from a few seconds to several weeks for the largest proteins, performed on a single Intel i5 processor (Table 1, Figure S3b). The longest time required to obtain the 'final' result in our benchmark with a complete description of ambiguity was 2 days (Table 1), a duration substantially shorter than the 'total' calculation time.

A typical MAGMA result will contain both confident (1 option) and ambiguous (multiple option) assignments. Ambiguity exists in regions where individual assignments can be exchanged and still satisfy a maximum number of restraints. In these regions, further information is required to resolve the assignment ambiguity. MAGMA contains a protocol to select candidates for point mutation that will yield the most significant number of new assignments per mutation made (Figure S2). In addition, the results can be re-scored against other experimental data. We have shown that commonly used chemical shift predictors are not sufficiently reliable at this time (Figure S5). By contrast, we anticipate re-scoring against PRE data will likely prove to be a powerful complementary restraint[19]. Similarly, signal intensities of cross-peaks in inter-methyl NMR experiments contain information about inter-methyl separations. To quantitatively access this information for use in assignments, data spanning a range of mixing times could be employed (54), which might require a prohibitively large acquisition time. Signal intensities from solution-state NOE data with a single mixing time used for this work was unable to reliably reduce ambiguity in the assignments (Figure S6). Finally we expect optimum performance from the graph matching algorithms where we include only inter-methyl restraints that can be confidently attributed to a unique pair of resonances and we do not allow for ambiguous restraints. We recommend recording high dimensional NMR spectra to increase the number of unambiguous experimental restraints. If this is not possible, MAGMA results can be re-scored to exclude any possibilities that are not consistent with any ambiguous restraints that were excluded from the calculation. Before re-scoring, MAGMA obtains more reliable information from inter-methyl distance restraints than any alternative method.

The choice of methyl labelling scheme for a protein can affect the success of the assignment (Figure 3, Supporting information S1.5). When working with leucines or valines, it is desirable to prepare a sample where both Leu-δ and Val-γ methyl pairs are labeled and perform an experiment to detect their proximity[8, 27]. This allows pairs of methyl groups to be combined into a single residue 'pseudo atom', which simplifies analysis (detailed methods, Supporting information S1.5). Moreover, it is highly desirable to distinguish leucine and valine residues[38, 46]. In general, increasing the number of distinguishable residue types is highly beneficial and incorporating Met-ε and Ala-β methyl groups is desirable[47], though for larger proteins, this could result in broader, potentially unresolvable resonances.

MAGMA will be freely available for academic users. When provided with inter-methyl connectivity information from experimental NMR data and a structural model, it will provide an assignment list that contains all possibilities, requiring no parameters to be specified by a user. On both the benchmark, and an independent test (Hsp90) we obtain 100% accuracy. While we cannot guarantee this in all future cases, the benchmarked accuracy is sufficiently high to provide a reliable assignment for *de novo* systems analysed using methyl-based NMR spectroscopy. The accuracy of MAGMA allows it be used to discriminate structures (Figure 6). Moreover, when combined with ligand/protein restraints, it can distinguish between possible ligand-binding modes (Figure 7). By removing the assignment bottleneck, MAGMA can enable routine application of methyl-based NMR experiments for studies of high molecular weight molecular machines.

DETAILED METHODS

Analysis of experimental NMR data. Inter-methyl restraints were obtained from NOE spectra (ATCase R_2 dimer, EIN[19], NTD of HSP90α), pre-analysed inter-methyl NOE restraints ($\alpha_7\alpha_7$[8], ubiquitin solution-state[13]) and data deposited in the BMRB and PDB (ubiquitin solid-state data, MsrB, MSG, and MBP)[12, 15, 26]. NOE spectra were available as 3D (H)CCH (ATCase R_2 dimer, HSP90α) or 4D HCCH (EIN1) heteronuclear multiple quantum correlation (HMQC) forms, and were analysed using Sparky[48] to obtain a list containing inter-methyl correlations, and signal-to-noise ratios. The following filters were applied by MAGMA to establish a list of inter-methyl distance restraints: **1)** Each inter-methyl resonance has a signal-to-noise ratio ≥ 2, where noise is given as the standard deviation of a region of the spectra devoid of signal. **2)** Inter-methyl resonances that overlap significantly with a diagonal peak were not considered. **3)** Inter-methyl resonances must be reciprocally observed from both starting and ending ^{1}H-^{13}C resonances. **4)** The signal to noise ratios of the two cross peaks must be similar, to within 10%. Where we did not have access to the raw data containing signal intensities, only condition 3 was applied (section S.5). For MAGMA operation, we recommend manual curation of inter-methyl restraints such that there are only confidently identified cross peaks, excluding ambiguous data (section S.5). We recommend acquiring 4D inter-methyl NMR spectra to substantially reduce the ambiguity in cross-peak assignment over 3D equivalents, which can be acquired expediently with non-uniform sampling. Moreover, both identifying the residue type of individual amino acids, and pairing the leu-$\delta_{1/2}$ and val-$\gamma_{1/2}$ residues using a short-mixing time experiment on a double methyl labeled sample is highly desirable for maximising confident assignments (Figure 3, section S1.5). The benchmark will be freely downloadable from http://baldwinlab.chem.ox.ac.uk/ .

Defining the data graph. Pairs of ^{1}H/ ^{13}C resonant frequencies for each methyl group are selected to define a set of vertices, and experimental inter-methyl distance restraints define the edges (Figure 1c). The vertices of the graph are 'labelled' according to the amino acid type. If both methyl groups within a single Leu-δ or Val-γ have been identified, then the distance restraints from each are combined to give a single vertex. If intermolecular NOEs between the ligand resonances and methyl resonances are available, the ligand resonances can be added to the set of data graph vertices,

and ligand/protein distance restraints are included as additional edges.

Defining the structure graph. Carbon atoms from relevant methyl groups in a structure are identified, and used to define a set of structure graph vertices, labelled by amino acid type. In the case where both methyl groups of Leu and Val residues can be combined into a single vertex in the data graph (see above), these are merged into a single pseudo-atom located at the average position of the two carbon atoms in the structure graph. Euclidean distances between all methyl carbon atoms are calculated. In the case where there are multiple copies of a protein, the shortest inter-carbon distances are taken. The inter-methyl distances that are below or equal to the threshold, determined as described below, define a set of edges of the structure graph (Fig 1d). When a protein structure contains a ligand of interest, the set of atoms of the ligand can be added as structure graph vertices. Euclidean distances measured between the methyl carbons and specified ligand atoms, which fall below the ligand-methyl threshold, define additional structure graph edges.

Setting the distance threshold too low can result in incorrect results, whereas setting it too high leads to increased heterogeneity and longer, possibly intractable calculation times as discussed in the text (Figure S3). We determined an empirical method to automatically determine the appropriate threshold to use that results in 100% accurate confident assignments both in our benchmark, and in the independent test case of HSP90α. MAGMA runs are initiated as described in detail below, at 10 Å. If a solution is found that explains all of the edges, the calculation is stopped, and the distance threshold is reduced in increments of 0.5 Å. This is repeated until the MCES is the same size as the number of restraints. MAGMA ran at 10 Å in all cases other than the ubiquitin datasets, where a shorter threshold was determined.

Matching of the data and structure graphs by MAGMA. MAGMA compares the structure and data graphs to obtain the mapping of their vertices (assignment) that results in the maximal overlap of edges between the two. If the data graph is disconnected, the number of independent, but internally connected data subgraphs is first determined, and a MAGMA run is performed independently on each, before they are recombined at the end. A data graph under consideration is tested to determine if it is subgraph isomorphic to the structure graph. If it is, the *VF2* algorithm[36] is run (faster). If not, the adapted McGregor[35] algorithm is executed (slower). The implementation of the *VF2* algorithm used in MAGMA is contained in the high performance graph library *python-igraph0.7*[49]. The adapted McGregor MCES algorithm in MAGMA is implemented in Python. The program reports all possible mappings of the two graphs that provide the maximal overlap, from which the assignment is obtained. Both algorithms will explore all possible assignment possibilities and provide the same final result, but will vary significantly in execution time.

The order in which data vertices are tested by the algorithm, and the order in which structure vertices are tested for each data vertex in the MCES algorithm significantly affect its efficiency. MAGMA optimises both. A 'first guess' order of matching is obtained by quantitatively comparing the local similarities of each pair of vertices in the structure and data graphs. An implementation of the Hungarian method, the Munkres algorithm[50] (Python package: munkres1.0.7.)

uses this information to generate an optimised mapping between the two. This determines which structure vertex will be the first to be tested for each data vertex, and is typically found to already be the correct assignment. The subsequent order of structure vertices to be tested against a data vertex follows an ordered list of their Jaccard similarity coefficients[51].

An initial data vertex (root) is chosen as a starting position and subsequent data graph vertices to be tested are ordered according to their adjacency to its predecessors. MAGMA performs a run where each data vertex is selected in turn as the root and the algorithm matches the structure vertices to each data vertex in the order described above. The first MCES found by the algorithm is returned. The root that provides the largest initial MCES is selected as the starting vertex, for a calculation that is allowed to run to completion. During the calculation, matching priorities are continuously updated to prioritise local adjacency relations in both the data and structure graphs (Table 1). These optimisations enable graphs to be matched efficiently by MAGMA. The optimisation protocol significantly reduces the total calculation time without introducing any approximation, and so the final result remains exact.

MAGMA returns the set of assignments that maximises the number of assigned distance restraints. When multiple structure vertices are assigned to a data vertex, a score given to each possibility that is the reciprocal of the total number of possible assignments for that data vertex (Figures 2-7). For a confident assignment, there is only one possibility of matching of a methyl resonance to a single CH_3 group in the structure, and the score is unity. Analyses and plotting of graphs were supported by the Python package *networkx-1.9*[52] and the *Mayavi* application[53].

Software implementation. MAGMA will be available for download and free for academic use. Input files for the benchmark will be provided with the download to facilitate operation of the program. A user needs to supply the program with a suitably formatted list containing the inter-methyl distance restraints (data graph edges), a structure file in a PDB format, and the labels to extract. The program returns an output file listing all assignment possibilities for each methyl resonance.

ASSOCIATED CONTENT

Supporting information. SI is available containing formal definitions and concepts from graph theory and how these are exploited by MAGMA. The protocol for simulated data, and methods for the performance comparison to MAP-XSII, and FLAMEnGO2.0, and chemical shift scoring are included with supporting figures S1-5. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Miura, T.; Fukami, T. A.; Hasegawa, K.; Ono, N.; Suda, A.; Shindo, H.; Yoon, D. O.; Kim, S. J.; Na, Y. J.; Aoki, Y.; Shimma, N.; Tsukuda, T.; Shiratori, Y., *Bioorg. Med. Chem. Lett.* **2011,** *21* (19), 5778-83.
2. Rosenzweig, R.; Kay, L. E., *Annu. Rev. Biochem.* **2014,** *83*, 291-315.
3. Barrett, P. J.; Chen, J.; Cho, M. K.; Kim, J. H.; Lu, Z.; Mathew, S.; Peng, D.; Song, Y.; Van Horn, W. D.; Zhuang, T.; Sonnichsen, F. D.; Sanders, C. R., *Biochemistry* **2013,** *52* (8), 1303-20.
4. Baldwin, A. J.; Kay, L. E., *Nat. Chem. Biol.* **2009,** *5* (11), 808-14.
5. Vijay-Kumar, S., Bugg, C.E., Cook, W.J., Structure of ubiquitin refined at 1.8 A resolution. *J. Mol. Biol.* **1987,** *194* (3), 531-544.
6. Shah, D. M.; Ab, E.; Diercks, T.; Hass, M. A. S.; van Nuland, N. A. J.; Siegal, G., *J. Med. Chem.* **2012,** *55* (23), 10786-10790.
7. Baldwin, A. J.; Hansen, D. F.; Vallurupalli, P.; Kay, L. E., *J. Am. Chem. Soc.* **2009,** *131* (33), 11939-11948.
8. Tugarinov, V.; Sprangers, R.; Kay, L. E., *J. Am. Chem. Soc.* **2007,** *129* (6), 1743-50.
9. Rosenzweig, R.; Moradi, S.; Zarrine-Afsar, A.; Glover, J. R.; Kay, L. E., *Science* **2013,** *339* (6123), 1080-3.
10.a Saio, T.; Guan, X.; Rossi, P.; Economou, A.; Kalodimos, C. G., *Science* **2014,** *344* (6184), 1250494.
11. Baldwin, A. J.; Walsh, P.; Hansen, D. F.; Hilton, G. R.; Benesch, J. L. P.; Sharpe, S.; Kay, L. E., *J. Am. Chem. Soc.* **2012,** *134* (37), 15343-15350.
12. Huber, M.; Hiller, S.; Schanda, P.; Ernst, M.; Bockmann, A.; Verel, R.; Meier, B. H., *Chemphyschem* **2011,** *12* (5), 915-918.
13. Chao, F. A.; Shi, L.; Masterson, L. R.; Veglia, G., *J. Magn. Reson.* **2012,** *214* (1), 103-10.
14. Kim, Y. K.; Shin, Y. J.; Lee, W. H.; Kim, H. Y.; Hwang, K. Y., *Mol. Microbiol.* **2009,** *72* (3), 699-709.
15. Lange, O. F.; Rossi, P.; Sgourakis, N. G.; Song, Y. F.; Lee, H. W.; Aramini, J. M.; Ertekin, A.; Xiao, R.; Acton, T. B.; Montelione, G. T.; Baker, D., *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (27), 10873-10878.
16. Gardner, K. H.; Konrat, R.; Rosen, M. K.; Kay, L. E., *J. Biomol. NMR* **1996,** *8* (3), 351-6.
17. Garrett, D. S.; Seok, Y. J.; Liao, D. I.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M., *Biochemistry* **1997,** *36* (9), 2517-2530.
18. Amero, C.; Asuncion Dura, M.; Noirclerc-Savoye, M.; Perollier, A.; Gallet, B.; Plevin, M. J.; Vernet, T.; Franzetti, B.; Boisbouvier, J., *J. Biomol. NMR* **2011,** *50* (3), 229-36.
19. Venditti, V.; Fawzi, N. L.; Clore, G. M., *J. Biomol. NMR* **2011,** *51* (3), 319-28.
20. Jin, L.; Stec, B.; Lipscomb, W. N.; Kantrowitz, E. R., *Proteins* **1999,** *37* (4), 729-42.
21. Poulsen, F. M.; Hoch, J. C.; Dobson, C. M., *Biochemistry* **1980,** *19* (12), 2597-607.
22. Velyvis, A.; Schachman, H. K.; Kay, L. E., *J. Am. Chem. Soc.* **2009,** *131* (45), 16534-43.
23. Chao, F. A.; Kim, J.; Xia, Y.; Milligan, M.; Rowe, N.; Veglia, G., *J. Magn. Reson.* **2014,** *245*, 17-23.
24. Xu, Y.; Matthews, S., *J. Biomol. NMR* **2013,** *55* (2), 179-87.
25. Duan, X.; Quiocho, F. A., *Biochemistry* **2002,** *41* (3), 706-12.
26. Tugarinov, V.; Choy, W. Y.; Orekhov, V. Y.; Kay, L. E., *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (3), 622-627.
27. Tugarinov, V.; Kay, L. E., *J. Am. Chem. Soc.* **2003,** *125* (45), 13868-78.
28. Forster, A.; Masters, E. I.; Whitby, F. G.; Robinson, H.; Hill, C. P., *Mol. Cell* **2005,** *18* (5), 589-99.
29. Strogatz, S. H., *Nature* **2001,** *410* (6825), 268-76.
30. Willett, P., *MATCH Commun. Math. Comput. Chem.* **2016,** (In Press).
31. Vuillon, L.; Lesieur, C., *Curr. Opin. Struct. Biol.* **2015,** *31*, 1-8.
32. Nature Reviews GeneticsParikshak, N. N.; Gandal, M. J.; Geschwind, D. H., *Nat. Rev. Genet* **2015,** *16* (8), 441-458.
33. Wiesner, S.; Sprangers, R., *Curr. Opin. Struct. Biol.* **2015,** *35*, 60-7.
34. Bunke, H. In *Graph matching: Theoretical foundations, algorithms, and applications*, Proc. Vision Interface, 2000; pp 82-88.
35. Mcgregor, J. J., *Softw Pract Exp* **1982,** *12* (1), 23-34.
36. Cordella, L. P.; Foggia, P.; Sansone, C.; Vento, M., *IEEE Trans. Pattern Anal. Mach. Intell.* **2004,** *26* (10), 1367-1372.

37.   Richardson, C. M.; Williamson, D. S.; Parratt, M. J.; Borgognoni, J.; Cansfield, A. D.; Dokurno, P.; Francis, G. L.; Howes, R.; Moore, J. D.; Murray, J. B.; Robertson, A.; Surgenor, A. E.; Torrance, C. *Bioorg. Med. Chem. Lett.* **2006,** *16* (5), 1353-7.
38.   Lichtenecker, R. J.; Coudevylle, N.; Konrat, R.; Schmid, W., *Chembiochem* **2013,** *14* (7), 818-21.
39.   Han, B.; Liu, Y. F.; Ginzinger, S. W.; Wishart, D. S., *J. Biomol. NMR* **2011,** *50* (1), 43-57.
40.   Sahakyan, A. B.; Vranken, W. F.; Cavalli, A.; Vendruscolo, M., *J. Biomol. NMR* **2011,** *50* (4), 331-346.
41.   Stieglitz, K.; Stec, B.; Baker, D. P.; Kantrowitz, E. R., *J. Mol. Biol.* **2004,** *341* (3), 853-868.
42.   Lowe, J.; Stock, D.; Jap, R.; Zwickl, P.; Baumeister, W.; Huber, R., *Science* **1995,** *268* (5210), 533-539.
43.   Howard, B. R.; Endrizzi, J. A.; Remington, S. J., *Biochemistry* **2000,** *39* (11), 3156-3168.
44.   Trepel, J.; Mollapour, M.; Giaccone, G.; Neckers, L., *Nat. Rev. Cancer* **2010,** *10* (8), 537-49.
45.   Stebbins, C. E.; Russo, A. A.; Schneider, C.; Rosen, N.; Hartl, F. U.; Pavletich, N. P., *Cell* **1997,** *89* (2), 239-50.
46.   Lichtenecker, R. J.; Weinhäupl, K.; Reuther, L.; Schörghuber, J.; Schmid, W.; Konrat, R., *J. Biomol. NMR* **2013,** *57* (3), 205-209.
47.   Proudfoot, A.; Frank, A. O.; Ruggiu, F.; Mamo, M.; Lingel, A., *J. Biomol. NMR* **2016,** *65* (1), 15-27.
48.   Goddard, T. D. K., D. G., SPARKY 3. *University of California, San Francisco*.
49.   Csardi, G.; Nepusz, T., *Inter. J. Comp. Syst.* **2006,** *1695* (5), 1-9.
50.   Munkres, J., *J. Soc. Ind. Appl. Math.* **1957,** *5* (1), 32-38.
51.   Levandowsky, M.; Winter, D., *Nature* **1971,** *234* (5323), 34-35.
52.   Hagberg, A., Schult DA., Swart, PJ. Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA, August 19-24, 2008.; Varoquaux G, J Millman, Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA, 2008; pp 11-15.
53.   Ramachandran, P.; Varoquaux, G., Mayavi: 3D Visualization of Scientific Data. *Comput Sci Eng* **2011,** *13* (2), 40-50.

54.   Sounier, R.; Blanchard, L.; Wu Z.; Boisbouvier J., *J. Am. Chem. Soc.* **2007,** *129* (3), 472-473.
55. Sprangers, R.; Kay, L. E., Nature **2007**, 445 (7128), 618-22