On the role of conformational dynamics in enzyme function

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Fa-An Chao

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisers

Dr. Gianluigi Veglia, Dr. Burckhard Seelig

December 2013

Acknowledgements

First of all, I would like to thank my advisors, Dr. Gianluigi Veglia and Dr. Burckhard Seelig. In Veglia's lab, I was given lots of space to develop my own approaches and use my talent to solve many problems in research. His support and constant excitement for research has inspired me to become a better scientist. At the same time, Dr. Seelig is a very good mentor in different aspects, including scientific research and daily life. He provided lots of valuable suggestions which guided me to the end of my PhD study.

I would also thank all the past and present members of both the Veglia group and the Seelig group who have made the last five years a great experience. Especially Larry Masterson, Lei Shi, Kaustubh Mote, and Jonggul Kim, I thank them for all the brilliant ideas, collaborations, and helpful discussions. I also want to thank Aleardo Morelli, John Haugner, Lewis Churchfield, and Leonardo Hagmann, who contributed a lot in the Ligase 10C project.

Finally, I would like to thank the people that have supported me outside the lab. My parents, grand-parents, brothers and sisters have always encouraged me to do what I want to do even when I decided to move all the way to USA.

Last but not least, I thank Grace for always being there with me.

Abstract

The inherent flexibility of biological macromolecules plays a critical role in their function. This is especially true in the case of enzymes, as these proteins must be able to alter their molecular structures in order to recognize substrates and release products. While they have been shown to be critical for understanding how enzymes function, the exact roles of these structural dynamics are still largely unknown. I have used Nuclear Magnetic Resonance (NMR) Spectroscopy to study the impact of these structural fluctuations for enzyme function in two dramatically different systems: RNA Ligase 10C, a novel enzyme derived via *in vitro* selection, and the ubiquitous eukaryotic signaling enzyme cAMP-dependent Protein Kinase A (PKA). Despite the variation in their structures and function, I have nonetheless been able to demonstrate the importance of the conformational dynamics of structural elements involved in both systems, suggesting that this is a universal phenomenon, essential for function in all enzymes.

Table of Contents

List of Tables

List of Figures

Chapter 1- Introduction

Correlation between conformational dynamics and enzyme function

It has been long recognized that motions inherent to all biological molecules are essential for their respective functions, including ligand binding, catalysis, and allostery (1-6). The time scales for these motions range from the very fast motions of bond vibration and side-chain rotation, to the slower dynamics of rearrangement of secondary structures, all the way down to the sluggish "breathing" of multiple domains for biological functions (7,8). While the existence of these motions has been well recognized, the exact roles of these movements in enzyme catalysis remain less understood. Questions as to how enzymes move and alter their structural elements to recognize substrates and release products are largely unanswered. Furthermore, the possible role these molecular movements may play in the chemical step of enzyme catalysis is left undefined (5,9).

In our attempts to answer the above questions, nuclear magnetic resonance (NMR) spectroscopy has been increasingly pushed to the forefront of the biophysical methods arsenal (1,10,11). Its ability to probe molecular movements ranging from the picosecond to the second time scales with site-specific resolution makes NMR uniquely suited to these types of investigations. There are many NMR experiments available for probing conformational dynamics at different time scales. At the fast time scale (picosecond to nanosecond), the spin-lattice (T_1) and spin-spin (T_2) relaxation times of a nuclear spin allow one to

probe conformational dynamics (11,12). (T_1: the time of returning the inverted nuclear spin population back to the equilibrium; T_2: the time of dephasing the magnetization in the transverse plane.) Using model-free analysis (13), the order parameters, indicators of rigidity, extracted from those relaxation times can be used as qualitative reporters of entropy in proteins, which have been shown in the binding of calmodulin to many different proteins (14,15). Another spectroscopic phenomenon that has been exploited more recently with NMR is chemical exchange, in which a chemical group exchanges between two different chemical environments on a microsecond to millisecond timescale. The Carr-Purcell-Meiboom-Gill (CPMG) is the most common one of these experiments, allowing chemical exchange rates to be probed and quantified, in favorable cases (16,17).

CPMG experiments have been used to investigate the roles of conformational dynamics along the catalytic cycle for many different enzymes (2,4,6,18). These investigations have illuminated the crucial impact of conformational transitions and molecular dynamics on enzyme function. For example, a study of the Escherichia coli dihydrofolate reductase (DHFR) from Wright's group established the relationship between the enzyme's conformational dynamics and reaction rates using exchange rates from CPMG experiments (2). At each catalytic step, this enzyme samples the conformations of both preceding and following intermediates based on the consistent chemical shift differences from titration experiments and CPMG experiments. After each catalytic step, not only the

relative populations of the enzyme's conformations are shifted, but also its conformational dynamics are also altered. These conformational transitions, from excited states to ground states, were shown to be the rate-limiting steps for important catalytic processes such as hydride transfer, cofactor release, and product release (**Figure 1**). This correlation between conformational dynamics and catalytic function was determined in other enzymes as well. In the case of the thermophilic-mesophilic pair of adenylate kinases, differences in catalytic efficiency at ambient temperature for each kinase were shown to result from differing conformational dynamics in the proteins (4). These and other enzymatic studies support the hypothesis that conformational dynamics govern the rates of important catalytic steps (3,6,18,19); two enzymes differing significantly in both form and function are currently under investigation to further examine this relationship. The results of these studies could lead to universal answers to the question of molecular dynamics and enzyme function.

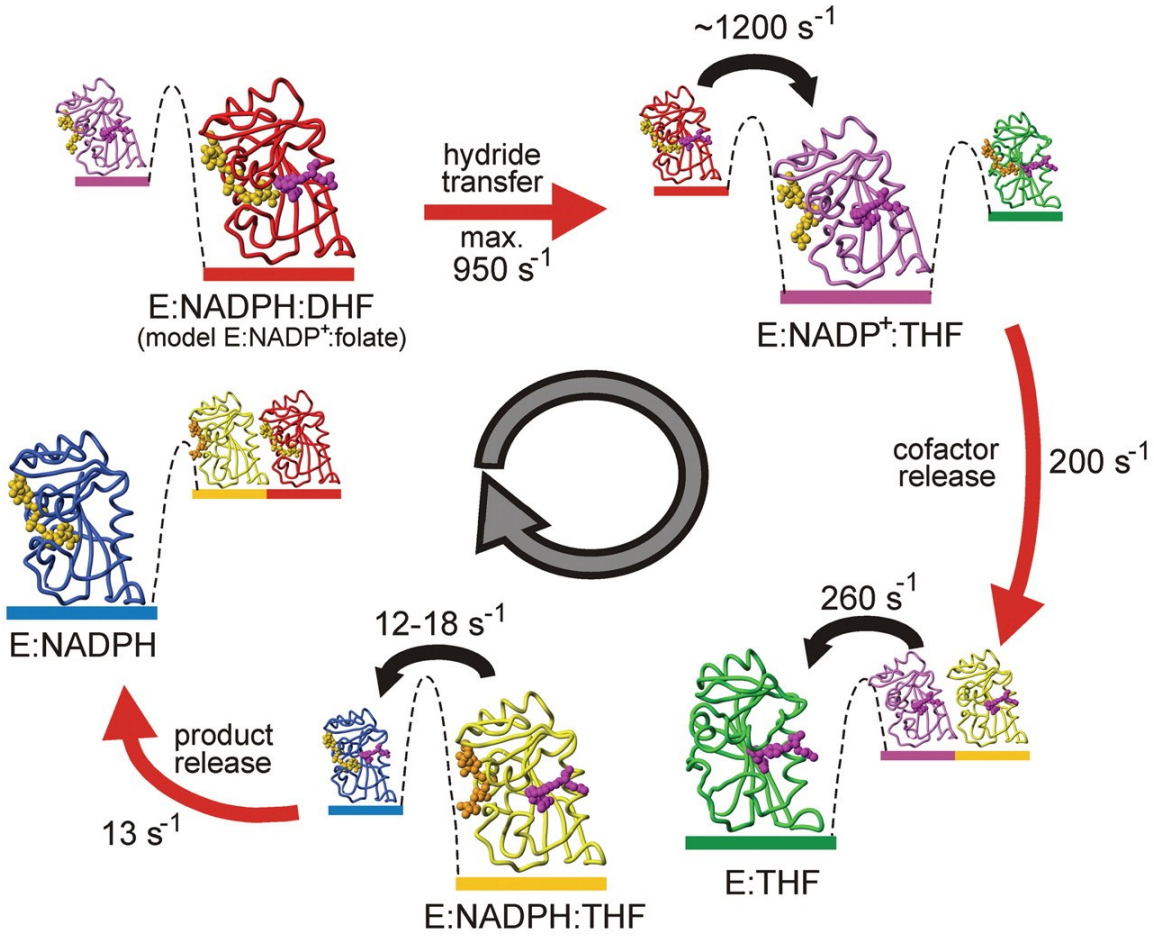


Figure 1. The dynamic energy landscape of DHFR catalysis. Ground state (larger) and higher energy (smaller) structures of each intermediate in the cycle, modeled on the published X-ray structures, with NADPH and $NADP^+$ shown in gold and substrate, product, and analogs shown in magenta. For each intermediate in the catalytic cycle, the higher energy conformations detected in the relaxation dispersion experiments resemble the ground-state conformations of adjacent intermediates; their interconversion rates, also obtained from the relaxation dispersion experiments, are shown with black arrows. Rate constants for the interconversion between the complexes, measured by pre-steady-state

enzyme kinetics are indicated with red arrows. Figure from Boehr et al (2).

RNA ligase 10C from directed evolution

RNA ligase 10C is an artificial enzyme evolved and selected by Dr. Seelig from a partially randomized protein library (20,21) (**Figure 2**). The in vitro evolutionary process began with a partially randomized DNA library (22) (upper left in **Figure 2A**) based on the DNA sequence of the stable DNA-binding domain of human retinoid X receptor (hRXRα) (**Figure 2C**). In order to make an mRNA-displayed protein library, an mRNA library was generated via in-vitro transcription, and then each mRNA was subsequently chemically ligated with a puromycin molecule (23). After in-vitro translation, an mRNA-displayed protein library (RNA-protein fusions) was generated (upper right in **Figure 2A**). At this point, the phenotypes (proteins) are connected to the genotypes (mRNA). Reverse transcription was performed to create cDNA templates to form cDNA-mRNA-protein complexes and to prepare for the selection step (lower right in **Figure 2A**). In vitro selection was based on the ability of the protein to ligate its RNA primer in preparation for reverse transcription with another piece of RNA chemically linked to the beads (lower left in **Figure 2A**). A new cDNA library from the selected sub-population was amplified by error-prone PCR (upper left in **Figure 2A**); the resulting cDNA was then used to generate another mRNA library for selection, starting the cycle over again. This cycle was repeated until most of the population demonstrated the desired activity. RNA ligase 10C, the result of this evolutionary process, can

catalyze an unusual reaction: the ligation of a 5' triphosphate group of RNA with a 3' hydroxyl group of RNA at 65 ^{0}C (**Figure 2B**). However, as neither the structure nor conformational dynamics of this artificial enzyme were available to aid in understanding the catalytic mechanism, we set out to acquire structural and dynamic data to allow us to understand its catalytic reaction and tune enzyme activity.

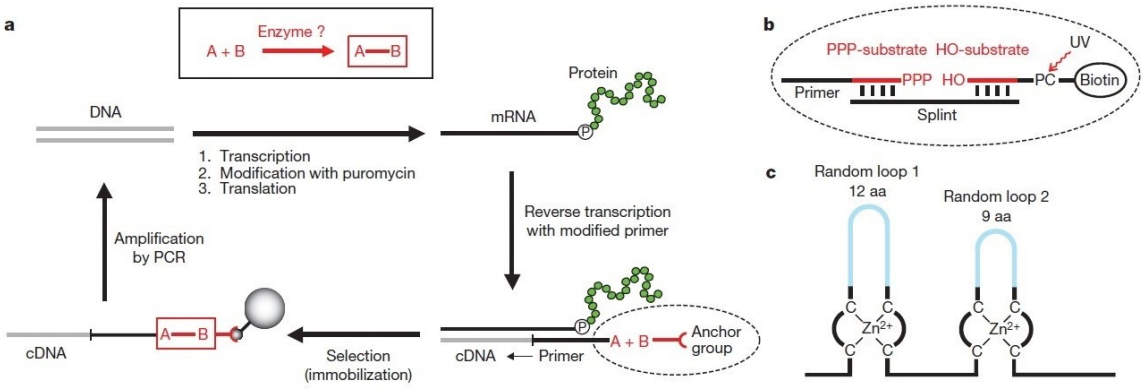


Figure 2. In vitro selection of enzymes by mRNA display. A) General selection scheme for enzymes that catalyze bond-forming reactions. A DNA library is transcribed into RNA, cross-linked to a 3'-puromycin oligonucleotide, and *in vitro* translated. The library of mRNA-displayed proteins is reverse transcribed with a primer bearing substrate A. Substrate B, which carries an anchor group, is added. Proteins that join A and B attach the anchor group to their encoding cDNA. Selected cDNA sequences are then amplified by PCR, and used as input for the next round. B) Selection of enzymes that perform template-dependent ligation of a 5'-triphosphate-activated RNA (PPP-substrate) to a second RNA with a 3'-hydroxyl (HO-substrate). The PPP-substrate is ligated to the primer and then used in the reverse transcription reaction. The cDNA of the catalytically

active molecules is immobilized on streptavidin-coated beads via biotin, washed, and released by ultraviolet (UV)-irradiation of the photocleavable linker (PC). C) The scaffolded library is based on a two zinc finger domain with two loop regions (light blue) that are replaced by segments of 12 or 9 random amino acids (aa). Figure from Seelig at al (20).

Protein kinase A from natural evolution

Protein kinase A, also called cAMP-dependent protein kinase, belongs to serine/threonine family of protein kinases (24). The holoenzyme consists of two regulatory R subunits and two catalytically active C subunits (PKA-C) (25); upon cAMP binding to the R subunit, the C subunits are released (**Figure 3**). PKA-C phosphorylates multiple substrates to modulate their functions, adopting three separate conformations along its catalytic cycle (3,26). The open or apo form of PKA-C binds ATP, resulting in a conformation transition to the binary form. The ternary or closed form of the enzyme results from substrate or inhibitor binding to this binary form. However, the crystal structures of these different forms of PKA-C can be superimposed (27), suggesting that the mechanism of function is not solely structurally based. Interestingly, previous reports from Veglia's group have found that the different states of PKA-C are correlated with significantly different backbone conformational dynamics (3,19,27) (**Figure 4**). However, the details of allosteric regulation of PKA-C are still obscure. Further investigation of the side-chain conformational dynamics using methyl-TROSY spectroscopy (28) would

complete the overall picture of this enzyme's allosteric regulation.

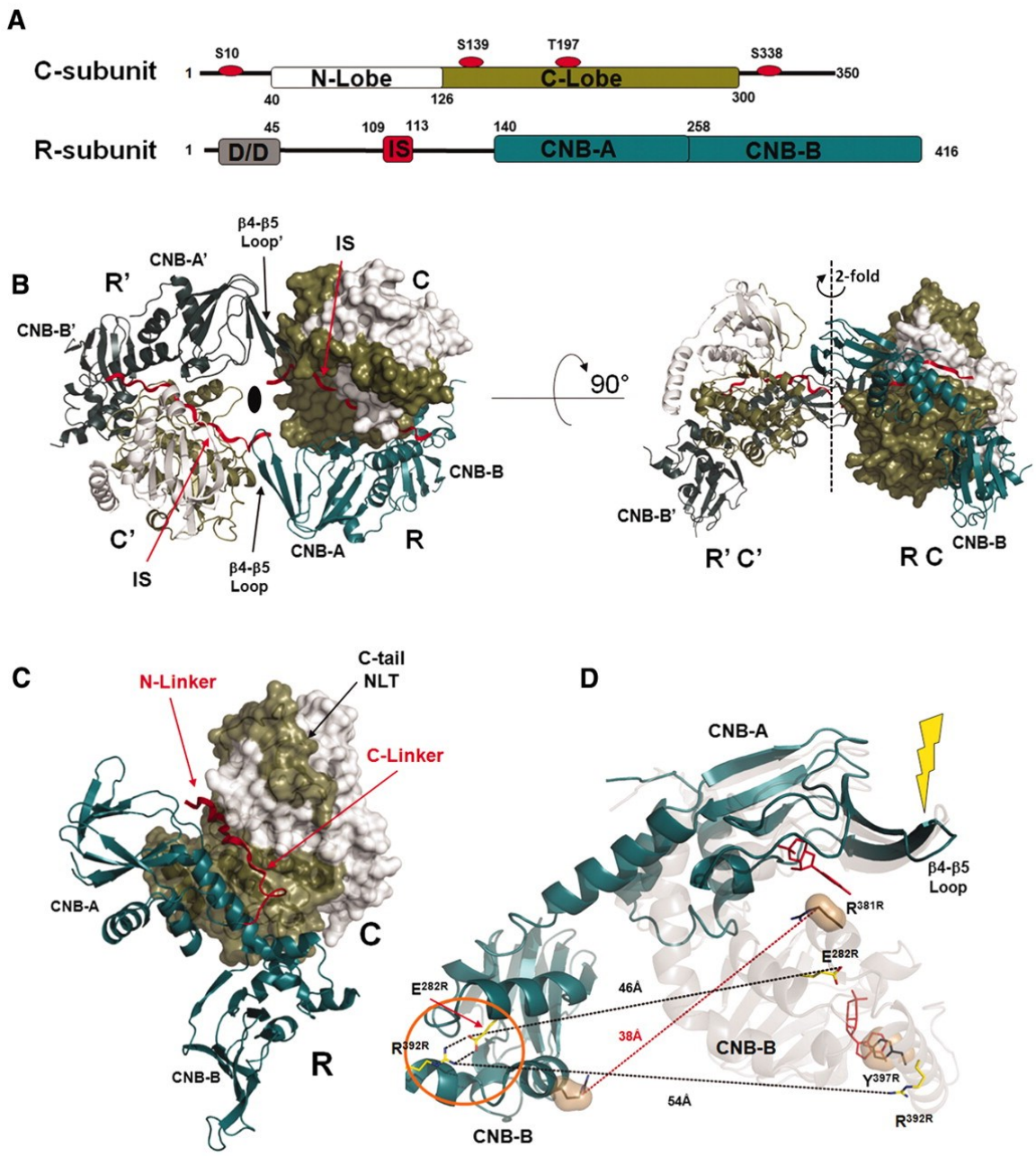


Figure 3. Overall view of the RIIβ2:C2 tetrameric holoenzyme. A) Domain organization and color coding of the R and C subunits. The four red spheres indicate the phosphorylation sites in C subunit. B) Structure of the $RII\beta_2$:C_2

tetrameric holoenzyme. The N lobe of the C subunit is white; the C lobe and C tail are tan. The N-linker segment of the R subunit that contains the IS (inhibitory site) and docks to the active-site cleft of the C subunit is shown as a red ribbon. One heterodimer is labeled as RC, and its twofold symmetry mate is labeled as R'C'. The twofold axis position is shown as a black oval (left) and dotted line (right). (**C**) One RIIβ:C heterodimeric holoenzyme structure in the $RII\beta_2:C_2$ tetrameric holoenzyme. (**D**) RIIβ undergoes dramatic conformational changes upon binding to the C subunit. RIIβ bound to the C subunit is shown on the left and bound to cAMP shown on the right (PDB ID code 1CX4, in black). The salt bridge Glu^{282R}-Arg^{392R} formed upon binding to the C subunit is shown in an orange circle. Figure from Zhang et al (25).

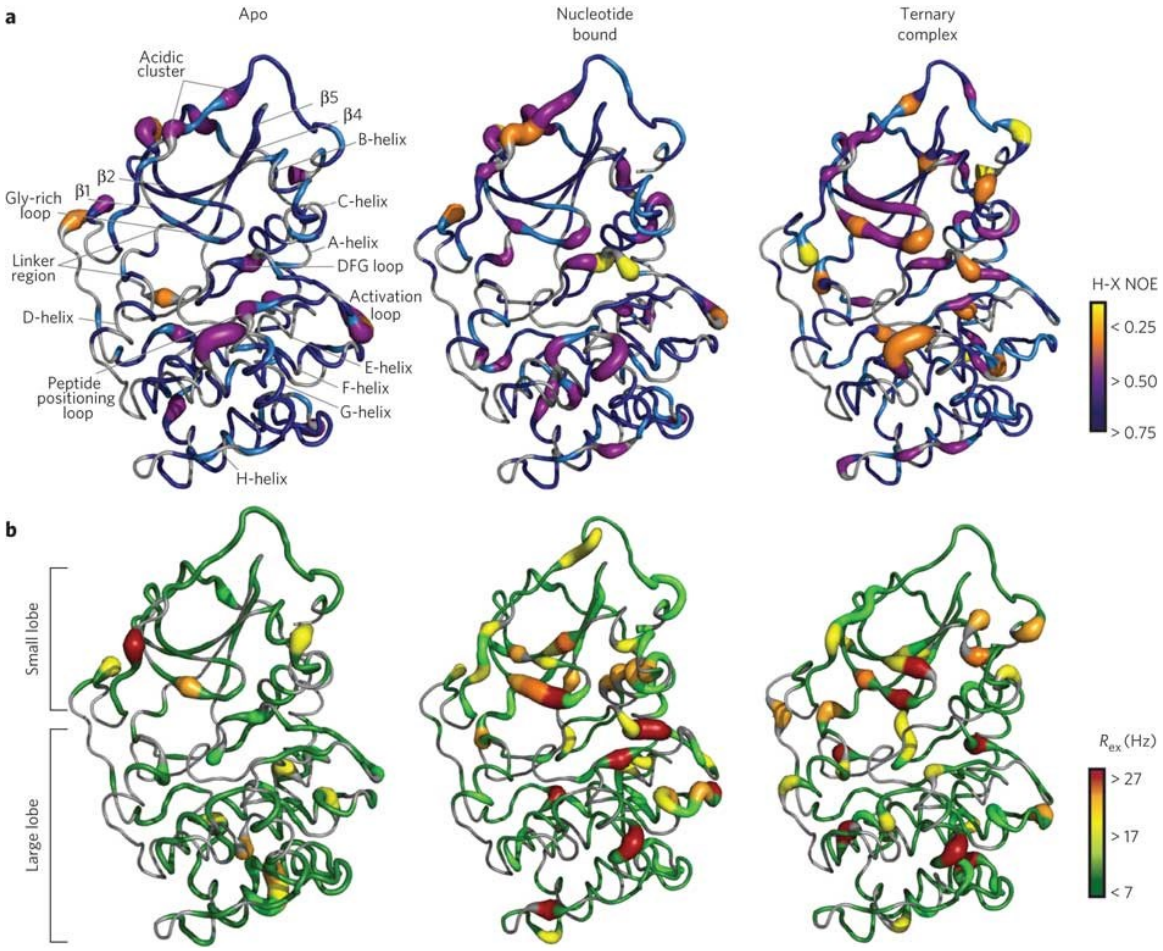


Figure 4. Mapping of the backbone amide dynamics of PKA-C from the apo to ternary complex. Structural elements and loops are indicated. Fast (A) and slow (B) dynamics are shown for PKA-C in the apo (left), binary complex (middle), and the ternary complex containing AMP-PNP and $PLN_{1–20}$ (right). Figure from Masterson et al (3).

Objectives

As has been previously stated, the importance of conformational dynamics in enzymatic functions is supported by many examples in current literature (2-7,14,15,18,19,29). Based on these reported phenomena, I hypothesized that enzyme activity can be regulated by tuning the conformational dynamics. To test this hypothesis, I have used NMR spectroscopy to probe the conformational dynamics of the two distinct enzymes described above, artificial RNA ligase 10C and protein kinase A. As both the structure and the catalytic mechanism of RNA ligase 10C are unknown, I also needed to determine the structure of this enzyme in solution using NMR methods. More specifically, I aimed to answer the following questions:

- How is the structure of RNA ligase 10C evolved from that of the template protein, the DNA-binding domain of hRXRα?
- What is the role of conformational dynamics in the activity of RNA ligase 10C?
- How are the enzyme functions of PKA regulated by ligands through changes in conformational dynamics?

By answering the above questions, we are better able to understand the detailed mechanism of conformational dynamics in enzyme function, an increased understanding that will lead to the design of better artificial enzymes and drugs to treat the many diseases resulting from enzyme dysfunction.

References

1. Boehr DD, Dyson HJ, Wright PE. An NMR perspective on enzyme dynamics. Chem Rev. 2006 08/01; 2013/12;106(8):3055-79.

2. Boehr DD, McElheny D, Dyson HJ, Wright PE. The dynamic energy landscape of dihydrofolate reductase catalysis. Science. 2006 Sep 15;313(5793):1638-42.

3. Masterson LR, Cheng C, Yu T, Tonelli M, Kornev A, Taylor SS, et al. Dynamics connect substrate recognition to catalysis in protein kinase A. Nat Chem Biol. 2010 Nov;6(11):821-8.

4. Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ, Kern D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat Struct Mol Biol. 2004 Oct;11(10):945-9.

5. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, et al. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science. 2011 Apr 8;332(6026):234-8.

6. Whittier SK, Hengge AC, Loria JP. Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. Science. 2013 Aug 23;341(6148):899-903.

7. Yon JM, Perahia D, Ghélis C. Conformational dynamics and enzyme activity. Biochimie. 1998 1;80(1):33-42.

8. McCammon JA, Harvey SC, editors. Dynamics of proteins and nucleic acids. Cambridge: Cambridge University Press; 1987.

9. Loveridge EJ, Behiry EM, Guo J, Allemann RK. Evidence that a â€˜dynamic knockoutâ€™ in escherichia coli dihydrofolate reductase does not affect the chemical step of catalysis. Nat Chem. 2012 print;4(4):292-7.

10. Ishima R, Torchia DA. Protein dynamics from NMR. Nat Struct Biol. 2000 Sep;7(9):740-3.

11. Palmer AG,3rd. NMR characterization of the dynamics of biomacromolecules. Chem Rev. 2004 Aug;104(8):3623-40.

12. Malcolm H. Levitt. Spin dynamics : Basics of nuclear magnetic resonance. second edition ed. John Wiley & Sons Ltd; 2007.

13. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. J Am Chem Soc. 1982 08/01; 2012/03;104(17):4546-59.
14. Marlow MS, Dogan J, Frederick KK, Valentine KG, Wand AJ. The role of conformational entropy in molecular recognition by calmodulin. Nat Chem Biol. 2010 May;6(5):352-8.
15. Frederick KK, Marlow MS, Valentine KG, Wand AJ. Conformational entropy in molecular recognition by proteins. Nature. 2007 Jul 19;448(7151):325-9.
16. Loria JP, Rance M, Palmer AG. A relaxation-compensated Carrâˆ'Purcellâˆ'Meiboomâˆ'Gill sequence for characterizing chemical exchange by NMR spectroscopy. J Am Chem Soc. 1999 03/01; 2013/12;121(10):2331-2.
17. Millet O, Loria JP, Kroenke CD, Pons M, Palmer AG. The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. J Am Chem Soc. 2000 03/01; 2013/12;122(12):2867-77.
18. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, et al. Intrinsic dynamics of an enzyme underlies catalysis. Nature. 2005 Nov 3;438(7064):117-21.
19. Masterson LR, Shi L, Metcalfe E, Gao J, Taylor SS, Veglia G. Dynamically committed, uncommitted, and quenched states encoded in protein kinase A revealed by NMR spectroscopy. Proc Natl Acad Sci U S A. 2011 Apr 26;108(17):6969-74.
20. Seelig B, Szostak JW. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature. 2007 Aug 16;448(7155):828-31.
21. Seelig B. mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. Nat Protoc. 2011 Apr;6(4):540-52.
22. Cho G, Keefe AD, Liu R, Wilson DS, Szostak JW. Constructing high complexity synthetic libraries of long ORFs using in vitro selection. J Mol Biol. 2000 Mar 24;297(2):309-19.

23. Kurz M, Gu K, Lohse PA. Psoralen photo-crosslinked mRNA-puromycin conjugates: A novel template for the rapid and facile preparation of mRNA-protein fusions. Nucleic Acids Res. 2000 Sep 15;28(18):E83.
24. Adams JA. Kinetic and catalytic mechanisms of protein kinases. Chem Rev. 2001 08/01; 2012/07;101(8):2271-90.
25. Zhang P, Smith-Nguyen EV, Keshwani MM, Deal MS, Kornev AP, Taylor SS. Structure and allostery of the PKA RIIbeta tetrameric holoenzyme. Science. 2012 Feb 10;335(6069):712-6.
26. Johnson DA, Akamine P, Radzio-Andzelm E, Madhusudan, Taylor SS. Dynamics of cAMP-dependent protein kinase. Chem Rev. 2001 08/01; 2013/12;101(8):2243-70.
27. Masterson LR, Cembran A, Shi L, Veglia G. Allostery and binding cooperativity of the catalytic subunit of protein kinase A by NMR spectroscopy and molecular dynamics simulations. Adv Protein Chem Struct Biol. 2012;87:363-89.
28. Sheppard D, Sprangers R, Tugarinov V. Experimental approaches for NMR studies of side-chain dynamics in high-molecular-weight proteins. Prog Nucl Magn Reson Spectrosc. 2010 Jan;56(1):1-45.
29. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature. 2007 Dec 6;450(7171):913-6.

Chapter 2- Structure and dynamics of a primordial catalytic fold generated by in vitro evolution

Reprinted with permission from:

Fa-An Chao, Aleardo Morelli, John C Haugner III, Lewis Churchfield, Leonardo N Hagmann, Lei Shi, Larry R Masterson, Ritimukta Sarangi, Gianluigi Veglia and Burckhard Seelig. Structure and dynamics of a primordial catalytic fold generated by in vitro evolution. Nature Chemical Biology 9, 81–83 (2013).

Engineering functional protein scaffolds capable of carrying out chemical catalysis is a major challenge in enzyme design. Starting from a non-catalytic protein scaffold, we recently generated a novel RNA ligase by *in vitro* directed evolution. This artificial enzyme lost its original fold and adopted an entirely novel structure with dramatically enhanced conformational dynamics, demonstrating that a primordial fold with suitable flexibility is sufficient to carry out enzymatic function.

2.1 Introduction

The known structures of naturally occurring proteins can be assigned to an apparently finite number of different fold families (1,2). Starting from an existing fold, divergent evolution through a combination of gene duplication and mutations is a common path for proteins to acquire new functions while retaining their original fold (3,4). However, the origin of those biological folds remains subject to debate (5,6). Only a few examples have been described in which new function acquisition is accompanied by a simultaneous change in the protein fold. Those examples have largely been generated by rational design or involve protein binders (7-13).

Recently, we created artificial RNA ligase enzymes by *in vitro* evolution (14,15). These enzymes catalyze the joining of a 5′-triphosphorylated RNA to the 3′-hydroxyl group of a second RNA, a reaction for which no natural enzyme catalysts have been found. We began with a small noncatalytic protein domain consisting of two zinc finger motifs from the DNA-binding domain of human retinoid X receptor (hRXRα) (16) (**Figure 1**). Two adjacent loops of this protein were randomized to generate a combinatorial library of mutants as input for the selection and evolution process (17). Although zinc fingers are common structural motifs, they are not known to take part in catalysis in natural proteins. In contrast, we isolated from the zinc finger library active enzymes that have rate accelerations of more than 2-million-fold (14). Sequence analysis of the artificial enzyme showed that several amino acids essential to maintaining zinc finger structure integrity were mutated or deleted, suggesting that the original scaffold may have been abandoned during the process of mutagenesis and evolution. The original hRXRα scaffold consisted of two loop-helix domains, each containing a zinc ion tetrahedrally coordinated by four cysteines (16). However, during evolution of the ligase enzyme, only half of the zinc-coordinating cysteines had been conserved. In the starting scaffold, two helices were packed perpendicularly to form the globular fold and build the hydrophobic core, and an

additional helix was located at the C terminus (**Figure 1b**). In the ligase enzyme, seven residues of the former DNA recognition helix and ten residues of the C-terminal helix were deleted from the original hRXRα scaffold.

2.2 Results

NMR structural analyses of the ligase 10C (Online Methods), chosen for its superior solubility and thermostability, revealed that the evolved ligase lost the original zinc finger scaffold, adopting an entirely new structure (**Figure 1b**). This new 3D structure still contained two zinc sites that constitute the folding core of the protein; however, the two Zn^{2+} ions were coordinated by several new ligands with a different register. The deletion of two N-terminal cysteines during directed evolution resulted in the concomitant rearrangement of the local geometry of the zinc-binding loop. Additionally, the short stretch of antiparallel β-sheet within the first zinc finger (Zn-I) was also deleted. The C-terminal loop-helix domains and the recognition helix of hRXRα responsible for binding to the DNA groove (18) were lost completely; the latter was replaced by an unstructured loop of twenty amino acids connecting the two new zinc fingers. The zinc fingers made up the most structured region, as demonstrated by the presence of short- and long-range NOE contacts. Moreover, several long-range NOEs indicated that the two metal-binding loops are in close proximity, whereas most of the protein presented only short-range NOE contacts (**Supplementary Results**, **Figure S1** and **Table S1**). The conformational ensemble resulting from simulated annealing calculations showed two well-defined regions (residues 17–35 and 49–69) with r.m.s. deviation from the average of less than 1 Å, whereas the large loop encompassing residues 36–48 was completely unstructured (r.m.s. deviation greater than 6 Å). The 3D structure of the enzyme was compounded by residual dipolar coupling measurements, which also helped to better define the local geometry around the zinc-binding sites (**Figures S2** and **S3**).

The two metal centers were responsible for the overall fold of the ligase. In the absence of Zn^{2+}, the NMR fingerprint spectrum of the enzyme showed broad and mostly unresolved resonances, typical of a molten globule. Titration of Zn^{2+} to the metal-free protein first saturated the C-terminal Zn^{2+}-binding site (Zn-II) and induced a substantial structural rearrangement with sharper and more dispersed resonances (**Figures S4** and **S5**). The transition between the unfolded and folded states of the ligase involved multiple intermediate species. For selected resonances, we could discern the presence of two distinct states in slow exchange in the NMR timescale. Complete saturation with Zn^{2+} funneled the enzyme into a more defined structure, with the complete resolution of fingerprint resonances showing only one population of peaks. Elemental analysis by inductively coupled plasma MS revealed 2.74 ± 0.01 equivalents (± s.d.) of bound zinc per ligase molecule. We were able to fit the thermocalorimetry data using models with two or more Zn^{2+}-binding sites; the fit, however, does not improve considerably with $n > 2$ (**Figure S6** and **Table S2**). Assigning two sites in accordance with the NMR titration data leads to one binding site, Zn-II, with higher affinity (Kd ~3 μM) and a second binding site, Zn-I, with lower affinity for Zn^{2+} (Kd ~93 μM). These values were further supported by the zinc concentration dependence of the enzyme activity, showing a steep drop in activity at concentrations below 100 μM Zn^{2+} (**Figure S7**). Notably, the ligase affinity for Zn^{2+} was substantially lower than those reported for natural zinc-containing proteins, which commonly have dissociation constants of 10^{-8} to10^{-13} M (19). Structure calculations were carried out in the absence of explicit Zn^{2+} ions to avoid conformational search bias and converged toward a structural ensemble with two distinct Zn^{2+}-binding sites: the tetracoordinated N-terminal site (Zn-I) with weaker binding affinity and the hexacoordinated C-terminal loop (Zn-II) with higher binding affinity (**Figures S2** and **S3**). Extended X-ray absorption fine structure (EXAFS) data corroborated these results, showing that both Zn^{2+} sites coordinated with two sulfur ligands (cysteines) with a Zn-S distance of 2.3 Å, and at least one site had four Zn-N or Zn-O ligands, whereas the other site had two to

four. These atom ligands can be either protein-based molecules or water molecules (**Figure S8**).

The directed evolution process that yielded the artificial enzyme was based only on product formation without structural constraints (14). As a result, the ligase enzyme evolved into a new structure with substantially increased conformational dynamics compared to the original DNA-binding scaffold (20). In fact, ligase 10C showed an overall increase in structural plasticity and malleability. Although the two zinc fingers had heteronuclear NOEs similar to those of the original scaffold, the loop region that replaced the recognition helix had much higher flexibility, with heteronuclear NOEs below 0.5. These data indicate augmentation of conformational dynamics in the ps-ns timescale supported by longitudinal (T_1) and transverse (T_2) nuclear spin-relaxation measurements as well as hydrogen-deuterium exchange data (**Figure 2a** and **Figures S9** and **S10a**). A distinct signature of the hRXRα structure is slow (ms-ms) conformational dynamics (20) (**Figure S10b**), which may be correlated with the protein's ability to optimize protein-DNA interactions. The *in vitro* evolution of hRXRα into the RNA ligase redistributed those conformational dynamics, particularly in the region N-terminal to the Zn^{2+}-binding site Zn-II (residues 46–53; **Figure 2b**).

To probe the substrate-binding surface, we carried out an NMR titration with a pseudosubstrate that lacked the 2′-hydroxyl group, preventing enzyme turnover (**Figure S11**). Chemical shift perturbation mapping of the ligase structure (**Figure 3a** and **Figure S12**) indicated that one of the highly perturbed regions in the substrate-bound form (residues 46–53) corresponds to high values of chemical exchange (slow conformational dynamics) in the substrate-free form (**Figure 2b**). Notably, most alanine mutations in this region decreased or completely obliterated the enzyme's activity (**Figure 3b**). Specifically, mutations E48A, Y50A and H51A abolished enzymatic activity, whereas the C47A and C53A mutations caused a 97% reduction in ligase function. These residues' high conservation among evolved ligase variants further demonstrated their importance (**Figure 3c**). The combined results suggest that this protein region (residues 46–53) is

important for substrate recognition and binding and may contain the active site of the enzyme. Four of those five mutation-sensitive residues (Cys47, Glu48, His51 and Cys53) are good potential metal ligands. Many natural enzymes, such as polymerases, that catalyze chemical reactions similar to the specific RNA ligation described here use a mechanism involving catalytic divalent metal ion cofactors, which are coordinated jointly by the nucleic acid substrates and active site residues of the enzyme (21). One may speculate that, upon forming the enzyme-substrate complex, some of the mutation-sensitive residues in ligase 10C are involved in binding additional Zn^{2+} ions that facilitate catalysis but are not bound by the protein alone. However, additional experiments studying the enzyme in complex with substrate are needed to elucidate the catalytic mechanism of our artificial enzyme.

The increased flexibility of the new ligase structure relative to the parent hRXRα could potentially originate from their different functional roles. hRXRα is a DNA binder and has been proposed to work through an induced-fit mechanism (16). In contrast, the RNA ligase has evolved to function as a catalyst. This role requires additional flexibility to optimize interactions with a target molecule and carry out chemical catalysis using transient interactions that occur in excited conformational states rather than through a stable, low-energy complex (22-24). This argument is supported by an independent directed evolution experiment in which the same hRXRα library yielded proteins that bind ATP and maintain the original, noncatalytic DNA-binding scaffold but have no catalytic function (17). In contrast, the evolution of the ligase enzyme resulted in a different structure and increased dynamics.

2.3 Discussion

Compared to natural enzymes, which evolved over billions of years, the laboratory-evolved ligase enzyme contains substantially fewer secondary structure elements, such as α-helices and β-strands, and instead has increased

flexibility. The complete reorganization of the starting scaffold during *in vitro* evolution may have led to the loss of these structural elements. This new structure has not been subjected to the extensive selection pressure that shaped contemporary enzymes during their natural evolution and can therefore be considered an early or primordial catalytic fold. Further evolution of this enzyme *in vitro* or inside a cell will allow us to explore whether incremental mutations lead to structural and dynamic properties more similar to natural enzymes. Although flexibility has been suggested to increase the probability of developing new functions (5), it also reduces overall protein stability—a trade-off that enzymes must balance during evolution.

This report describes what is to our knowledge the first new protein structure emerging simultaneously with a new enzymatic function. This ligase evolved in the absence of selection pressure to maintain the protein's original function (DNA binding). Would proteins evolving in nature also more readily adopt new folds and functions if they were freed from having to maintain their original function? Although the search for such examples in nature is still ongoing, the simplified environment of *in vitro* evolution enables us to generate precedents and study basic principles of complex natural evolution. Finally, *in vitro* directed evolution has the potential to produce new biocatalysts for a wide range of applications. The unique structure of the artificial ligase enzyme demonstrates that this approach can successfully generate new enzymes without being limited to known biological folds (25).

2.4 Supplementary information

All chemical compounds used in this study were purchased from Sigma-Aldrich unless noted otherwise, of molecular biology grade and certified for the absence of RNases when used for ligation reactions.

Sequence of RNA ligase 10C. MGAPVPYPDPLEPRGGKHICAICGNNA EDYKHTDMDLTYTDRDYKNCESYHKCSDLCQYCRYQKDLAIHHQ HHHGGSMGMSGSGTGY

All ligase protein preparations consisted of the sequence above except for point mutations in the case of ligase mutants. Note that the sequence HHQHHH functions similarly to a His_6 tag.

Expression and purification of ^{15}N-labeled ligase protein for NMR studies. Ligase samples were expressed in *Escherichia coli* BL21-DE3 Rosetta strain cells (Novagen). Cells were grown in LB medium with 36 µg/mL kanamycin overnight at 37 °C. This culture was then used to inoculate 1 l of LB medium containing 36 µg/mL kanamycin. The cultures were grown to a $D_{600\ nm}$ of 0.6–0.8 at 37 °C, spun down and resuspended in M9 minimal medium (50 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM NaCl, 2 mM $MgSO_4$, 1 mg/L thiamine, 1 mg/L biotin, 60 µM $ZnSO_4$, 10 g/L dextrose, 1 g/L $^{15}NH_4Cl$ and 36 µg/mL kanamycin, pH = 7.3). Cultures were shaken for 1 h at 37 °C, induced with 1 mM IPTG and shaken overnight at room temperature before being spun down and stored at −20 °C. Frozen cell pellets were resuspended in lysis buffer (20 mM HEPES, 400 mM NaCl, 100 µM $ZnCl_2$, 100 mg/L Triton X-100, 5 mM β-mercaptoethanol, pH = 7.4) and lysed using a S-450D Digital Sonifier (Branson). Cell debris was removed by centrifugation, and the His_6-tagged ligase protein was purified by affinity chromatography using Ni-NTA Superflow resin (Qiagen). The protein was eluted with acidic elution buffer (20 mM NaOAc, 400 mM NaCl, 0.1 mM $ZnCl_2$, 100 mg/L Triton X-100, 5 mM β-mercaptoethanol, pH = 4.5) into 1 M HEPES at pH 7.5 and immediately mixed to adjust the pH. Protein purification was evaluated by SDS-PAGE on Ready Gel precast gels (Bio-Rad). Elution fractions containing ligase

protein were concentrated under high pressure in a stirred-cell concentrator unit with a 5,000-MWCO Ultracel Ultrafiltration cellulose membrane (Millipore) and were dialyzed into FPLC buffer (20 mM HEPES, 150 mM NaCl, 0.1 mM $ZnCl_2$ and 0.5 mM β-mercaptoethanol, pH = 7.5).

Monomer ligase protein was isolated by size-exclusion chromatography using the AKTA FPLC system (GE Healthcare) equipped with a 10 mm × 300 mm column (Tricorn) and Superdex 75 resin (GE Healthcare). The separation was carried out in FPLC buffer. Fractions containing monomer protein were pooled and concentrated using 10,000-MWCO Ultra-4 Centrifugal Filter units (Millipore). Purity was assessed by SDS-PAGE gel (**Fig. S13**).

Expression and purification of ^{15}N/^{13}C-labeled ligase samples for NMR studies. Ligase samples were expressed in *E. coli* BL21-DE3 Rosetta strain cells (Novagen). Cells were grown in LB medium with 36 μg/mL kanamycin overnight at 37 °C, spun down and resuspended in M9 minimal medium (contents as described above, except with 2 g/L [^{13}C] dextrose). The resuspended cells were used to inoculate 100 mL of M9 minimal medium and were grown to a $D_{600\ nm}$ of 0.6 at 37 °C, at which time the culture was used to inoculate 900 mL of M9 minimal medium. The 1-l culture was grown to a $D_{600\ nm}$ of 1.0 at 37 °C, induced with 1 mM IPTG and shaken overnight at 37 °C before being spun down and stored at −20 °C. The ^{15}N,^{13}C-labeled protein was purified in the same manner as the ^{15}N-labeled protein samples.

Expression of selectively labeled ligase protein for NMR studies. Ligase samples were expressed in *E. coli* BL21-DE3 Rosetta strain cells (Novagen). Cells were grown in LB medium with 36 μg/mL kanamycin overnight at 37 °C, spun down and used to inoculate 1 l of selectively labeled M9 medium (40 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM NaCl, 1 mM $MgSO_4$, 50 μM $CaCl_2$, essential vitamins and minerals and 36 μg/mL kanamycin, pH 7.0). To the medium was also added 250 mg of a single ^{15}N-labeled amino acid (cysteine, leucine, lysine or tyrosine), 600 mg of the remaining 19 unlabeled amino acids and, except when labeling [^{15}N] cysteine, one of the following additional amino acid

supplements: 900 mg glutamine, asparagine and arginine when labeling [^{15}N] lysine; 900 mg valine and isoleucine when labeling [^{15}N] leucine; and 900 mg phenylalanine, tryptophan, alanine, serine, glycine and cysteine when labeling [^{15}N] tyrosine. Cultures were grown to a $D_{600\ nm}$ of 1.0 at 37 °C, induced with 1 mM IPTG and shaken for 6 h at 37 °C before being spun down and stored at −20 °C. Selectively labeled protein was purified in the same manner as the ^{15}N-labeled protein samples.

Generation of ligase mutants. Ligase mutants were obtained by site-directed mutagenesis (QuikChange Lightning, Agilent). Plasmid DNA was purified using the QIAprep Spin Miniprep kit (Qiagen). The ligase mutants were verified by DNA sequencing. The primer sequences used to generate the indicated mutations in the ligase were designed in accordance with the QuikChange Primer Design tool (Agilent) and were as follows:

K45A F: 5′-CTACACCGATCGAGACTACGCGAATTGTGAGAGCTACC-3′
K45A R: 5′-GGTAGCTCTCACAATTCGCGTAGTCTCGATCGGTGTAG-3′
N46A F: 5′-CCGATCGAGACTACAAGGCTTGTGAGAGCTACCATAAGTG-3′
N46A R: 5′-CACTTATGGTAGCTCTCACAAGCCTTGTAGTCTCGATCGG-3′
C47A F: 5′-CCGATCGAGACTACAAGAATGCTGAGAGCTACCATAA-3′
C47A R: 5′-TTATGGTAGCTCTCAGCATTCTTGTAGTCTCGATCGG-3′
E48A F: 5′-GACTACAAGAATTGTGCGAGCTACCATAAGTGCTCGG-3′
E48A R: 5′-CCGAGCACTTATGGTAGCTCGCACAATTCTTGTAGTC-3′
S49A F: 5′-AGACTACAAGAATTGTGAGGCCTACCATAAGTGCTCGGAC-3′
S49A R: 5′-GTCCGAGCACTTATGGTAGGCCTCACAATTCTTGTAGTCT-3′
Y50A F: 5′-CTACAAGAATTGTGAGAGCGCCCATAAGTGCTCGGACTTGTG-3′
Y50A R: 5′-CACAAGTCCGAGCACTTATGGGCGCTCTCACAATTCTTGTAG-3′
H51A F: 5′-CTACAAGAATTGTGAGAGCTACGCTAAGTGCTCGGACTTGTG-3′
H51A R: 5′-CACAAGTCCGAGCACTTAGCGTAGCTCTCACAATTCTTGTAG-3′
K52A F: 5′-ACAAGAATTGTGAGAGCTACCATGCGTGCTCGGACTTGTGC-3′
K52A R: 5′-GCACAAGTCCGAGCACGCATGGTAGCTCTCACAATTCTTGT-3′
C53A F: 5′-GTGAGAGCTACCATAAGGCCTCGGACTTGTGCCAGT-3′

C53A R: 5′-ACTGGCACAAGTCCGAGGCCTTATGGTAGCTCTCAC-3′
S54A F: 5′-GTGAGAGCTACCATAAGTGCGCGGACTTGTG-3′
S54A R: 5′-CACAAGTCCGCGCACTTATGGTAGCTCTCAC-3′

Expression and purification of ligase mutants. Ligase mutants were expressed in *E. coli* BL21-DE3 Rosetta cells (Novagen). Cells were cultured in 1 l of LB medium with 36 μg/mL kanamycin to a $D_{600\ nm}$ of 0.8–1.0 at 37 °C. Cultures were induced with 1 mM IPTG and shaken for 6 h at 37 °C before being spun down and stored at −20 °C. Ligase mutant proteins were purified by Ni-NTA affinity chromatography in the same manner as described for the ^{15}N-labeled protein samples.

Analysis of metal content by ICP-MS. Ligase 10C was purified as described for the ^{15}N-labeled protein samples and then dialyzed three times against buffer (100 mM NaCl, 10 mM β-mercaptoethanol and 20 mM TrisHCl at pH 7.5; pretreated with Chelex 100 beads (Bio-Rad) for 2 h and filtered) at a ratio of 1/1,000. The metal content of 14 μM protein was measured by ICP MS (Thermo Scientific XSERIES 2 ICP-MS with ESI PC3 Peltier-cooled spray chamber at the Department of Earth Sciences, University of Minnesota).

Ligase activity assay for zinc dependence. Ligase 10C was purified as reported previously (14). Zinc was removed from ligase 10C by treatment with ion-exchange resin (Chelex 100, Bio-Rad). Ligase 10C (5 μM) was incubated with 20 μM HO-substrate, 10 μM ^{32}P-labeled PPP-substrate base-paired to splint, 20 mM HEPES (pH 7.5), 150 mM NaCl, 500 μM β-mercaptoethanol and the concentrations of $ZnCl_2$ as indicated in **Figure S7** for 6 h at room temperature. The ligation reactions were quenched with 20 mM EDTA in 8 M urea, heated to 95 °C for 4 min and separated by denaturing PAGE gel. The gel was analyzed using a GE Healthcare (Amersham Bioscience) Phosphorimager and ImageQuant software (Amersham Bioscience).

Ligase activity assay of 10C and alanine mutants. Ligase 10C (5 μM) (or alanine mutant) was incubated for 6 h at room temperature in the presence of 20 μM HO-substrate, 10 μM ^{32}P-labeled PPP-substrate base-paired to splint, 24 mM

HEPES (pH 7.5), 130 mM NaCl, 100 μM β-mercaptoethanol and 120 μM $ZnCl_2$. The ligation reactions were quenched with 20 mM EDTA and 8 M urea, heated to 95 °C for 4 min and separated by denaturing PAGE gel. The gel was analyzed using a GE Healthcare (Amersham Bioscience) Phosphorimager and ImageQuant software (Amersham Bioscience).

Resonance assignments. All NMR data were collected at the University of Minnesota NMR Center. NMR spectra were acquired at 298 K on a Bruker spectrometer equipped with a cryoprobe at 700 MHz and a Varian spectrometer at 600 MHz. The samples were in a buffer of 150 mM NaCl, 20 mM HEPES and 10 mM β-mercaptoethanol, pH 7.5. Moreover, all protein samples were saturated with $ZnCl_2$ by observing changes in HSQC spectra before other NMR experiments. Triple resonance spectra such as CBCA(CO)NH and HNCACB (26-28) were used to assign peaks on ^{15}N-HSQC. All resonances in these two three-dimensional spectra and ^{15}N-HSQC were picked and fed into the PISTACHIO program (National Magnetic Resonance Facility, Madison, Wisconsin, USA) (29) to obtain preliminary assignments. Final complete assignments were done by manual checks and searches. Carbonyl groups and others side chain carbons were assigned by HNCO and C(CO)NH-TOCSY (30); side chain protons were assigned by ^{15}N-NOESY-HSQC, ^{15}N-TOCSY-HSQC and HC(CO)NH-TOCSY experiments (30), with mixing times of 150 ms, 60 ms and 12 ms, respectively.

Distance restraints. All proton distance restraints were determined from the cross-peak intensities in the NOESY spectra by calibration with HN(*i*)Hα(*i*–1) distances located at the C-terminal region (31), whose helix propensity was shown by chemical shift index and $^{3}J_{HNH\alpha}$ coupling values (31,32). The cross-peaks from HN(*i*)Hα(*i*–1) distances in that region were categorized as medium NOEs, so the intensities of other cross-peaks smaller than this intensity range were defined as weak NOEs, and those larger than this range belonged to strong NOEs. The upper bounds of distance restraints of strong, medium and weak NOEs were given as 2.9 Å, 3.5 Å and 5 Å, respectively, and lower bounds were set to 1.8 Å in all cases. Starting from unambiguously assigned NOEs at the

beginning of calculation, miscalibrated NOEs were adjusted, and then ambiguously assigned NOEs were gradually added into the restraint table during iterative calculation.

Torsion angle restraints. Backbone ϕ angle restraints were acquired from the HNHA experiment, and the quantitative 3*J*HNHα coupling values were calculated from the intensity ratios of cross-peaks to diagonal peaks and corrected by 3.7% to account for relaxation (33). The correction is proportional to the rotational correlation time of the protein (3 ns), which was measured from one-dimensional TRACT experiments (34). The ϕ angle of residue *i* with J-coupling larger than 8.5 Hz was restrained from −160° to −80°, and that with J-coupling smaller than 6 Hz was restrained from −90° to −40°. Moreover, the ψ angle restraints were derived from the ^{15}N-NOESY data. If the intensity ratio of the HN(*i*)Hα(*i*) cross-peak to the HN(*i*)Hα(*i*–1) peak is smaller than one, the ψ(*i*-1) is restrained from 20° to 220°; otherwise, the ψ(*i*–1) is restrained from 80° to −140° (35,36).

RDC measurement. The stability of several alignment media for ligase 10C was tested. Using 5% neutral and negatively charged acrylamide gel was first attempted, but only weak residual dipolar couplings (absolute values <5 Hz) were obtained. Additionally, the samples precipitated in both DMPC-D7PC and DMPC-D6PC bicelle preparations. We also tested the liquid crystalline medium formed by cetylpyridinium chloride (CPCl) and 1-hexanol but had poor results in terms of sample stability. The sample was finally aligned in the other liquid crystalline medium made by the mixture of $C_{12}E_5$ (5% alkylpoly(ethylene glycol)) and 1-hexanol (r = 0.85) (37). The residual dipolar couplings of amide groups were obtained by measuring the splitting difference between a decoupling HSQC peak and a TROSY peak in isotropic solution and anisotropic medium.

Structure calculations. Simulating annealing protocols were performed in the XPLOR package (38). An extended structure was first generated, and the initial temperature was set at 3,500 K, then the temperature was cooled down to 0 K with 15,000 steps. The structure with the lowest energy was used for refinement with the initial temperature of 5,000 K and 30,000 steps. The resulting structure

was further refined with RDC data after optimization of the parameters Da and Rh. The angle restraints of the zinc coordination geometry were based on ideal geometries derived from X-ray data (39,40), which are in quantitative agreement with the EXAFS experiments. Distances derived from EXAFS have previously been used as restraints in NMR refinement (41,42). Here, we report a structural ensemble of 20 conformers. The PROCHECK statistics show that 76.4% of residues are in most favored regions and that 21.1% of residues are in allowed regions.

Zn K-edge EXAFS. Ligase 10C protein was fully saturated with excess Zn^{2+} and then dialyzed to remove excess Zn^{2+}. The final protein sample was 1.39 mM in 15 mM Tris, pH 7.5 and 112.5 mM NaCl. Glycerol (20% v/v) was added to the protein samples to form a glass required for the EXAFS experiments. The Zn K-edge X-ray absorption spectra of ligase 10C were measured at the Stanford Synchrotron Radiation Lightsource (SSRL) on the 16-pole, 2-T Wiggler beamline 9-3 under standard ring conditions of 3 GeV and ~200-mA ring current. A Si(220) double-crystal monochromator was used for energy selection. Another optical component used for the experiment was a cylindrical Rh-coated bent focusing mirror. Spectra were collected in the fully tuned configuration of the monochromator. The solution samples were immediately frozen after preparation and stored under liquid N_2 until measurement. During data collection, the samples were maintained at a constant temperature of ~6 K using an Oxford Instruments CF 1208 liquid helium cryostat. Data were measured to k = 16 $Å^{-1}$ by using a Canberra Ge 100-element monolith detector. Internal energy calibration was accomplished by simultaneous measurement of the absorption of a Zn foil placed between two ionization chambers situated after the sample. The first inflection point of the foil spectrum was fixed at 9,660.7 eV. Data presented here are a 15-scan average. The data were processed by fitting a second-order polynomial to the pre-edge region and subtracting this from the entire spectrum as background. A five-region spline of orders 2, 3, 3, 3 and 3 was used to model the smoothly decaying post-edge region. The data were normalized by

subtracting the cubic spline and assigning the edge jump to 1.0 at 9,680 eV using the Pyspline program (43). Theoretical EXAFS signals $\chi(k)$ were calculated by using FEFF (Macintosh version 8.4) (44-46). The initial model was based on the $Zn(Cys)_2(His)_2$ active site in a zinc finger protein (Protein Data Bank code 1MEY) (47). On the basis of the preliminary fits, the models were modified to accommodate a six-coordinate active site (four Zn-N/O and two Zn-S(Cys)). The theoretical models were fit to the data using EXAFSPAK (48). The structural parameters varied during the fitting process were the bond distance (R) and the bond variance σ^2, related to the Debye-Waller factor resulting from thermal motion and static disorder of the absorbing and scattering atoms. The nonstructural parameter E_0 (the energy at which $k = 0$) was also allowed to vary but was restricted to a common value for every component in a given fit. Coordination numbers were systematically varied in the course of the fit but were fixed within a given fit.

2.5 References

1. Chothia C. Proteins. one thousand families for the molecular biologist. Nature. 1992 Jun 18;357(6379):543-4.
2. Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. J Mol Biol. 1995 Apr 7;247(4):536-40.
3. Ohno S, editor. Evolution by gene duplication. New York: Springer-Verlag; 1971.
4. Chothia C, Gough J, Vogel C, Teichmann SA. Evolution of the protein repertoire. Science. 2003 Jun 13;300(5626):1701-3.
5. James LC, Tawfik DS. Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. Trends Biochem Sci. 2003 Jul;28(7):361-8.
6. Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science. 2009 Apr 10;324(5924):203-7.
7. Cordes MH, Walsh NP, McKnight CJ, Sauer RT. Evolution of a protein fold in vitro. Science. 1999 Apr 9;284(5412):325-8.
8. Kaplan J, DeGrado WF. De novo design of catalytic proteins. Proc Natl Acad Sci U S A. 2004 Aug 10;101(32):11566-70.
9. Tuinstra RL, Peterson FC, Kutlesa S, Elgin ES, Kron MA, Volkman BF. Interconversion between two unrelated protein folds in the lymphotactin native state. Proc Natl Acad Sci U S A. 2008 Apr 1;105(13):5057-62.
10. Bryan PN, Orban J. Proteins that switch folds. Curr Opin Struct Biol. 2010 Aug;20(4):482-8.
11. Smith BA, Hecht MH. Novel proteins: From fold to function. Curr Opin Chem Biol. 2011 Jun;15(3):421-6.
12. Keefe AD, Szostak JW. Functional proteins from a random-sequence library. Nature. 2001 Apr 5;410(6829):715-8.

13. Mansy SS, Zhang J, Kummerle R, Nilsson M, Chou JJ, Szostak JW, et al. Structure and evolutionary analysis of a non-biological ATP-binding protein. J Mol Biol. 2007 Aug 10;371(2):501-13.
14. Seelig B, Szostak JW. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature. 2007 Aug 16;448(7155):828-31.
15. Seelig B. mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. Nat Protoc. 2011 Apr;6(4):540-52.
16. Holmbeck SM, Foster MP, Casimiro DR, Sem DS, Dyson HJ, Wright PE. High-resolution solution structure of the retinoid X receptor DNA-binding domain. J Mol Biol. 1998 Aug 14;281(2):271-84.
17. Cho GS, Szostak JW. Directed evolution of ATP binding proteins from a zinc finger domain by using mRNA display. Chem Biol. 2006 Feb;13(2):139-47.
18. Zhao Q, Chasse SA, Devarakonda S, Sierk ML, Ahvazi B, Rastinejad F. Structural basis of RXR-DNA interactions. J Mol Biol. 2000 Feb 18;296(2):509-20.
19. Maret W, Li Y. Coordination dynamics of zinc in proteins. Chem Rev. 2009 Oct;109(10):4682-707.
20. van Tilborg PJ, Czisch M, Mulder FA, Folkers GE, Bonvin AM, Nair M, et al. Changes in dynamical behavior of the retinoid X receptor DNA-binding domain upon binding to a 14 base-pair DNA half site. Biochemistry. 2000 Aug 1;39(30):8747-57.
21. Yang W, Lee JY, Nowotny M. Making and breaking nucleic acids: Two-Mg2+-ion catalysis and substrate specificity. Mol Cell. 2006 Apr 7;22(1):5-13.
22. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, et al. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science. 2011 Apr 8;332(6026):234-8.
23. Baldwin AJ, Kay LE. NMR spectroscopy brings invisible protein states into focus. Nat Chem Biol. 2009 Nov;5(11):808-14.
24. Henzler-Wildman K, Kern D. Dynamic personalities of proteins. Nature. 2007 Dec 13;450(7172):964-72.

25. Golynskiy MV, Seelig B. De novo enzymes: From computational design to mRNA display. Trends Biotechnol. 2010 Jul;28(7):340-5.
26. Grzesiek S, Bax A. Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. Journal of Magnetic Resonance (1969). 1992 2/1;96(2):432-40.
27. Muhandiram DR, Kay LE. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. Journal of Magnetic Resonance, Series B. 1994 3;103(3):203-16.
28. Wittekind M, Mueller L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. Journal of Magnetic Resonance, Series B. 1993 4;101(2):201-5.
29. Eghbalnia HR, Bahrami A, Tonelli M, Hallenga K, Markley JL. High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection. J Am Chem Soc. 2005 Sep 14;127(36):12528-36.
30. Grzesiek S, Anglister J, Bax A. Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. Journal of Magnetic Resonance, Series B. 1993 2;101(1):114-9.
31. Wuthrich K. NMR of proteins and nucleic acids. . 1986.
32. Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol. 1991 Nov 20;222(2):311-33.
33. Vuister GW, Bax A. Quantitative J correlation: A new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins. J Am Chem Soc. 1993 08/01;115(17):7772-7.
34. Lee D, Hilty C, Wider G, Wuthrich K. Effective rotational correlation times of proteins from NMR relaxation interference. J Magn Reson. 2006 Jan;178(1):72-6.
35. Gagne SM, Tsuda S, Li MX, Chandra M, Smillie LB, Sykes BD. Quantification of the calcium-induced secondary structural changes in the regulatory domain of troponin-C. Protein Sci. 1994 Nov;3(11):1961-74.

36. Wang Y, Zhao S, Somerville RL, Jardetzky O. Solution structure of the DNA-binding domain of the TyrR protein of haemophilus influenzae. Protein Sci. 2001 Mar;10(3):592-8.
37. Rückert M, Otting G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. J Am Chem Soc. 2000 08/01;122(32):7793-7.
38. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The xplor-NIH NMR molecular structure determination package. J Magn Reson. 2003 Jan;160(1):65-73.
39. Alberts IL, Nadassy K, Wodak SJ. Analysis of zinc binding sites in protein crystal structures. Protein Sci. 1998 Aug;7(8):1700-16.
40. Viles JH, Patel SU, Mitchell JB, Moody CM, Justice DE, Uppenbrink J, et al. Design, synthesis and structure of a zinc finger with an artificial beta-turn. J Mol Biol. 1998 Jun 19;279(4):973-86.
41. Ohlenschlager O, Seiboth T, Zengerling H, Briese L, Marchanka A, Ramachandran R, et al. Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. Oncogene. 2006 Sep 28;25(44):5953-9.
42. Banci L, Bertini I, Del Conte R, Mangani S, Meyer-Klaucke W. X-ray absorption and NMR spectroscopic studies of CopZ, a copper chaperone in bacillus subtilis: The coordination properties of the copper ion. Biochemistry. 2003 Mar 4;42(8):2467-74.
43. Tenderholt A, editor. Pyspline. Stanford: Stanford University; 2007.
44. Mustre de Leon J, Rehr JJ, Zabinsky SI, Albers RC. Ab initio curved-wave x-ray-absorption fine structure. Phys Rev B Condens Matter. 1991 Sep 1;44(9):4146-56.
45. Rehr JJ, Mustre dL, Zabinsky SI, Albers RC. Theoretical x-ray absorption fine structure standards. J Am Chem Soc. 1991 07/01; 2013/11;113(14):5135-40.
46. Rehr JJ, Albers RC. Theoretical approaches to x-ray absorption fine structure. Rev Mod Phys. 2000 07/01;72(3):621-54.

47. Kim CA, Berg JM. A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA. Nat Struct Biol. 1996 Nov;3(11):940-5.
48. George GN, editor. EXAFSPAK and EDG-FIT. Menlo Park: Stanford Synchrotron Radiation Lightsource; 2000.

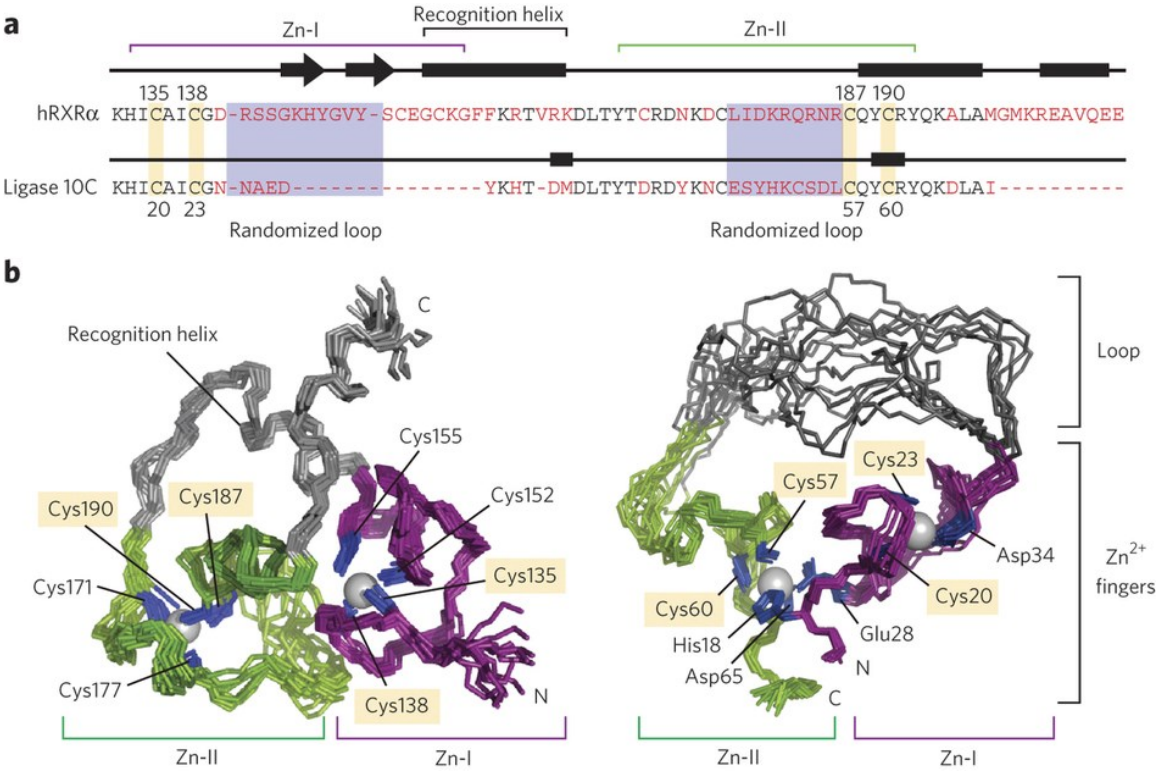


Figure 1. Changes in primary sequence and 3D structure upon directed evolution of the hRXRα scaffold to the ligase enzyme 10C. (**a**) Comparison of the primary sequences of hRXRα (residues 132–208) and the artificially evolved ligase (residues 17–68). The two zinc finger regions are highlighted with purple and green brackets. Red letters denote residues not conserved between the two sequences. (**b**) 3D structure of hRXRα (left) and NMR ensemble of ligase 10C (right; flexible termini omitted for clarity). Although both proteins contain two zinc fingers, the overall structures are substantially different. Only two zinc-coordinating cysteines of each zinc finger in hRXRα are still coordinating zinc in ligase 10C (highlighted in yellow; also shown in **a**), whereas all other ligands differ in the two structures. Zinc-coordinating residues are labeled and shown in blue. In contrast to that in hRXRα, zinc finger Zn-II in ligase 10C comprises residues of both N- and C-terminal sequences, imposing a cyclic structure to the enzyme. Notably, the new ligase lost both helical domains of hRXRα (gray), replacing the recognition helix with a long unstructured loop (gray).

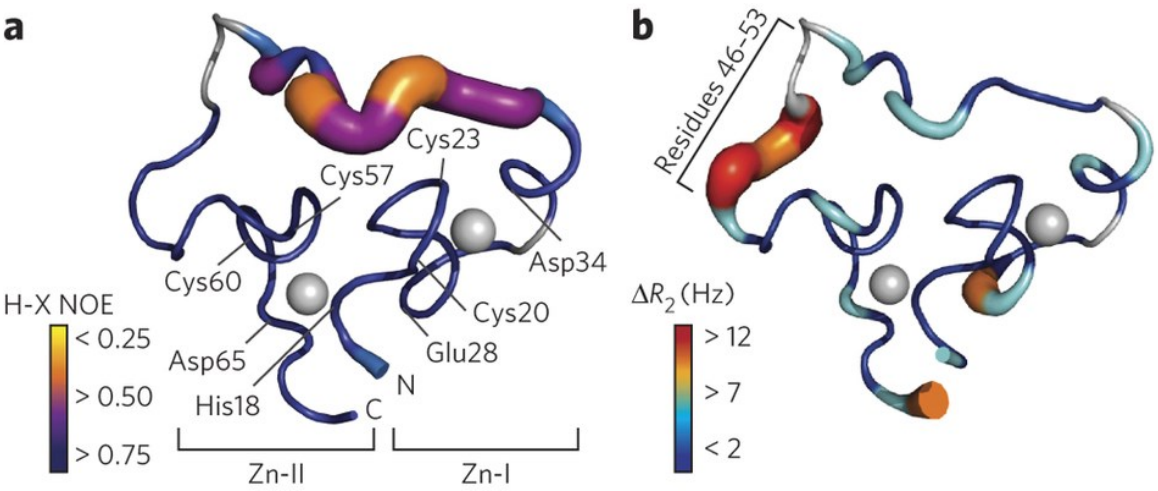


Figure 2. Conformational dynamics of ligase enzyme 10C. (**a**) Mapping of heteronuclear (H-X) NOEs (proxy for fast dynamics on a ps-ns timescale) on ligase 10C. Zinc-coordinating residues are labeled. (**b**) Mapping of exchange rates (R_{ex} and ΔR_2 (difference between R_2 = 5 ms and R_2 = 0.2 ms)) obtained from relaxation dispersion measurements as proxy for slow dynamics (μs-ms timescale). Color gradient and thickness of backbone indicate that the fast dynamics are located mostly in the unstructured loop, whereas the slow dynamics are located mostly in the region N-terminal to the Zn-II site (residues 46–53) and are potentially correlated to catalytic activity.

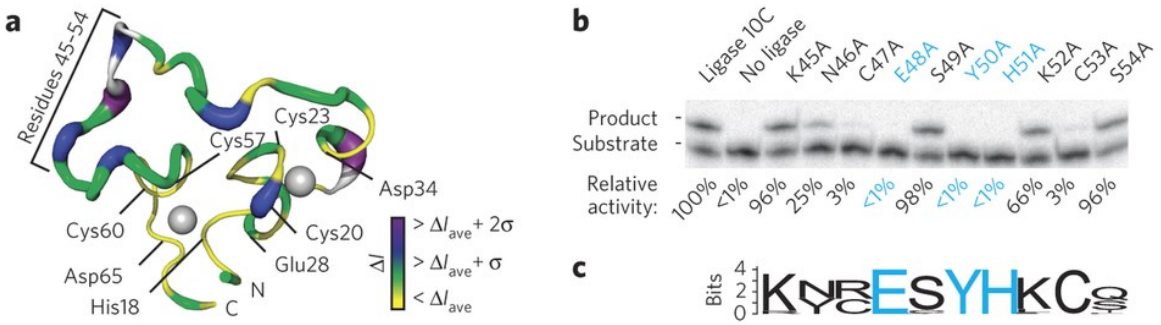


Figure 3. Substrate-binding surface of ligase 10C probed by NMR and alanine scanning. (**a**) Mapping of NMR chemical shift perturbations (intensity changes, Δ*I*) for ligase 10C shows regions affected by formation of the complex with RNA substrate. The most perturbed regions are indicated by thicker lines and darker colors. Zinc-coordinating residues are labeled. (**b**) Activity assay of ligase 10C and alanine mutants by gel shift. Ligation activity is normalized to the activity of ligase 10C and represents the mean value from two independent experiments. Residues with activity below detection limit are shown in blue. (**c**) Sequence conservation analysis by WebLogo (V2.8.2) of ligase residues 45–54 from 49 enzyme sequences generated by directed evolution. Residues Glu48, Tyr50 and His51 (blue), which completely lost activity during alanine scanning (described above), were conserved among all sequences. The two residues with 97% reduced activity of their alanine mutant were either conserved (Cys53) or had one alternative amino acid (Cys47). None of the other residues in this region were conserved.

	Protein
NMR distance and dihedral constraints	
Distance constraints	
Total NOE	354
Intra-residue	106
Inter-residue	248
Sequential (|*i* – *j*| = 1)	162
Medium-range (|*i* – *j*| < 4)	34
Long-range (|*i* – *j*| > 5)	52
Intermolecular	0
Hydrogen bonds	0
Total dihedral angle restraints	34
ϕ	15
ψ	19
Total RDCs	26
Q (%)	14.6
Structure statistics	
Violations (mean and s.d.)	
Distance constraints (Å)	0.1 (0.01)
Dihedral angle constraints (º)	1.3 (0.4)
Max. distance constraint violation (Å)	0.8 (0.4)
Max. dihedral angle violation (º)	5.7 (1.6)
Deviations from idealized geometry	
Bond lengths (Å)	0.008
Bond angles (º)	1.0
Impropers (º)	0.5
Average pairwise r.m.s. deviation** (Å)	
Heavy	1.4
Backbone	0.8

Table S1. Summary of NMR structural statistics of 20 conformers. The RMSD of the structural ensemble is calculated within well-structured regions (residues 17-35 and 49-69).

	Average value	Standard deviation
K_d1 (μM)	3.0	0.6
ΔH1 (kcal/mole)	122.9	14.8
ΔS1 (cal/mole/°)	437.7	49.7
K_d2 (μM)	92.8	8.9
ΔH2 (kcal/mole)	-123.7	13.3
ΔS2 (cal/mole/°)	-396.3	45.0

Table S2. Zinc binding determined by Isothermal Titration Calorimetry. After completely removing the Zn^{2+} from the protein with Chelex 100 chelating ion exchange resin, 5 μM ligase enzyme was slowly titrated with 400 μM $ZnCl_2$ solution, and the heat release was monitored by ITC. All samples contained at 150 mM NaCl, 20 mM HEPES, 10 mM β-mercaptoethanol, and pH 7.5, and the data were fitted with a sequential two-site binding model. The values represent the average of three measurements.

Coordination/Path	R(Å)[a]	σ^2(Å^2)[b]	E_0 (eV)	*F* [c]
4 Zn-N	2.00	731		
2 Zn-S	2.30	453		
6 Zn-C	3.00	1077		
12 Zn-C-N	3.09	1077[d]	-12.3	0.18
6 Zn-C	4.16	169		
6 Zn-C-N	4.19	169[d]		
6 Zn-C-N	4.30	169 [d]		

[a]The estimated standard deviations for the distances are in the order of ± 0.02 Å. [b]The σ^2 values are multiplied by 10^5. [c]Error is given by $\Sigma[(\chi_{obsd} - \chi_{calcd})^2 k^6]/\Sigma[(\chi_{obsd})^2 k^6]$. [d]The σ^2 value for the the Zn-C (single scattering) and Zn-C-N (multiple scattering) paths were linked to be the same value.

Table S3. EXAFS least squares fitting results for ligase 10C.

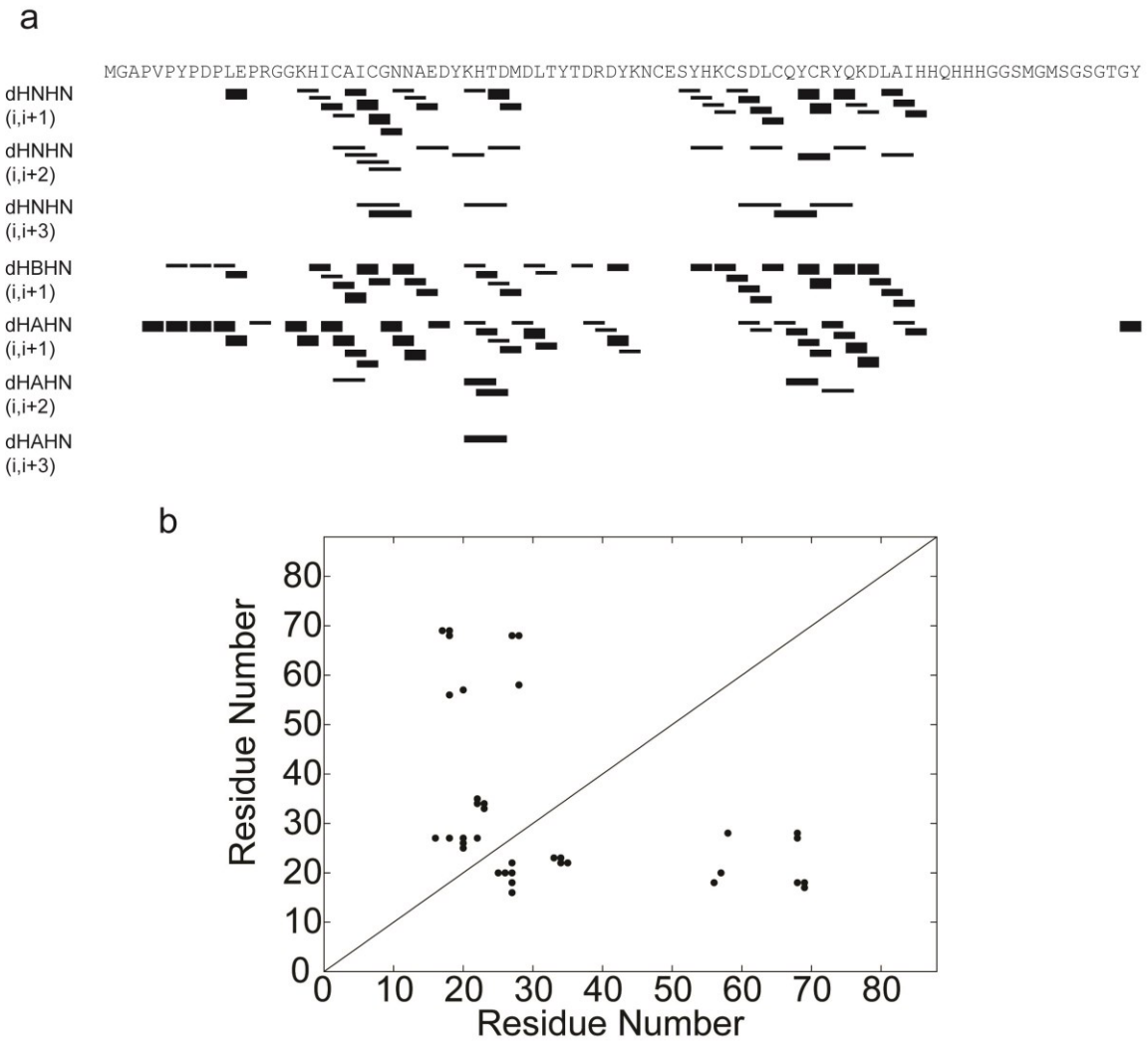


Figure S1. Representative distance restraints observed from NOESY spectra. (**a**) Bars represent the existence of NOE signals between residues with the thickness corresponding to NOE intensity. (**b**) Cross peaks between residues in the primary sequence illustrate long-range NOEs between residues.

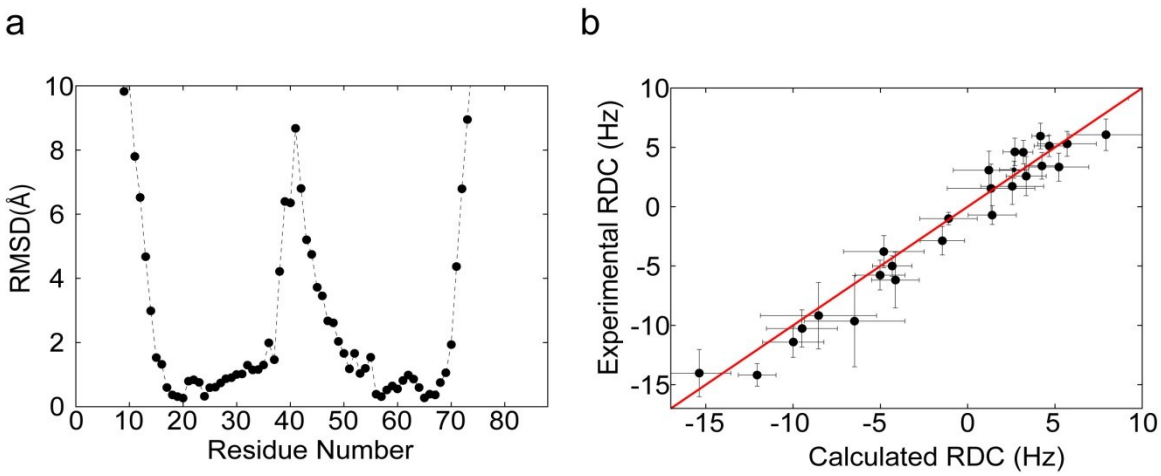


Figure S2. Quality of the structural ensemble of 20 conformers is shown by RMSD and RDC data. (**a**) Average backbone RMSD of the ensemble of the proposed structural model. (**b**) Correlation between experimental RDC values and average back-calculated RDC values from the ensemble.

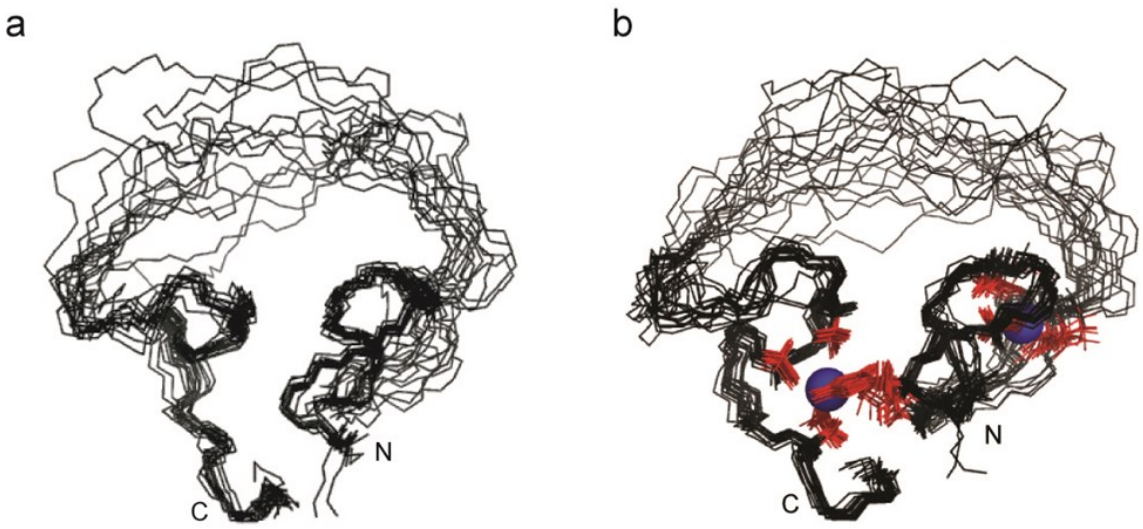


Figure S3. The structural ensembles are calculated before and after incorporating zinc ions into restraints. (**a**) Ensemble of 20 lowest energy conformers (residues 16-69) from 100 structures, including RDC refinement but without Zn^{2+} coordination information. (**b**) Ensemble of 20 lowest energy conformers (residues 16-69) obtained with Zn^{2+} coordination information. The two Zn^{2+} are shown as blue spheres, and the side chains involved in the coordination are marked as red (H18, E28(OD2), C57, C60, D65(OD1), and C20, C23, D34(OD1)). Both zinc binding sites contain a single water molecule each in the coordination (not shown). The backbone RMSD between the ensemble without Zn^{2+} and with Zn^{2+} is about 0.52 Å at the well-structured regions (residues 17-35 and 49-69).

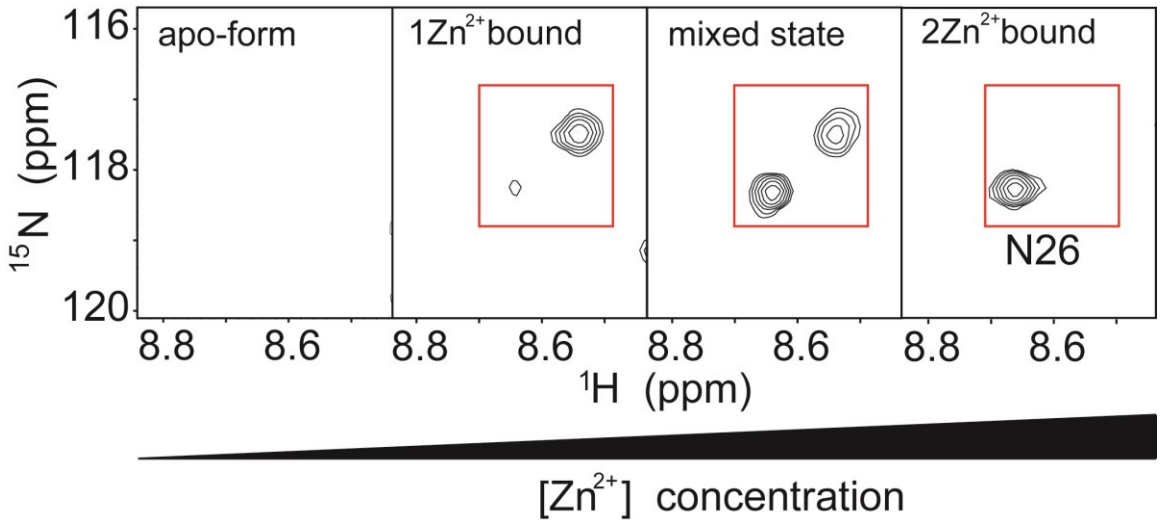


Figure S4. Zn^{2+} titration into ^{15}N-labeled ligase 10C monitored by NMR. A selected region of HSQC spectra recorded during Zn^{2+} titration is shown. Residues of partially Zn^{2+}-saturated sample displayed slow exchange on the NMR time scale between forms bound to one or two Zn^{2+}.

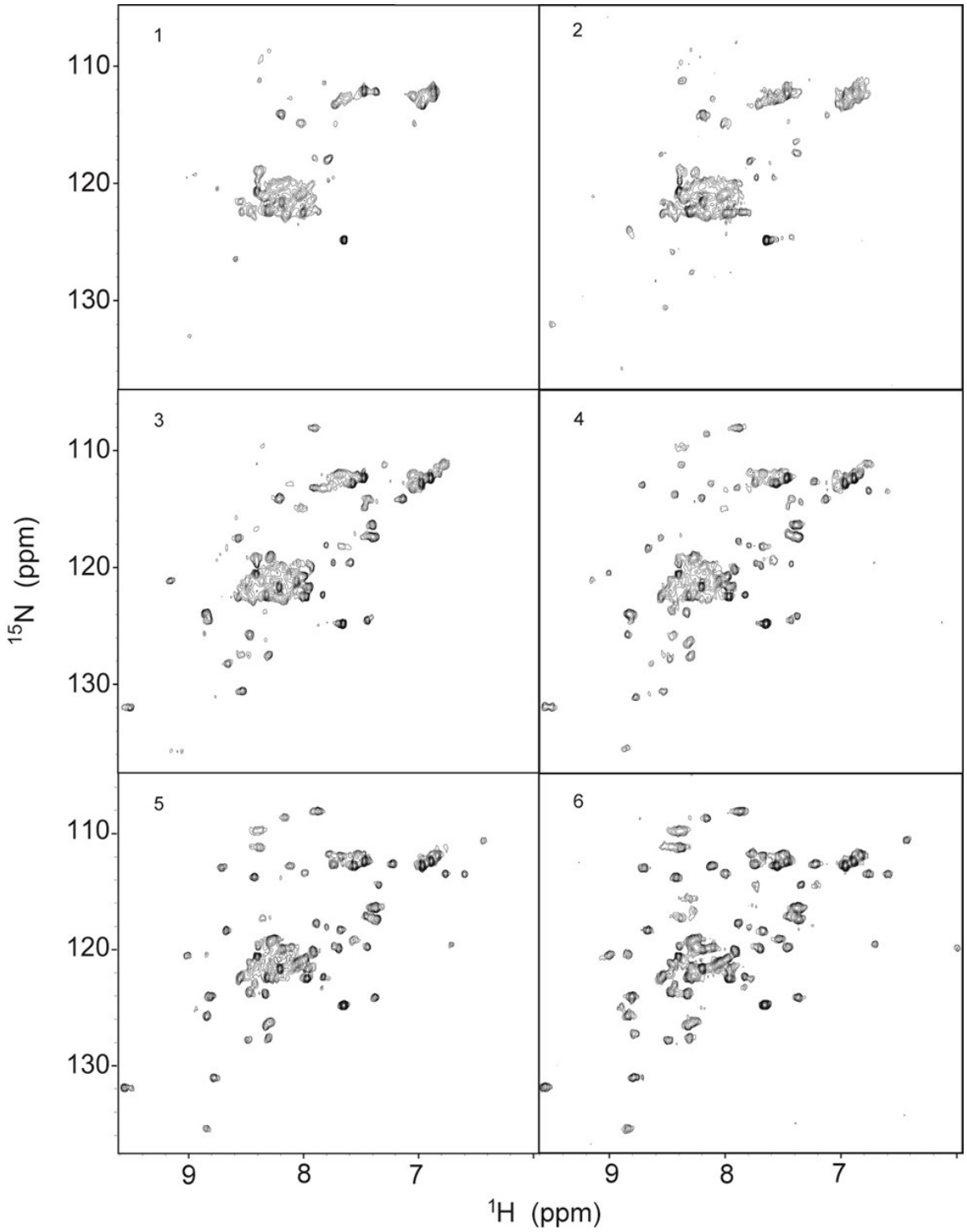

1
2
3
4
5
6
110
120
130
9
8
7
^{15}N (ppm)
^{1}H (ppm)

Figure S5. HSQC spectra recorded during zinc titration. Molar ratios of ligase 10C to zinc were: 1) 10C:Zn=1:0, 2) 10C:Zn=1:1, 3) 10C:Zn=1:2, 4) 10C:Zn=1:3, 5) 10C:Zn=1:4, 6) 10C:Zn=1:6.

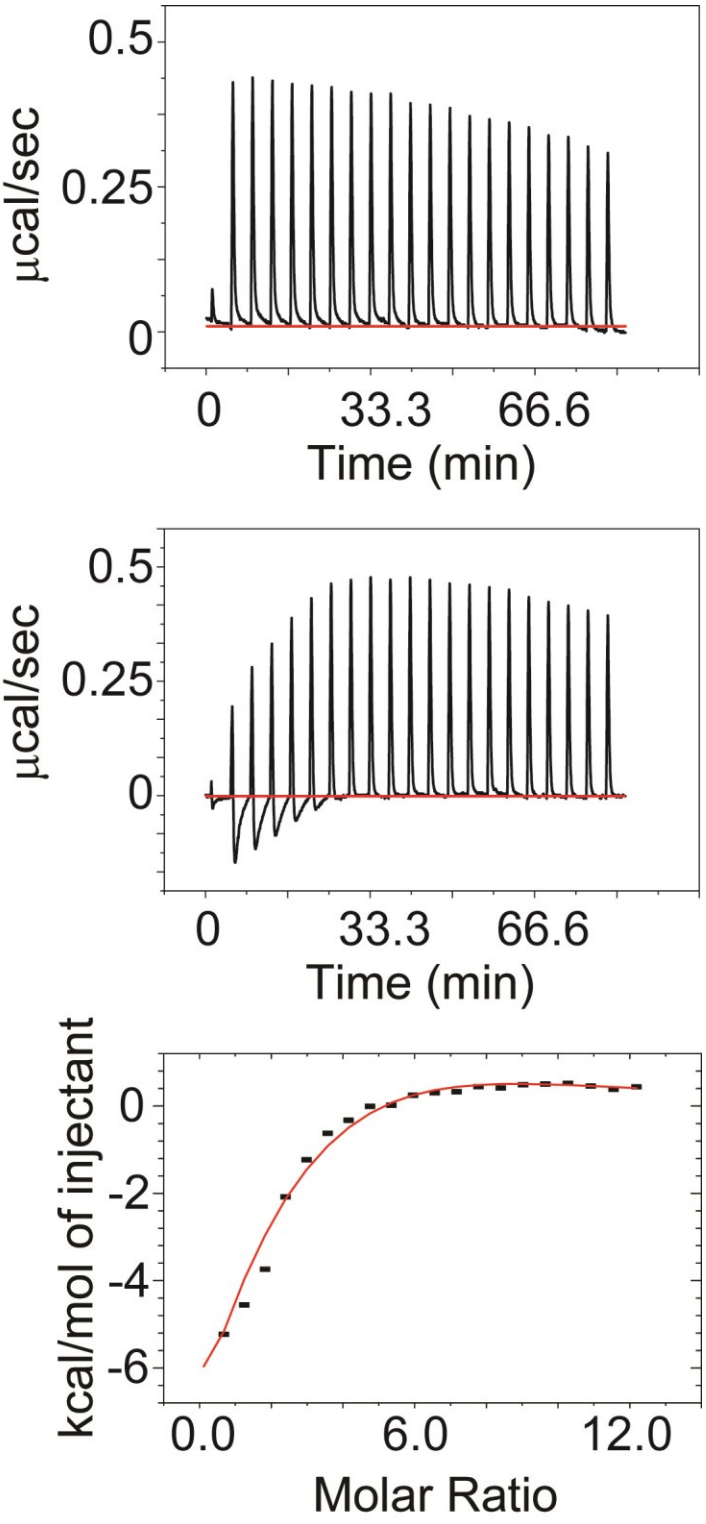

0.5
0.25
0
µcal/sec
0
33.3
66.6
Time (min)
0.5
0.25
0
µcal/sec
0
33.3
66.6
Time (min)
0
-2
-4
-6
kcal/mol of injectant
0.0
6.0
12.0
Molar Ratio

Figure S6. Zn^{2+} titration into the ligase enzyme monitored by ITC. Samples contained 5 μM ligase 10C, 150 mM NaCl, 20 mM HEPES, 10 mM β-mercaptoethanol, pH 7.5 and were measured by Isothermal Titration Calorimetry using a MicroCal VP-ITC instrument (GE Healthcare). (**a**) The graph represents the raw data for the blank titration (buffer without ligase). (**b**) The graph represents raw data for the Zn^{2+} titration of ligase 10C. (**c**) The figure shows the heat release of the Zn^{2+} titration of ligase 10C after subtracting the blank titration. The data is fit to a model of two binding sites. The data can be fitted to models with two or more Zn^{2+} binding sites, however, the fit does not improve significantly with n>2. The Zn^{2+} titration was carried out in triplicate and the errors are summarized in the **Table S2**.

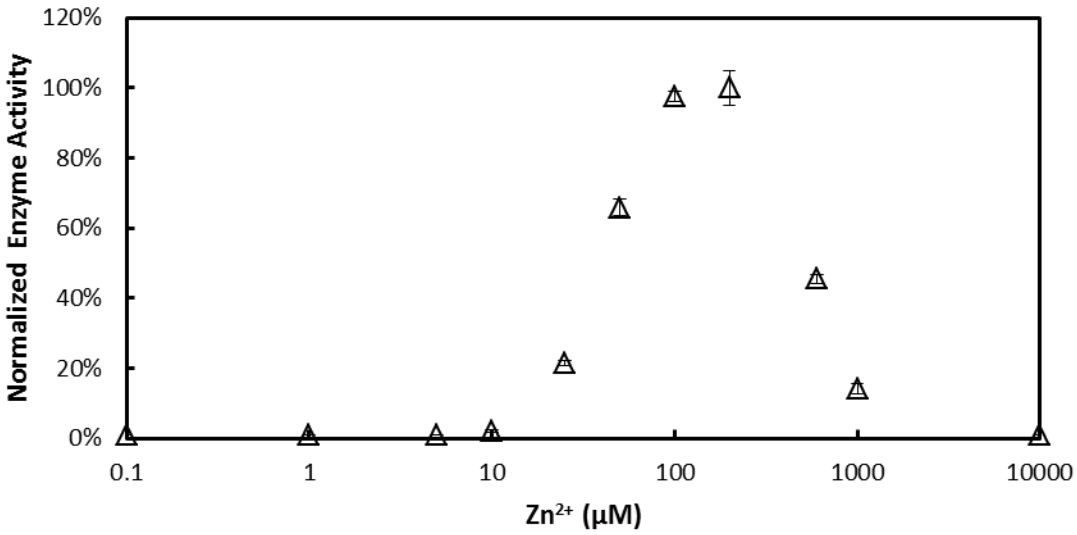
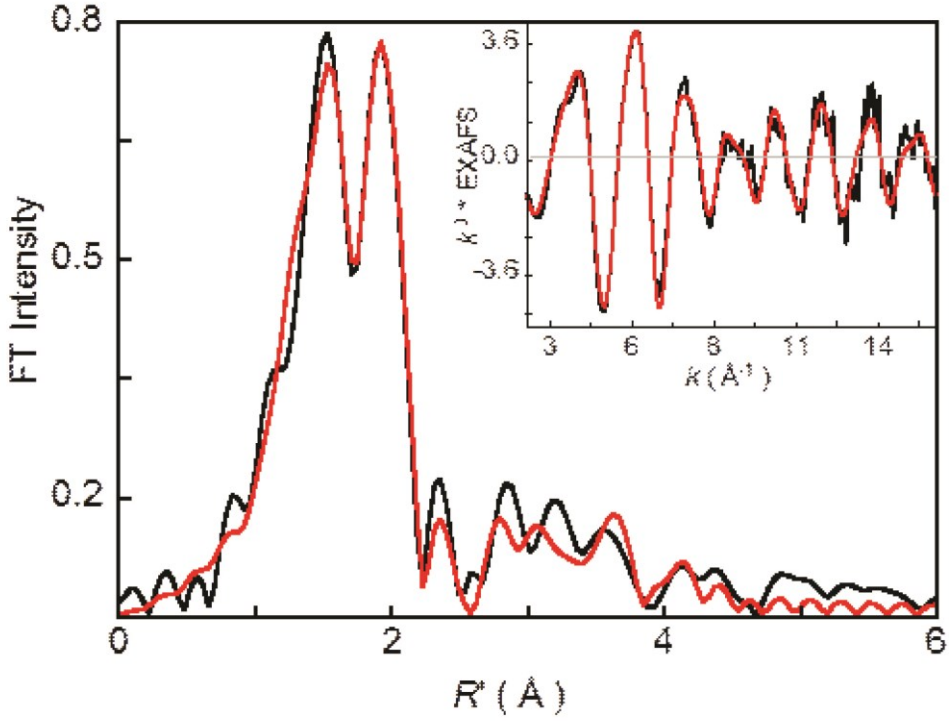


Figure S7. Zinc dependence of ligase activity. The maximum activity was observed at 145 µM zinc. Towards lower zinc concentrations the activity sharply drops, matching the expected behavior predicted from the dissociation constants measured by Thermocalorimetry. Towards higher zinc concentrations, the activity also decreases but more slowly. One possible explanation is that Zn^{2+} at high concentrations might also bind to additional sites with lower affinity thereby reducing the activity. Error bars represent one standard deviation. Ligation activity for samples at 0.1, 1, 5 and 10,000 µM $ZnCl_2$ was below the detection limit of 1%.

a


0.8
0.5
0.2
FT Intensity
0
2
4
6
R' (Å)
3.6
0.0
-3.6
k³ * EXAFS
3
6
8
11
14
k (Å⁻¹)

b

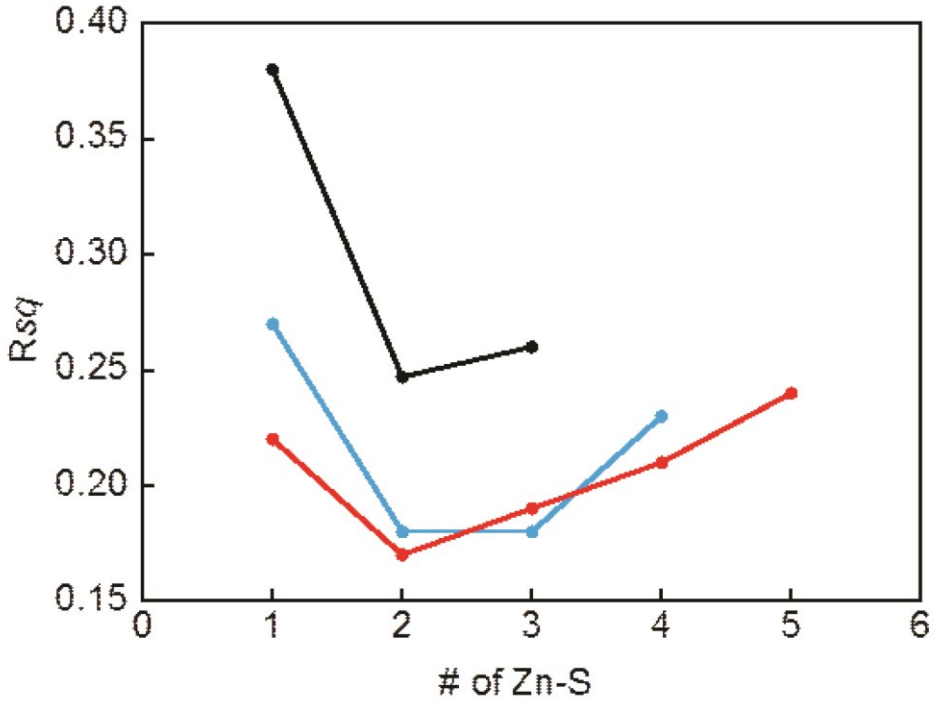

0.40
0.35
0.30
0.25
0.20
0.15
Rsq
0
1
2
3
4
5
6
# of Zn-S

Figure S8. Analysis of zinc coordination by EXAFS spectroscopy (Extended X-ray Absorption Fine Structure). (**a**) The k^3 weighted Zn K-edge EXAFS (inset) and their corresponding non-phase shift corrected Fourier transforms for ligase 10C are presented. The experimental data are shown as black lines and the fit as red lines. The best-fit parameters are given in **Table S3**. While a coordination with four ligands is most commonly observed for zinc ions, a coordination geometry including six ligands has been observed in natural proteins numerous times. The first shell coordination number was varied from four-coordinate to sixcoordinate. In each case the number of Zn-S and Zn-N/O components was systematically varied to obtain the best *F* value. These fits show that the data are most consistent with a sixcoordinate site with 2 Zn-S and 4 Zn-N/O components. A 1:1 occupation of the two sites modeled from NMR analysis would have resulted in best-fit with 3 Zn-N/O and 2 Zn-S coordination. However, the process of dialysis (removal of excess Zn is necessary for EXAFS experiments) may lead to stripping of some Zn from the weakly bound N-terminal site. This leads to an increase in the number of six-coordinate sites over four-coordinate sites in the protein and results in a best-fit first shell with more than 3 Zn-N/O paths.
The EXAFS data are dominated by first shell Zn-N/O and Zn-S, while second and third shells are significantly weaker. The second and third shells were fit with single (Zn-C) and multiple-scattering (Zn-C-N) theoretical paths generated using a representative Zn-N(His) model. These weak features are due to a combination of single and multiple scattering from the amino acid ligands. The multiple scattering features are different from characteristic Zn-N(His)n^3 or ZnS(Cys)n^4 systems due to interference between second shell components of Cys and His ligands. Note that standard deviations in bond distances obtained from EXAFSPAK assume the use of raw, low-noise data. Although the data quality presented here are quite high, it is important to note that in the presence of several single and multiple scattering paths, the choice of a specific path to represent an average of multiple paths will also affect the standard deviations. Typically second shell paths have errors of the order 0.05 to 0.1 Å. Furthermore

these standard deviations do not reflect the fact EXAFS analysis typically underestimates bond distances (relative to crystallography). The protein samples used for EXAFS analysis were extensively dialyzed and had no extraneous source of sulfur, precluding non-protein based Zn-S ligation.

A visual inspection of the FEFF fit presented here shows that the first peak (corresponding to Zn-N/O paths) in the Fourier Transform is a slightly poorer fit relative to the second peak (corresponding to Zn-S paths). In an attempt to improve the fits and to differentiate between 3 Zn-N/O and 4 Zn-N/O fits, split first shell fits were performed. Significant statistical improvement was not observed.

(**b**) The R_{sq} values ($\Sigma[(\chi_{obsd} - \chi_{calcd})^2 k^6]/\Sigma[(\chi_{obsd})^2 k^6]$) for four- to six- coordinate first shell fits are presented as a function of increasing number of Zn-S ligands with concomitant decrease in the number of Zn-N/O ligands. (---) four-coordinate, (---) five-coordinate. (---) six-coordinate. The R_{sq} of the four-coordinate fit is significantly worse than that of the five- or six- coordinate fits.

Note that although the best R_{sq} value is obtained with 4 Zn –N/O and 2 Zn-S ligands, the five coordinate fits with either 3 Zn-N/O and 2 Zn-S or 2 Zn-N/O and 3 Zn-S ligands also have reasonably low R_{sq} values. For the 2 Zn-N/O and 3 Zn-S fit to be correct, the two Zn sites need to have 2 Zn-S and 4 Zn-S ligands, respectively. Such a structure is ruled out by NMR data, which show that the sites do not have more than two S-based ligands. Since the first shell coordination number error can be up to 20%, it is difficult to differentiate between the 4 Zn-O/N and 3 Zn/O fits with a high level of statistical confidence. Note that both the 4 Zn-N/O and 3 Zn-N/O fits indicate that the high Zn-affinity site is six-coordinate. Six-coordinate Zn sites account for at least 11% of all Zn sites in biology based on NMR and crystallography studies. EXAFS studies with cysteine ligands are typically limited to four- and five-coordinate sites. Studies have been performed on six-coordinate sites, but typically with all light atom ligands. In general, a comparison of total EXAFS intensity can give an insight into the coordination number but the presence of two different first shell ligands (N/O and

S) modulates the EXAFS data strongly, making an accurate comparison of EXAFS data between two systems with different coordination numbers difficult. In such a situation, the error in first shell coordination number determination can be greater than 20%. Since the EXAFS data are best fit with between 3 and 4 light atom ligands, the higher error indicates that the second site can be between tetra and septa-coordinate. Since, a seven coordinate site has no biological precedence, the second site is between tetra- and hexa-coordinate.

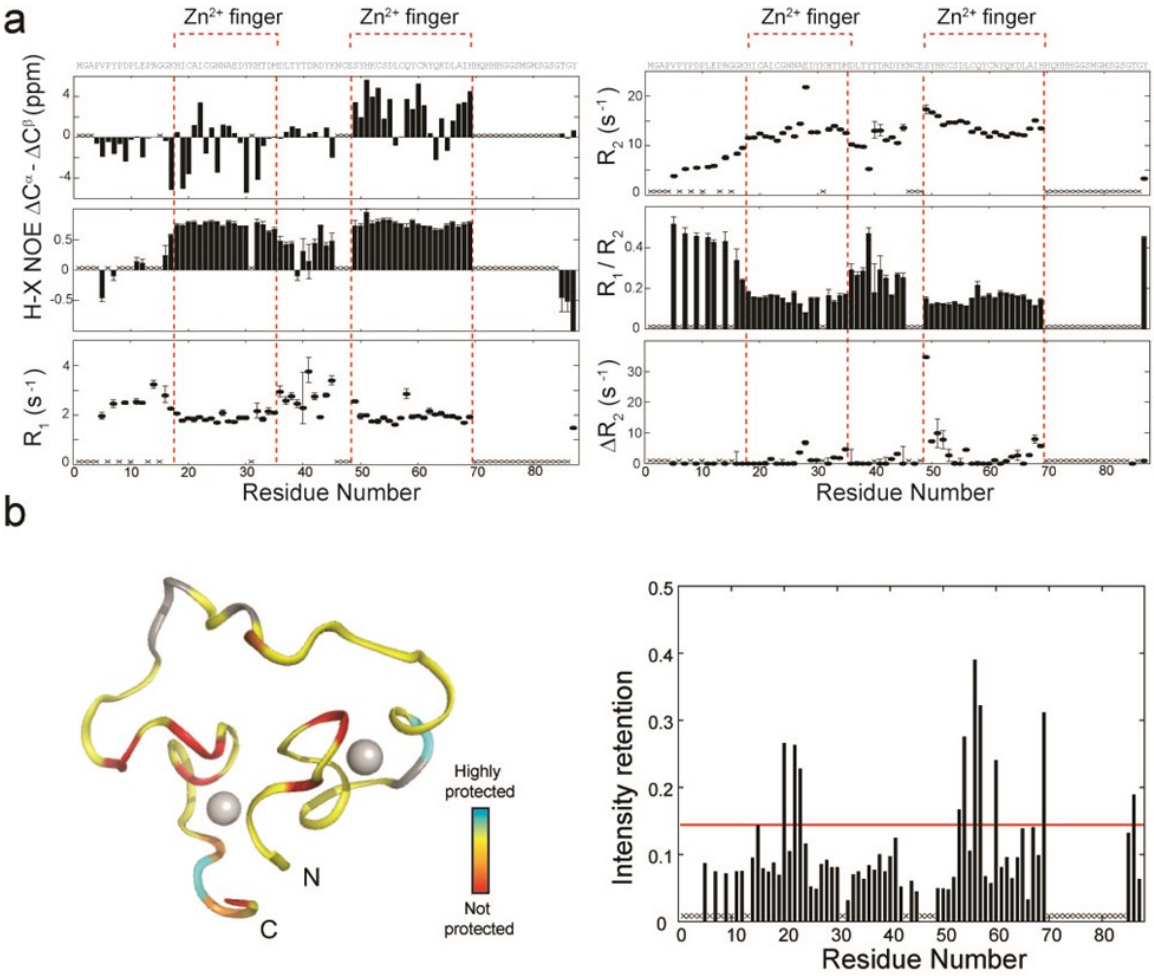


Figure S9. Fast, intermediate, and slow dynamics are probed by NMR experiments. (**a**) Chemical shift indexes (ΔCα-ΔCβ), steady-state NOE, longitudinal relaxation rates (R_1), transverse relaxation rates (R_2), and R_1/R_2 ratios as determined by NMR spectroscopy (unassigned residues are marked with an "X"). The errors are estimated by the signal-to-noise (H-X NOE), standard deviations of the fitting (R_1, R_2, and R_1/R_2), or duplicate experiments (ΔR_2). The two Zn^{2+} fingers are highlighted with dashed red lines. (**b**) The decrease of peak intensities due to H/D exchange was mapped onto one NMR conformer (residues 16-69 are displayed). Before the experiment, a reference HSQC spectrum of the sample in water was obtained. The sample was lyophilized and dissolved in the same volume of 80% D_2O. After 6 minutes, the

HSQC spectrum was acquired and compared with the initial spectrum to monitor solvent exposed amide groups. The solid red line in the diagram represents the average intensity retention plus 2σ.

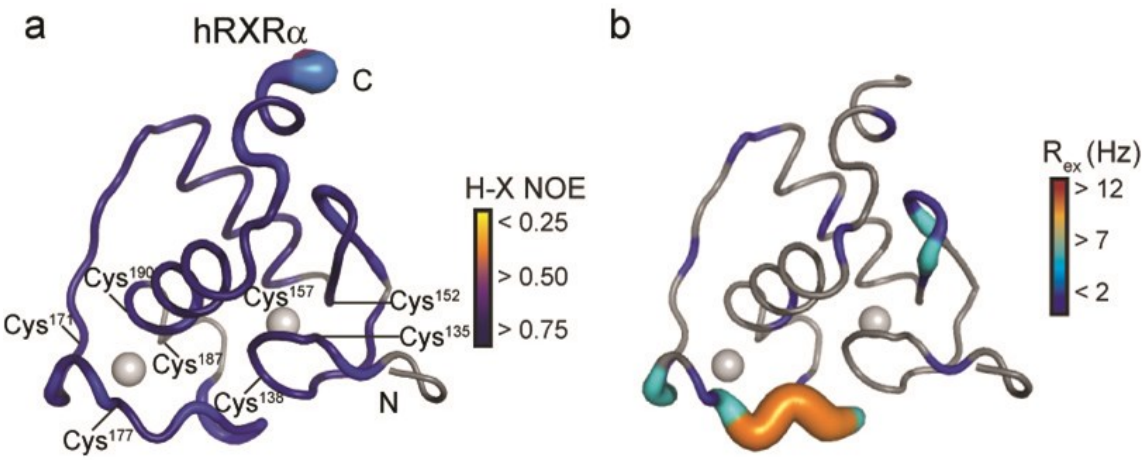


Figure S10. Comparison of the motions of the DNA binding domain, hRXRα. (**a**) Mapping of the heteronuclear NOEs (proxy for fast dynamics on a picosecond-nanosecond time scale) on representative structures of hRXRα. (**b**) Mapping of the exchange rates (R_{ex}) obtained from relaxation dispersion measurements as a proxy for slow dynamics (microsecond-millisecond time scale).

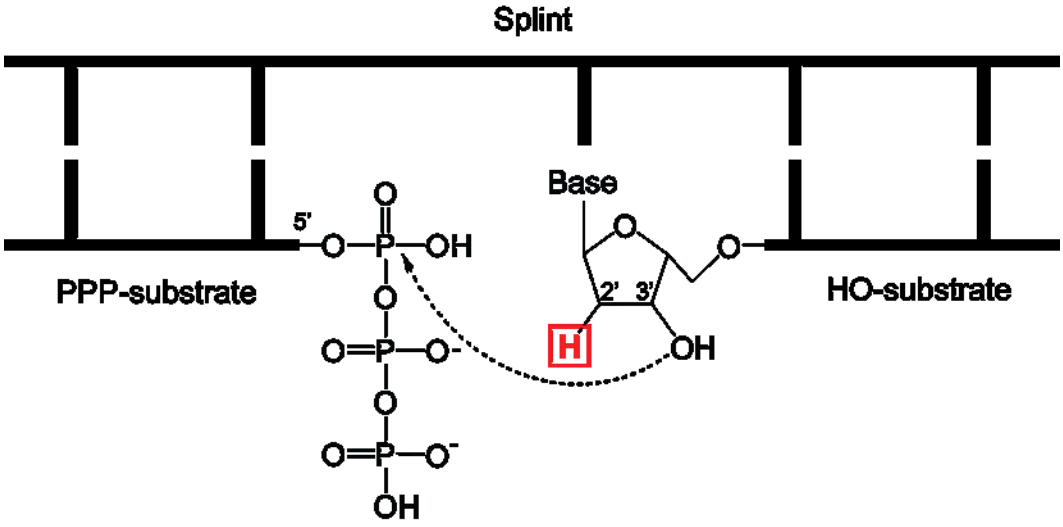


Figure S11. Chemical structure of inactive ligation substrate. Substitution of the 2'-hydroxyl group of the terminal nucleotide in the HO-substrate with a 2'-deoxy modification (red box) results in inactivation of the ligation reaction. Ligation of active substrates occurs between a 5'-triphosphorylated RNA (PPP-substrate) and the 3'-hydroxyl group of the second RNA (HO-substrate) while both RNAs are base-paired to a complementary oligonucleotide (splint). The dashed arrow symbolizes the proposed bond formation.

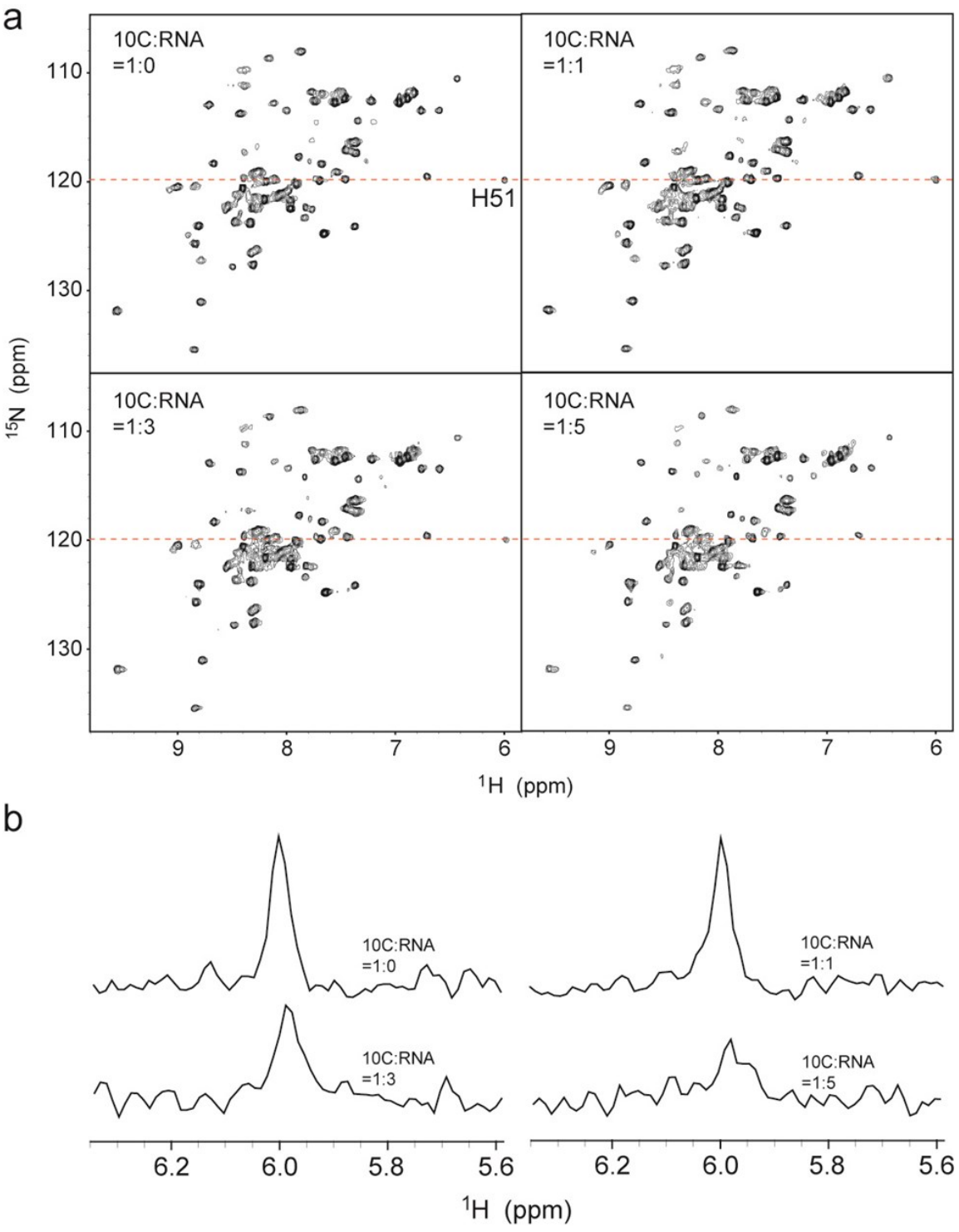


Figure S12. Titration of RNA substrate to ligase 10C monitored by NMR spectroscopy. (**a**) The ligase enzyme (300μM) was titrated with the inactive RNA

ligand in 150 mM NaCl, 20 mM HEPES, 10 mM β-mercaptoethanol, and pH 7.5. The HSQC spectra during the titration showed no significant changes in chemical shifts. (**b**) Slices of a selected peak (H51) in HSQC spectra during ligand titration showed significant line-broadening.

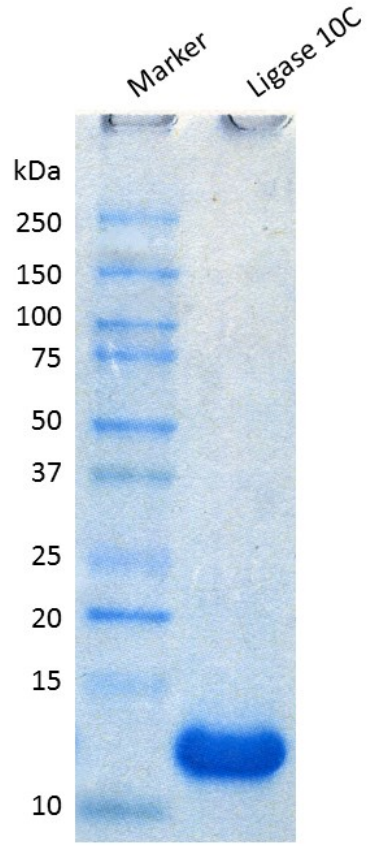


Figure S13. SDS-PAGE gel of purified ligase 10C. NuPAGE 4-12% Bis-Tris gel (Invitrogen) Coomassie stained of ligase 10C purified by nickel affinity chromatography and size exclusion chromatography and 10-250 kDA ladder P7703S (New England Biolabs) used as a marker.

Chapter 3- FLAMEnGO: A fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data

Reprinted with permission from:

Fa-An Chao, Lei Shi, Larry R. Masterson, Gianluigi Veglia. FLAMEnGO: A fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. J Magn Reson. 2012 Jan;214(1):103-10.

Building on a recent method by Matthews and co-workers, we developed a new and efficient algorithm to assign methyl resonances from sparse and ambiguous NMR data. The new algorithm (FLAMEnGO: Fuzzy Logic Assignment of MEthyl GrOups) uses Monte Carlo sampling in conjunction with *fuzzy logic* to obtain the assignment of methyl resonances at high fidelity. Furthermore, we demonstrate that the inclusion of paramagnetic relaxation enhancement (PRE) data in the assignment strategy increases the percentage of correct assignments with sparse NOE data. Using synthetic tests and experimental data we show that this new approach provides up to ~80% correct assignments with only 30% of methyl–methyl NOE data. In the experimental case of ubiquitin, PRE data from two spin labeled sites improve the percentage of assigned methyl groups up to ~91%. This new strategy promises to further expand methyl group NMR spectroscopy to very large macromolecular systems.

3.1 Introduction

Traditionally, the structural elucidation of large proteins and protein complexes at high resolution has been the realm of X-ray crystallography. However, this technique does not provide an atomic view of the molecular motions. Also, the presence of conformational transitions can only be inferred from the B-factors, rather than directly probed. On the other hand, conventional NMR approaches based on the main chain spectroscopy are not sensitive enough to study large macromolecular complexes.

In the past few years, selective methyl group labeling techniques in highly deuterated proteins, in concert with TROSY spectroscopy, enabled the analysis of large protein complexes up to 1 MDa (1). A significant problem for this approach is the assignment of the methyl $^{1}H/^{13}C$ HMQC (methyl-TROSY) spectrum. Ideally, one can utilize methyl 'out-and-back' experiments, mapping all of the methyl groups and linking them to the protein backbone (2,3). However, this strategy necessitates the assignment of the backbone nitrogen, C^{α}, or C' resonances, which is problematic to achieve for large systems. Moreover, the pulse sequences utilized for these experiments require high-level of deuteration and are generally very insensitive, due to fast T_2 relaxation. When the macromolecular complexes are sufficiently large or in the presence of conformational dynamics (broad resonances and overlap), this approach fails. In favorable cases, it is possible to use a *divide and conquer* strategy (4), where smaller fragments of the proteins or isolated components of the complex are expressed individually and the resonance assignment is transferred from the sub-spectra to the spectrum of the intact protein or protein complex (5,6). Site specific mutagenesis has also been used to assign specific methyl resonances (6,7). Though, the latter is very time-consuming and prone to generation of non-native folds.

To overcome these hurdles, Matthews and co-workers introduced an automated assignment procedure for methyl group assignment (8), which compares the

experimental chemical shifts and NOE contacts with those back-calculated from an X-ray crystal structure. The procedure requires that the number of experimental NOEs is at least 50% of those back-calculated (8). Another stringent condition is the unambiguous mapping of the NOE data onto the methyl-TROSY spectrum, which requires 4D F_2–^{13}C, F_3–^{13}C-edited NOESY experiments. Under these conditions, the approach leads to >90% correct assignments of methyl resonances for small and large systems (8). However, large systems often display fewer NOE cross-peaks than those predicted from the X-ray structures. The lack of complete NOE networks significantly deteriorates the performance of this procedure (8). In addition, resonance overlap makes it difficult to accurately map all of the NOE data to the donor resonances in the methyl-TROSY spectrum. While the use of a 4D NOESY spectrum alleviates this issue, this route is not robust enough for larger macromolecular systems, which have limited solubility and lower sensitivity. Here, we present a new automated assignment algorithm, FLAMEnGO (Fuzzy Logic Assignments of MEthyl GrOups), which has high tolerance for sparse NOE information (as low as 30%), and enables the use of ambiguous methyl–methyl NOEs through the combination of Monte Carlo sampling and *fuzzy logic (9)*. FLAMEnGO can incorporate 3D NOESY data from amide-methyl contacts, 4D NOESY data, as well as paramagnetic relaxation enhancements (PREs). The latter represents a crucial aid for unambiguous assignment of the NOEs in large systems and in the presence of sparse data. Using only 30% of synthetic NOE data, we show that FLAMEnGO is able to assign ~70% of the methyl resonances of maltose binding protein (MBP) and cutinase. Finally, for experimental data acquired on ubiquitin, FLAMEnGO can achieve ~80% of the methyl assignments with only 30% of the NOE data, and up to ~91% when PRE data are included.

3.2 Theoretical basis of the algorithm

3.2.1 Global score function

The architecture of FLAMEnGO is illustrated in **Figure 1**. As for the previously proposed method (8), our algorithm is based on a global score function that estimates the agreement between experimental and simulated NOE contacts. As an input, our algorithm requires a X-ray crystal structure, a peak picking of the 2D methyl-TROSY spectrum, and experimental NOESY data of the system under examination (**Figure 1A**). A seed assignment is given to the 2D methyl-TROSY spectrum and no assignment is necessary for the NOESY spectrum. The global score function is defined as:

$$G(x) = \max_{a \in A}\{Match_{total}\}$$

where *a* is an assignment from the 2D methyl-TROSY spectrum and *A* is the set of all possible assignments of the methyl-TROSY spectrum , *x* is the NOE distance cutoff. The function *G*(*x*) is used to find the maximum of a total matching function ($Match_{total}$), which is a linear combination of NOEs ($Match_{NOE}$), chemical shift ($Match_{CS}$), and PRE ($Match_{PRE}$) terms:

$$Match_{total} = Match_{NOE} + Match_{CS} + Match_{PRE}$$

Similar to the approach by Matthews and co-workers, a scaling factor is used in the $Match_{CS}$ term (here, we used 0.2), while all other terms are not scaled.

3.2.2 NOE matching and fuzzy logic

The NOE matching function compares the expected NOE contacts obtained from the crystal structure with the experimentally determined NOEs. The program first simulates all NOE contacts among methyl groups from a crystal structure using a given NOE distance cutoff (*x*), where contacts are considered to be any methyl pair less than *x*. Based on this NOE contact information, we use the experimental methyl group chemical shifts from the methyl-TROSY spectrum and a seed assignment to identify the expected NOESY cross peaks ($S_{NOE}(x, a)$), i.e., simulated NOESY spectrum. Note that we did not calculate the intensity of the NOEs, since we are only concerned about the number and position of expected

NOE cross peaks from a 3D structure and the donor peaks in a methyl-TROSY spectrum.

In our algorithm, both the NOE distance cutoff and the peak assignment are regarded as variables, which are optimized to generate an optimal cutoff and the best-fit assignment. To compare the simulated and experimental spectrum (*E*), the algorithm matches the simulated NOE peak (p_2) to the closest experimental NOE peak (p_1). The NOE matching function ($Match_{NOE}$) is then defined as:

$$Match_{NOE}(S_{NOE}(x,a),E) = \sum_{p_1 \in E} \max_{p_2 \in S_{NOE}(x,a)} e^{-0.5 \cdot D^2(p_1,p_2)} \qquad \text{(E1)}$$

where $D(p_1, p_2)$ represents the distance between the experimental $p_1(x_1, y_1, z_1)$ and simulated $p_2(x_2, y_2, z_2)$ peaks. $D(p_1, p_2)$ is estimated from the full width at half maximum (FWHM), assuming Gaussian shaped peak fitting with FWHM LW_x, LW_y and LW_z for the *x*, *y*, and *z* dimensions:

$$D(p_1,p_2) = 2\sqrt{2ln2} \cdot \left[\left(\frac{x_1 - x_2}{LW_x}\right)^2 + \left(\frac{y_1 - y_2}{LW_y}\right)^2 + \left(\frac{z_1 - z_2}{LW_z}\right)^2 \right] \qquad \text{(E2)}$$

The procedure starts with an initial seed assignment, which is swapped iteratively using a Monte Carlo sampling step (10). The best-fit assignment is achieved when the match score function reaches its maximum. To compare the simulated spectra with experimental ones after each assignment swap, we used *fuzzy logic (9)*. *Fuzzy logic* accounts for ambiguous NOE information by providing a range of values between a perfect match ('100%' or 'yes') and a complete mismatch ('0%' or 'no'), rather than binary responses (such as 'yes' or 'no'). The *fuzzy logic* step allows the program to compare and score objects (*i.e.*, spectra) based on measurable criteria (*i.e.*, chemical shift differences), estimating a percentage of confidence.

Assuming a Gaussian line shape for the experimental NOE peaks, the FWHM is measured and converted in to a standard deviation of the average peak position. Then, the chemical shift difference between an experimental cross peak and a simulated peak will be divided by the standard deviation (E2). Using Eq. (E1), the program screens all simulated NOE cross peaks (p_2) to match the given experimental peak (p_1) by maximizing the Gaussian function. Thus, the smaller

the difference between simulated and experimental chemical shifts, the higher the score of the Gaussian function. A score between 1 (complete match) and 0 (no match) from the Gaussian function will be returned for each peak (p_1) in the experimental NOESY spectrum (*E*), and the program sums all scores to provide a match score ($Match_{NOE}(S_{NOE}(x, a),E)$) for the experimental data and the simulated ones.

3.2.3 Predicted chemical shift information

As an option, we also included a matching function for chemical shifts predicted from the X-ray structure using software such as CH3*Shift (11)* or SHIFTX2 (12). For the chemical shifts, the target function ($Match_{CS}$) for all of the methyl groups (*i*) is defined as:

$$Match_{CS} = -c \cdot \sum_i \frac{|\delta_i^r - \delta_i^p|}{\sigma_i} \quad \text{(E3)}$$

where δ_i^r represents the experimental chemical shifts, δ_i^p the predicted chemical shifts, σ_i the predicted error, and *c* is a scaling constant set to 0.2. The index, *i*, represents each methyl resonance. This option is very similar to that introduced by Matthews and co-workers (8), and minimizes the differences between experimental and predicted chemical shifts. For methyl-TROSY peaks that do not have NOESY cross-peaks with other resonances, this term can increase the percent of correct assignments.

3.2.4 PRE information

Finally, we included an additional function that uses PRE data ($Match_{PRE}$). The PRE data can be implemented either qualitatively or quantitatively. For *qualitative* PRE data, we used the following convention:

$$Match_{PRE} = \sum_i r_i \begin{cases} r_i = 0 \;\; if\ the\ resonance\ i\ is\ NOT\ in\ its\ PRE\ range \\ r_i = 1 \;\; if\ the\ resonance\ i\ is\ in\ its\ PRE\ range \end{cases} \quad \text{(E4)}$$

where r_i is a restraint imposed by PRE on each methyl resonance *i*. Methyl resonances are classified into three groups based on the quenching effects:

unaffected (<20% reduction in signal intensity), slightly quenched (20–80%), and strongly quenched (above 80%). At the same time, all methyl groups in the structure are ranked based on their distance to the spin label and grouped into three categories in analogy with the quenching patterns. For all methyl resonance, $Match_{PRE}$ (E4) is used to score the agreement between experimental data and distance range. In other words, if a resonance in the strongly quenched group is assigned to a methyl group within close distance to the spin label, r_i is set to 1, indicating a good agreement of the assignment with PRE restraints. Otherwise, a penalty is added by setting r_i to 0.

Furthermore, our algorithm can also use a more quantitative interpretation of the PRE data, with explicit distance restraints between each individual methyl group and the spin label (13). If *quantitative* PRE data are available, the distance between a methyl group and a spin label can be calculated using the following equations (13):

$$\frac{I_{ox}}{I_{red}} = \frac{\exp(-2R_2^{ox}\tau)}{\exp(-2R_2^{red}\tau)} \cdot \left(\frac{R_2^{red}}{R_2^{ox}}\right)^2 \quad \text{(E5)}$$

$$d = \left[\frac{\beta}{R_{2,PRE}}\left(4\tau_c \frac{3\tau_c}{1+\omega_H^2\tau_c^2}\right)\right]^{1/6} \quad \text{(E6)}$$

where I_{red} is the peak intensity in the methyl-TROSY spectrum of the protein with the spin label in the reduced state and I_{ox} is the peak intensity in the oxidized state. The value τ is the time for transfer of magnetization between ^1H and ^{13}C spins, $\beta = 1.23 \times 10^{-44}$ m^6 s^{-2}, ω_H is Larmor frequency of protons (rad s^{-1}), and τ_c is the rotational correlation time of the protein (s). R_2^{ox} can be calculated from Eq. (E5) if R_2^{red} is known, and then $R_{2,PRE} = R_2^{ox} - R_2^{red}$ (13). Finally, the distance d can be obtained, and an additional 2 Å uncertainty is included to form a PRE range for a particular methyl resonance (13). Again, the same convention (Eq. (E4)) is used: if the resonance is assigned to a methyl group which is within $d - 2$ and $d + 2$ of the spin label, r_i is set to 1; otherwise a penalty is added by setting r_i to be 0.

3.2.5 Determination of the optimal NOE distance cutoff

An important step in the entire protocol is the optimization of the NOE distance cutoff (x) to calculate the NOESY data from the X-ray structure. Smaller NOE distance cutoffs are unable to take full advantage of the information in the NOE data, while larger cutoffs introduce more uncertainties, resulting in lower accuracy of the methyl assignment. In our approach, we optimize the cutoff by carrying out multiple assignment calculations with different distance cutoff values. Thus, the optimal distance cutoff is obtained when the *global score* function reaches a plateau. As a result, the simulated data with the optimal cutoff can utilize most of experimental data without overfitting. A schematic of the algorithm is reported in **Figure 1B**. For a given NOE distance cutoff value, the total matching function is maximized using the Metropolis Monte Carlo method (10) by swapping the initial seed assignment on the experimental methyl-TROSY spectrum. The maximal value is called a *global score*. In the Monte Carlo sampling an annealing 'temperature' T defined as the total number of average cross-peaks plus PRE restraints is reduced to 1 in a ~1 million swapping steps. At each step, the assignments on two methyl resonances from the same residue-type (and the same prochirality) are randomly chosen and exchanged. If the value of the target function increases or decreases within a random value, the swapping step is accepted; otherwise, it is rejected and a new pair of assignments is chosen and swapped. Depending on the system size, the number of steps can be adjusted to reach convergence. To determine the confidence for each assignment, we repeated the calculations at the optimal cutoff several times. The most probable assignment is chosen as the final assignment. The mathematical proof for the optimization of the NOE distance cutoff is provided in the Supporting Information.

3.3 Materials and methods

Sample preparation

The BL21(DE3) competent cells were transformed with plasmids containing the ubiquitin sequence, and inoculated into 1 l LB medium supplemented with 100 μg ampicillin. Upon reaching OD_{600} of ~1, the cells were harvested and transferred into 250 ml 100% deuterated M9 medium containing 1 g $^{15}NH_4Cl$, 4 g deuterated glucose, 70 mg/L methyl labeled α-ketoisovalerate, and 90 mg/L methyl labeled α-ketobutyrate. After 1 h of incubation at 37 °C, 1 mM IPTG was added to induce protein over-expression. The culture was harvested after 5 h of induction at 37 °C and stored at −20 °C. The frozen cell pellet was lysed by sonication in 50 mM sodium acetate buffer at pH 5.0 and centrifuged at 45,000 g at 4 °C. The supernatant was loaded into a P11 cation exchange column (WHATMAN) and eluted with a gradient of 0–1 M NaCl. The pooled fractions containing ubiquitin were further purified by size-exclusion chromatography using a Sephacryl S-200 resin (GE) with 100 mM phosphate buffer (pH 7.0). The fractions were concentrated and the sample was then dialyzed in NMR buffer (20 mM phosphate buffer, 1 mM NaN_3, and pH 6.5), and concentrated to ~1.5 mM for the NOESY experiments.

The K48C and G75C mutants of ubiquitin were generated using a QuikChange kit from Stratagene. Expression and purification were performed as described above. A fivefold excess of MTSSL was added to the mutant protein dissolved in NMR buffer at 25 °C for 4 h. The free MTSSL was dialyzed out in NMR buffer at room temperature. The final samples were concentrated to ~1.5 mM for PRE measurements.

NMR spectroscopy

A time-shared 3D HMQC-NOESY-HMQC was acquired on a Varian VNMRS instrument operating at a ^1H larmor frequency of 600 MHz. A mixing time of 800 ms was used based on the build-up in 2D planes at various mixing times (150–1000 ms). The spectrum was acquired using a spectral width of 10,000 (3500/2200) Hz for ^1H ($^{13}C/^{15}N$). The indirect dimensions were acquired with 128 increments in the carbon/nitrogen time-shared dimension, and 42 increments in the carbon dimension. For data processing, the number of points in the ^{13}C

dimension was doubled by linear prediction and all dimensions were zero-filled. Before MTSSL spin labeling, the integrity of the mutant ubiquitin samples were confirmed by [^{1}H, ^{15}N]-HSQC and [^{1}H, ^{13}C]-HMQC, which showed no changes in the methyl resonances and negligible differences in the amide fingerprint (see **Figure S1**). After MTSSL labeling, a [^{1}H, ^{13}C]-HMQC spectrum was acquired, the sample was then reduced in the presence of a 10-fold excess of DTT, and the spectrum was reacquired (**Figure S2**).

3.4 Results

To test the performance of the algorithm in the case of sparse NOE data, we ran initial tests with two different proteins, whose structures have been determined at high resolution by X-ray and NMR: maltose binding protein (MBP) (PDB: 1DMB, BMRB: 7114) and cutinase (PDB: 1CEX, BMRB: 4101) (**Figure 2A** and **B**). For the calculations, the stereospecificity of Leu and Val methyl groups in both proteins was assumed to be known. We generated two sets of sparse NOE data for both MBP and cutinase, with a distance cutoff of 7 Å, and randomly eliminated 70% of the back-calculated NOEs. Using the algorithm from Matthews and co-workers (8), we found that the optimal distance cutoffs were 6.1 and 5.4 Å for MBP and cutinase, respectively. As an output, we obtained 53% of correct methyl assignments for MBP (**Figure 2C**) and 41% for cutinase (**Figure 2D**). In contrast, FLAMEnGO found an optimal distance cutoff of 7 Å for both cases, and resulted in ~85% and ~76% correct assignments for MBP and cutinase, respectively, with an improvement of ~35% in accuracy with respect to the original approach. A higher accuracy in the determination of the NOE distance cutoff corresponds to a better tolerance for sparse NOE data. To evaluate the original algorithm against FLAMEnGO, we again calculated an assignment with the original script using the correct cutoff (7 Å). While the accuracy of the resulting assignment improved to 77% for MBP and 47% for cutinase, it was still found to be significantly lower than that obtained with FLAMEnGO. Since the

same cutoff is used in this case, the improved results from our algorithm over the original are due to the more efficient sampling accessible by the Monte Carlo algorithm built in FLAMEnGO.

To test the algorithm in the presence of spectral overlap, we simulated 3D F_2–^{13}C, F_3–^{13}C-edited NOESY experiments for both MBP and cutinase using their X-ray structures with a 7 Å NOE distance cutoff. For the original method, we took the NOE data set and randomly eliminated 30% of the data. If an NOE cross-peak fell within 0.01 ppm from ^{1}H and 0.1 ppm from the ^{13}C frequencies in the methyl-TROSY spectrum, we assigned it to that donor resonance. Note that more than one donor resonance can be matched. In this case the donor is chosen arbitrarily. Under these conditions, the accuracy of the assignment using the original algorithm decreased by 55% and 20% for both MBP and cutinase, respectively (**Figure 2E** and **F**). In contrast, FLAMEnGO provided ~97% and ~87% accuracy for MBP and cutinase, respectively, for cross-peaks falling within 0.025 ppm for ^{1}H and 0.4 ppm for ^{13}C frequency. Therefore, the performance of the original algorithm decreases with the increase of the uncertainty of the match between the donor and the cross-peak (**Figure 2E** and **F**).

Moreover, we tested the performance of FLAMEnGO using 4D NOESY, PRE, and amide-methyl NOEs as inputs. For both MBP and cutinase, we back-calculated 3D amide-methyl and 4D methyl–methyl NOESY spectra, containing only 30% of the NOE contacts measurable in the X-ray structures. The line widths of ^{1}H, ^{13}C, ^{15}N in the simulated 3D NOESY data were set to 0.025 ppm, 0.4 ppm, and 0.5 ppm, respectively. For the simulated 4D NOESY data 0.8 ppm and 0.1 ppm for both ^{1}H and ^{13}C were used, respectively. We obtained ~10% improvement in the assignments of both MBP and cutinase when 4D data sets were used (**Figure 3A**). Subsequently, we tested the effects of PRE data. For MBP, we simulated the effect of nitroxide spin labels engineered at positions 145 and 306 and back-calculated three groups of PRE effects: methyl groups within 15 Å of the spin label are strongly quenched, methyl groups between 15 and 35 Å are slightly quenched, and those beyond 35 Å are unaffected (14). When

the simulated PRE data were included in the calculations, the percentage of correct assignment reached ~86% for the 3D dataset and ~91% for the 4D dataset for MBP (**Figure 3A**). Most strikingly, if the assignment of amide chemical shifts and amide-methyl NOE information were both included in the calculation, the percentage of correct assignment increased up to ~93% for MBP and ~87% for cutinase (**Figure 3B**). It is worthy to note that only 15% of the simulated 3D amide-methyl NOESY data were used.

Furthermore, we evaluated the probability of correct assignments by repeating the calculation five times with the same optimized NOE distance cutoff and choosing the assignment with the highest occurrence frequency (OF). In this way, we filtered out low probability assignments using an OF of at least 0.8. The percent of correct assignments within this filtered subset (*i.e.*, % correct with OF $\geqslant$ 0.8, *y*-axis of **Figure 3C** and **D**) was >95% and >85% for MBP and cutinase, respectively. Adding more experimental information (*i.e.*, PREs or 4D NOESY data) significantly increased the number of assignments with an OF $\geqslant$ 0.8 (**Figure 3C** and **D**, *x*-axis). Therefore, additional experimental data provides higher confidence in the assignments, while the % correct with OF $\geqslant$ 0.8 did not change significantly (**Figure 3C** and **D**, *y*-axis). When the 3D NOESY data with and without PRE information were used for cutinase, a ~10% drop occurs in the accuracy of the assignment with OF $\geqslant$ 0.8 (**Figure 3C**). However, as shown in **Figure 3A**, there is an increase in the average number of correct assignments. This is due to the fact that there are more overall assignments with an OF $\geqslant$ 0.8.

We also compared the performance of FLAMEnGO with structures of MBP derived from X-ray crystallography and solution NMR (all atom RMSD ~2 Å). Using the synthetic experimental NOE data from the crystal structure, the protocol gave a lower accuracy when different conformers were used (the difference was about 20%). However, the latter can be alleviated by providing additional stereospecific restraints. This demonstrated that relatively large

structural deviations can be tolerated by this approach (**Figure 4**) and highlights the importance of providing stereospecific information to the algorithm.

Finally, we tested the performance of the two algorithms with experimental data using a ubiquitin sample, which was uniformly ^{2}H, ^{15}N-labeled and selectively $^{1}H/^{13}C$-ILV methyl-labeled. We acquired a 3D (F_2–^{13}C, F_3–$^{13}C/^{15}N$-edited) time-shared NOESY at a mixing time of 800 ms. Since the original algorithm cannot handle amide-methyl NOEs, only methyl–methyl NOE data were used to provide a fair comparison. The experimental NOE data set displayed approximately 45% of the correlations expected from the X-ray crystal structure (PDB: 1UBQ, NOE distance cutoff = 5.9 Å). Since this number is likely to be smaller for large proteins, we reduced the data set to 30% of expected peaks. Ignoring the stereospecificity, FLAMEnGO provided ~70% correct assignments (**Figure S3**), while the original method provided only 39% correct assignments.

To test the effect of PRE data, we expressed and purified two mutants of ubiquitin (K48C and G75C) and engineered two MTSSL spin labels. The [^{1}H, ^{15}N]-HSQC and [^{1}H, ^{13}C]-HMQC of the two mutants were essentially unaffected by the mutations (**Figure S1**). The PRE effects were estimated qualitatively using the intensity of the resonances from methyl-TROSY experiments (**Figure S2**). Using data corresponding to 45%, 30% or 15% of the expected methyl–methyl NOE cross peaks, as well as stereospecific and residue-type information, FLAMEnGO generated ~85% correct assignments (**Figure 5B**) without PREs and ~91% with PREs. Remarkably, the inclusion of amide-methyl NOE cross peaks increased the correct assignments up to ~94% without PREs, and ~100% with PREs (**Figure 5D**).

3.5 Discussion

Our approach builds on the previous work from Matthews and co-workers (8) for the assignment of methyl groups. A significant improvement with respect of the original approach is obtained with the *fuzzy logic* step and enhanced sampling by Monte Carlo method. The use of *fuzzy logic* allows the comparison of highly

ambiguous NOEs obtained from 3D experiments and is more tolerant for sparse data. While the Matthews' approach does not perform well with sparse NOEs (less than 50% than expected), our algorithm performs well even in the presence of 30% of the NOE data. Sparse and ambiguous data sets are expected for large macromolecular systems or for situations in which conformational dynamics reduce the sensitivity of the NMR experiments. Therefore, this implementation will extend the methyl-based NMR spectroscopy approach to larger systems. Our program handles additional information such as PRE data, amide-methyl NOE data, and 4D NOESY data to improve assignment accuracy and convergence of the sampling.

As for the Matthews approach (8), our algorithm is sensitive to residue-type and stereospecific information available for the methyl groups. For instance, when applied to ubiquitin with the complete experimental NOESY data, but without residue-type information, FLAMEnGO provides ~82% of the correct assignment; in contrast to only ~55% of the correct assignment when stereospecific information is missing. Nonetheless, residue-type and stereospecific assignment can be easily obtained with selective labeling schemes (15,16), reducing the sampling space for the algorithm and increasing its convergence. For ubiquitin, the algorithm generated ~94% correct assignments when residue type information was provided.

What is the impact of the implementation of chemical shifts predicted from the X-ray structures? The protocol from Matthews and co-workers (8) relies on predicted chemical shifts as a crucial step in the assignment procedure. We found that the inclusion of predicted chemical shifts is important when NOE data are sparse. However, the inclusion of predicted chemical shifts may decrease the overall accuracy of the assignments, since discrepancies between calculated and experimental chemical shifts can be large. Furthermore, if the chemical shifts are not referenced correctly, it is possible to introduce severe errors in the assignment procedure.

As mentioned, the determination of optimal cutoffs for the NOE spectra simulated from crystal structures is very important. The NOE cutoff is optimized using several runs, arraying its value to obtain the minimal cutoff value, which accounts for most of the experimental NOE data. Although larger NOE cutoffs might appear optimal for taking into account long-range NOE information, they can also introduce more uncertainties during the assignment swapping steps of the algorithm. Long-range information can also be introduced in the algorithm using PRE effects, which we demonstrated to increase the accuracy of the assignment protocol dramatically. In fact, PRE data can provide long-range distance restraints that cannot be obtained from NOE, resulting in improvements of the assignment accuracy by up to 15% (**Figures 3A, C, 5B,** and **D**).

Finally, the density of the NOE network influences the performance of the algorithm: the higher the number of NOE cross-peaks, the more accurate is the assignment. Moreover, pushing the limits on the range of the experimental NOE conveys more information to the program to assign methyl groups. Thus, the mixing time in the NOESY experiment should be optimized to provide such information. In addition to Leu, Ile(δ1), and Val, new biosynthetic strategies have been introduced for producing proteins with $^{13}CH_3$ labeling at Ala, Met, Ile(γ2), and Thr (17-19). The combination of these labeling schemes will provide more dense clusters of methyl–methyl NOE network, which should dramatically increase the performance of our approach.

3.6 Conclusion

In summary, we propose a new assignment strategy that is very tolerant to sparse and ambiguous NOE data. We demonstrated that the inclusion of either qualitative or quantitative PRE data, or amide-methyl NOE data, dramatically increases the convergence of the algorithm to assignment accuracy greater than 90%. These aspects make our approach more applicable to larger macromolecular systems, where sparse information due to size and intrinsic dynamics reduces the performance of NMR experiments.

3.7 Mathematical proof of the algorithm:

In order to demonstrate the validity of the choice of optimal NOE distance cutoff (x_{opt}), we define a simplified global score function *G(x),* which depends only on the modified $Match_{NOE}$:

$$G(x) = \max_{a \in A}\{match[S_{NOE}(x), E(a)]\}$$

where $S_{NOE}(x)$ is the back-calculated NOE from the X-ray structure at the cutoff *x*, *a* is an assignment of the methyl groups in the 2D methyl-TROSY spectrum, *E(a)* are the experimental NOE contacts which are unambiguously assigned by *a*, *A* is the set of all possible assignments of the methyl-TROSY spectrum, and *match* indicates a matching function used to compare experimental and simulated data, which is defined as the number of experimental NOEs that match the simulated data. To obtain x_{opt}, we first define *d* as the longest distance determined in the protein by the experimental NOEs. Therefore, the following two statements must be proved:

(1) *d* is the upper bound of the following expression:

$$d' = \min\{x \in R^{+} | G(x) = \max(G(x))\}$$

where d' represents the minimal value of x at the maximal value of the global score function (*G(x)*) (**Figure S3**). To minimize the uncertainty in the assignment, we need to define another function:

(2) $|B(x)| = |\{a \in A | match[S_{NOE}(x), E(a)] = G(x)\}|$

describes the number of possible assignments that completely match the simulated data with the experimental ones and |*B(x)*| is a non-decreasing function for $x \geq d'$.

Proof:

Based on the definition of the *match* function and $S_{NOE}(x)$ contains $S_{NOE}(y)$, where *y* is another possible value of NOE cutoff and $x > y$, *G(x)* is a non-decreasing function. The total number of cross-peaks in the experimental NOE data *P* is the upper bound of *G(x)* (*i.e.*, *P = max(G(x))*) for all the cutoffs *x*. Since there are a NOE cutoff distance *d* and the correct assignment for which all the

experimental NOEs can match the simulated ones, *P* is the lower bound of *G(d);* so *G(d)* is exactly *P*.

Thus, $d' = min\{x \in R^+ | G(x)=P\} \leq d$, and $G(x) = P$ for all $x \geq d'$. Moreover, for $x \geq d'$, $B(x) = \{a \in A | match[S_{NOE}(x), E(a)]=P)\}$. Because $S_{NOE}(x)$ contains $S_{NOE}(y)$ if $x > y$, $B(x)$ contains $B(y)$ if $x > y \geq d'$.

Although the global score function is a non-decreasing function, it will not increase without limit, and it will stop increasing when all of the experimental data are matched. Ideally, we would like to use a NOE distance cutoff that includes all of the observed NOEs. However, values of the NOE cutoffs larger than the optimal one will also increase the number of possible solutions (assignments), complicating the overall analysis and leading to uncertainty. The latter is exacerbated when the experimental data are sparse (**Figure S3**). Practically, a few NOEs with very long distances may be created by conformational dynamics or uniquely slow relaxation. These NOEs are sacrificed to prevent excessive uncertainty being introduced into the calculation. Therefore, the optimal NOE distance cutoff is determined when the slope of the global score function starts approaching to zero.

Further explanation of Fuzzy logic step:

In the script from the Matthews' group, all NOE data are assumed to be unambiguous, so the calculation can be carried out by classical logic, which is fast and powerful as compared to *fuzzy logic*. Therefore, the NOESY spectrum needs to be pre-assigned before performing the calculation, but this will introduce significant errors when the data are ambiguous. However, if we directly compare the experimental NOESY spectrum with the simulated one without assigning the NOESY spectrum, we can bypass the above problem. As long as the simulated spectrum can be used to explain the data in the experimental spectrum, we can regard it as a complete match. In a different perspective, this approach implicitly assigns an experimental NOE cross peak to a NOE contact between any pair of methyl groups if those two methyl groups are within a given distance cutoff (*x*)

and their assigned chemical shifts (*a*) can simulate the same cross peak. But, inevitable experimental errors occur in the experimental data, such as peak distortion and peak overlap, which prevent us from using this approach.
By including *fuzzy logic*, FLAMEnGO compares the positions of experimental peaks and simulated ones, and uses chemical shift differences and line widths to calculate the confidence of matching. Assuming chemical shift perturbations due to the experimental errors are relatively small compared to the line widths, the real peak position shall be very close to the theoretical peak position. The program can still compare the experimental data with the simulated ones based on peak positions. Thus, for each experimental cross peak, FLAMEnGO assigns a percentage of confidence to any pair of methyl groups according to their distance and their assigned chemical shifts. If a pair of methyl groups has a distance within the given cutoff, and their simulated cross peak is very close to the real experimental peak, the program will assign the experimental peak with a high confidence value to the NOE contact between this pair. The program will pick the highest confidence value for each experimental cross peak, and summate all the highest confidence values together to provide a value for $Match_{NOE}$. Finally, when the given NOE distance cutoff (*x*) is large enough and the best-fit assignment (*a*) is found, the program will return the highest match value for the two spectra.

3.8 References

1. Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007 Feb 8;445(7128):618-22.

2. Tugarinov V, Kay LE. Side chain assignments of ile delta 1 methyl groups in high molecular weight proteins: An application to a 46 ns tumbling molecule. J Am Chem Soc. 2003 May 14;125(19):5701-6.

3. Tugarinov V, Kay LE. Ile, leu, and val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc. 2003 11/01;125(45):13868-78.

4. Velyvis A, Schachman HK, Kay LE. Application of methyl-TROSY NMR to test allosteric models describing effects of nucleotide binding to aspartate transcarbamoylase. J Mol Biol. 2009 Apr 3;387(3):540-7.

5. Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, et al. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell. 2007 Nov 16;131(4):756-69.

6. Kato H, van Ingen H, Zhou BR, Feng H, Bustin M, Kay LE, et al. Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. Proc Natl Acad Sci U S A. 2011 Jul 26;108(30):12283-8.

7. Amero C, Asuncion Dura M, Noirclerc-Savoye M, Perollier A, Gallet B, Plevin MJ, et al. A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. J Biomol NMR. 2011 Jul;50(3):229-36.

8. Xu Y, Liu M, Simpson PJ, Isaacson R, Cota E, Marchant J, et al. Automated assignment in selectively methyl-labeled proteins. J Am Chem Soc. 2009 Jul 15;131(27):9480-1.

9. Dubois D, Prade H. An introduction to fuzzy systems. Clinica Chimica Acta. 1998 2/9;270(1):3-29.

10. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys. 1953;21(6):1087-92.

11. Sahakyan AB, Vranken WF, Cavalli A, Vendruscolo M. Structure-based prediction of methyl chemical shifts in proteins. J Biomol NMR. 2011 Aug;50(4):331-46.
12. Han B, Liu Y, Ginzinger SW, Wishart DS. SHIFTX2: Significantly improved protein chemical shift prediction. J Biomol NMR. 2011 May;50(1):43-57.
13. Religa TL, Sprangers R, Kay LE. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. Science. 2010 Apr 2;328(5974):98-102.
14. Gaponenko V, Howarth JW, Columbus L, Gasmi-Seabrook G, Yuan J, Hubbell WL, et al. Protein global fold determination using site-directed spin and isotope labeling. Protein Sci. 2000 Feb;9(2):302-9.
15. Lian L, Middleton DA. Labelling approaches for protein structural studies by solution-state and solid-state NMR. Prog Nucl Magn Reson Spectrosc. 2001 10/19;39(3):171-90.
16. Plevin MJ, Hamelin O, Boisbouvier J, Gans P. A simple biosynthetic method for stereospecific resonance assignment of prochiral methyl groups in proteins. J Biomol NMR. 2011 Feb;49(2):61-7.
17. Fischer M, Kloiber K, Hausler J, Ledolter K, Konrat R, Schmid W. Synthesis of a 13C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. Chembiochem. 2007 Apr 16;8(6):610-2.
18. Ayala I, Sounier R, Use N, Gans P, Boisbouvier J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. J Biomol NMR. 2009 Feb;43(2):111-9.
19. Ruschak AM, Kay LE. Methyl groups as probes of supra-molecular structure, dynamics and function. J Biomol NMR. 2010 Jan;46(1):75-87.

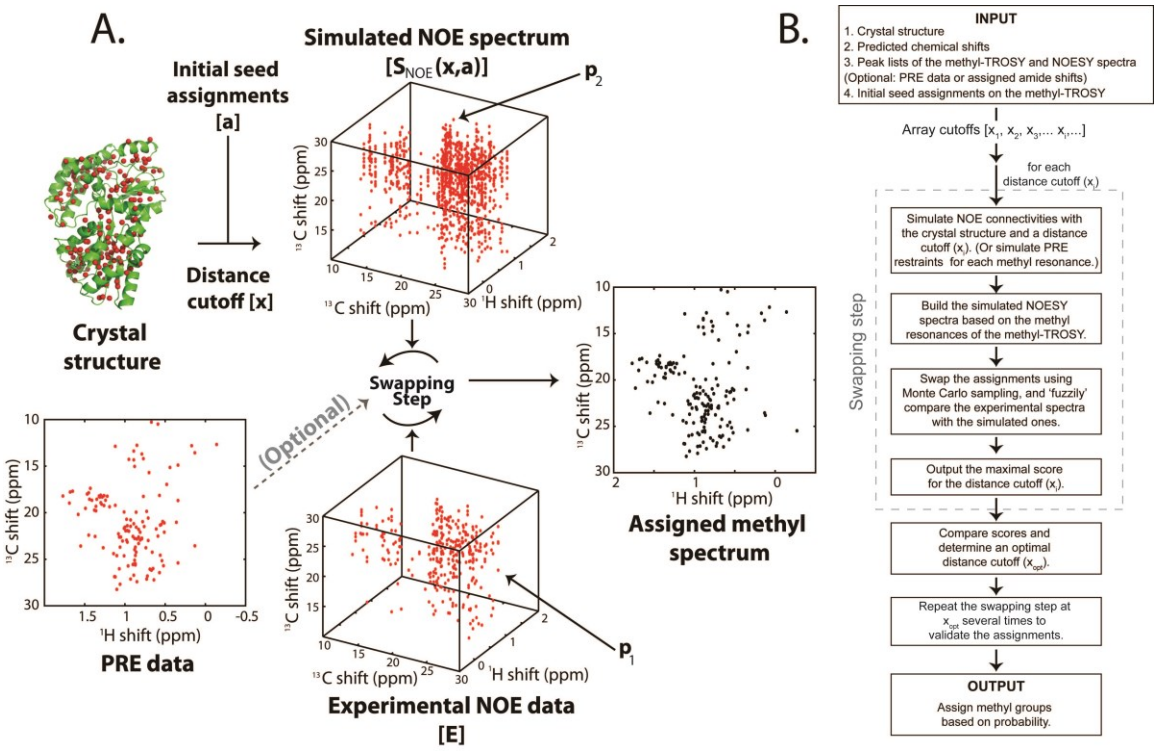


Figure 1. Outline of the autoassignment procedure. **A)** A graphical illustration of the algorithm shows that an X-ray crystal structure, experimental NOE data, and (optional) PRE data are required as input information. **B)** A flowchart of the FLAMEnGO algorithm.

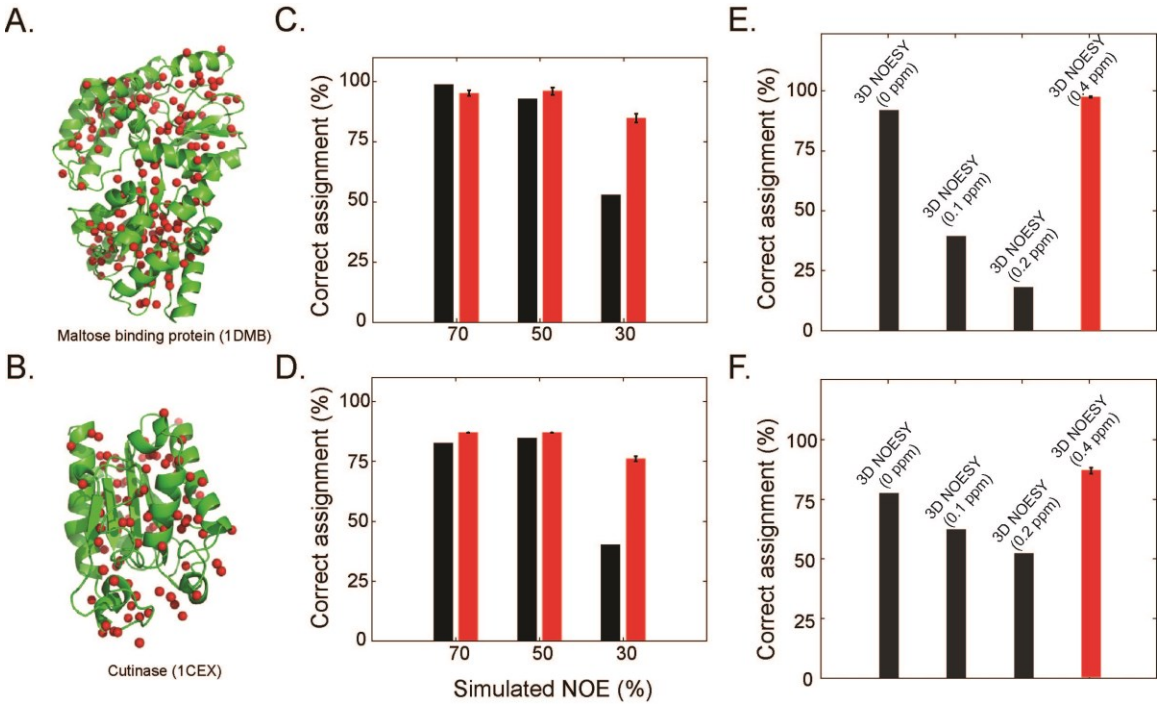


Figure 2. Comparison of the performances of FLAMEnGO (red bars) and the original algorithm by Matthews and co-workers (black bars). Crystal structures of MBP (42 kDa, panel **A**) and cutinase (22 kDa, panel **B**). Methyl groups of Ala, Ile(δ1), Leu, and Val are marked as red spheres, for a total of 166 (92) methyl groups in MBP (cutinase). All unambiguous NOE contacts were simulated using the crystal structures and 7 Å NOE distance cutoff (panels **C** and **D**). All of the calculations repeated three times. The error bars indicate the standard deviations. Panels **E** and **F** show the percentage of correct assignment as a function of the tolerance, i.e. chemical shift difference between NOE cross-peak and peak donor.

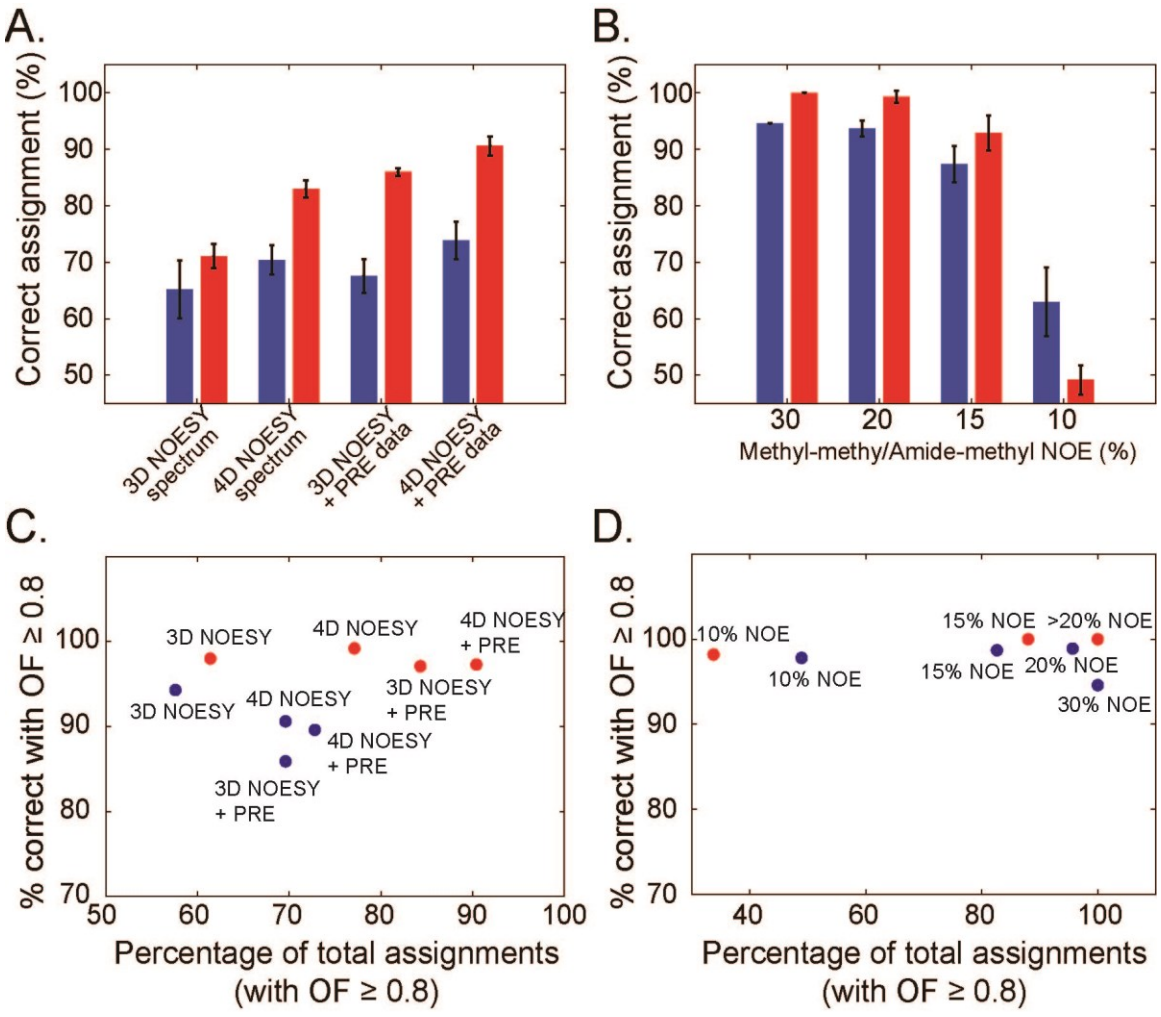


Figure 3. Effects of different input data on FLAMEnGO's performance. Red bars indicate the calculations on MBP, while the blue bars indicate the calculations on cutinase. Only sparse and ambiguous data were used in the calculations. Data in **A** and **B** are based on five independent calculations. The error bars reflect the standard deviation among the different runs. The final assignments reported in panels **C** and **D** were based on the highest occurrence frequency (OF) for each assignment. For MBP (red), S145C and S306C were chosen as spin labeled sites, while for cutinase (blue), S54C and S135C were chosen.

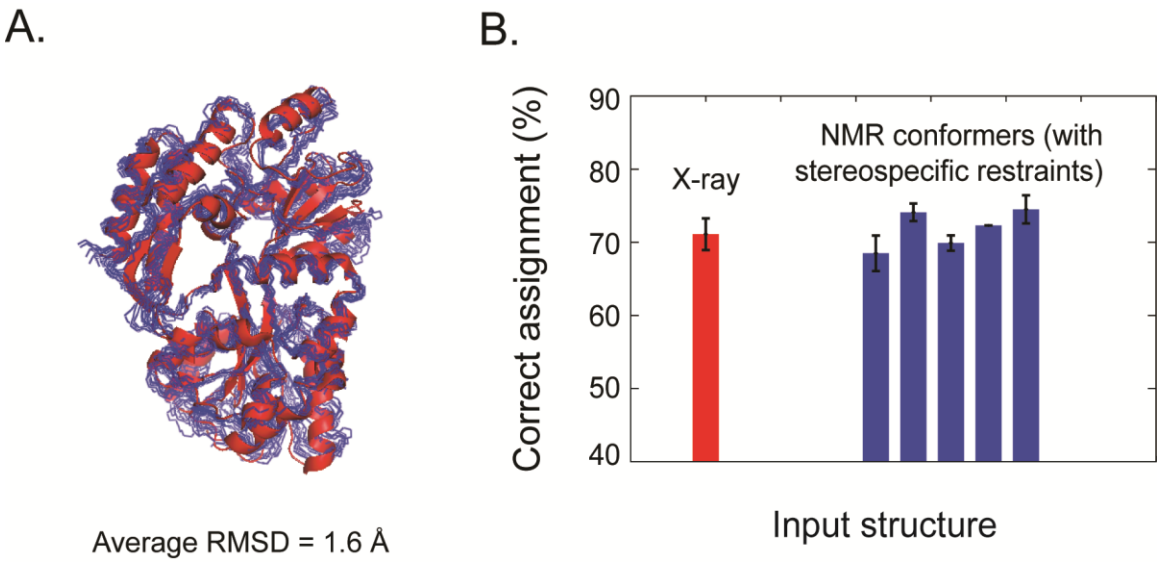


Figure 4. Performance of FLAMEnGO using the ensemble of different conformers of MBP. **A)** Overlay of a crystal structure (1DMB) and the solution structure ensemble (2H25) of MBP. The average all-atom RMSD between the crystal structure and each NMR conformer is up to 2.2 Å. **B)** FLAMEnGO calculations for 5 selected NMR conformers. In addition to the seed assignments, we used residue-type information, stereospecific restraints, and 30% of the methyl-methyl NOE data as an input. Three independent calculations were carried out with an average of ~72% correct assignment.

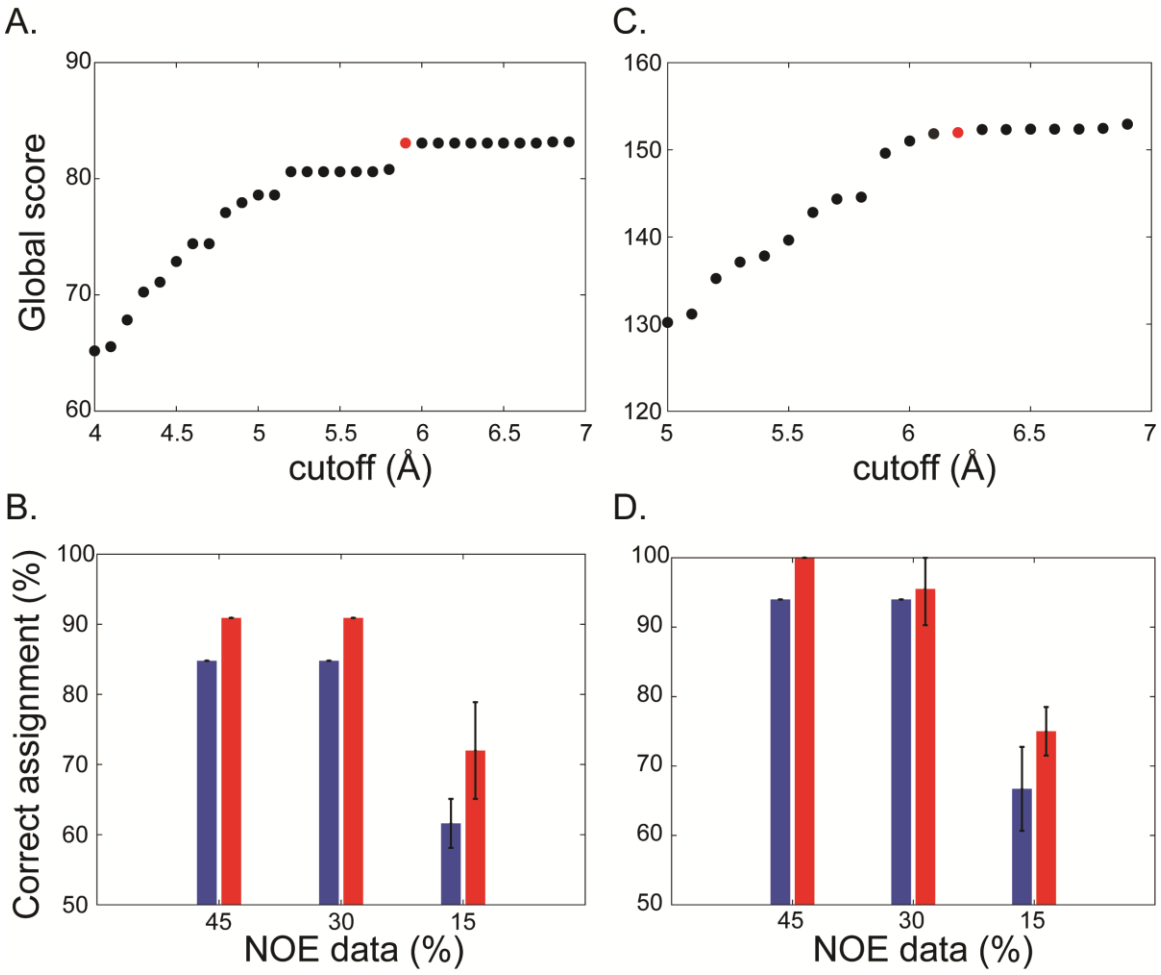


Figure 5. Automated assignment of experimental data from ubiquitin using FLAMEnGO. Percent assignment with NOE data only is reported using blue bars, while with NOE and PRE data is reported using red bars. Three independent calculations were carried out. **A)** Plot of the global score function versus the NOE distance cutoff. The optimal NOE distance cutoff (5.9 Å) is chosen at the curve plateau (red point). **B)** Percentage assignment obtained with only methyl-methyl NOE data (~45% of the expected NOE cross-peaks). **C)** Plot of the global function versus NOE distance cutoff for both amide-methyl and methyl-methyl NOESY data. **D)** Plot of the percentage of correct assignment using both amide-methyl and methyl-methyl NOESY experiments and NOE distance cutoff of 6.2 Å.

				Matthews' script		FLAMEnGO	
PDB code (BMRB code)	mol. wt. (kDa)	total number of methyls	average number NOEs per methyl group	% of correct assignments ignoring stereospecificity	% of correct assignments	% of correct assignments ignoring stereospecificity (standard deviation)	% of correct assignments (standard deviation)
1DMB (7114)	41.9	166	2	13.3	13.3	75.9 (3.3)	71.1 (2.2)
1CEX (4101)	22.3	92	1.5	23.9	17.4	70.4 (2.9)	63.9 (4.6)
1JV4 (4340)	19.7	61	1.1	27.9	19.7	60.1 (1.9)	44.8 (1.9)
1FIL (4082)	15	64	2	32.8	26.6	75 (4.7)	68.2 (4.8)
4I1B * (1061)	17.4	62	1.4	32.3	25.8	39.8 (3.7)	33.9 (0)
1BYL (4786)	14	64	1.4	21.9	10.9	62.5 (0)	40.6 (0)
1PEY (5899)	14.2	60	1.7	31.7	21.7	88.3 (0)	85 (2.9)
1RX4 (5741)	19.3	68	1.4	33.8	29.4	74 (5.9)	66.2 (10.2)
1LZG (4562)	14.7	46	1.3	56.5	37	90.3 (6.4)	79.2 (7.5)
1P7T **	82.8	270	1.5	12.6	7.4	62 (0.2)	51.9 (0)

Table S1. Methyl groups of various model proteins are assigned by the original script and FLAMEnGO. Only 30% of 3D methyl-methyl NOESY spectra are synthesized with a NOE distance cutoff 7Å for each test, residue-type information is provided, and all Ala, Ile(δ1), Leu, and Val methyl groups are assumed to be labeled. * In this highly symmetric protein, the accuracy of the assignment can be improved to ~56% (or ~65% when the stereospecificity is ignored) by PRE restraints from two spin labels. ** In this large system, only Ile(δ1), Leu, and Val methyl groups are assumed to be labeled, and the 4D NOESY spectrum is used to help the converge. The assignment is obtained from *J. Am. Chem. Soc.*, **2003**, 125 (45), 13868-13878.

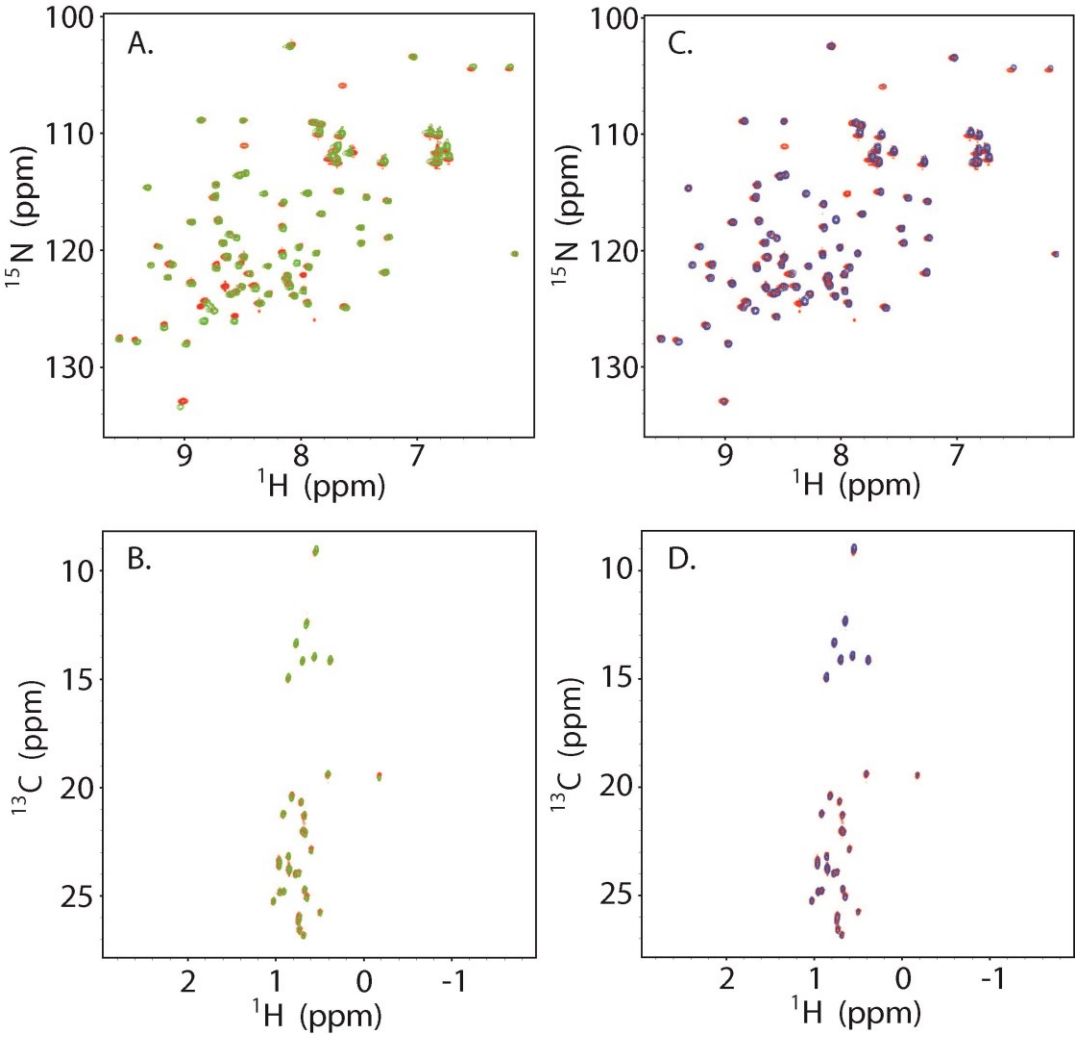


Figure S1. Effects of two mutations (K48C and G75C) on the amide and methyl fingerprints of ubiquitin. Panels **A)** and **B)** show the overlay of the amide HSQC spectra and the methyl-TROSY spectra for the mutant K48C (green) and the wild-type (red). Panels **C)** and **D)** represent the overlay of the amide HSQC spectra and the methyl-TROSY spectra for the mutant G75C (blue) and the wild-type (red).

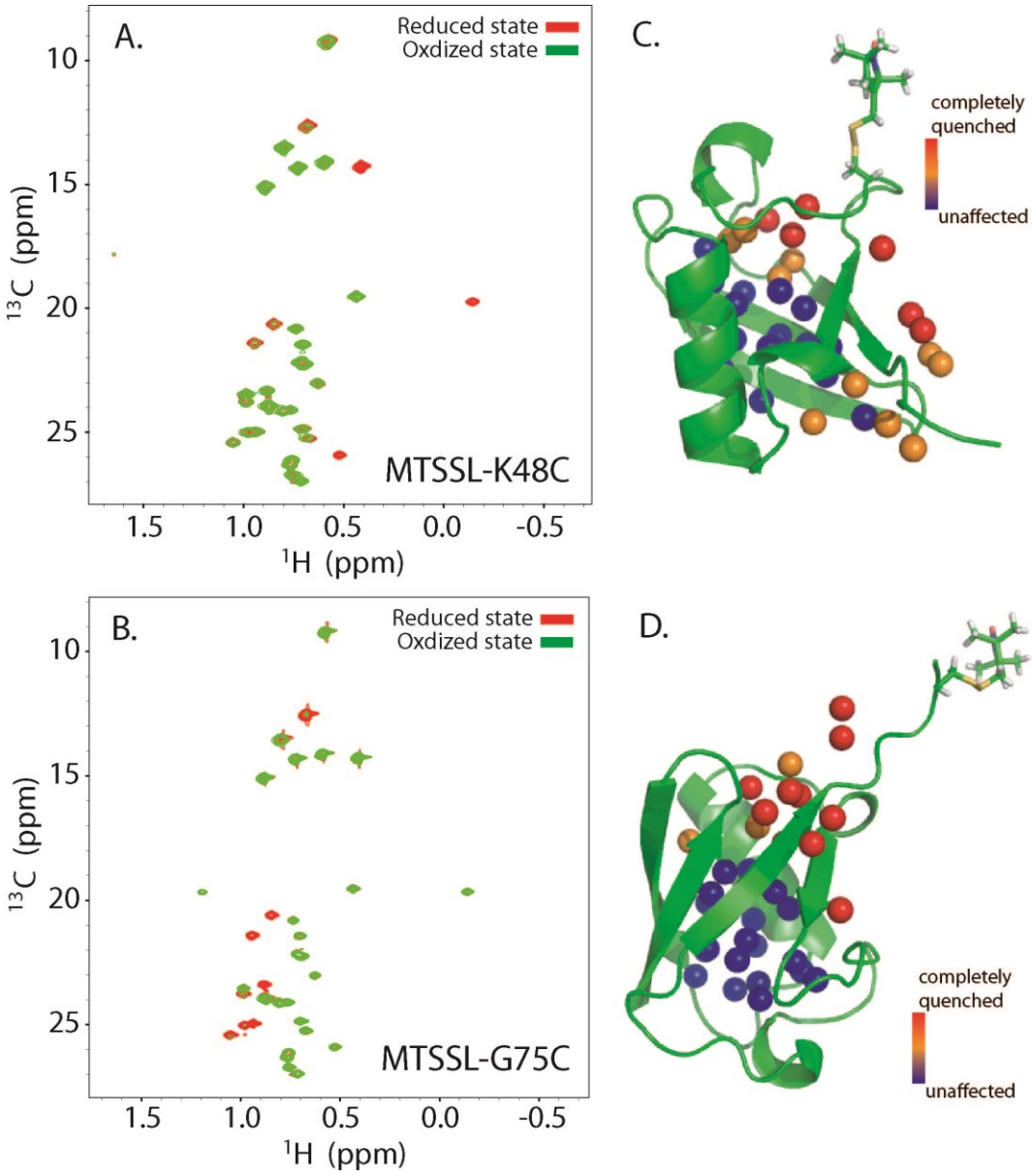


Figure S2. Effects of paramagnetic spin labels on the methyl fingerprint of ubiquitin. Panels **A)** and **B)** show the overlay of the methyl-TROSY spectra for the oxidized (green) and reduced state (red) of the MTSSL labeled ubiquitin. The two labels were placed at positions K48C and G75C, respectively. Mapping of the PRE effects for the two mutants on the X-ray structures of ubiquitin (Panel **C** and **D)**.

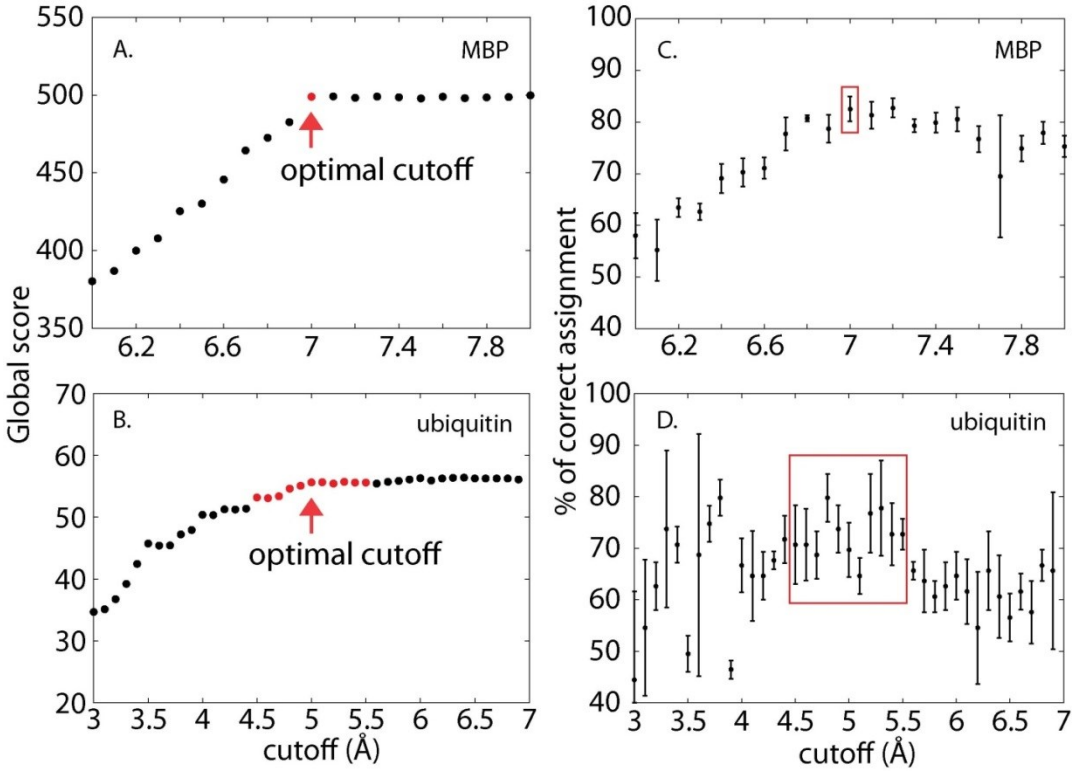


Figure S3. Comparison of the methyl assignments for MBP (synthetic data) and ubiquitin (experimental data) using different values of NOE distance cutoffs. The global score function levels off or increases slightly near the optimal cutoff (**A** and **B)**. The most accurate assignment is provided at the optimal cutoff value (panels **C** and **D**). Approximately, 30% of the methyl-methyl NOE data were used for both ubiquitin and MBP.

Chapter 4- Structure-based assignment of methyl groups by FLAMEnGO 2.0: Application to C-subunit of Protein Kinase A

The paper is in preparation.

Methyl-TROSY spectroscopy is the one of the most powerful NMR techniques developed for atomic resolution structural analysis of large molecular systems. One significant shortcoming of this approach is the assignment of the methyl resonances, which relies on COSY-type transfers from methyl group resonances to the protein backbone. Particularly, these experiments do not perform well for fast relaxing resonances, i.e. larger systems. On the other hand, the use of brute force via extensive mutagenesis of methyl-bearing amino acids is not practical and may result in small changes in the methyl fingerprint that make the assignment more cumbersome. When the X-ray structure is available, it is possible to use a combination of NOE experiments and computational methods to carry out site-specific assignments. Previously, we demonstrated that this approach can be improved by including Monte Carlo sampling and a fuzzy logic algorithm (i.e., Fuzzy Logic Assignment of Methyl Group or FLAMEnGO). Here, we present an improved version of FLAMEnGO and demonstrate its remarkable performance on the C-subunit of the cAMP-dependent protein kinase A (PKA-C). Using rather sparse and ambiguous NOESY, PRE, and TOCSY data we obtained a nearly complete assignment (~98%) of methyl groups with ~80% accuracy. This enhanced approach can be used standalone or to assist in resonance assignment when conventional through-bond experiments are available.

4.1 Introduction

Recent development of methyl-TROSY based methodology on selectively protonated methyl groups in a highly deuterated background has significantly expanded the molecular weight range in which quantitative NMR studies can be performed (1-3). However, such studies require chemical shift assignments to provide site-specific probes of structure, conformational dynamics, and molecular interactions. Previous assignment strategies have involved correlating the backbone carbon chemical shifts with methyl chemical shifts, either through a TOCSY or COSY type transfer scheme (4,5). For molecules too large for backbone assignments, a “divide and conquer” approach has been used in which the molecule is separated into smaller fragments that are amenable to routine assignment experiments (6). The remaining unassigned residues are then identified through any combination of analysis of NOE correlations and PRE with available crystal structures and site-specific mutational analysis. However, this approach is time consuming and it is not always possible to obtain backbone assignments in the very large systems that methyl-TROSY is optimized for. To address this, we proposed a new methodology based on a previous effort from the Matthews laboratory (7), incorporating both NOEs and PREs, and driven by a new algorithm based on Monte Carlo sampling and fuzzy logic, i.e., FLAMEnGO (8). This method does not require backbone assignments and utilizes a pre-existing structure (typically X-ray derived structure) and spatial correlations in the form of methyl-methyl NOEs as well as PRE mapping using site-directed spin labeling. An optimal NOE distance cutoff is determined, and the best-fit assignment is back-calculated from the input structure. Here, we report the performance of an improved version of our program FLAMEnGO 2.0 on the methyl group assignments of the C subunit of the cAMP-dependent protein kinase A (PKA- C). Additionally, we compare our results with previously established methodologies for methyl group assignments.

Protein Kinase A is a ubiquitous eukaryotic kinase that is responsible for signaling a variety of cellular processes. Our previous work showed that

conformational dynamics are prominent in this enzyme and are activated upon ligand binding and retained by substrate binding (9-11). Therefore, PKA-C represents a challenging system for classical through-bond correlation spectroscopy. Combining NOESY and TOCSY data with PRE restraints, the application of FLAMEnGO 2.0 resulted in 100% assignment of methyl groups of the Ile residues, 95% of the Leu and 100% of the Val residues. The main advantage of this new implementation is the generation of accurate assignments using sparse and ambiguous NMR data together with unambiguous information to assign the methyl groups of large systems.

4.2 Material and methods

Protein Expression and Sample Preparation. The expression and purification of PKA-C and RIIα(R213K) subunit were carried out as described previously (12,13). Briefly, cell pellets of catalytic subunit and RIIα(R213K) were combined and lysed in buffer A (30mM Mops, 200μM ATP, 15mM $MgCl_2$, 5mM 2-mercaptoethanol, pH 8). After removing cell debris by centrifugation, the supernatant was mixed with HIS-Select Ni^{2+} affinity gel (SIGMA, 1ml resin/1L culture) at 4 ^{0}C overnight. The flow-through after overnight incubation was collected and buffer B (buffer A containing 25mM KCl) was used to wash the Ni^{2+} resin. Finally, PKA-C was eluted using elution buffer A (buffer B containing 1mM cAMP), and the R subunit was eluted using elution buffer B (elution buffer A containing 250mM imidazole). The different phosphorylated isoforms of PKA-C were further separated by HiTrap SP column (GE) with a gradient from Buffer A (20 mM KH_2PO_4, pH 6.5) to 30% Buffer B (20 mM KH_2PO_4, 1.0 M KCl, pH 6.5) with a flow rate of 2.0 mL/min The ternary complex PKA-C/AMPPNP/PKI_{5-24} was prepared using a PKI_{5-24} to PKA-C molar ratio of 1.2, 12mM of AMPPNP in NMR buffer (20mM KH_2PO_4, 90mM KCl, 60mM $MgCl_2$, 10mM DTT, pH 6.5). The final concentration of PKA-C is about 0.5mM. Stock solution of 100mM Pefabloc is prepared in water, and the effective concentration (~0.5mM) is used in the final NMR sample.

NMR spectroscopy. All of the NMR experiments were carried out on Bruker Advance III spectrometers operating at 850 or 900 MHz ^{1}H frequency. The temperature was held constant at 27 ºC. For the residue-type assignment, WT PKA-C was expressed using selectively labeled M9 medium containing [^{15}N,^{13}C]-Leu or [^{15}N,^{13}C]- Val. The constant-time ^{13}C-^{1}H HMQC spectra were performed to provide residue-type assignments. For the PRE restraints, PKA-C mutants (K16C and I244C) were expressed in ^{15}N labeled M9 medium supplemented with 70mg/L of $^{13}C_{me}$ α-ketobutyrate and 90mg/L of $^{13}C_{me}$ α-ketoisovalerate. The ~200uM of samples were reacted with (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate, MTSSL, (10 times the protein concentration) and incubated at 4 ^{0}C for 3 hours. The MTSSL excess was removed by diafiltration for the spectra of the oxidized samples, and then the spectra of the reduced samples will be recorded after adding 10mM DTT into the samples. The intensity ratios of resonances from oxidized samples to those from reduced samples were used as PRE restraints in the calculations.

For the NOESY restraints, the wild-type PKA-C was expressed in ^{15}N labeled deuterated M9 medium (80% $^{2}H_2O$ and 100% ^{2}H-glucose) supplemented with 70mg/L of 2-ketobutyric acid-4-^{13}C,3,3-d_2 and 90mg/L of 2-keto-3-(methyl-d_3)-butyric acid-4-^{13}C. The final concentration of the enzyme was ~0.5 mM and experiments were performed at 27 ºC. The optimal mixing time (0.25 second) was chosen after analyzing the 2D NOESY spectra with different mixing times (0.1- 0.5 seconds). The 3D ^{13}C-^{13}C-^{1}H HMQC-NOESY-HMQC was collected with 32 scans using 2048 (proton), 80 (carbon), and 120 (carbon) complex points. For the TOCSY restraints, we used the same sample utilized for the NOESY experiments. The 3D TOCSY-HMQC experiment correlates the frequency of the methine proton with the frequencies of the methyl groups within the same residue. The experiment was carried out using 50 ms mixing time with 16 scans per FID. The total number of complex points were 2048 (proton), 90 (carbon), and 136 (proton). For the backbone triple resonance and HMCM(CG)CBCA

experiments, the sample was expressed in ^{15}N labeled deuterated M9 medium (80% $^{2}H_2O$ and 100% $^{2}H_7$,$^{13}C_6$-glucose supplemented with 70mg/L of 2-ketobutyric acid-$^{13}C_4$,3,3-d_2 and 90mg/L of 2-keto-3-methyl-d_3-3-d_1-^{13}C-butyric acid. The final concentration of the sample was ~0.8 mM. The TROSY-based HNCA and HN(CO)CA experiments were collected with 24 scans, 2048 (proton), 70 (nitrogen), and 80 (carbon) complex points and the TROSY-based HN(CA)CB and HN(CO)CACB experiments were collected with 40 scans, 2048 (proton), 70 (nitrogen) and 100 (carbon) complex points (14). The 3D methyl "out-and-back" HMCM(CG)CBCA experiment was performed with 32 scans, 2048 (proton), 80 (methyl carbon) and 100 (Cβ and Cα) complex points (4).

4.3 Theoretical basis of the algorithm

Previously we had developed a new algorithm, FLAMEnGO (8), for the assignment of methyl groups from sparse and ambiguous data. First a "dummy" assignment on the methyl-TROSY spectra is applied and the expected peak list of the 3D NOESY is generated from the expected distances given by a crystal structure with a given NOE cutoff distance. This simulated 3D NOESY peak list is compared to the experimental 3D NOESY peak list. The closeness between simulated and experimental peak lists is scored using a global scoring function:

$$G(x) = \max_{a \in A}\{Match_{total}\}$$

in which the maximum value of a matching function among all possible assignments "A" with a given NOE cutoff distance "x" is chosen as the global score. This maximum value is found by performing a Monte Carlo sampling scheme (15) on all possible assignment possibilities, selecting the one that provides the maximum global score value. The matching function is given as a summation of matching functions for various experimental restraints:

$$Match_{Total} = Match_{NOE} + Match_{CS} + Match_{PRE} + Match_{TOCSY}$$

in which each of the possible experimental constraints (NOE, chemical shift, PRE, and TOCSY) are scored for each NOE cutoff distance (x) and assignment

(a). As the details regarding most of the matching function had already been discussed, here we will discuss only the improvements and additions that have been implemented in FLAMEnGO 2.0.

This new version of the FLAMEnGO algorithm contains several improvements that are summarized in the following synopsis.

1. *Higher tolerance for the missing data.* Due to exchange broadening, some methyl group resonances may be missing in the methyl-TROSY spectrum. In addition, some PRE data cannot be included unambiguously in the calculation due to peak overlapping. While in the old version of the program these resonances lacking information are treated in the same way as the resonances with information, in this new implementation the missing resonances or those resonances without PRE information will be assigned arbitrarily to any methyl groups when other restraints are satisfied.

$$\mathrm{match}_{\mathrm{total}} = \sum_{\mathrm{m}\in\mathrm{A};\ \mathrm{r}\in\mathrm{B}} \mathrm{match}_{\mathrm{m,r}}$$

where A is the set of all observable resonances, and B is the set of all input restraints.

2. *Improved assignment of Val and Leu resonances by incorporation of TOCSY data.* TOCSY data are now used as additional restraints to limit the possible assignments sampled by the algorithm. These restraints are quite powerful since they are independent of structural models used in the calculation. For our TOCSY experiments (**Figure S1**), we utilized a sample that contains an additional methine proton in all Val and Leu residues in a highly deuterated background. Therefore, a pair of methyl resonances originating from the same residue will share the same methine resonance. This kind of restraint significantly reduces the sampling space. The function to be maximized for the TOCSY data is:

$$match_{TOCSY} = \sum_{(m_1,m_2)\in C} e^{-4ln2\cdot|f(m_1)-f(m_2)/LW|}$$

where C is the set of all pairs of methyl resonances assigned to the same residues, f(*m*) is the chemical shift of the methine proton correlated to a given methyl group, *m*, and LW is the linewidth of the methine proton shift.

3. *Improved version of fuzzy logic* (16). Due to experimental error, a given cross peak in the NOESY spectrum may have proton and carbon chemical shifts that differ slightly from those of the donor methyl groups. These slight differences in chemical shifts were affecting the confidence score in the previous version of the algorithm, even though these peaks were located in well-resolved regions of the spectrum and no other possibilities were associated with a specific pair. In general, the confidence score for a cross peak located in a well-resolved region of the methyl-TROSY should be higher than a peak located in crowded regions. To take this into account, we modified the matching function for the NOEs:

$$match_{NOE} = \sum_{p_1 \in E} \begin{cases} 1; if\ |\{p_2 \in S | D(p_1, p_2) \leq 1\}| = 1 \\ max_{p_2 \in S} e^{-4ln2 \cdot D^2(p_1, p_2)}; otherwise \end{cases}$$

where E is the peak positions of experimental data, S is the peak positions of simulated data, and

$$D(p_1, p_2) = \left[\left(\frac{(x_1 - x_2)}{LW_x} \right)^2 + \left(\frac{(y_1 - y_2)}{LW_y} \right)^2 + \left(\frac{(z_1 - z_2)}{LW_z} \right)^2 \right]^{\frac{1}{2}}$$

where $p_i = (\mathbf{x_i}, \mathbf{y_i}, \mathbf{z_i})$ and LW_j is the linewidth in the j dimension.

4.4 Results

For ease of use, a GUI interface of the FLAMEnGO was built to allow users to incorporate several different types of NMR restraints and to adjust parameters during the calculation (**Figure 1A**). In the beginning window, the essential input files contain arbitrarily assigned chemical shifts, predicted chemical shifts, structural coordinates, and the assignment swap file. The optional NMR restraints include NOE data (methyl-methyl or amide-methyl NOE data), PRE data (qualitative or quantitative data), TOCSY data (identifying spin systems),

and assigned amide shifts (only required for incorporation of amide-methyl NOE data). Moreover, the pre-assigned chemical shifts can be included in the calculation by fixing assignments of those methyl groups in the file of arbitrarily assigned chemical shifts. The important parameters, number of Monte Carlo sampling steps, the range and the interval for NOE distance cutoffs, and labeled amino acid types, may be adjusted in the setup tab. After starting the calculation, the program will plot a global score for each NOE distance cutoff (**Figure 1B**). In this auto-assignment algorithm, repeated calculations are required for two different purposes, 1) to refine the global score curve and 2) to provide the statistical analysis in the final assignment. In the first case, since the global score function is non-decreasing, an uneven curve will result from insufficient sampling, which will be a common case for large protein systems with huge sampling space. Sometimes, multiple runs of calculations are required to smooth the curve of the global scores. In the second case, once the user determine the optimal NOE distance cutoff by picking a data point on the global score curve, a small window will show up for the number of repeated calculations in order to generate the final probability-based assignment file (**Figure 1B**).

The new version of algorithm was first tested with simulated data from maltose binding protein (MBP). The purpose of this test was two-fold. The first was to demonstrate the accuracy of the new algorithm with sparse and ambiguous NOE data, as well as the way in which additional experimental restraints improve accuracy with regard to the previous algorithm. The second part of this test was to observe whether the new algorithm would tolerate these structural deviations. In this test the crystal structure of MBP (1DMB) was used to generate the simulated experimental constraints. The 3D NOESY peak list was simulated using the crystal structure (1DMB), chemical shift assignment (BMRB accession #7114), and a given distance cutoff (7Å). Sparse NOE data was generated by randomly removing 70% of the cross peaks. 3D TOCSY data was created using the same assignment deposited in the BMRB (7114). The effects of spin labels at positions S145C and S306C were back-calculated assuming that methyl groups

within 15Å of the spin label were completely quenched, those between 15 and 35Å were partially quenched, and the rest were not affected. Different combinations of data sets were input into versions 1.0 and 2.0 of FLAMEnGO and the results were compared. In general, we found that the new implementations allow the algorithm to converge faster than the old version during the Monte Carlo sampling (**Figure S2A**). To observe how well FLAMEnGO 2.0 would tolerate structural deviations, different conformers from the solution NMR structure of MBP (2H25) were used as the input structures for FLAMEnGO, and the synthetic NMR data from the crystal structure was used as the simulated experimental restraints. It was shown that the structural deviations deteriorate the accuracy of the best-fit assignment with just NOE restraints; however, the decrease in accuracy can be compensated for by the inclusion of other structural data, such as PRE and TOCSY restraints. (**Figure S2B**)

Once we were satisfied with the results from the test, the assignment of the PKA-C methyl groups proceeded. We first added the proton atoms to the crystal structure (1ATP) using CHARMM19 software package. To predict the chemical shifts of the methyl groups in the methyl-TROSY spectra, we fed the structure into the prediction program CH3Shift (17). As some deviation from the experimental and the predicted methyl chemical shifts is expected, the predicted chemical shifts were referenced to the experimental shifts. The referencing choice was determined by optimizing the chemical shift matching function as previously described. We show that no referencing of predicted shifts was required (**Figure 2A**). Residue-type assignments were performed by selective amino acid labeling and identification with constant time HMQC.

Three sets of experimental restraints, PRE, NOESY, and TOCSY, were utilized in the calculation of the optimal assignment. Long-range distance PRE restraints were obtained by spin labeling the two PKA-C mutants, K16C and I244C, with MTSSL. The intensity ratios of resonances from oxidized samples to those from reduced samples provide distance restraints of methyl groups with respect to the positions of two spins labels. The average position of the two spin labels in the

crystal structure was calculated using XPLOR-NIH (18). Short-range distance restraints were obtained from the 3D HMQC-NOESY-HMQC data on methyl groups. To restrain two methyl groups to the same residue for Val and Leu, a 3D TOCSY-HMQC experiment was performed. By restraining pairs of methyl resonances originating from the same residues, the sampling space can be significantly reduced. Finally two assignments given by mutagenesis, I244 and I315, were fixed during the calculation.

Assignment with FLAMEnGO was performed initially by considering each experimental restraint separately, evaluating their performance, and then combining each together to increase the accuracy of the assignment. Considering PRE data alone, it was demonstrated that additional restraints would be necessary to generate any reasonable confident assignment (**Figure 2B**). When only NOE restraints are considered, there is a drastic improvement in the accuracy of the assignment; however, the assignment does not independently satisfy the experimental PRE restraints (92% percent of the assignment). Once PRE and NOE data are combined in FLAMEnGO, the overall precision slightly decreases (**Figure 2B**), but most of the experimental PRE restraints are well satisfied (99% of the assignment). The incorporation of TOCSY restraints with PRE and NOE data significantly boosts the precision of the assignment. (**Figure 2C**) To calculate the final statistics and assignment, 100 randomized calculations with all the experimental data were performed at the optimized NOE cutoff value (10.2 Å). The ten sets of assignments with highest scores were used to determine the final assignment. The result shows that 92% of methyl groups are assigned and 80% of the assignments have above 90% confidence.

Site-directed mutagenesis on V191 and L82C was performed to assess the overall accuracy of the assignment. Both methyl groups of V191 are assigned with 100% confidence, but both methyl groups of L82 are unassigned. The ^{13}C HMQC spectrum of V191C is completely consistent with the assignment (**Figure S3A**), but the identification of L82 in the L82C mutant was ambiguous (**Figure S3B**) without other data sets, such as TOCSY data and residue-type

assignments. The result shows that the two resonances from L82 were mis-assigned to L27, and the mis-assignments had a 70% confidence level. However, upon closer inspection this mis-assignment was not surprising given the lack of direct experimental restraints for L27. In order to test the effect of the correct L82 assignment on the new calculation and the resulting assignment, the same calculations were performed after fixing V191 and L82. The resulting assignment has no change for those assignments with a high confidence level (>90%), suggesting that the L82 from mutagenesis is consistent with the experimental data. Including all the above assignments from mutagenesis, 100% of methyl groups in ILE residues, 100% of methyl groups in VAL residues, and 95% of methyl groups in LEU residues are assigned (**Figure 3**) using FLAMEnGO.

To establish the quality and accuracy of the assignment of PKA-C from FLAMEnGO, comparison with previously established methodology was performed. To compare with the established “through-bond” approach the methyl “out-and-back”, HNCA, HN(CO)CA, HN(CO)CACB, and HNCACB (4,14) (**Figure 3**) were performed and methyl cross peaks were compared to the amide backbone assignment In total, two mis-assignments (10%) are found in the methyl groups of ILE residues, seven mis-assignments (18%) are identified in the methyl groups of VAL residues, and ten mis-assignments (17%) are determined in the methyl groups of LEU (**Table 1**). After comparison, the average accuracy of the FLAMEnGO was determined to be above 80%. To understand the source of these mis-assignments, the structural positions of the residues were investigated. We observed that many mis-assigned methyl groups, such as V226 and V255, are found to be spatially very close, and the assignments are simply swapped. In fact, about 73% of the total mis-assignments follow this behavior (**Figure 5**). Other mis-assignments are primarily located at the surface of the protein, likely due to the scarcity of NOE restraints. In general, the accuracy of the assignments obtained by the new algorithm is comparable with that of assignments reported previously.

4.5 Discussion

The fundamental difference between FLAMEnGO and previous work is the approach by which the simulated and experimental NOE is considered. In FLAMEnGO, an arbitrary assignment is first placed on the methyl-TROSY spectra and, using the crystal structure as a guide, a 3D NOESY spectrum is simulated for a given NOE distance (the cutoff value). This simulated data is compared to the experimental data and the two are matched together using the matching function for the entire assignment set. This process is repeated for different possible assignments by Monte Carlo sampling (15) and the one with the highest matching score is selected as the global scoring value for a particular NOE cutoff value. In contrast, the algorithm proposed by Matthews and co-workers (7) dictates that the experimental NOE contacts determined by the experimental spectrum are used to match the NOE contacts extracted from a given crystal structure. The process is done with the empirical score functions and in much restricted sampling space. This critical difference explains why even the newest version of MAP-XS (19) requires over 50% of the theoretical NOE crosspeaks to be present in the experimental data, while FLAMEnGO can proceed with assignment with much sparser data but require more computational power.

The new version of FLAMEnGO provides significant improvements over the previous implementation: better handling of missing resonances, incorporation of TOCSY based restraints, and significantly better fuzzy logic algorithm. Incorporation of PRE restraints was structured so the quenching pattern rather than the absence of resonances was considered for the calculation. This implementation considers that often many resonances are not present in the methyl-TROSY spectra due to exchange broadening. Any missing resonances are excluded from the calculation, and are rather randomly assigned to methyl groups left out. Improvement in the assignment of Leu and Val residues was performed by utilizing TOCSY based restraints. By measuring the correlation of

the Leu/Val methyl groups to the shared methine proton, resonances from the same residue could be linked together during the calculation. Finally, the new version of fuzzy logic is used to consider the local resolution of the methyl-TROSY spectrum when assigning cross peaks to donor resonances. Higher confidence will be obtained when the cross peaks are assigned to the donor resonances in the well-resolved region of the methyl-TROSY spectrum. The above improvement is apparent on our application on PKA-C, which contains experimental errors typical of difficult systems. Initial tests demonstrated that the combination of all experimental restraints with FLAMEnGO 2.0 marked a significant improvement in the assignment accuracy.

To demonstrate the efficacy of the program, FLAMEnGO 2.0 was applied to the catalytic subunit of Protein Kinase A. This ubiquitous 350 residue eukaryotic kinase is responsible for regulation of a myriad of cellular processes and previous work in our laboratory had established that significant conformation dynamics are present in this enzyme, proving to be a difficult system for study with NMR. Hence, this system would be a very good experimental test of the robustness of FLAMEnGO 2.0. For the assignment of PKA-C, a combination of 3D NOESY, PRE, and TOCSY restraints were utilized, each with their own experimental challenges we address here. In the case of assigning methyl groups in PKA-C, the NOESY experiment was the least sensitive, which may be due to different dynamics and relaxation rates of methyl groups. Only 26% of the predicted NOE cross peaks were measured with 7 days of experimental time with a mixing time of 250ms. However, as compared with PRE and TOCSY data, NOESY data are essential to resolve the nearby methyl groups in the protein structure. The optimized NOE cutoff distance for the global score function was determined to be 10.2 Å, which would be unlikely reflective on the experimental distance. A possible explanation is the existence of a deviation between the crystal structure and the structure in the solution-state, which is to be expected. However tests show that FLAMEnGO is well capable for compensating for such deviations. For long-distance restraints, qualitative PRE data determined by

intensity ratios of two methyl-TROSY spectra with reduced and oxidized samples proved to be the most sensitive of the experiments. The only drawback is the poor stability and integrity of the spin-labeled mutant proteins. Among all 7 mutants (K16C, K81C, L82C, V191C, F239C, I244C, and I315C) engineered for PRE data, only two mutant proteins (K16C and I244C) generated useful restraints, but several mutant proteins (L82C, V191C, I244C, and I315C) ultimately provided assignments through site-directed mutagenesis. The total PRE restraints contain 78% of the simulated data if assuming no peak overlapping ambiguity. Finally, the TOCSY experiment links two methyl groups in the same Leu and Val residue to their shared methine proton and is shown to be more sensitive than the NOESY experiment, which is probably due to the lack of multiple relays in the magnetization transfer. Although the sample used for the NOESY experiment is not optimal for methyl-TROSY (the presence of a neighboring methine proton induces extra dipolar relaxation and spin flips of the methyl groups), the sample could generate additional TOCSY data which proved to significantly improve assignment accurary. 97% of the expected TOCSY cross peaks were measured from 4 days of experimental time.

The assignment from FLAMEnGO was compared to the results from the “through-bond” approach. When possible mis-assignments were identified and then analyzed. Mistaken assignments were mapped onto the crystal structure (Fig. 4) and it was found that assignments from residues spatially very close often were swapped. This is unsurprising since the overall NOE data is very sparse, 26% of the total expected NOE contacts. Moreover, in the current assignments, there are methyl groups from a few pairs of LEU residues assigned on only half the total simulations performed. Even if they are either confirmed or not identified as wrong assignments by the through-bond experiments, the program lacks confidence to assign them to a specific residue within a given pair of residues. This may be ameliorated by a site-directed mutagenesis strategy which was recently published in which a conservative mutation is performed on the unknown residue. Other mis-assignments were found on the surface of the

protein, most likely due to the lack of richness of NOE crosspeaks that would be found for surface residues. Overall the sparseness of the NOE data can cause mild spatial ambiguity in methyl assignments.

As compared to the “through-bond” approach the experimental data used in the calculation are much more sparse and ambiguous but we demonstrate that it provides assignments of comparable accuracy. Previously work using the “through-bond” approach for large proteins was unable to obtain complete assignment despite the richness in data and was often complimented with “through-space” restraints such as NOESY, PRE and PCS for near completion. The sparseness of the data is reflected in the total number and time required for the experiments. The total experimental time for FLAMEnGO was only about 12 days, but all experiments required for the through-bond approach required at least 26 days. FLAMEnGO can also be combined with restraints from other experiments (site-directed mutagenesis and “through-bond” restraints) to help increase the accuracy of the algorithm. Despite the richness of the “through-bond” approach, it is by no means error-proof due to the high degeneracy of chemical shifts and the requirement of successful amide detected triple resonance experiments for Cα and Cβ chemical shifts, both of which become increasingly uncertain as the complexity of the system increases. This common challenge was evident with PKA-C. Despite the high quality of data from the methyl “out and back” experiment, not all the residues were observed due to broadening of the Cα and Cβ on either the amide or methyl detected experiments. Incomplete back exchange of amide protons in our system further complicated the observation of Cα and Cβ chemical shifts. On our application to PKA-C we show that FLAMEnGO is able to generate a robust set of methyl assignment that contains these difficulties, bypass them and is complimentary to the “through-bond” approach.

The greatest advantage of FLAMEnGO lies within the fact that an assignment can be made by utilizing extremely sparse ambiguous data. We anticipate that FLAMEnGO will be able to provide confident methyl group assignments when the

traditional approaches fail, further expanding the capabilities of methyl-TROSY spectroscopy for large molecular systems.

4.6 Conclusion

When the structure is available, this algorithm provides an alternative way to assign methyl groups with sparse and ambiguous data, and can be used to complete partial assignments provided by other approaches, and can validate the consistency between the structure and experimental restraints. Especially, when methyl out-and-back and triple resonance experiments fail in some large and dynamic systems, this through-space approach will be a more practical choice. In this demonstration, using only 26% NOE, 78% PRE, and 98% TOCSY data, this algorithm can provide near complete methyl assignments (98%) of PKA-C with above 80% accuracy in general. Above half of mis-assignments are simply the swaps of methyl groups with nearby methyl groups. Thus, in terms of relative positions of methyl groups, the assignments in the PKA-C structure are very accurate. Further improvements in the accuracy can be achieved by including either more restraints or more pre-assignments from mutagenesis. Finally, when through-bond experiments are available, the methyl assignments acquired by this algorithm can serve as seed assignments to assist the backbone assignments in the larger proteins.

4.7 References

1. Religa TL, Sprangers R, Kay LE. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. Science. 2010 Apr 2;328(5974):98-102.

2. Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature. 2007 Feb 8;445(7128):618-22.

3. Tugarinov V, Kay LE. Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. Chembiochem. 2005 Sep;6(9):1567-77.

4. Tugarinov V, Kay LE. Ile, leu, and val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc. 2003 11/01;125(45):13868-78.

5. Yang D, Zheng Y, Liu D, Wyss DF. Sequence-specific assignments of methyl groups in high-molecular weight proteins. J Am Chem Soc. 2004 Mar 31;126(12):3710-1.

6. Velyvis A, Schachman HK, Kay LE. Assignment of ile, leu, and val methyl correlations in supra-molecular systems: An application to aspartate transcarbamoylase. J Am Chem Soc. 2009 Nov 18;131(45):16534-43.

7. Xu Y, Liu M, Simpson PJ, Isaacson R, Cota E, Marchant J, et al. Automated assignment in selectively methyl-labeled proteins. J Am Chem Soc. 2009 Jul 15;131(27):9480-1.

8. Chao F, Shi L, Masterson LR, Veglia G. FLAMEnGO: A fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. Journal of Magnetic Resonance. 2012 1;214(0):103-10.

9. Masterson LR, Cheng C, Yu T, Tonelli M, Kornev A, Taylor SS, et al. Dynamics connect substrate recognition to catalysis in protein kinase A. Nat Chem Biol. 2010 Nov;6(11):821-8.

10. Masterson LR, Shi L, Metcalfe E, Gao J, Taylor SS, Veglia G. Dynamically committed, uncommitted, and quenched states encoded in protein kinase A

revealed by NMR spectroscopy. Proc Natl Acad Sci U S A. 2011 Apr 26;108(17):6969-74.
11. Masterson LR, Cembran A, Shi L, Veglia G. Allostery and binding cooperativity of the catalytic subunit of protein kinase A by NMR spectroscopy and molecular dynamics simulations. Adv Protein Chem Struct Biol. 2012;87:363-89.
12. Bastidas AC, Deal MS, Steichen JM, Keshwani MM, Guo Y, Taylor SS. Role of N-terminal myristylation in the structure and regulation of cAMP-dependent protein kinase. J Mol Biol. 2012 Sep 14;422(2):215-29.
13. Hemmer W, McGlone M, Taylor SS. Recombinant strategies for rapid purification of catalytic subunits of cAMP-dependent protein kinase. Anal Biochem. 1997 Feb 15;245(2):115-22.
14. Muhandiram DR, Kay LE. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. Journal of Magnetic Resonance, Series B. 1994 3;103(3):203-16.
15. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. J Chem Phys. 1953;21(6):1087-92.
16. Dubois D, Prade H. An introduction to fuzzy systems. Clinica Chimica Acta. 1998 2/9;270(1):3-29.
17. Sahakyan AB, Vranken WF, Cavalli A, Vendruscolo M. Structure-based prediction of methyl chemical shifts in proteins. J Biomol NMR. 2011 Aug;50(4):331-46.
18. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The xplor-NIH NMR molecular structure determination package. J Magn Reson. 2003 Jan;160(1):65-73.
19. Xu Y, Matthews S. MAP-XSII: An improved program for the automatic assignment of methyl resonances in large proteins. J Biomol NMR. 2013 Feb;55(2):179-87.

		>90% confidence			<90% confidence		
	Assigned methyl groups	Confirmed	Unconfirmed	Wrong	Confirmed	Unconfirmed	Wrong
ILE (21)	100%	90%	0%	10%	0%	0%	0%
LEU (64)	95%	31%	25%	7%	13%	15%	10%
VAL (40)	100%	68%	5%	18%	8%	1%	0%

Table 1. The statistics of confirmation of the final assignment using conventional through-bond experiments.

100% of ILE methyl groups, 95% of LEU methyl groups, and 100% of VAL methyl groups are assigned. The assignments are separated into high confidence (>90%) and low confidence (<90%). Using methyl out-and-back and triple resonance experiments, methyl assignments will be confirmed if sequential walk can be done. If the through-bond experiments don't support the methyl assignment, it will be identified as a wrong assignment. However, if no information is available or the data are not clear, the methyl assignment will be unconfirmed.

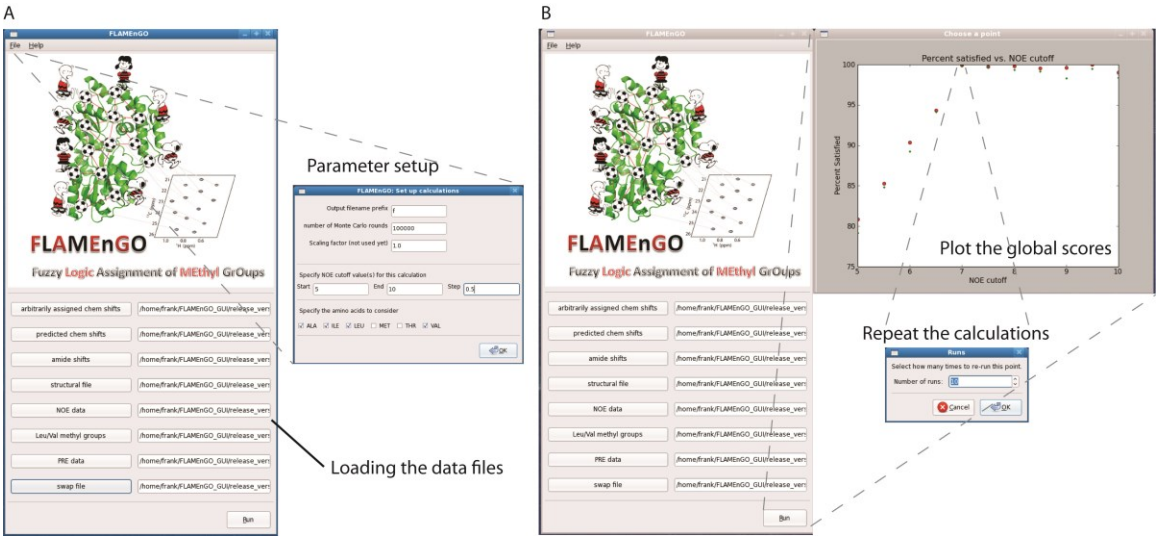


Figure 1. The GUI interface of the new version of FLAMEnGO.

A) The interface allows the user to load various NMR restraints and to input different parameters for calculations. The setup tab allows one to optimize the parameters for the overall calculation

B) The program will plot the global score for each NOE distance cutoff, and the plot can be refined by multiple runs of calculations. The optimal NOE distance cutoff can be chosen for repeated calculations, which can then generate probability-based assignments.

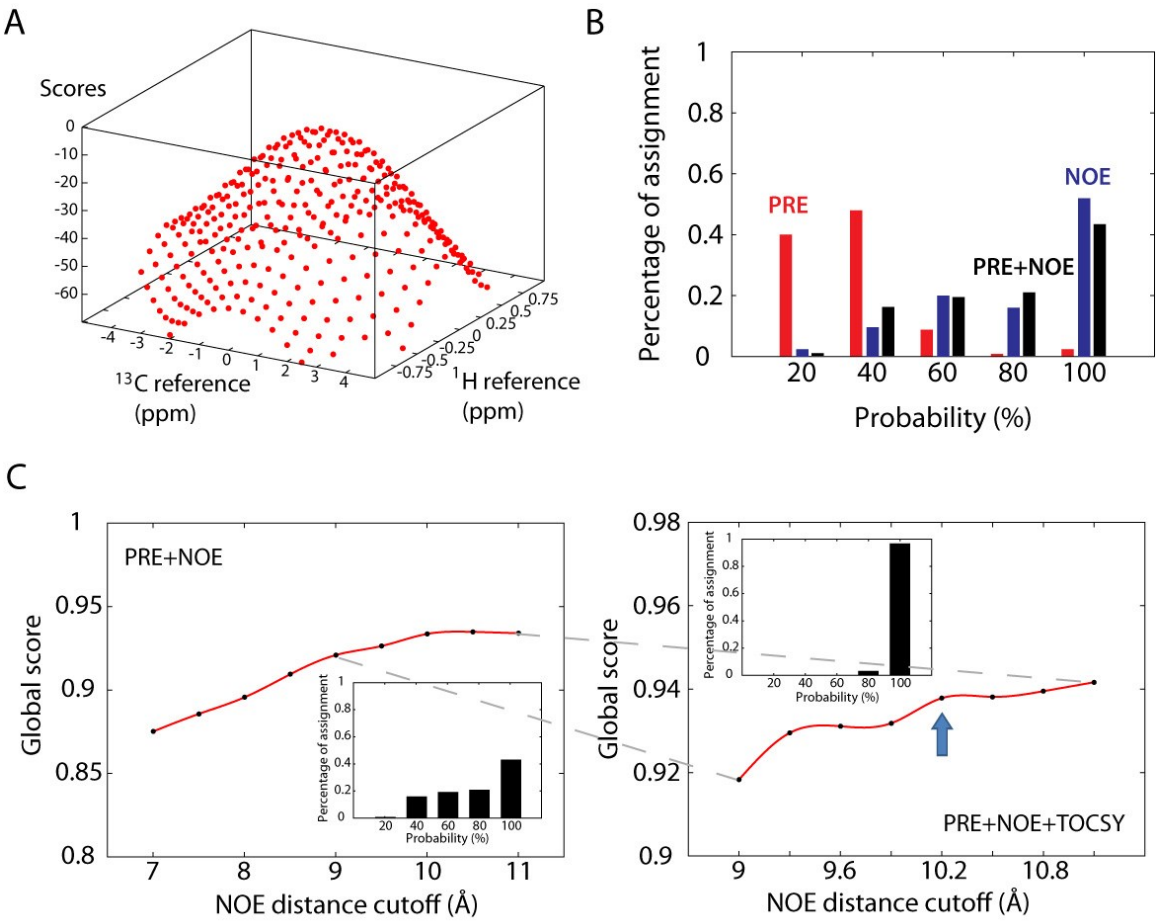


Figure 2. Application of the new version of FLAMEnGO to the catalytic subunit of protein kinase A.

A) The surface of scores by matching experimental chemical shifts with predicted chemical shifts with different reference shifts is presented. No experimental restraint is used.

B) Different combinations of experimental restraints are used for calculations. Each data set is fed into five different calculations, and then the probability assignments are determined. The result from PRE data is marked by red color, the result from NOE data is marked by blue color, and the final result from both PRE and NOE data is marked by black color.

C) The optimal NOE distance cutoff is determined by arraying different NOE distance cutoffs. In the calculations using PRE and NOE data with a rough

interval (left), the optimal cutoff is 10Å and the statistics of assignment precision is shown in the insert. In the calculations using all experimental data (PRE, NOE, and TOCSY) with a fine interval (right), the optimal cutoff is 10.2Å and the statistics of assignment precision is shown in the insert.

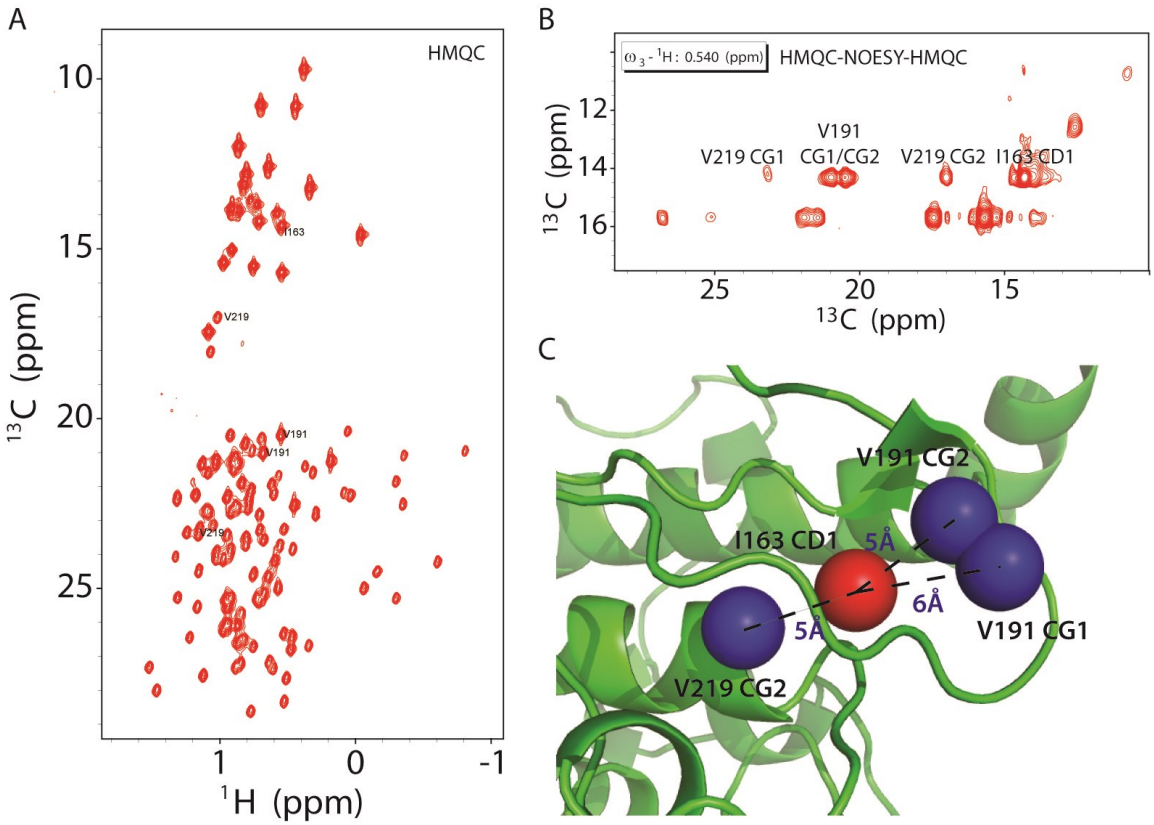


Figure 3. Representation of methyl assignments of PKA-C determined by 10 assignments with highest global scores from 100 independent calculations.

A) The methyl-TROSY spectrum of PKA-C is shown with partial resonances labeled with assignments.

B) A slice of 3D HMQC-NOESY-HMQC spectrum of PKA-C shows assigned cross peaks of the assigned I163 methyl group (δ1).

C) The assigned cross peaks of the I163 methyl group (δ1) correspond to the short-distance restraints, which are presented in the crystal structure (1ATP).

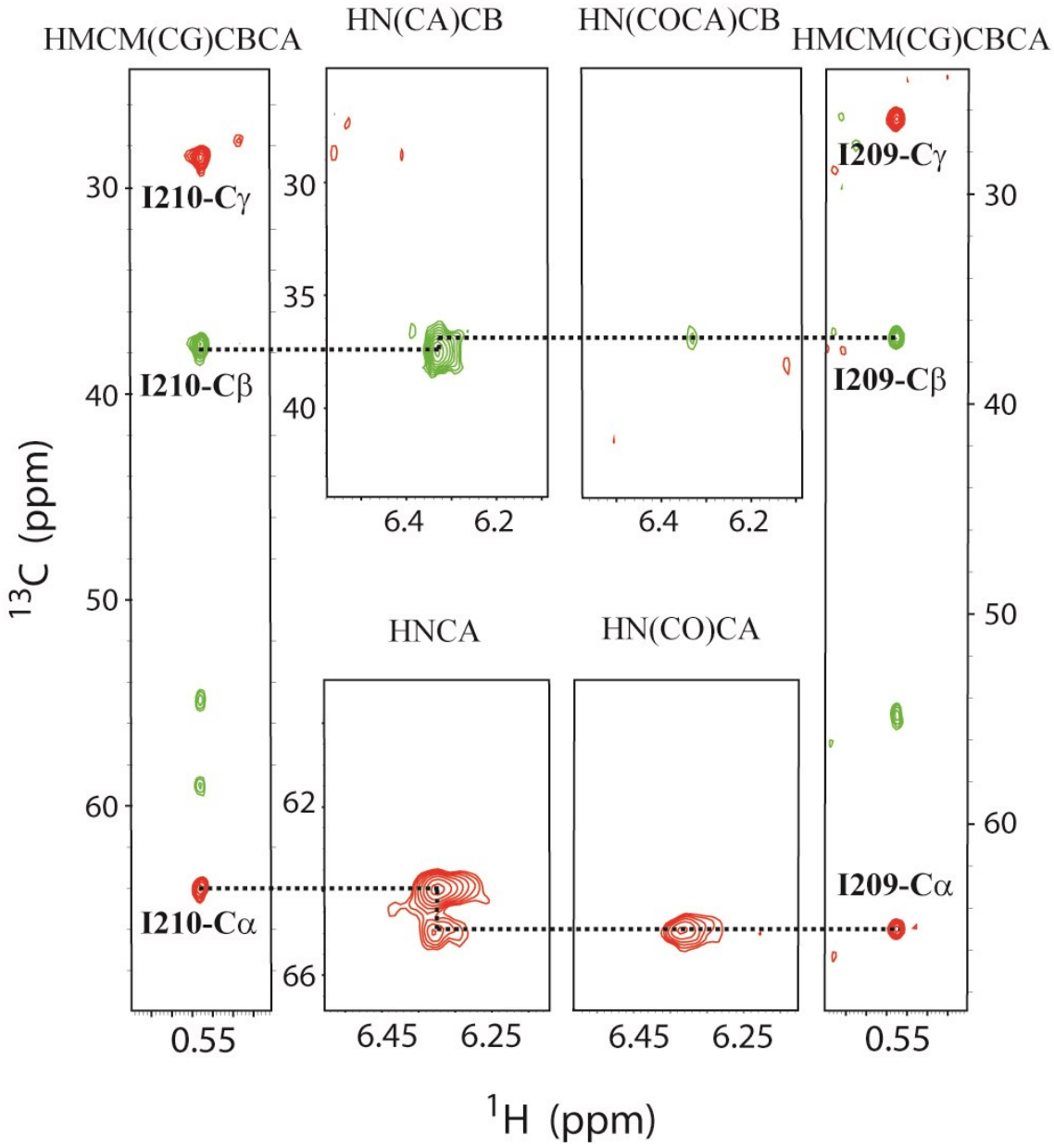


Figure 4. Methyl assignments of I210 and I209 from FLAMEnGO confirmed by sequential walks using triple resonance and methyl out-and-back experiments.

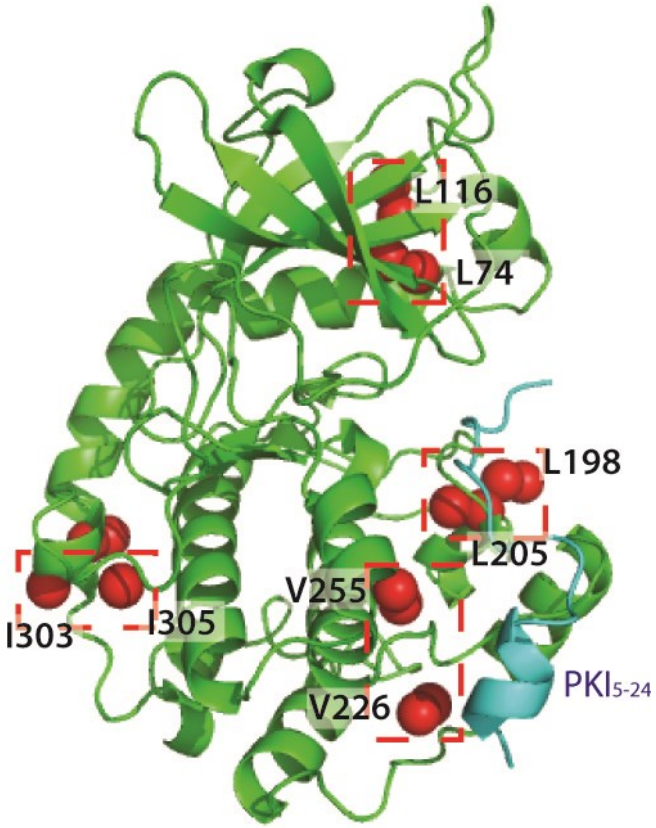


Figure 5. Distribution of mis-assigned methyl groups in PKA-C (1ATP).

The mis-assigned methyl groups are marked as red balls, and the ones in the red circle are the ones mis-assigned to the nearby methyl groups. The light blue peptide is the PKI peptide (5~24).

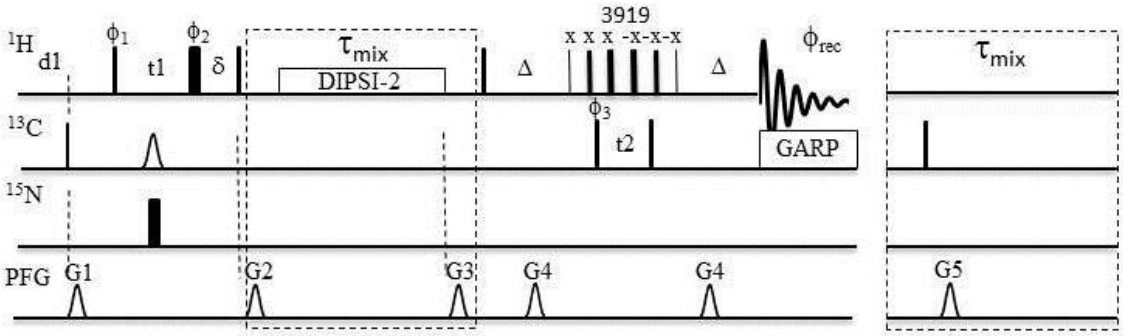


Figure S1. Pulse sequences of 3D TOCSY-HMQC and NOESY-HMQC (the part in left dashed box is replaced by that in right one) with 3919 for water suppression.

The narrow and wide bars represent 90° and 180° hard pulses, respectively. The ^{13}C shaped pulse is 500 μs long smoothed chirp adiabatic pulse of 60 KHz inversion bandwidth. δ = 500 μs, Δ = 3.7 ms, the separation of pulses in 3919 is 80 μs, and the mixing time τ are 70 ms and 150 ms for TOCSY-HMQC and NOESY-HMQC, respectively. ϕ_1 = (x x –x –x), ϕ_2 = (x x x x –x –x –x –x), ϕ_3 = (x –x), ϕ_{rec} = (x –x –x x). Quadrature detections in t1 and t2 dimensions are achieved by increase ϕ_1 and ϕ_3 by 90°, respectively. PFGs are G1 = (1 ms, 20%); G2 = (1 ms, 15%); G3 = (1 ms, 30%), G4 = (1 ms, 60%), G5 = (3 ms, 30%).

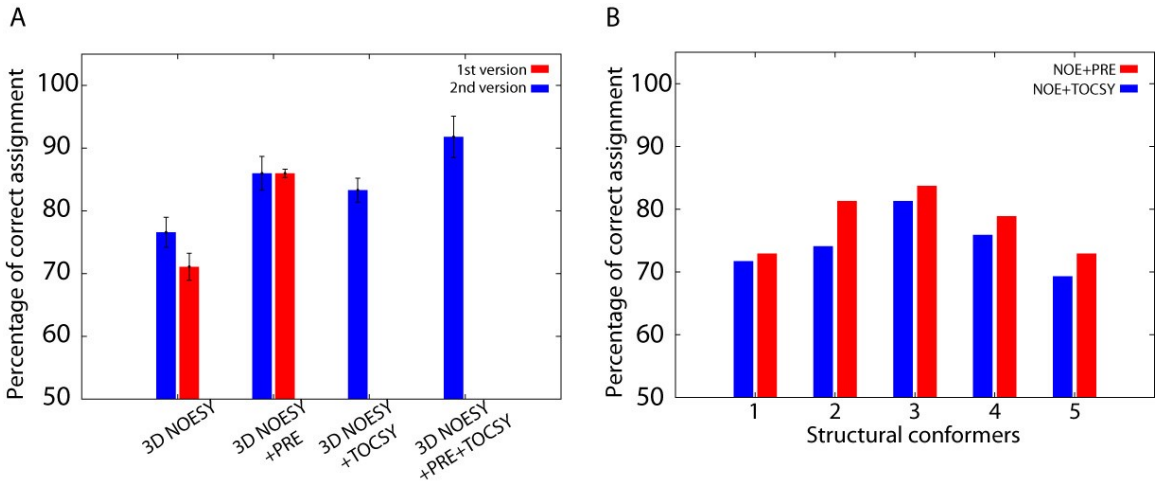


Figure S2. Tests of the new version of FLAMEnGO with simulated data.

The data are simulated from the crystal structure of maltose binding protein (1DMB), BMRB assignments (7114), and the solution structural ensemble (2H25). The simulation procedures are described in the previous report and in the text.

A) The accuracy of calculated assignments from different version of algorithms and from different data sets is determined by 5 independent calculations. The blue bars represent the results from the newest version of FLAMEnGO, while the red ones represent the results from the original version.

B) The calculation of methyl assignments is done using the newest version of FLAMEnGO and different structural conformers in 2H25 as input structures. The blue bars represent the results from NOE and TOCSY data, while the red ones represent the results from NOE and PRE data.

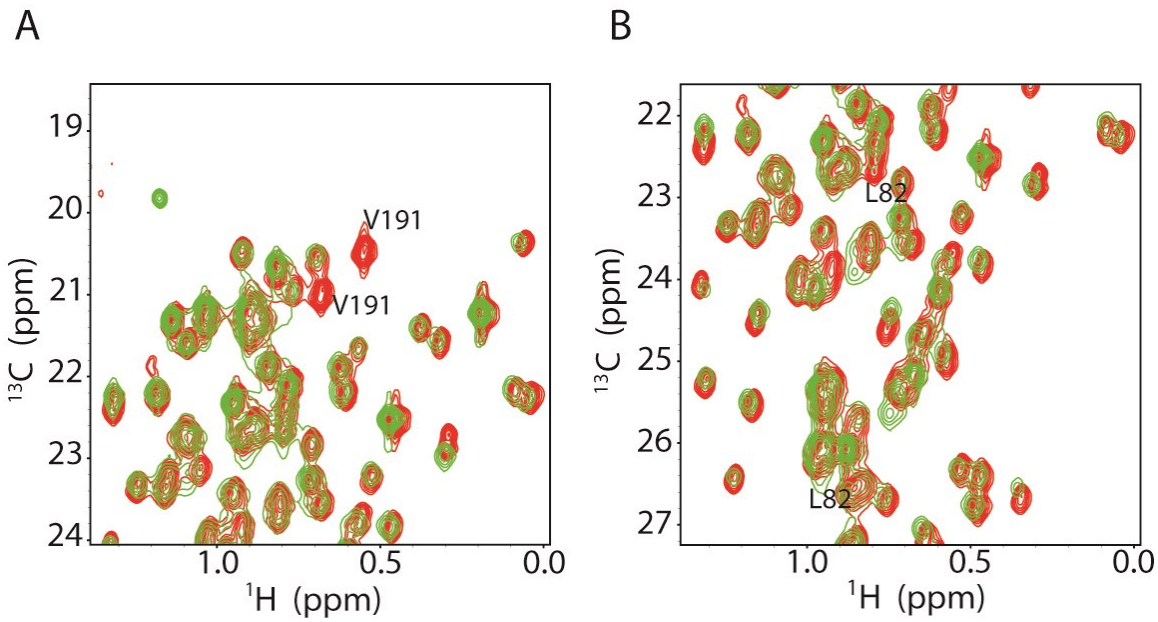


Figure S3. Validation of methyl assignments of FLAMEnGO by mutagenesis.

A) The methyl-TROSY spectrum of the wild-type PKA-C (red) is overlaid with the one of the mutant V191C (green). Two missing methyl resonances from V191 are labeled in the overlaid spectra.

B) The methyl-TROSY spectrum of the wild-type PKA-C (red) is overlaid with the one of the mutant L82C (green). Two missing methyl resonances from L82 are labeled in the overlaid spectra.

Chapter 5- Tracing intramolecular allostery through hydrophobic spines in protein kinase A

The paper is in preparation.

Nuclear magnetic resonance (NMR) relaxation data of the methyl groups in protein kinase A (PKA) provide side-chain information as to how PKA transmits allosteric signals in the molecule. Here, we report the experimental data that allows us to trace allostery in PKA and supporting the previously proposed hydrophobic spines in this kinase family. Our NMR relaxation data detect the changes in conformational dynamics upon ligand binding, which shows that the significantly quenched side-chain dynamics occur on the catalytic spine (C spine) and regulatory spine (R spine) on the fast and slow time scales respectively. The results also demonstrate our abilities to identify allostery in this highly regulated biomolecule, as our data strongly suggest that the alteration of conformational dynamics is the major mechanism for the allosteric effects of inhibitor binding.

5.1 Introduction

Protein phosphorylation is one of the most important post-translational modifications in cells, and plays important roles in signal transduction and cancer development in various tissues (1,2). The catalytic subunit of PKA (PKA-C) was the first protein to have its structure determined by X-ray crystallography in the entire protein kinase family (3). However, in order to fully characterize the regulation of PKA-C activity and its interaction with various substrates, not only structural data, but also information regarding conformational dynamics is required to complete a full picture of the thermodynamic process (4). The changes in the average structures are related to the enthalpic term in the thermodynamic process due to the changes in the number of hydrogen bonds, in torsion angles, in salt bridges etc. On the other hand, the changes in conformational dynamics are related to the entropic term through the changes in the conformational space sampled by the protein. The crystal structures of PKA-C complexes bound to different substrates have been previously determined. The only structural difference between these different complexes is the angle between the large and the small lobes (5). However, the above structural information is not enough to elucidate the allosteric regulation of PKA-C.

The conformational dynamics of PKA complexes with various peptides have been extensively studied by solution-state NMR, and a strong correlation between backbone conformational dynamics and its functions has been identified (5-7). When PKA-C binds to the nucleotide substrate to form a binary complex, both fast and slow conformational dynamics are tuned to allow the binding of a second substrate, such as phospholamban (PLN) (6). In fact, the incorporation of the second substrate into the binary complex to form a ternary complex only re-distributes the overall conformational dynamics without quenching them significantly. On the other hand, the binding of the protein kinase inhibitor (PKI) peptide will quench conformational dynamics and inhibit the enzymatic activity of

PKA-C (5,7). All the above microscopic observations are consistent with the more macroscopic data obtained by isothermal titration calorimetry (ITC), as well as changes in melting temperatures (5,7). For example: the loss of conformational dynamics upon binding of the PKI_{5-24} peptide matches the decrease in entropy during the binding process identified by ITC. However, no clear allosteric pathway can be identified through the backbone dynamics of the protein.

In previous studies, two conserved hydrophobic patches were identified by analyzing structural homology among different proteins within the kinase family. There are two hydrophobic spines, R spine and C spine; the integrity of R spine (L106, L95, F185, and Y164) is required for the activity of the enzyme, and the C spine (A70, V57, L173, L172, L227, I174, M128, and M231) is required for orientating the nucleotide substrate (8-10). Both spines are anchored to the αF-helix, a hub which positions several important catalytic and regulatory elements. The αF-helix also anchors the catalytic loop, which can position the D166 and K168 residues involved in phosphoryl transfer (9). Therefore, regulatory allostery may be transmitted through the hydrophobic patch to regulate the enzymatic activity. In order to probe allostery in the hydrophobic core, methyl groups within the enzyme are ideal candidates to detect signal transmission within the hydrophobic core of the kinase.

In this report, we use the newest version of FLAMEnGO (11) to successfully assign 98% of methyl resonances in PKA-C using minimal and affordable NMR data, and we confirm the assignments further using several through-bond experiments. We then demonstrate that the data from side-chain relaxation experiments can be used to trace the allosteric signals in PKA-C. Our experimental data shows that nucleotide binding has relatively small effects on the overall dynamics in PKA-C. However, the inhibitor can lock the kinase in the closed state by quenching dynamics in the R and C spines on two different time scales. These experimental data support the important roles of these two hydrophobic spines in protein allostery, and other potentially important allosteric

effects are also revealed. Most significantly, the data suggest that the alteration of conformational dynamics is the primary contributor to the allosteric effects of inhibitor binding.

5.2 Results

Methyl group assignment of PKA-C by FLAMEnGO

In order to tracing allostery by probing side-chain dynamics, the first step was to obtain NMR assignments of methyl resonances. Wild-type PKA-C are expressed in the selectively (^{13}C labeled Leu or Val) labeled medium. The constant-time ^{13}C-^{1}H HMQC spectra were acquired to provide residue-type assignments. Long distance restraints, such as two sets of PRE restraints, were obtained by spin labeling the two PKA-C mutants, K16C and I244C. The intensity ratios of resonances from oxidized samples to those from reduced samples provided distance restraints of methyl groups with respect to the positions of two spins labels (MTSLs), which were simulated in the crystal structure by XPLOR-NIH (12). Short distance restraints from the 3D HMQC-NOESY-HMQC constrained methyl groups to one another within a certain distance. Finally, the TOCSY-HMQC data of the same sample correlated the methyl frequencies of Val and Leu residues with the frequencies of the methine protons at the same residues. Using only 26% NOE, 78% PRE, and 98% TOCSY data (as compared with the simulated data), nearly complete methyl assignments (98%) in the closed form of PKA-C were obtained by a new version of FLAMEnGO without the need of amide assignments (**Fig.1**). After the assignments were subsequently confirmed by mutagenesis and through-bond experiments (13), ~20% of methyl assignments were corrected, but most of the above mis-assignments were simply due to the swap of assignments with nearby methyl groups, which may be due to ambiguous distance restraints from NOE and PRE data. Moreover, the methyl assignments were also used to accelerate the assignment processes of amide resonances.

Chemical shift perturbation of methyl groups in PKA-C traces allostery

The methyl assignments for closed PKA-C were transferred to those of apo PKA-C and binary PKA-C by tracing the changes in chemical shifts during ligand titration. Comparing the differences in chemical shifts of three different forms of PKA-C, the chemical shift perturbations were mapped to the primary sequence to identify the effects of substrate binding. However, no obvious correlation between the substrate binding sites and the perturbed residues was observed in the primary sequence (**Fig. S1**), which may result from the fact that the positions of backbone atoms are not good representatives of the ones from methyl groups. In order to follow the allostery in the 3 dimensional space, the perturbations were mapped onto the crystal structure (**Fig. 2**). On the structure, the strong perturbations was observed around and only limited to the binding site when the nucleotide was bound to apo PKA-C. But, after PKI_{5-24} peptides were bound to binary PKA-C, the strong perturbations were propagated from the binding site of the inhibitor to the small lobe of PKA-C through the nucleotide-binding pocket. Moreover, during the titration experiments by NMR, most of the resonances experienced the intermediate exchange during the nucleotide binding, but most of them experienced the slow exchange during the inhibitor binding (**Fig. S2** and **S3**). The results are consistent with the fact that the inhibitor is a stronger binder as compared to the nucleotide.

Order parameters (S^2) of methyl groups in PKA-C identify C spine

Although interesting chemical shift perturbations were observed during titration experiments, the perturbations contained mixed information of both changes in average conformations and in conformational dynamics. In order to better understand the roles of conformational dynamics in the allosteric regulation of PKA-C, relaxation experiments were required for measuring fast (ps-ns) and slow (μs-ms) conformational dynamics. In order to characterize the fast dynamics (ps-ns), the R_2-like deuterium relaxation, R_1 carbon relaxation, and R_{1rho} carbon relaxation experiments (14,15) were performed in the same sample. Then, two different approaches were used to generate order parameters with a given global

rotational correlation time. The global rotational correlation time was provided by dynamic light scattering (**Tab. S1**). Thus, the order parameters from two different analyses were compared to confirm the confidence of the measurement.
In the relaxation measurement, AMPPCP replaced AMPPNP for the nucleotide analog to increase the stability in the NMR sample of the binary form. Both NMR relaxation data and dynamic light scattering data reported consistent global rotational correlation times (31ns for the apo form, 29ns for the binary form, and 28ns for the closed form). Moreover, both approaches for S^2 determination provided highly correlated S^2 values in all three cases (upper panels in **Fig. 3A**, **B**, and **C**), which proved that the measurement was accurate. Because deuterium relaxation was less sensitive to additional relaxation due to external protons, only S^2 values determined by deuterium relaxation were used for the following analysis. The overall order parameters of methyl groups were affected more significantly when the inhibitor is bound to the PKA-C as compared to the binding of the nucleotide (**Fig. S4**). The obtained order parameters from three different forms were mapped onto the structures to see the changes of side chain dynamics (lower panels in **Fig. 3C**, **D**, and **F**). In the previous structural homology analysis, the C spine (A70, V57, L173, I174, L172, L227, M128, and M231) is completed by the binding of the nucleotide (10). However, we found a hydrophobic patch in the C spine changing the rigidity when the inhibitor was bound to PKA-C. The hydrophobic patch (V57, L173, L172, and L227) started from the large lobe and ended at the small lobe through the nucleotide-binding pocket. The patch was rigidified when the inhibitor was bound, but the rigid patch was not complete when only the nucleotide was bound (**Fig. 4**), which is consistent with our chemical shift perturbation data.

Relaxation dispersion experiments on methyl groups in PKA-C identify R spine

In order to probe the slow dynamics (μs-ms) of methyl groups in PKA-C, ^1H CPMG experiments were performed on the same sample with methyl groups labeling as $^{13}CHD_2$ in two different fields (700MHz and 850MHz) (16). Assuming

the two-state exchange model, GUARDD (17) was used to analyze the dispersion curves, and generated populations, Δω, and k_{ex} for each methyl group experiencing chemical exchange. Because the different α values of curve fitting are widely distributed in the protein, the global fitting is not possible and the reliable Δω values are not available (18). The extracted R_{ex} was used as a probe for conformational exchange and mapped onto the structure of PKA-C. We found that the overall conformational exchanges re-distribute upon the nucleotide binding (**Fig. 5A** and **B**). However, when the inhibitor was bound to the binary form of PKA-C, the overall chemical exchanges decreased significantly (**Fig. 5B** and **C**). In the previous structural homology analysis, the integrity of R spine (L106, L95, F185, and Y164) in the kinase family is required for the kinase activity (10). In our data, interestingly, the binding of the inhibitor quenched conformational exchanges on methyl groups (L95, V98, L103, and V104) within or around the R spine (**Fig. 6**), which is also consistent with our chemical shift perturbation data. Moreover, many methyl groups in both the large lobe and the small lobe also showed less conformational dynamics when the inhibitor was bound to PKA-C. Thus, the allosteric effects of PKI_{5-24} on the PKA-C propagated to both large and small lobes by quenching slow conformational dynamics.

Other interesting results are also observed in the activation loop and catalytic loop, playing important roles in the kinase activity and catalysis (9,19). Methyl groups from I150, L167, and L227 displayed large relaxation dispersion in the apo form, but the dispersion was dramatically suppressed by the binding of the nucleotide (**Fig. S5**). The side chains from these three residues form an anchoring region between catalytic loop, αE helix, and αF helix. It suggests that the binding of the nucleotide can position the catalytic loop for the catalysis and transmit the allostery to αE helix and αF helix. On the other hand, the binding of the inhibitor further locks both activation loop and catalytic loop onto αE helix and αF helix respectively (**Fig. S6** and **S7**). Thus, it is another clear evidence of the allosteric effect of PKI_{5-24} by locking the kinase into the closed state.

5.3 Discussion

Using sparse and ambiguous distance restraints from through-space experiments, such as NOE and PRE, the new version of FLAMEnGO was able to assign 98% of methyl resonances with very high spacial accuracy. As compared with other available algorithms, this version of FLAMEnGO can utilize various experimental data, such as PRE and TOCSY data, and tolerate multiple types of ambiguities and other problems in the data, such as overlapped resonances and lack of stereospecific assignments. Given limited information, this program provides the best solution, allowing the user to determine other experiments that might improve the accuracy of the assignments. For example, in our case we used through-bond experiments and site-directed mutagenesis to correct the mis-assignments. Moreover, the methyl assignments determined from the through-space approach were able to assist in assigning amide resonances using triple resonance experiments as well as methyl out-and-back experiments (13).

With the confident resonance assignment we were able to probe conformational dynamics in PKA-C using relaxation experiments. In previous structural studies, two hydrophobic spines in the kinase family of PKA were identified important for intramolecular communications, catalytic functions, and activity regulation (8-10). One of these is called the R spine, which is intact in all active kinases and usually formed as a result of phosphorylation of the kinases' activation loop (10). The other is a non-continuous hydrophobic spine called the C spine, which is made continuous by the incorporation of the nucleotide (10). Based on our experimental data, both spins are rigidified in different time scales by the binding of the inhibitor, which shows that the internal conformational dynamics of the enzyme are quenched. This significant decrease in conformational dynamics is also consistent with the ITC data from previous reports (5,7), suggesting that

inhibitor binding is enthalpically driven and not entropically favorable. Although the binding of the inhibitor shifts the conformational population to the closed state, crystal structure data shows that the internal structural organization of the enzyme does not change (5). All the above data suggest that the intramolecular allosteric effect of the inhibitor is primarily transmitted via altered conformational dynamics within the enzyme.

When the nucleotide alone was bound to PKA-C, a similar, though less pronounced, allosteric effect was also observed. In the absence of the nucleotide, methyl groups located at L167 in the catalytic loop exhibited significant conformational exchange as probed by relaxation dispersion experiments. The methyl groups (I150 and L227) anchoring those at L167 had a similarly high amplitude of conformational exchange. In the presence of the nucleotide, the conformational exchange of methyl groups at L167, I150, and L227 was dramatically suppressed. The data suggest that the conformational population of the catalytic loop was pushed toward the active conformation, and the functional groups on the catalytic loop were oriented so as to coordinate with the phosphate groups in the nucleotide. To sum up, conformational dynamics are required to sample the active conformations of the PKA-C, and these active conformations are selected by various ligands, a conclusion which supports similar findings observed in other enzymes.

5.4 Experimental methods

- **Sample preparation:**

In order to assign methyl groups (CH_3) in PKA-C, the protein sample for NOESY and TOCSY experiments was expressed in ^{15}N labeled deuterated M9 medium (80% $^{2}H_2O$ and 100% ^{2}H-glucose) supplemented with 70mg/L of 2-Ketobutyric acid-4-^{13}C,3,3-d_2 and 90mg/L of 2-Keto-3-(methyl-d_3)-butyric acid-4-^{13}C. In substrate titration experiments or PRE data, the wild-type or mutant PKA-C was expressed in ^{15}N labeled M9 medium supplemented with 70mg/L of $^{13}C_{me}$ α-

ketobutyrate and 90mg/L of $^{13}C_{me}$ α-ketoisovalerate. For the purpose of relaxation experiments on methyl groups (CHD_2), the PKA-C was expressed in ^{15}N labeled deuterated M9 medium (80% $^{2}H_2O$ and 100% ^{2}H-glucose) supplemented with 70mg/L of 2-Ketobutyric acid-4-^{13}C,3,3,4,4-d_4 and 90mg/L of 2-Keto-3-(methyl-^{13}C,d_2)-butyric acid-4-^{13}C,d_2. The purification procedures for PKA-C combined both affinity chromatography and ion exchange chromatography, which have been described before (20,21).

- **Methyl resonance assignment:**

The closed form of methyl-labeled PKA-C, AMPPNP, and PKI_{5-24} were prepared for all experiments for resonance assignment, the molar ratio of PKI_{5-24} to PKA-C was set to 1.2, and 12mM of AMPPNP was used in NMR buffer (20mM KH2PO4, 90mM KCl, 60mM MgCl2, 10mM DTT, pH 6.5). Stock solution of 100mM Pefabloc was prepared in the water, and the effective concentration (~0.5mM) was used in the final NMR sample. The 3D ^{13}C-^{13}C-^{1}H HMQC-NOESY-HMQC was collected with 250ms as the mixing time, 32 scans, 2048 (proton), 80 (carbon), and 120 (carbon) complex points in the Bruker 900MHz. The ^{1}H-^{13}C-^{1}H TOCSY-HMQC experiment was carried out using 50ms as the mixing time and 16 scans, and 2048 (proton), 90 (carbon), and 136 (proton) complex points in the Bruker 900MHz. The PRE experiments required that PKA-C mutants (K16C and I244C) were labeled with MTSL (10 times the protein concentration). The spectra of the oxidized samples were first obtained, and then the spectra of the reduced samples were recorded after adding 10mM DTT into the samples. The intensity ratios of resonances from oxidized samples to those from reduced samples generated PRE restraints. Combining NOE, PRE, and TOCSY data, the resonances were assigned by a new version of FLAMEnGO, and the assignments were confirmed by mutagenesis and through-bond experiments.

- **Substrate titration experiments:**

The buffer conditions for the NMR sample were 20mM KH_2PO_4, 90KCl, 10mM $MgCl_2$, 10mM DTT, 1mM NaN_3, pH6.5. The dilute methyl-labeled PKA-C sample was concentrated to 400uM in the volume of 250ul. The sample was first titrated

with AMPPNP (0.1, 0.2, 0.6, 1, 2, 3, 12mM), and then titrated with PKI_{5-24} (100, 200, 300, 400uM). The whole titration experiments were monitored by ^{13}C-HMQC.

- **Relaxation experiments:**

The buffer conditions for the NMR sample were 20mM KH_2PO_4, 90mM KCl, 10mM $MgCl_2$, 10mM DTT, 1mM NaN_3, pH6.5. The dilute methyl-labeled PKA-C sample was concentrated to 250uM for the apo form (PKA-C), 230uM for the binary form (PAK-C and 10mM AMPPCP), and 200uM for the closed form (PKA-C, 10mM AMPPNP, and 400uM PKI_{5-24}) in the volume of 300ul. All the samples were checked by dynamic light scattering to confirm the sample conditions and determine the global rotational correlation time (τ_c). Because of the isotopomer (CHD_2) using in this labeling scheme, both carbon and deuterium relaxation experiments could be done on the same sample in the Bruker 850MHz. The relaxation delays for R_2-like deuterium relaxation experiment were arrayed with the maximal relaxation time of 5ms. The relaxation delays for R_1 carbon relaxation experiment were arrayed with the maximal relaxation time of 800ms. The relaxation delays for R_{1rho} carbon relaxation experiment were arrayed with the maximal relaxation time of 80ms. The ^1H CPMG experiments were also done in the same sample in the Bruker 850MHz and 700MHz. The constant time is 40ms and the maximal CPMG pulse frequency is 1000Hz.

- **Data analysis:**

Two approaches were used to cross-validate the order parameters acquired from relaxation data. One was the approximated order parameters from the transverse deuterium relaxation rates, and the other was the order parameters obtained by fitting all deuterium and carbon relaxation data with model-free analysis.

Deuterium relaxation approximation (14,22):

$$R(D_+) \approx \left(\frac{1}{80}\right)(2\pi QCC)^2 S^2 \tau_c \quad \ldots[1]$$

where QCC is the quadrupolar coupling constant (167k). The order parameters were calculated directly using transverse deuterium relaxation rates and the global rotational correlation times based on the above equation.

Model-free analysis (15,22):

$$R(D_+) = \left(\frac{1}{80}\right)(2\pi QCC)^2\left(9J(0) + 15J(w_D) + 6J(2w_D)\right) \quad \ldots[2]$$

$$R_1\left({}^{13}C\right) = 2d_{CH}^2[3J(w_C) + J(w_H - w_C) + 6J(w_H + w_C)] + 4d_{CD}^2[3J(w_C) + J(w_D - w_C) + 6J(w_D + w_C)] + 6C^2[J'(w_C)]$$

…[3]

$$R_2\left({}^{13}C\right) = d_{CH}^2[4J(0) + 3J(w_C) + J(w_H - w_C) + 6J(w_H) + 6J(w_H + w_C)] + 2d_{CD}^2[4J(0) + 3J(w_C) + J(w_D - w_C) + 6J(w_D) + 6J(w_D + w_C)] + C^2[4J'(0) + 3J'(w_C)]$$

…[4]

$$J(w) = \frac{1}{9}S^2\left({}^{\tau_c}/_{1 + (w\tau_c)^2}\right) + (1 - \frac{1}{9}S^2)({}^{\tau'}/_{1 + (w\tau')^2})$$

$$J'(w) = S^2\left({}^{\tau_c}/_{1 + (w\tau_c)^2}\right) + (1 - S^2)({}^{\tau'}/_{1 + (w\tau')^2})$$

where $d_{CH}^2 = \frac{1}{20}(\frac{\mu_0}{4\pi})^2 \frac{h^2\gamma_H^2\gamma_C^2}{r_{CH}^6}$, $d_{CD}^2 = \frac{2}{15}(\frac{\mu_0}{4\pi})^2 \frac{h^2\gamma_D^2\gamma_C^2}{r_{CD}^6}$, and $C^2 = \frac{1}{45}w_C^2\Delta\sigma_C^2$, assuming that $r_{CH} = r_{CD} = 1.135$Å, and $\Delta\sigma_C^2$ of 17.8 ppm, 30.0 ppm, 25.2 ppm for Ile, Leu, and Val, respectively. The home-made script was used to analyze the relaxation data. Given a global rotational correlation time τ_c, the theoretical relaxation values could be calculated by the given S^2 and local correlation time τ_f. The best matched S^2 and τ_f could be determined by minimizing the difference between experimental values and theoretical ones. The relaxation contribution of intermediate exchange in the carbon R_2 rates were also taken into account.

Relaxation dispersion data analysis:

^{1}H CPMG data at two different fields (700MHz and 850MHz) were used to fit the model of two-site exchange using the software, GUARRD (17), which could extract populations, Δω, and k_{ex} for each methyl group.

5.5 References

1. Manning G, Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man. Trends Biochem Sci. 2002 10/1;27(10):514-20.

2. Ortutay C, Väliaho J, Stenberg K, Vihinen M. KinMutBase: A registry of disease-causing mutations in protein kinase domains. Hum Mutat. 2005;25(5):435-42.

3. Zheng J, Knighton DR, ten Eyck LF, Karlsson R, Xuong N, Taylor SS, et al. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. Biochemistry. 1993 Mar 9;32(9):2154-61.

4. Tzeng S, Kalodimos CG. Protein dynamics and allostery: An NMR view. Curr Opin Struct Biol. 2011 2;21(1):62-7.

5. Masterson LR, Cembran A, Shi L, Veglia G. Allostery and binding cooperativity of the catalytic subunit of protein kinase A by NMR spectroscopy and molecular dynamics simulations. Adv Protein Chem Struct Biol. 2012;87:363-89.

6. Masterson LR, Cheng C, Yu T, Tonelli M, Kornev A, Taylor SS, et al. Dynamics connect substrate recognition to catalysis in protein kinase A. Nat Chem Biol. 2010 Nov;6(11):821-8.

7. Masterson LR, Shi L, Metcalfe E, Gao J, Taylor SS, Veglia G. Dynamically committed, uncommitted, and quenched states encoded in protein kinase A revealed by NMR spectroscopy. Proc Natl Acad Sci U S A. 2011 Apr 26;108(17):6969-74.

8. Kornev AP, Haste NM, Taylor SS, Eyck LF. Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. Proc Natl Acad Sci U S A. 2006 Nov 21;103(47):17783-8.

9. Kornev AP, Taylor SS, Ten Eyck LF. A helix scaffold for the assembly of active protein kinases. Proc Natl Acad Sci U S A. 2008 Sep 23;105(38):14377-82.

10. Kornev AP, Taylor SS. Defining the conserved internal architecture of a protein kinase. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 2010 3;1804(3):440-4.
11. Chao F, Shi L, Masterson LR, Veglia G. FLAMEnGO: A fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. Journal of Magnetic Resonance. 2012 1;214(0):103-10.
12. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The xplor-NIH NMR molecular structure determination package. J Magn Reson. 2003 Jan;160(1):65-73.
13. Tugarinov V, Kay LE. Ile, leu, and val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc. 2003 11/01;125(45):13868-78.
14. Tugarinov V, Ollerenshaw JE, Kay LE. Probing side-chain dynamics in high molecular weight proteins by deuterium NMR spin relaxation: An application to an 82-kDa enzyme. J Am Chem Soc. 2005 Jun 8;127(22):8214-25.
15. Tugarinov V, Kay LE. Quantitative 13C and 2H NMR relaxation studies of the 723-residue enzyme malate synthase G reveal a dynamic binding interface. Biochemistry. 2005 Dec 13;44(49):15970-7.
16. Baldwin AJ, Religa TL, Hansen DF, Bouvignies G, Kay LE. 13CHD2 methyl group probes of millisecond time scale exchange in proteins by 1H relaxation dispersion: An application to proteasome gating residue dynamics. J Am Chem Soc. 2010 Aug 18;132(32):10992-5.
17. Kleckner IR, Foster MP. GUARDD: User-friendly MATLAB software for rigorous analysis of CPMG RD NMR data. J Biomol NMR. 2012 Jan;52(1):11-22.
18. Millet O, Loria JP, Kroenke CD, Pons M, Palmer AG. The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. J Am Chem Soc. 2000 03/01; 2013/12;122(12):2867-77.

19. Johnson DA, Akamine P, Radzio-Andzelm E, Madhusudan, Taylor SS. Dynamics of cAMP-dependent protein kinase. Chem Rev. 2001 08/01; 2013/12;101(8):2243-70.
20. Bastidas AC, Deal MS, Steichen JM, Keshwani MM, Guo Y, Taylor SS. Role of N-terminal myristylation in the structure and regulation of cAMP-dependent protein kinase. J Mol Biol. 2012 Sep 14;422(2):215-29.
21. Hemmer W, McGlone M, Taylor SS. Recombinant strategies for rapid purification of catalytic subunits of cAMP-dependent protein kinase. Anal Biochem. 1997 Feb 15;245(2):115-22.
22. Sheppard D, Sprangers R, Tugarinov V. Experimental approaches for NMR studies of side-chain dynamics in high-molecular-weight proteins. Prog Nucl Magn Reson Spectrosc. 2010 Jan;56(1):1-45.

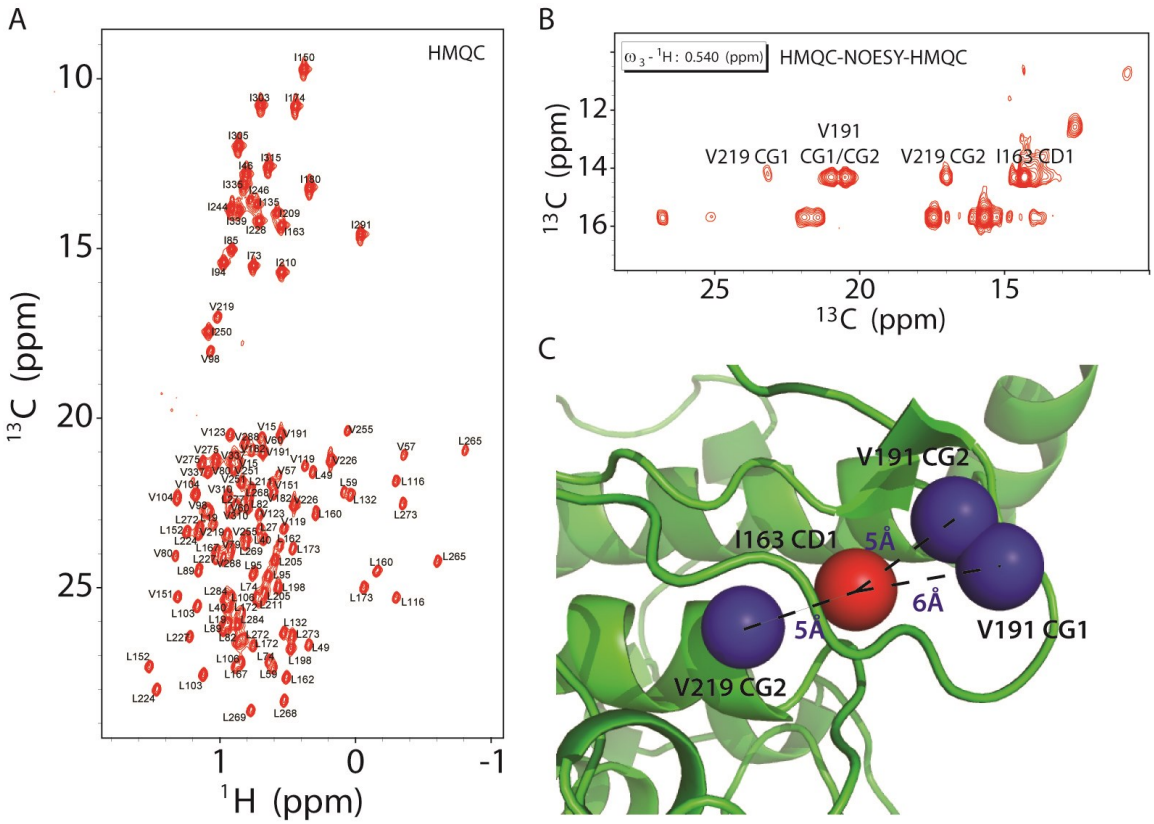


Figure 1. The final methyl assignment of the closed form of PKA-C is determined by 10 results with highest global scores from 100 independent calculations using FLAMEnGO.

D) The methyl-TROSY spectrum of PKA-C is shown with assignments.

E) A slice of 3D HMQC-NOESY-HMQC spectrum of PKA-C shows assigned cross peaks of the assigned I163 methyl group (δ1).

F) The assigned cross peaks of the I163 methyl group (δ1) correspond to the short-distance restraints, which are presented in the crystal structure (1ATP).

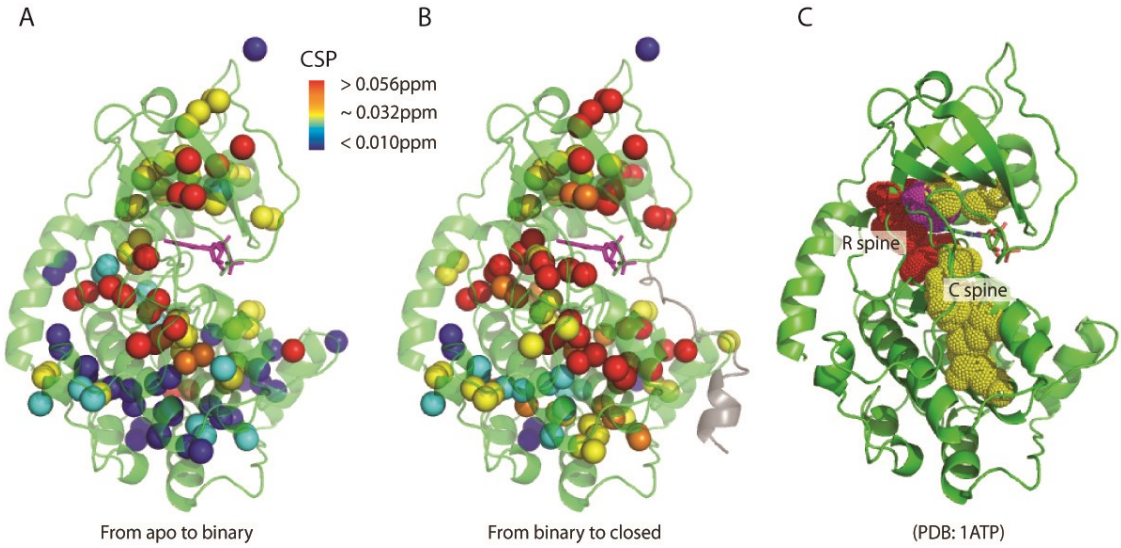


Figure 2. Chemical shift perturbations of methyl groups are mapped onto the crystal structure of PKA-C (1ATP).

The perturbation data from apo form to binary form (AMPPNP bound form) of PKA-C are shown in (A), and the perturbation data from binary form (AMPPNP bound form) to closed form (AMPPNP and PKI_{5-24} bound form) are shown in (B). The hydrophobic spines are shown in (C) with R spine colored red and C spine colored yellow.

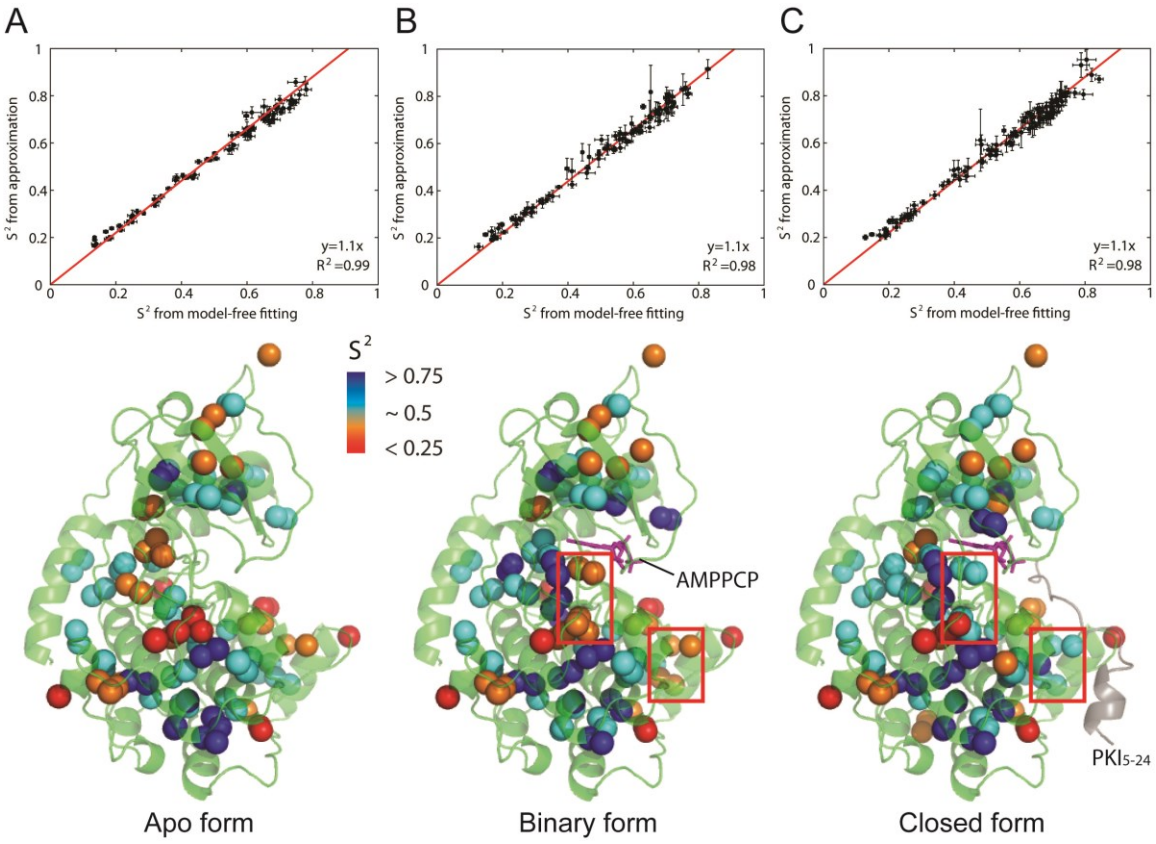


Figure 3. Experimental measurements of order parameters on the three different forms of PKA-C.

The order parameters from deuterium relaxation data are compared with the ones from the model-free analysis using all relaxation data. The correlation between S^2 values from two different approaches is plotted for the apo form (A), the binary form (B), and the closed form (C) of PKA-C in the upper panels. The order parameters from deuterium relaxation data of apo, binary, and closed PKA-C are subsequently mapped onto the structure in the below panels. The red squares contain important perturbations due to the binding of PKI_{5-24}.

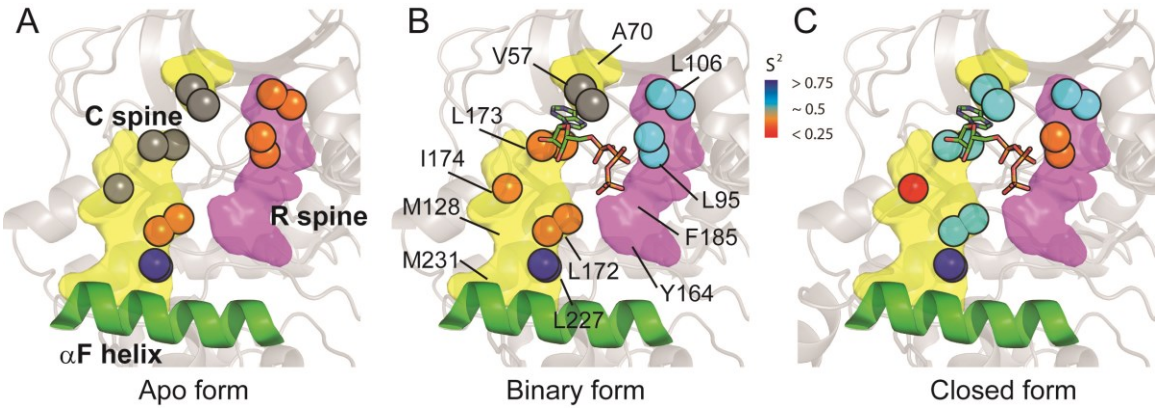


Figure 4. Effects of inhibitor binding on the hydrophobic spines in the fast time scale (ps-ns).

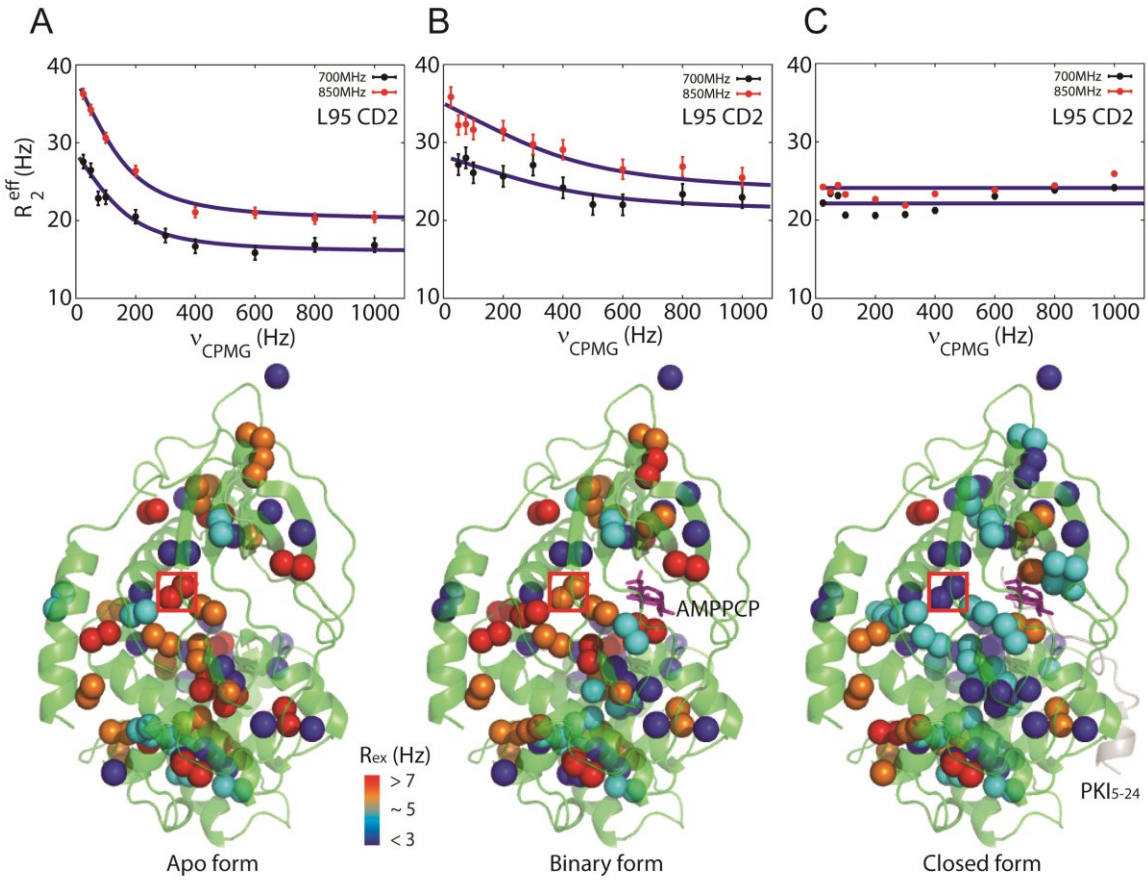


Figure 5. Experimental measurements of R_{ex} due to chemical exchanges on the three different forms of PKA-C.

The curve fitting of CPMG data at the L95 methyl group (δ2) is shown in the upper panels for apo PKA-C (A), binary PKA-C (B), and closed PKA-C (C), and the extracted R_{ex} values are subsequently mapped onto the structure in the lower panels. The red squares on the structure mark positions of L95 methyl groups.

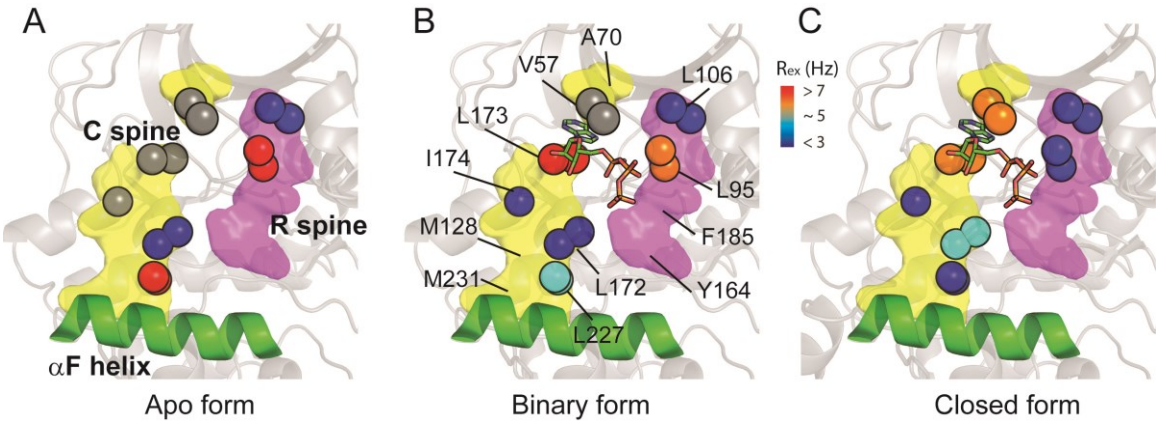


Figure 6. Effects of inhibitor binding on the hydrophobic spines in the slow time scale (μs-ms).

Record	Sample Name	T (°C)	Z-Ave (r.nm)	PdI	Pk 1 Mean Size (r.nm)	Pk 2 Mean Size (r.nm)	Pk 3 Mean Size (r.nm)
1	apo_PKA_250uM	25	3.287	0.036	3.385	0	0
2	apo_PKA_250uM	25	3.284	0.069	3.381	0	0
3	apo_PKA_250uM	25	3.256	0.026	3.333	0	0
4	apo_PKA_250uM	25	3.245	0.042	3.356	0	0
5	apo_PKA_250uM	25	3.24	0.05	3.353	0	0
6	apo_PKA_250uM	25	3.226	0.05	3.331	0	0
1	binary_PKA_230uM	25	3.192	0.072	3.345	0	0
2	binary_PKA_230uM	25	3.195	0.091	3.256	0	0
3	binary_PKA_230uM	25	3.232	0.098	3.256	0	0
4	binary_PKA_230uM	25	3.222	0.13	3.165	0	0
5	binary_PKA_230uM	25	3.141	0.069	3.243	0	0
6	binary_PKA_230uM	25	3.174	0.051	3.317	0	0
1	closed_PKA_200uM	25	3.114	0.103	3.14	0	0
2	closed_PKA_200uM	25	3.104	0.079	3.168	0	0
3	closed_PKA_200uM	25	3.13	0.102	3.177	0	0
4	closed_PKA_200uM	25	3.161	0.135	3.171	1772	0
5	closed_PKA_200uM	25	3.148	0.119	3.171	2279	0
6	closed_PKA_200uM	25	3.133	0.106	3.221	0	0

Table S1. The dynamic light scattering data of three different forms of PKA-C. The global rotational correlation times (31ns for apo form, 29ns for binary form, and 28ns for closed form) were determined by the average values of hydrodynamic radiuses using the Stoke-Einstein equation.

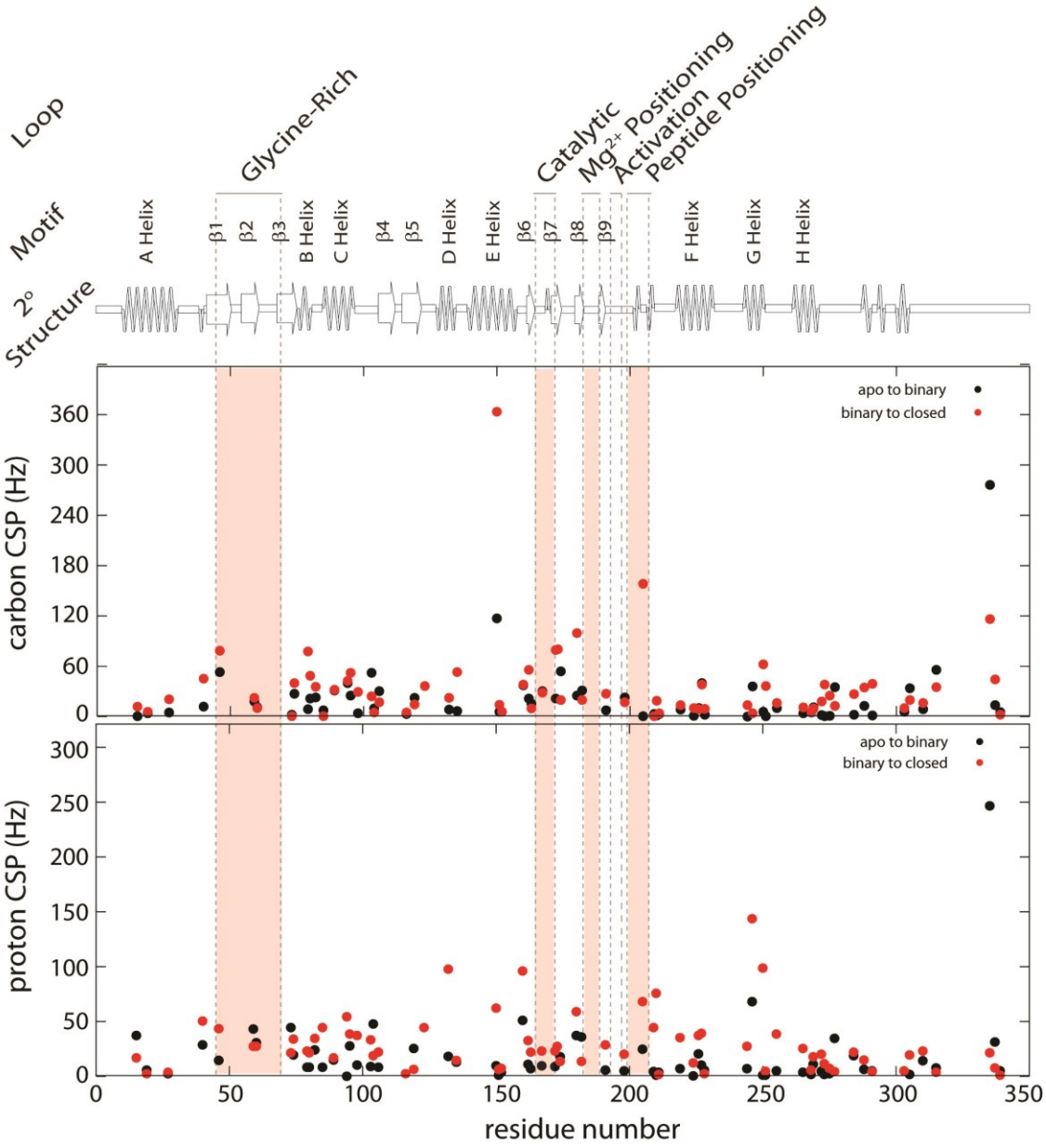


Figure S1. Chemical shift perturbations of methyl groups are mapped onto the primary sequence of PKA-C.

The perturbation data from apo form to binary form (AMPPNP bound form) of PKA-C are colored black, and the perturbation data from binary form (AMPPNP bound form) to closed form (AMPPNP and PKI_{5-24} bound form) are colored red.

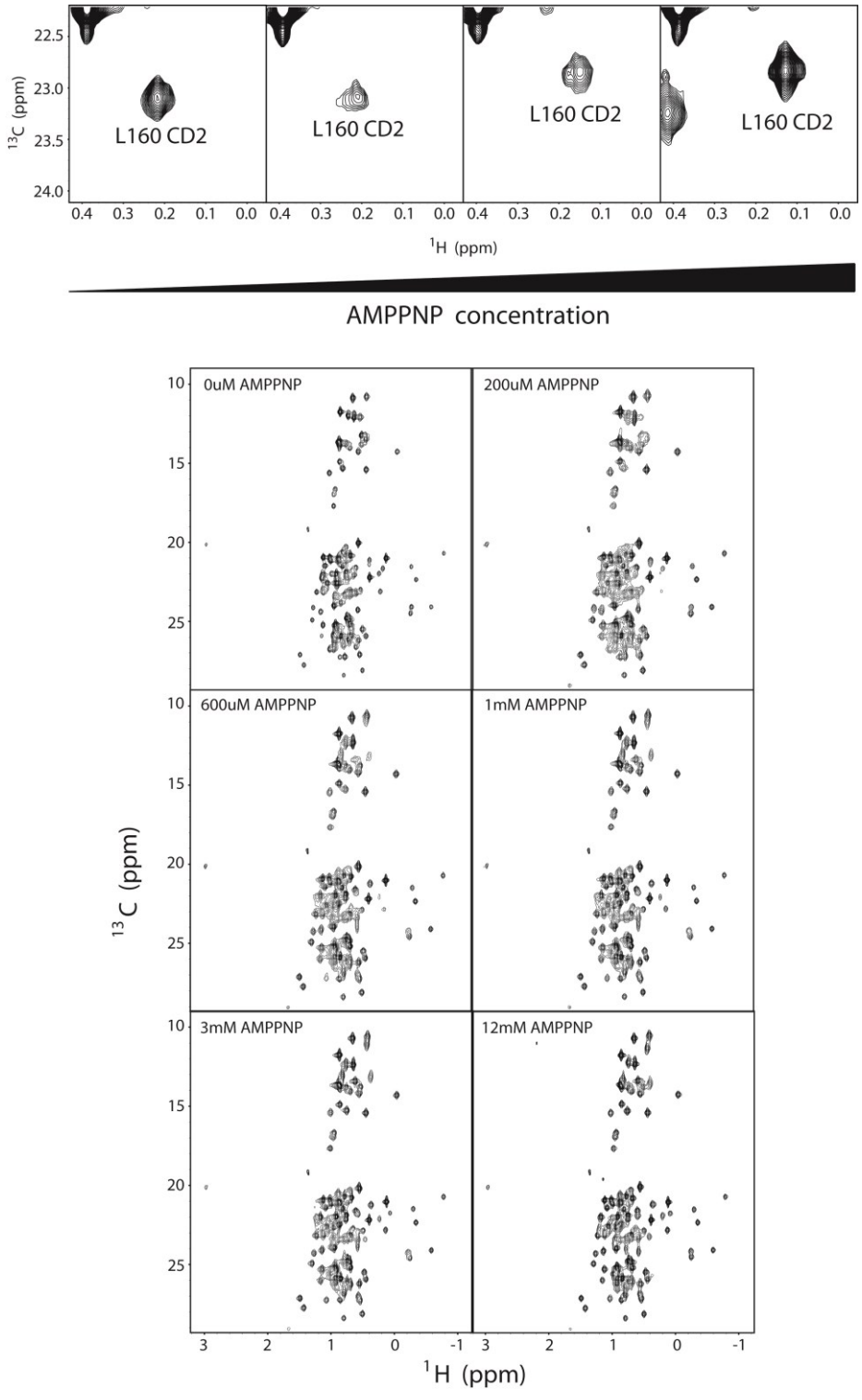


Figure S2. Nucleotide titration experiment monitored by methyl resonances from

PKA-C.

The upper panel represents the intermediate exchange of one methyl resonance from L160 during the titration. The bottom panel represents the overall changes in the ^{13}C-HMQC spectrum during the titration.

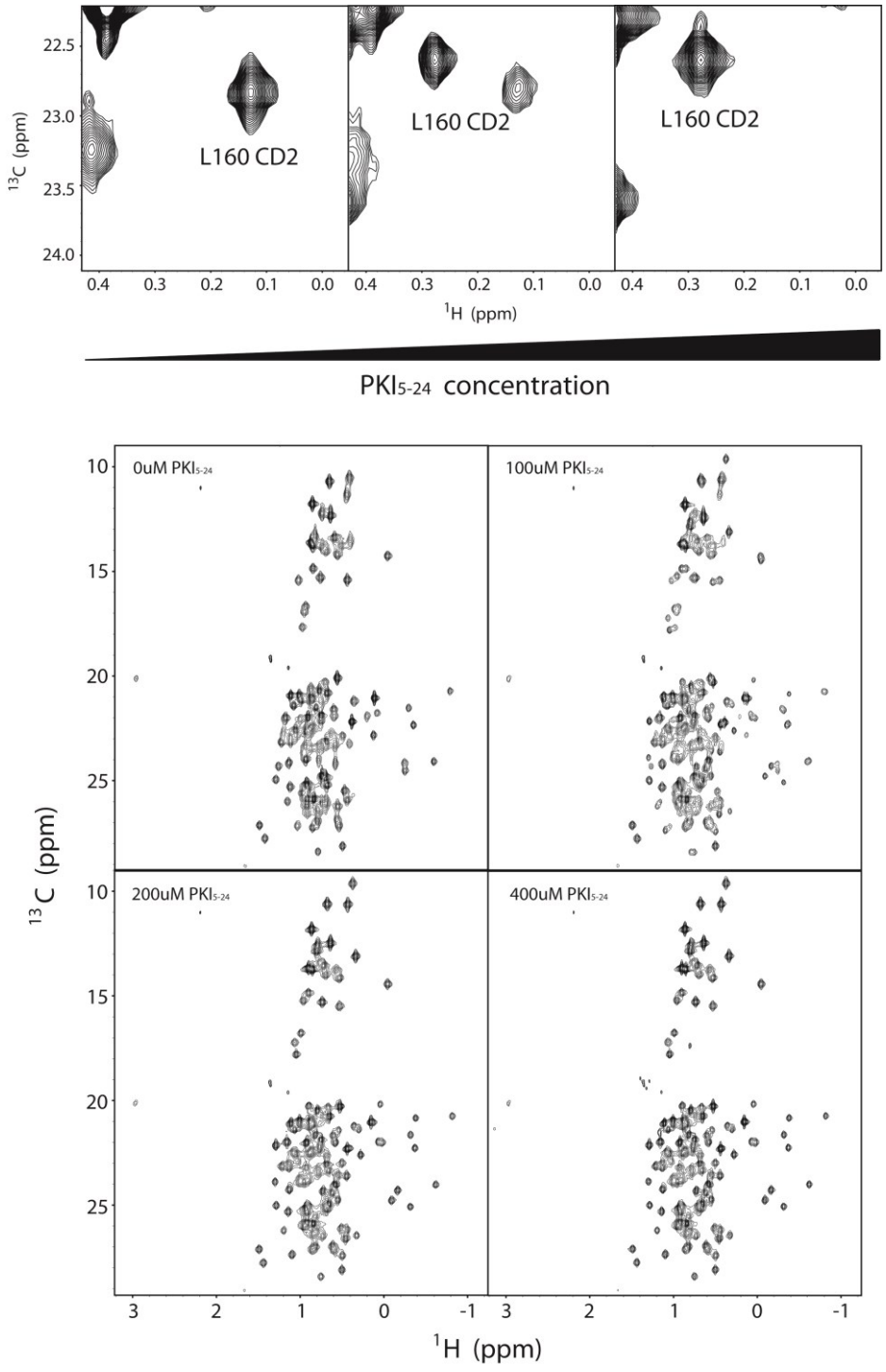


Figure S3. Inhibitor titration experiment monitored by methyl resonances from PKA-C.

The upper panel represents the slow exchange of one methyl resonance from L160 during the titration. The bottom panel represents the overall changes in the ^{13}C-HMQC spectrum during the titration.

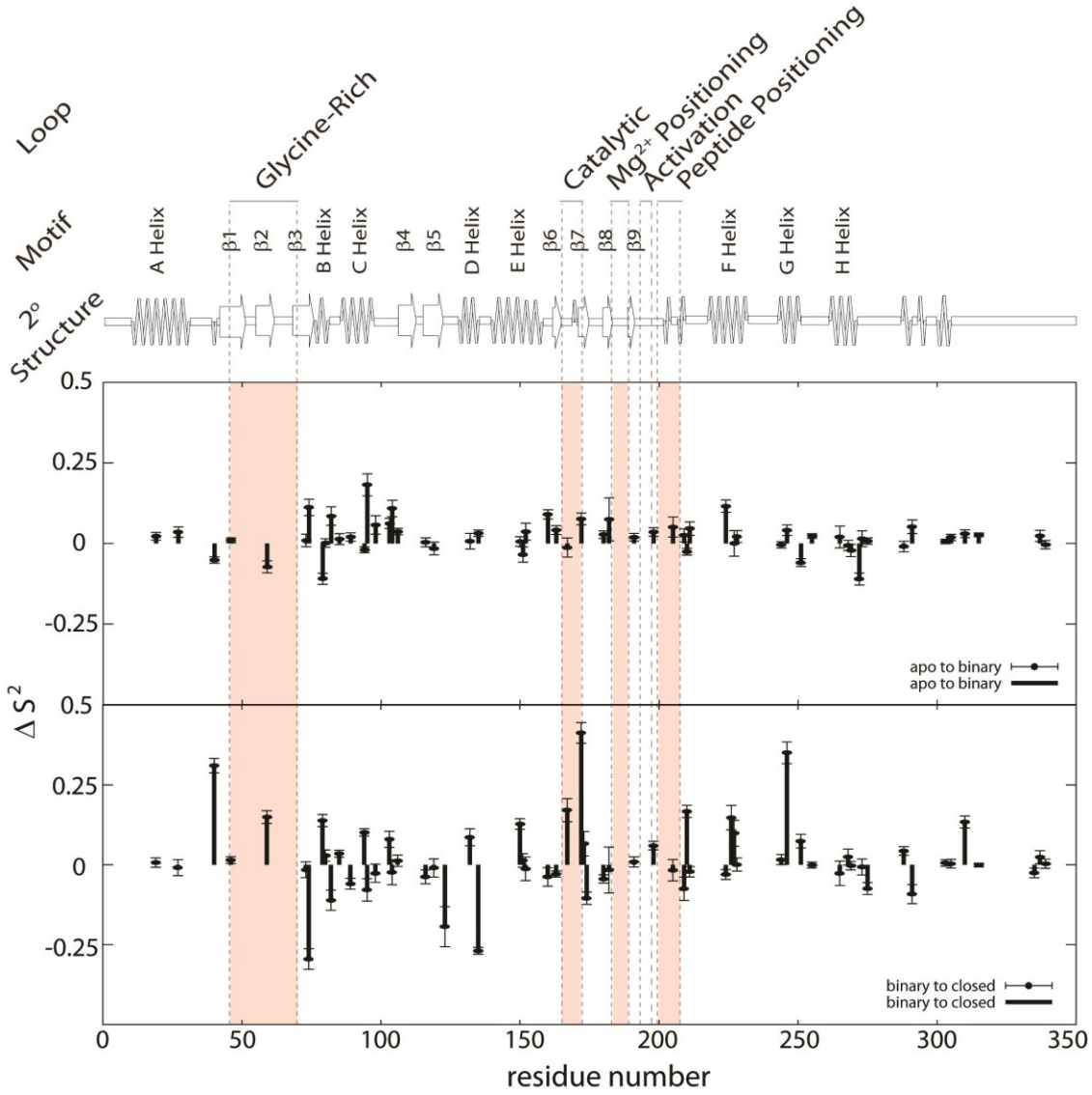


Figure S4. Differences in order parameters of methyl groups are mapped onto the primary sequence of PKA-C.

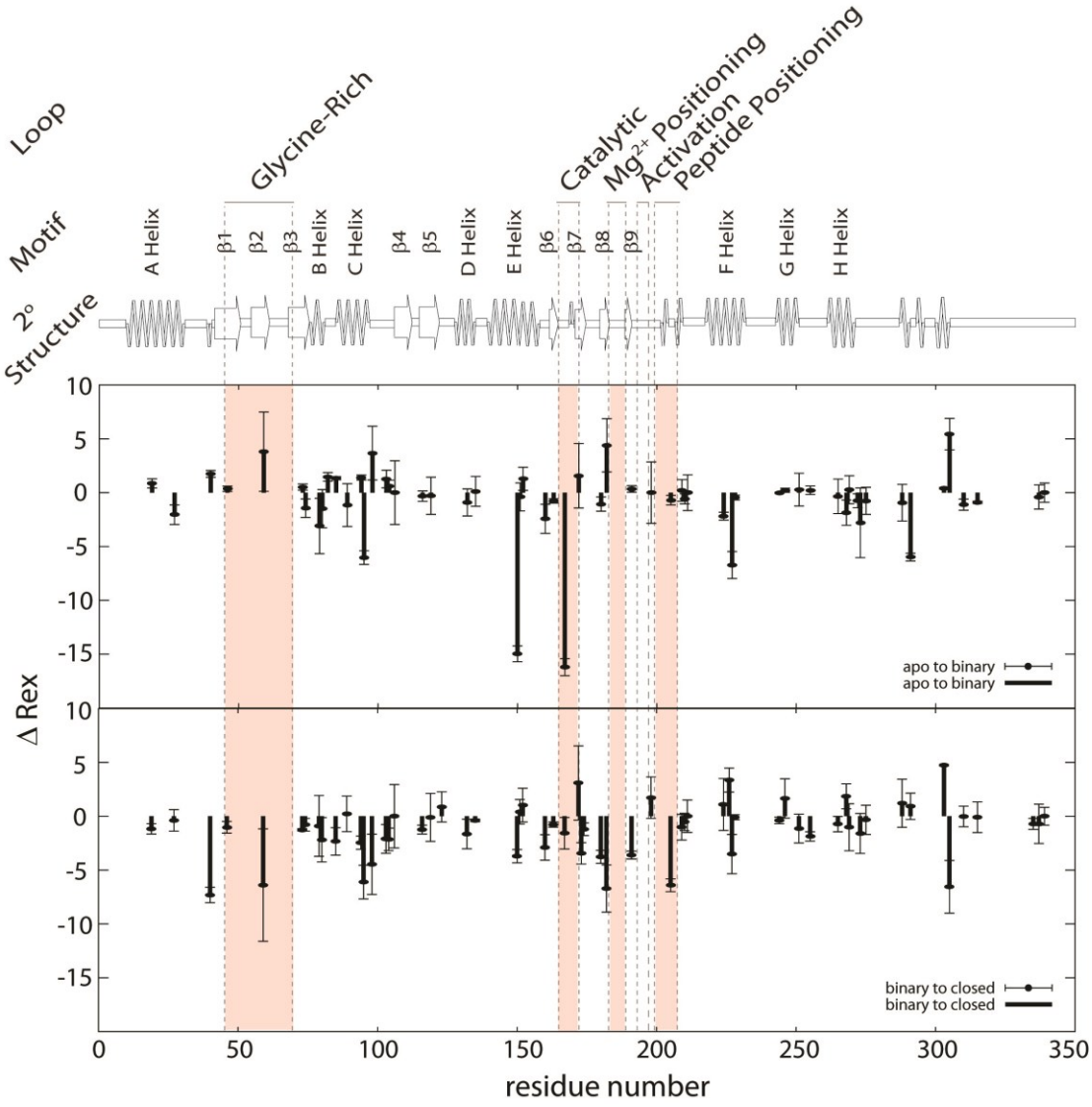


Figure S5. Differences in relaxation rates of methyl groups are mapped onto the primary sequence of PKA-C.

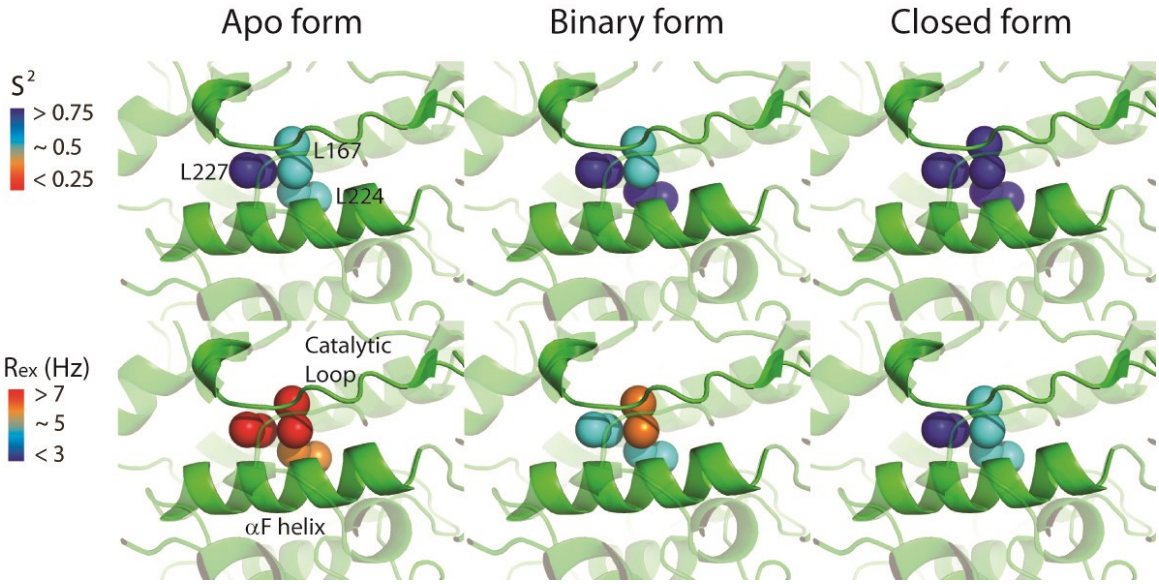


Figure S6. Catalytic loop anchoring with the changes in the side-chain dynamics during substrate binding.

The upper panels and the lower panels represent the fast (ps-ns) and slow (μs-ms) dynamics of methyl groups in the anchoring region between the catalytic loop and the F helix in different forms of PKA-C.

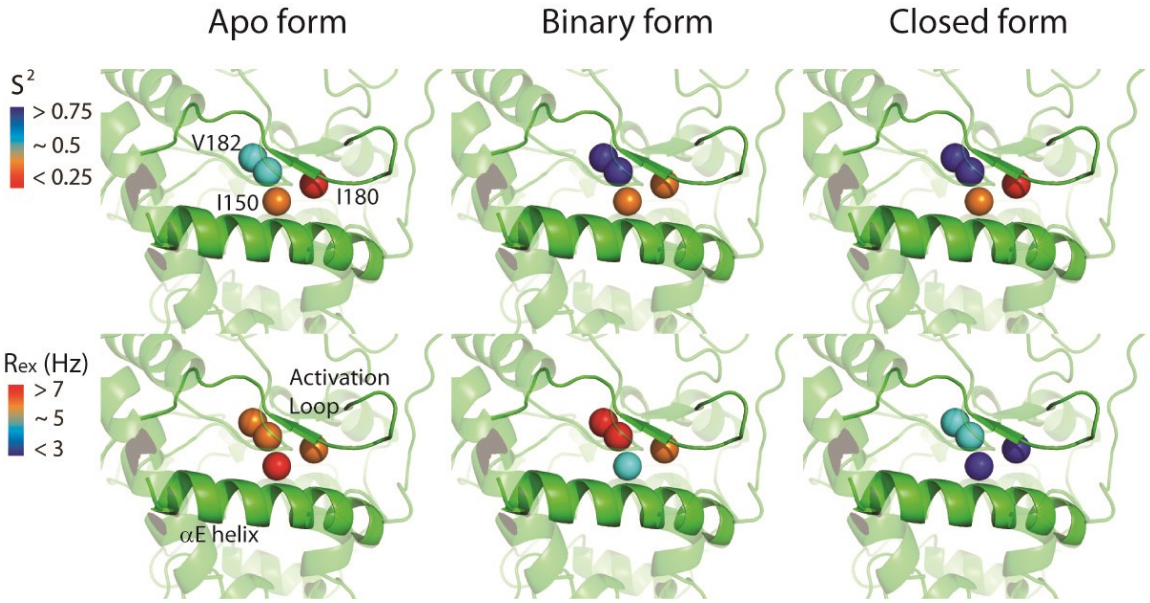


Figure S7. Activation loop anchoring with the changes in the side-chain dynamics during substrate binding.

The upper panels and the lower panels represent the fast (ps-ns) and slow (μs-ms) dynamics of methyl groups in the anchoring region between the activation loop and the E helix in different forms of PKA-C.

Chapter 6- Conclusions

Conformational dynamics have been proved to be important in enzyme catalysis in several cases, including dihydrofolate reductase, cyclophilin A, protein tyrosine phosphatase, etc (1-9). I hypothesize that every enzyme is potentially tunable by changing conformational dynamics. In order to validate this concept, I studied two intrinsically different enzymes, RNA ligase 10C and protein kinase A, by NMR spectroscopy.

Here, I summarize my conclusions based on the above-described studies:

- A primordial fold with suitable flexibility is sufficient to carry out enzymatic function.
- Backbone conformational dynamics are correlated with the activity of the artificial enzyme.
- Allosteric signals are transmitted through the hydrophobic spines in PKA-C upon ligand binding.
- The inhibitor of PKA-C not only shifts the conformational population toward the closed state, but also quenches the conformational dynamics of the enzyme.
- Methyl groups investigated via NMR spectroscopy are sensitive probes for conformational dynamics in the hydrophobic core of proteins.

Based on the experimental data on these two enzymes, I conclude that regulation of enzyme functions through the changes in conformational dynamics is likely a universal phenomenon. Therefore, the changes in conformational dynamics can be used to regulate enzyme functions in cells as well as to control the catalytic activity of artificial enzymes. Further improvement in understanding the roles of conformational dynamics in enzyme function will be of great benefit in developing new drug targets and improving artificial and therapeutic enzyme design.

References

1. Bhabha G, Lee J, Ekiert DC, Gam J, Wilson IA, Dyson HJ, et al. A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science. 2011 Apr 8;332(6026):234-8.
2. Boehr DD, McElheny D, Dyson HJ, Wright PE. The dynamic energy landscape of dihydrofolate reductase catalysis. Science. 2006 Sep 15;313(5793):1638-42.
3. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, et al. Intrinsic dynamics of an enzyme underlies catalysis. Nature. 2005 Nov 3;438(7064):117-21.
4. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature. 2007 Dec 6;450(7171):913-6.
5. Masterson LR, Cheng C, Yu T, Tonelli M, Kornev A, Taylor SS, et al. Dynamics connect substrate recognition to catalysis in protein kinase A. Nat Chem Biol. 2010 Nov;6(11):821-8.
6. Masterson LR, Shi L, Metcalfe E, Gao J, Taylor SS, Veglia G. Dynamically committed, uncommitted, and quenched states encoded in protein kinase A revealed by NMR spectroscopy. Proc Natl Acad Sci U S A. 2011 Apr 26;108(17):6969-74.
7. Masterson LR, Cembran A, Shi L, Veglia G. Allostery and binding cooperativity of the catalytic subunit of protein kinase A by NMR spectroscopy and molecular dynamics simulations. Adv Protein Chem Struct Biol. 2012;87:363-89.
8. Whittier SK, Hengge AC, Loria JP. Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. Science. 2013 Aug 23;341(6148):899-903.
9. Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ, Kern D. Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat Struct Mol Biol. 2004 Oct;11(10):945-9.

Bibliography

Adams, J. A. (2001). Kinetic and catalytic mechanisms of protein kinases. Chemical Reviews, 101(8), 2271-2290. doi:10.1021/cr000230w

Alberts, I. L., Nadassy, K., & Wodak, S. J. (1998). Analysis of zinc binding sites in protein crystal structures. Protein Science : A Publication of the Protein Society, 7(8), 1700-1716. doi:10.1002/pro.5560070805

Amero, C., Asuncion Dura, M., Noirclerc-Savoye, M., Perollier, A., Gallet, B., Plevin, M. J., . . . Boisbouvier, J. (2011). A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. Journal of Biomolecular NMR, 50(3), 229-236. doi:10.1007/s10858-011-9513-5; 10.1007/s10858-011-9513-5

Ayala, I., Sounier, R., Use, N., Gans, P., & Boisbouvier, J. (2009). An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. Journal of Biomolecular NMR, 43(2), 111-119. doi:10.1007/s10858-008-9294-7

Baldwin, A. J., & Kay, L. E. (2009). NMR spectroscopy brings invisible protein states into focus. Nature Chemical Biology, 5(11), 808-814. doi:10.1038/nchembio.238; 10.1038/nchembio.238

Baldwin, A. J., Religa, T. L., Hansen, D. F., Bouvignies, G., & Kay, L. E. (2010). 13CHD2 methyl group probes of millisecond time scale exchange in proteins by 1H relaxation dispersion: An application to proteasome gating residue dynamics. Journal of the American Chemical Society, 132(32), 10992-10995. doi:10.1021/ja104578n; 10.1021/ja104578n

Banci, L., Bertini, I., Del Conte, R., Mangani, S., & Meyer-Klaucke, W. (2003). X-ray absorption and NMR spectroscopic studies of CopZ, a copper chaperone in bacillus subtilis: The coordination properties of the copper ion. Biochemistry, 42(8), 2467-2474. doi:10.1021/bi0205810

Bastidas, A. C., Deal, M. S., Steichen, J. M., Keshwani, M. M., Guo, Y., & Taylor, S. S. (2012). Role of N-terminal myristylation in the structure and regulation of cAMP-dependent protein kinase. Journal of Molecular Biology, 422(2), 215-229. doi:10.1016/j.jmb.2012.05.021

Bhabha, G., Lee, J., Ekiert, D. C., Gam, J., Wilson, I. A., Dyson, H. J., . . . Wright, P. E. (2011). A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science (New York, N.Y.), 332(6026), 234-238. doi:10.1126/science.1198542; 10.1126/science.1198542

Bhabha, G., Lee, J., Ekiert, D. C., Gam, J., Wilson, I. A., Dyson, H. J., . . . Wright, P. E. (2011). A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. Science (New York, N.Y.), 332(6026), 234-238. doi:10.1126/science.1198542; 10.1126/science.1198542

Boehr, D. D., McElheny, D., Dyson, H. J., & Wright, P. E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. Science (New York, N.Y.), 313(5793), 1638-1642. doi:10.1126/science.1130258

Boehr, D. D., Dyson, H. J., & Wright, P. E. (2006). An NMR perspective on enzyme dynamics. Chemical Reviews, 106(8), 3055-3079. doi:10.1021/cr050312q

Bryan, P. N., & Orban, J. (2010). Proteins that switch folds. Current Opinion in Structural Biology, 20(4), 482-488. doi:10.1016/j.sbi.2010.06.002; 10.1016/j.sbi.2010.06.002

Chao, F., Shi, L., Masterson, L. R., & Veglia, G. (2012). FLAMEnGO: A fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. Journal of Magnetic Resonance, 214(0), 103-110. doi:http://dx.doi.org/10.1016/j.jmr.2011.10.008

Cho, G., Keefe, A. D., Liu, R., Wilson, D. S., & Szostak, J. W. (2000). Constructing high complexity synthetic libraries of long ORFs using in vitro selection. Journal of Molecular Biology, 297(2), 309-319. doi:10.1006/jmbi.2000.3571

Cho, G. S., & Szostak, J. W. (2006). Directed evolution of ATP binding proteins from a zinc finger domain by using mRNA display. Chemistry & Biology, 13(2), 139-147. doi:10.1016/j.chembiol.2005.10.015

Chothia, C. (1992). Proteins. one thousand families for the molecular biologist. Nature, 357(6379), 543-544. doi:10.1038/357543a0

Chothia, C., Gough, J., Vogel, C., & Teichmann, S. A. (2003). Evolution of the protein repertoire. Science (New York, N.Y.), 300(5626), 1701-1703. doi:10.1126/science.1085371

Cordes, M. H., Walsh, N. P., McKnight, C. J., & Sauer, R. T. (1999). Evolution of a protein fold in vitro. Science (New York, N.Y.), 284(5412), 325-328.

Dubois, D., & Prade, H. (1998). An introduction to fuzzy systems. Clinica Chimica Acta, 270(1), 3-29. doi:http://dx.doi.org/10.1016/S0009-8981(97)00232-5

Eghbalnia, H. R., Bahrami, A., Tonelli, M., Hallenga, K., & Markley, J. L. (2005). High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection. Journal of the American Chemical Society, 127(36), 12528-12536. doi:10.1021/ja052120i

Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., . . . Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. Nature, 438(7064), 117-121. doi:10.1038/nature04105

Fischer, M., Kloiber, K., Hausler, J., Ledolter, K., Konrat, R., & Schmid, W. (2007). Synthesis of a 13C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. Chembiochem : A European Journal of Chemical Biology, 8(6), 610-612. doi:10.1002/cbic.200600551

Frederick, K. K., Marlow, M. S., Valentine, K. G., & Wand, A. J. (2007). Conformational entropy in molecular recognition by proteins. Nature, 448(7151), 325-329. doi:10.1038/nature05959

Gagne, S. M., Tsuda, S., Li, M. X., Chandra, M., Smillie, L. B., & Sykes, B. D. (1994). Quantification of the calcium-induced secondary structural changes in the regulatory domain of troponin-C. Protein Science : A Publication of the Protein Society, 3(11), 1961-1974. doi:10.1002/pro.5560031108

Gaponenko, V., Howarth, J. W., Columbus, L., Gasmi-Seabrook, G., Yuan, J., Hubbell, W. L., & Rosevear, P. R. (2000). Protein global fold determination using site-directed spin and isotope labeling. Protein Science : A Publication of the Protein Society, 9(2), 302-309. doi:10.1110/ps.9.2.302

Gelis, I., Bonvin, A. M., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., . . . Kalodimos, C. G. (2007). Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell, 131(4), 756-769. doi:10.1016/j.cell.2007.09.039

George, G. N. (Ed.). (2000). EXAFSPAK and EDG-FIT. Menlo Park: Stanford Synchrotron Radiation Lightsource.

Golynskiy, M. V., & Seelig, B. (2010). De novo enzymes: From computational design to mRNA display. Trends in Biotechnology, 28(7), 340-345. doi:10.1016/j.tibtech.2010.04.003; 10.1016/j.tibtech.2010.04.003

Grzesiek, S., Anglister, J., & Bax, A. (1993). Correlation of backbone amide and aliphatic side-chain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. Journal of Magnetic Resonance, Series B, 101(1), 114-119. doi:DOI: 10.1006/jmrb.1993.1019

Grzesiek, S., & Bax, A. (1992). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. Journal of Magnetic Resonance (1969), 96(2), 432-440. doi:DOI: 10.1016/0022-2364(92)90099-S

Han, B., Liu, Y., Ginzinger, S. W., & Wishart, D. S. (2011). SHIFTX2: Significantly improved protein chemical shift prediction. Journal of Biomolecular NMR, 50(1), 43-57. doi:10.1007/s10858-011-9478-4; 10.1007/s10858-011-9478-4

Hemmer, W., McGlone, M., & Taylor, S. S. (1997). Recombinant strategies for rapid purification of catalytic subunits of cAMP-dependent protein kinase. Analytical Biochemistry, 245(2), 115-122. doi:10.1006/abio.1996.9952

Henzler-Wildman, K., & Kern, D. (2007). Dynamic personalities of proteins. Nature, 450(7172), 964-972. doi:10.1038/nature06522

Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., & Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. Nature, 450(7171), 913-916. doi:10.1038/nature06407

Holmbeck, S. M., Foster, M. P., Casimiro, D. R., Sem, D. S., Dyson, H. J., & Wright, P. E. (1998). High-resolution solution structure of the retinoid X receptor DNA-binding domain. Journal of Molecular Biology, 281(2), 271-284. doi:10.1006/jmbi.1998.1908

Ishima, R., & Torchia, D. A. (2000). Protein dynamics from NMR. Nature Structural Biology, 7(9), 740-743. doi:10.1038/78963

James, L. C., & Tawfik, D. S. (2003). Conformational diversity and protein evolution--a 60-year-old hypothesis revisited. Trends in Biochemical Sciences, 28(7), 361-368. doi:10.1016/S0968-0004(03)00135-X

Johnson, D. A., Akamine, P., Radzio-Andzelm, E., Madhusudan, & Taylor, S. S. (2001). Dynamics of cAMP-dependent protein kinase. Chemical Reviews, 101(8), 2243-2270. doi:10.1021/cr000226k

Kaplan, J., & DeGrado, W. F. (2004). De novo design of catalytic proteins. Proceedings of the National Academy of Sciences of the United States of America, 101(32), 11566-11570. doi:10.1073/pnas.0404387101

Kato, H., van Ingen, H., Zhou, B. R., Feng, H., Bustin, M., Kay, L. E., & Bai, Y. (2011). Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR. Proceedings of the

National Academy of Sciences of the United States of America, 108(30), 12283-12288. doi:10.1073/pnas.1105848108; 10.1073/pnas.1105848108

Keefe, A. D., & Szostak, J. W. (2001). Functional proteins from a random-sequence library. Nature, 410(6829), 715-718. doi:10.1038/35070613

Kim, C. A., & Berg, J. M. (1996). A 2.2 A resolution crystal structure of a designed zinc finger protein bound to DNA. Nature Structural Biology, 3(11), 940-945.

Kleckner, I. R., & Foster, M. P. (2012). GUARDD: User-friendly MATLAB software for rigorous analysis of CPMG RD NMR data. Journal of Biomolecular NMR, 52(1), 11-22. doi:10.1007/s10858-011-9589-y; 10.1007/s10858-011-9589-y

Kornev, A. P., Haste, N. M., Taylor, S. S., & Eyck, L. F. (2006). Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. Proceedings of the National Academy of Sciences of the United States of America, 103(47), 17783-17788. doi:10.1073/pnas.0607656103

Kornev, A. P., Taylor, S. S., & Ten Eyck, L. F. (2008). A helix scaffold for the assembly of active protein kinases. Proceedings of the National Academy of Sciences of the United States of America, 105(38), 14377-14382. doi:10.1073/pnas.0807988105; 10.1073/pnas.0807988105

Kornev, A. P., & Taylor, S. S. (2010). Defining the conserved internal architecture of a protein kinase. Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics, 1804(3), 440-444. doi:10.1016/j.bbapap.2009.10.017

Kurz, M., Gu, K., & Lohse, P. A. (2000). Psoralen photo-crosslinked mRNA-puromycin conjugates: A novel template for the rapid and facile preparation of mRNA-protein fusions. Nucleic Acids Research, 28(18), E83.

Lee, D., Hilty, C., Wider, G., & Wuthrich, K. (2006). Effective rotational correlation times of proteins from NMR relaxation interference. Journal of Magnetic Resonance (San Diego, Calif.: 1997), 178(1), 72-76. doi:10.1016/j.jmr.2005.08.014

Lian, L., & Middleton, D. A. (2001). Labelling approaches for protein structural studies by solution-state and solid-state NMR. Progress in Nuclear Magnetic Resonance Spectroscopy, 39(3), 171-190. doi:DOI: 10.1016/S0079-6565(01)00034-6

Lipari, G., & Szabo, A. (1982). Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. Journal of the American Chemical Society, 104(17), 4546-4559. doi:10.1021/ja00381a009

Loria, J. P., Rance, M., & Palmer, A. G. (1999). A relaxation-compensated Carrâˆ'Purcellâˆ'Meiboomâˆ'Gill sequence for characterizing chemical exchange by NMR spectroscopy. Journal of the American Chemical Society, 121(10), 2331-2332. doi:10.1021/ja983961a

Loveridge, E. J., Behiry, E. M., Guo, J., & Allemann, R. K. (2012). Evidence that a â€˜dynamic knockoutâ€™ in escherichia coli dihydrofolate reductase does not affect the chemical step of catalysis. Nat Chem, 4(4), 292-297. Retrieved from http://dx.doi.org/10.1038/nchem.1296

Malcolm H. Levitt. (2007). Spin dynamics : Basics of nuclear magnetic resonance (second edition ed.) John Wiley & Sons Ltd.

Manning, G., Plowman, G. D., Hunter, T., & Sudarsanam, S. (2002). Evolution of protein kinase signaling from yeast to man. Trends in Biochemical Sciences, 27(10), 514-520. doi:http://dx.doi.org/10.1016/S0968-0004(02)02179-5

Mansy, S. S., Zhang, J., Kummerle, R., Nilsson, M., Chou, J. J., Szostak, J. W., & Chaput, J. C. (2007). Structure and evolutionary analysis of a non-biological ATP-binding protein. Journal of Molecular Biology, 371(2), 501-513. doi:10.1016/j.jmb.2007.05.062

Maret, W., & Li, Y. (2009). Coordination dynamics of zinc in proteins. Chemical Reviews, 109(10), 4682-4707. doi:10.1021/cr800556u; 10.1021/cr800556u

Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G., & Wand, A. J. (2010). The role of conformational entropy in molecular recognition by calmodulin. Nature Chemical Biology, 6(5), 352-358. doi:10.1038/nchembio.347; 10.1038/nchembio.347

Masterson, L. R., Cembran, A., Shi, L., & Veglia, G. (2012). Allostery and binding cooperativity of the catalytic subunit of protein kinase A by NMR spectroscopy and molecular dynamics simulations. Advances in Protein Chemistry and Structural Biology, 87, 363-389. doi:10.1016/B978-0-12-398312-1.00012-3

Masterson, L. R., Cheng, C., Yu, T., Tonelli, M., Kornev, A., Taylor, S. S., & Veglia, G. (2010). Dynamics connect substrate recognition to catalysis in protein kinase A. Nature Chemical Biology, 6(11), 821-828. doi:10.1038/nchembio.452

Masterson, L. R., Mascioni, A., Traaseth, N. J., Taylor, S. S., & Veglia, G. (2008). Allosteric cooperativity in protein kinase A. Proceedings of the National Academy of Sciences of the United States of America, 105(2), 506-511. doi:10.1073/pnas.0709214104

Masterson, L. R., Shi, L., Metcalfe, E., Gao, J., Taylor, S. S., & Veglia, G. (2011). Dynamically committed, uncommitted, and quenched states encoded in protein kinase A revealed by NMR spectroscopy. Proceedings of the National Academy of Sciences of the United States of America, 108(17), 6969-6974. doi:10.1073/pnas.1102701108; 10.1073/pnas.1102701108

Masterson, L. R., Yu, T., Shi, L., Wang, Y., Gustavsson, M., Mueller, M. M., & Veglia, G. (2011). cAMP-dependent protein kinase A selects the excited state of the membrane substrate phospholamban. Journal of Molecular Biology, 412(2), 155-164. doi:10.1016/j.jmb.2011.06.041

McCammon, J. A., & Harvey, S. C. (Eds.). (1987). Dynamics of proteins and nucleic acids. Cambridge: Cambridge University Press.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. The Journal of Chemical Physics, 21(6), 1087-1092. doi:http://dx.doi.org/10.1063/1.1699114

Millet, O., Loria, J. P., Kroenke, C. D., Pons, M., & Palmer, A. G. (2000). The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. Journal of the American Chemical Society, 122(12), 2867-2877. doi:10.1021/ja993511y

Muhandiram, D. R., & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. Journal of

Magnetic Resonance, Series B, 103(3), 203-216. doi:DOI: 10.1006/jmrb.1994.1032

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology, 247(4), 536-540. doi:10.1006/jmbi.1995.0159

Mustre de Leon, J., Rehr, J. J., Zabinsky, S. I., & Albers, R. C. (1991). Ab initio curved-wave x-ray-absorption fine structure. Physical Review.B, Condensed Matter, 44(9), 4146-4156.

Ohlenschlager, O., Seiboth, T., Zengerling, H., Briese, L., Marchanka, A., Ramachandran, R., . . . Gorlach, M. (2006). Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. Oncogene, 25(44), 5953-5959. doi:10.1038/sj.onc.1209584

Ohno, S. (Ed.). (1971). Evolution by gene duplication. New York: Springer-Verlag.

Ortutay, C., Väliaho, J., Stenberg, K., & Vihinen, M. (2005). KinMutBase: A registry of disease-causing mutations in protein kinase domains. Human Mutation, 25(5), 435-442. doi:10.1002/humu.20166

Palmer, A. G.,3rd. (2004). NMR characterization of the dynamics of biomacromolecules. Chemical Reviews, 104(8), 3623-3640. doi:10.1021/cr030413t

Plevin, M. J., Hamelin, O., Boisbouvier, J., & Gans, P. (2011). A simple biosynthetic method for stereospecific resonance assignment of prochiral methyl

groups in proteins. Journal of Biomolecular NMR, 49(2), 61-67. doi:10.1007/s10858-010-9463-3

Rehr, J. J., & Albers, R. C. (2000). Theoretical approaches to x-ray absorption fine structure. Reviews of Modern Physics, 72(3), 621-654. Retrieved from http://link.aps.org/doi/10.1103/RevModPhys.72.621

Rehr, J. J., Mustre, d. L., Zabinsky, S. I., & Albers, R. C. (1991). Theoretical x-ray absorption fine structure standards. Journal of the American Chemical Society, 113(14), 5135-5140. doi:10.1021/ja00014a001

Religa, T. L., Sprangers, R., & Kay, L. E. (2010). Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. Science (New York, N.Y.), 328(5974), 98-102. doi:10.1126/science.1184991

Rückert, M., & Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. Journal of the American Chemical Society, 122(32), 7793-7797. doi:10.1021/ja001068h

Ruschak, A. M., & Kay, L. E. (2010). Methyl groups as probes of supra-molecular structure, dynamics and function. Journal of Biomolecular NMR, 46(1), 75-87. doi:10.1007/s10858-009-9376-1; 10.1007/s10858-009-9376-1

Sahakyan, A. B., Vranken, W. F., Cavalli, A., & Vendruscolo, M. (2011). Structure-based prediction of methyl chemical shifts in proteins. Journal of Biomolecular NMR, 50(4), 331-346. doi:10.1007/s10858-011-9524-2

Schwieters, C. D., Kuszewski, J. J., Tjandra, N., & Clore, G. M. (2003). The xplor-NIH NMR molecular structure determination package. Journal of Magnetic Resonance (San Diego, Calif.: 1997), 160(1), 65-73.

Schwieters, C. D., Kuszewski, J. J., Tjandra, N., & Clore, G. M. (2003). The xplor-NIH NMR molecular structure determination package. Journal of Magnetic Resonance (San Diego, Calif.: 1997), 160(1), 65-73.

Seelig, B. (2011). mRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. Nature Protocols, 6(4), 540-552. doi:10.1038/nprot.2011.312; 10.1038/nprot.2011.312

Seelig, B., & Szostak, J. W. (2007). Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. Nature, 448(7155), 828-831. doi:10.1038/nature06032

Sheppard, D., Sprangers, R., & Tugarinov, V. (2010). Experimental approaches for NMR studies of side-chain dynamics in high-molecular-weight proteins. Progress in Nuclear Magnetic Resonance Spectroscopy, 56(1), 1-45. doi:10.1016/j.pnmrs.2009.07.004; 10.1016/j.pnmrs.2009.07.004

Smith, B. A., & Hecht, M. H. (2011). Novel proteins: From fold to function. Current Opinion in Chemical Biology, 15(3), 421-426. doi:10.1016/j.cbpa.2011.03.006; 10.1016/j.cbpa.2011.03.006

Sprangers, R., & Kay, L. E. (2007). Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature, 445(7128), 618-622. doi:10.1038/nature05512

Tenderholt, A. (Ed.). (2007). Pyspline. Stanford: Stanford University.

Tokuriki, N., & Tawfik, D. S. (2009). Protein dynamism and evolvability. Science (New York, N.Y.), 324(5924), 203-207. doi:10.1126/science.1169375; 10.1126/science.1169375

Tugarinov, V., & Kay, L. E. (2003). Side chain assignments of ile delta 1 methyl groups in high molecular weight proteins: An application to a 46 ns tumbling molecule. Journal of the American Chemical Society, 125(19), 5701-5706. doi:10.1021/ja021452+

Tugarinov, V., & Kay, L. E. (2005). Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. Chembiochem : A European Journal of Chemical Biology, 6(9), 1567-1577. doi:10.1002/cbic.200500110

Tugarinov, V., & Kay, L. E. (2005). Quantitative 13C and 2H NMR relaxation studies of the 723-residue enzyme malate synthase G reveal a dynamic binding interface. Biochemistry, 44(49), 15970-15977. doi:10.1021/bi0519809

Tugarinov, V., Ollerenshaw, J. E., & Kay, L. E. (2005). Probing side-chain dynamics in high molecular weight proteins by deuterium NMR spin relaxation: An application to an 82-kDa enzyme. Journal of the American Chemical Society, 127(22), 8214-8225. doi:10.1021/ja0508830

Tugarinov, V., & Kay, L. E. (2003). Ile, leu, and val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. Journal of the American Chemical Society, 125(45), 13868-13878. doi:10.1021/ja030345s

Tuinstra, R. L., Peterson, F. C., Kutlesa, S., Elgin, E. S., Kron, M. A., & Volkman, B. F. (2008). Interconversion between two unrelated protein folds in the

lymphotactin native state. Proceedings of the National Academy of Sciences of the United States of America, 105(13), 5057-5062. doi:10.1073/pnas.0709518105; 10.1073/pnas.0709518105

Tzeng, S., & Kalodimos, C. G. (2011). Protein dynamics and allostery: An NMR view. Current Opinion in Structural Biology, 21(1), 62-67. doi:http://dx.doi.org/10.1016/j.sbi.2010.10.007

van Tilborg, P. J., Czisch, M., Mulder, F. A., Folkers, G. E., Bonvin, A. M., Nair, M., . . . Kaptein, R. (2000). Changes in dynamical behavior of the retinoid X receptor DNA-binding domain upon binding to a 14 base-pair DNA half site. Biochemistry, 39(30), 8747-8757.

Velyvis, A., Schachman, H. K., & Kay, L. E. (2009). Application of methyl-TROSY NMR to test allosteric models describing effects of nucleotide binding to aspartate transcarbamoylase. Journal of Molecular Biology, 387(3), 540-547. doi:10.1016/j.jmb.2009.01.066; 10.1016/j.jmb.2009.01.066

Velyvis, A., Schachman, H. K., & Kay, L. E. (2009). Assignment of ile, leu, and val methyl correlations in supra-molecular systems: An application to aspartate transcarbamoylase. Journal of the American Chemical Society, 131(45), 16534-16543. doi:10.1021/ja906978r

Viles, J. H., Patel, S. U., Mitchell, J. B., Moody, C. M., Justice, D. E., Uppenbrink, J., . . . Thornton, J. M. (1998). Design, synthesis and structure of a zinc finger with an artificial beta-turn. Journal of Molecular Biology, 279(4), 973-986. doi:10.1006/jmbi.1998.1764

Vuister, G. W., & Bax, A. (1993). Quantitative J correlation: A new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-

enriched proteins. Journal of the American Chemical Society, 115(17), 7772-7777. doi:10.1021/ja00070a024

Wang, Y., Zhao, S., Somerville, R. L., & Jardetzky, O. (2001). Solution structure of the DNA-binding domain of the TyrR protein of haemophilus influenzae. Protein Science : A Publication of the Protein Society, 10(3), 592-598. doi:10.1110/ps.45301

Whittier, S. K., Hengge, A. C., & Loria, J. P. (2013). Conformational motions regulate phosphoryl transfer in related protein tyrosine phosphatases. Science (New York, N.Y.), 341(6148), 899-903. doi:10.1126/science.1241735; 10.1126/science.1241735

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. Journal of Molecular Biology, 222(2), 311-333.

Wittekind, M., & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. Journal of Magnetic Resonance, Series B, 101(2), 201-205. doi:DOI: 10.1006/jmrb.1993.1033

Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., & Kern, D. (2004). Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nature Structural & Molecular Biology, 11(10), 945-949. doi:10.1038/nsmb821

Wuthrich, K. (1986). NMR of proteins and nucleic acids.

Xu, Y., Liu, M., Simpson, P. J., Isaacson, R., Cota, E., Marchant, J., . . . Matthews, S. (2009). Automated assignment in selectively methyl-labeled proteins. Journal of the American Chemical Society, 131(27), 9480-9481. doi:10.1021/ja9020233

Xu, Y., & Matthews, S. (2013). MAP-XSII: An improved program for the automatic assignment of methyl resonances in large proteins. Journal of Biomolecular NMR, 55(2), 179-187. doi:10.1007/s10858-012-9700-z; 10.1007/s10858-012-9700-z

Yang, D., Zheng, Y., Liu, D., & Wyss, D. F. (2004). Sequence-specific assignments of methyl groups in high-molecular weight proteins. Journal of the American Chemical Society, 126(12), 3710-3711. doi:10.1021/ja039102q

Yang, W., Lee, J. Y., & Nowotny, M. (2006). Making and breaking nucleic acids: Two-Mg2+-ion catalysis and substrate specificity. Molecular Cell, 22(1), 5-13. doi:10.1016/j.molcel.2006.03.013

Yon, J. M., Perahia, D., & Ghélis, C. (1998). Conformational dynamics and enzyme activity. Biochimie, 80(1), 33-42. doi:http://dx.doi.org/10.1016/S0300-9084(98)80054-0

Zhang, P., Smith-Nguyen, E. V., Keshwani, M. M., Deal, M. S., Kornev, A. P., & Taylor, S. S. (2012). Structure and allostery of the PKA RIIbeta tetrameric holoenzyme. Science (New York, N.Y.), 335(6069), 712-716. doi:10.1126/science.1213979; 10.1126/science.1213979

Zhao, Q., Chasse, S. A., Devarakonda, S., Sierk, M. L., Ahvazi, B., & Rastinejad, F. (2000). Structural basis of RXR-DNA interactions. Journal of Molecular Biology, 296(2), 509-520. doi:10.1006/jmbi.1999.3457

Zheng, J., Knighton, D. R., ten Eyck, L. F., Karlsson, R., Xuong, N., Taylor, S. S., & Sowadski, J. M. (1993). Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. Biochemistry, 32(9), 2154-2161.