White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/117255/

Version: Accepted Version

Article:
Perras, Frédéric A, Marion, Dominique, Boisbouvier, Jérôme et al. (2017) Observation of CH···π Interactions between Methyl and Carbonyl Groups in Proteins. Angewandte Chemie International Edition. pp. 1-5. ISSN: 1433-7851

https://doi.org/10.1002/anie.201702626

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Observation of CH··· π interactions between methyl and carbonyl groups in proteins

Frédéric A Perras,‡[a] Dominique Marion,[b-d] Jérôme Boisbouvier,*[b-d] David L Bryce,*[a] and Michael J Plevin*[e]

Abstract: Protein structure and function is dependent on myriad non-covalent interactions. Direct detection, and characterization, of these weak interactions in large biomolecules, such as proteins, is experimentally challenging. Here, we report the measurement of long-range "through space" scalar couplings between methyl and backbone carbonyl groups in proteins, the first such observation of this phenomenon. These *J* couplings are indicative of the presence of non-covalent C−H··· π hydrogen bond-like interactions involving the amide π network. Experimentally-detected scalar couplings are corroborated by a natural bond orbital analysis, which reveals the orbital nature of the interaction and the origins of the through-space *J* couplings. The experimental observation of this type of CH··· π interaction adds a new dimension to the study of protein structure, function and dynamics by NMR spectroscopy.

The structure, stability, and function of proteins, and other biomacromolecules, are underpinned by the formation and disruption of a variety of non-covalent interactions. The hydrogen bond is a well-studied non-covalent interaction that is prevalent in biochemistry. In proteins, the canonical hydrogen bonds that are formed between backbone amide and carbonyl groups underpin the formation of α-helix and β-sheet secondary structure elements. In recent years, the importance of weaker varieties of the hydrogen bond has grown.[1-3] The current definition of a hydrogen bond includes interactions that involve soft acid and/or base moieties,[4] such as C−H···A (A = N, O, or S), D−H··· π (D = N or O) and C−H··· π. These weaker interactions have been shown to exist in nature: C_α−H···O=C hydrogen bonds are often observed in β-sheet structures in proteins[5,6] while C−H···O=C hydrogen bonds involving aliphatic donor groups have been reported in informatics surveys of protein 3D structures.[7,8] C−H··· π interactions, where the donor group interacts with the π orbitals of a conjugated moiety, are also known to be a common structural feature with up to 15 % of aromatic groups in proteins participating in this type of interaction.[9-11] Similarly to aromatic groups, the amide N−C=O network also forms a π-conjugated system that can participate in weak non-covalent interactions. For example, $n \rightarrow \pi^*$ interactions involving electron donation from the lone pair of an oxygen atom to the unoccupied anti-bonding π^* orbital of a carbonyl carbon atom have a strong presence in proteins.[12,13]

Non-covalent electronic interactions can be conveniently studied by characterizing the *J* coupling interaction wherein nuclear spin-spin coupling is mediated by intervening electrons.[14] For example, canonical hydrogen bonds,[15] C_α−H···O=C hydrogen bonds,[6] C−H···π interactions,[11] and even van der Waals' interactions[16] have been shown to lead to measurable *J* coupling interactions, bringing the electronic significance of these interactions to a stronger experimental footing. Here, using density functional theory (DFT) and solution NMR spectroscopy, we demonstrate that the backbone N-C=O network can act as an acceptor in CH··· π interactions in proteins.

DFT calculations using a model ethane-glycinamide heterodimer predict the existence of *J* couplings on the order of tens of millihertz between carbon and proton nuclei involved in C_{methyl}−H··· π_{CO} hydrogen bond-like interactions (Figure 1). The magnitude of the $^{n\pi}J_{HMeCO}$ coupling is dependent on the distance, *d*, between the methyl carbon and the centre of the C=O bond. The coupling constant increases to a maximum at a distance of approximately 3.7 Å before decreasing at longer distances.

DFT-calculated $^{n\pi}J_{HMeCO}$ coupling constants are fairly insensitive to the angle that the C-H bond forms with the carbonyl group (i.e., the angle ϕ) but are, however, quite sensitive to the orientation with respect to the amide plane (i.e., the angle θ). A larger *J* coupling constant is calculated when the methyl group is oriented away from the normal of the amide plane, which is somewhat surprising given that one would expect that a C−H··· π interaction would be strongest with the CH bond oriented directly above the carbonyl π-bonding orbital, as per other C−H··· π interactions.[3,11] This is likely caused by a larger contribution from a C−H···O=C hydrogen bond-type interaction.

In order to gain a greater insight into the origins of this non-covalent interaction, a natural bonding orbital (NBO) analysis was used.[17] The NBO interaction leading to the largest stabilizing effect is a $\pi_{CO} \rightarrow \sigma^*_{CH}$ interaction, which corresponds to a C−H··· π hydrogen bond. Contributions from $\sigma_{CO} \rightarrow \sigma^*_{CH}$ and $\sigma_{CH} \rightarrow \pi^*_{CO}$ NBO interactions also occur but are significantly weaker (Supplementary Table S1 and Figure S1). The $\pi_{CO} \rightarrow \sigma^*_{CH}$ and $\sigma_{CO} \rightarrow \sigma^*_{CH}$ NBO interactions are illustrated in Figure 2.

[a] Dr F Perras, Prof Dr DL Bryce
Department of Chemistry and Biomolecular Sciences
University of Ottawa
10 Marie Curie St, Ottawa, Ontario (Canada)
E-mail: david.bryce@uottawa.ca
[b] Dr D Marion, Dr J Boisbouvier
Université Grenoble Alpes, Institut de Biologie Structurale (IBS),
71 Avenue des Martyrs, 38044, Grenoble Cedex (France)
[c] Dr D Marion, Dr J Boisbouvier
CNRS, IBS, 38027 Grenoble Cedex (France)
[d] Dr D Marion, Dr J Boisbouvier
CEAEA, IBS, 38027 Grenoble Cedex (France)
E-mail: jerome.boisbouvier@ibs.fr
[e] Dr M J Plevin
Department of Biology, University of York
Wentworth Way, York, YO10 5DD (United Kingdom)
E-mail: michael.plevin@york.ac.uk
[‡] Current address: US Department of Energy, Ames Laboratory,
213 Spedding Hall, Ames, IA, 50011 (USA)

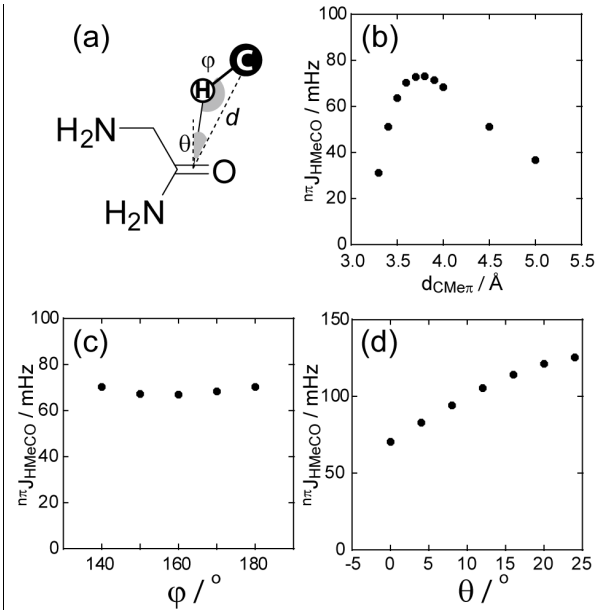


Figure 1. DFT analysis of the $^{n\pi}J_{HMeCO}$ coupling constants using a glycinamide-ethane model. The definitions of the distances and angles are shown in (a) and the calculated $^{n\pi}J_{HMeCO}$ coupling constants are plotted as a function of the C_{Me}-CO distance (b) as well as the angles ϕ (c) and θ (d). The default values of *d*, ϕ and θ (d) used for the calculations are 3.6 Å, 180°, and 0°, respectively. Rapid rotation of the methyl group was taken into consideration by averaging the coupling constants calculated for each of the three proton positions.

When the intermolecular $^{n\pi}J_{HMeCO}$ coupling is decomposed into contributions arising from individual natural localized molecular orbitals (NLMO)[18,19] the largest contributions arise from the σ_{CH} and σ_{CO} NLMOs. This would suggest that the $\sigma_{CO} \rightarrow \sigma^*_{CH}$ interaction would contribute the most to the $^{n\pi}J_{HMeCO}$ coupling whereas the most energetically important contribution is the C−H$\cdots\pi$ hydrogen bond. This is also evident when the *J* coupling and the energy of the $\sigma_{CO} \rightarrow \sigma^*_{CH}$ hydrogen bond component are plotted as a function of the distance between the methyl and the carbonyl groups. The size of the $\sigma_{CO} \rightarrow \sigma^*_{CH}$ interaction is largest when the $^{n\pi}J_{HMeCO}$ coupling constant is largest (Figure 2 and Supplementary Figure S1).

In order to experimentally characterise these weak interactions, we modified a ^{1}H-^{13}C heteronuclear multiple quantum coherence (HMQC) NMR experiment to allow the detection of weak long-range *J* couplings (see Supplementary Material). Given the small magnitude of the DFT-predicted $^{n\pi}J_{HMeCO}$ couplings, the optimal transfer delay is governed by the transverse relaxation rates of the donor group protons rather than by the magnitude of the coupling. To extend the transverse relaxation time (T_2) of these protons, two [*U*-^{2}H,^{13}C]-labeled samples of the protein ubiquitin were prepared with specific incorporation of (^{12}C^{1}H$_3$)-isotopomers in either Ile-δ_1 or Leu/Val-pro*S* methyl groups (details on the isotopic labeling and pulse sequence are provided online with supplementary materials). This isotopic labeling scheme preserves favorable relaxation properties, arising from intra methyl interference effects,[20] and allows for the use of transfer delays of several hundred milliseconds without suffering from truncation effects due to efficient intra-residue transfer. To further increase sensitivity, we modified the standard NMR signal acquisition approach such that ^{1}H detection was initiated before the end of the transfer delay (Supplementary Figure S2). Detecting both sides of the echo gives a theoretical improvement in the signal-to-noise ratio (S/N) of $\sqrt{2}$.

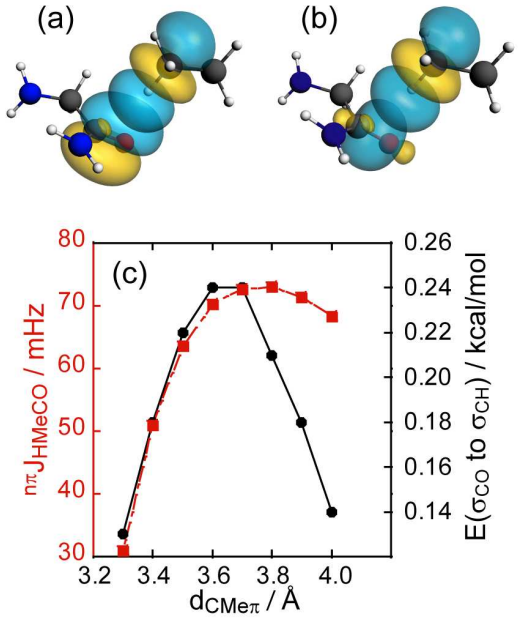


Figure 2. Results from an NBO analysis are shown. Orbitals corresponding to (a) the dominant $\pi_{CO} \rightarrow \sigma^*_{CH}$ NBO interaction and (b) the $\sigma_{CO} \rightarrow \sigma^*_{CH}$ NBO interaction. (c) Comparison of the calculated $^{n\pi}J_{HMeCO}$ coupling constants as a function of distance is plotted (red, primary y-axis) alongside the $\sigma_{CO} \rightarrow \sigma^*_{CH}$ NBO interaction energy (black, secondary y-axis).

Long-range 2D ($^1H_{Me}$,$^{13}C_{CO}$) HMQC spectra revealed a number of cross-peaks that could only arise due to through-space $^{n\pi}J_{HMeCO}$ couplings (Figure 3). To assign the observed C_{Me}−H$\cdots\pi_{CO}$ correlations, we recorded 3D HNCO NMR experiments using the same ubiquitin samples and identical spectral widths and resolution (Figure 4). $^{n\pi}J_{HMeCO}$ coupled nuclei were assigned by comparing the measured ^{1}H and ^{13}C resonance frequencies with assigned resonance frequencies of methyl and carbonyl nuclei in ubiquitin. Internuclear distances extracted from the 3D structure of ubiquitin (PDB code: 1UBQ)[21] were used to exclude nuclei separated by more than 6.0 Å. In situations when more than one carbonyl acceptor met chemical shift and distance criteria, the group with the largest DFT-predicted $^{n\pi}J_{HMeCO}$ coupling was selected (Supplementary Table 2). There are several candidate carbonyl acceptors for the peak at 175.0 ppm associated with Ile-44. It was not possible to distinguish Leu-43 or Gln-49 using DFT results alone. Of the two candidates, the carbonyl carbon of Leu-43 has the shortest internuclear distance (*d* = 2.8 Å) and a more favorable geometry. To exclude the possibility that these signals arise from a weak through-bond ^6J coupling between Ile-44-δ_1 and Leu-43-CO nuclei, we conducted DFT calculations using a model in which

the Ile-44 sidechain was replaced by a methane molecule (situated at the δ_1-position). These analyses showed that a through-space $^{n\pi}J_{HMeCO}$ coupling could yield the cross-peak seen in Figure 3 and that the carbonyl group Leu-43 was the most likely acceptor (Supplementary Table 2).

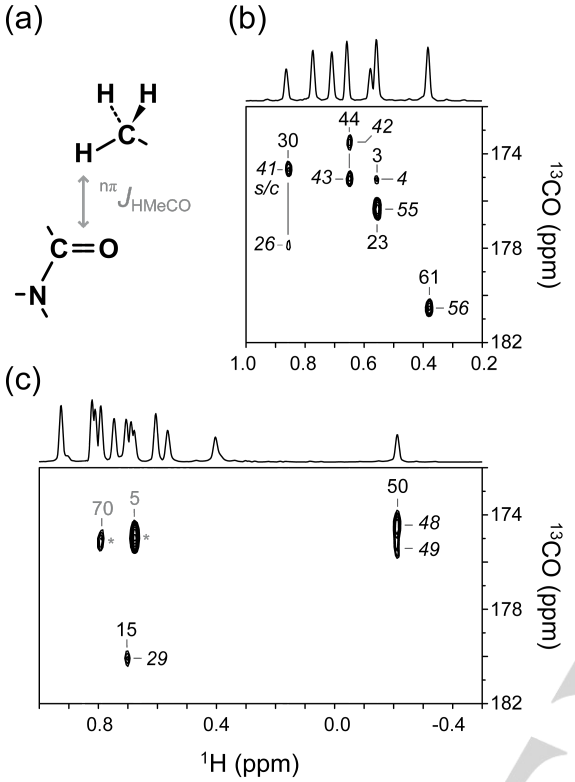


Figure 3. Detection of $^{n\pi}J_{HMeCO}$ couplings in proteins. (a) Schematic showing the through-space $^{n\pi}J_{HMeCO}$ coupling detected between methyl protons and backbone carbonyl carbon; (b) carbonyl region of a 2D ^{1}H-^{13}C HMQC spectrum recorded of [*U*-^{2}H,^{13}C,^{15}N], Ile-[δ_1-^{12}C^{1}H$_3$]-labeled ubiquitin. "s/c" indicates side chain carbonyl acceptor (c) carbonyl region of a 2D ^{1}H-^{13}C HMQC spectrum recorded of [*U*-^{2}H,^{13}C,^{15}N], Leu/Val-[^{12}C^{1}H$_3$]pro*S*-labeled ubiquitin. Asterisks indicate cross peaks arising from through-bond ^5J couplings. Above each data set is the 1D projection of a 2D (^{1}H,^{13}C) SOFAST-HMQC spectrum of a [^{13}C^{1}H$_3$]-methyl labeled ubiquitin sample.

Two cross-peaks were observed at ^{1}H resonance frequencies corresponding to Val-Cγ_2 methyl group protons. The corresponding ^{13}C chemical shifts indicated that these cross peaks originated from long-range 5-bond *J* couplings between a Valine Cγ_2 donor group and the carbonyl group of the preceding residue (indicated by asterisks in Figure 3).

All DFT-calculated *J* coupling constants are positive with the exception of that for the Ile-61/Leu-56 pair. Since this spin pair features a geometry most similar to a C−H···O=C hydrogen bond (Supplementary Figures S4 and S5), the *J* coupling is expected to have an opposite sign since the σ^*_{CH} NBO would interact with a lobe of opposite sign on the σ_{CO} NBO (Supplementary Figure S5), thus inverting the Fermi contact contribution to *J*.[19]

To evaluate the magnitude of the $^{n\pi}J$ couplings associated $C_{Me}H\cdots\pi_{CO}$ interactions, a data set was acquired with the ^{13}C carrier frequency set at 140 ppm and ^{13}C spectral width of 200 ppm. In this experiment magnetization can be transferred across $^{n\pi}J$ couplings associated with CH/π interactions that involve either aromatic or carbonyl acceptor groups. The intensities of correlations that correspond to $C_{Me}H\cdots\pi_{CO}$ interactions are weaker than those corresponding to $C_{Me}H\cdots\pi_{aromatic}$ interactions (Supplementary Figure S6), as predicted from the DFT data presented here. Moreover, the intensities of correlations arising from $C_{Me}H\cdots\pi_{CO}$ interactions involving Ile-δ_1 methyl group donors and backbone carbonyl acceptors correlate well with DFT calculated scalar couplings (Supplementary Figure S6).

Analysis of the distribution of the $C_{Me}H\cdots\pi_{CO}$ interactions observed in ubiquitin revealed no apparent relationship between the detection of a C−H···π interaction and the carbonyl donor group being involved in a canonical hydrogen bond,[22] nor a preference for a particular type of secondary structure (Supplementary Figure S7). All CO acceptors identified here reside in peptide groups with average or above average backbone order parameters (Supplementary Figure S7).[23] However, more variability in the dynamics of the methyl donor groups is evident (Supplementary Figure S8).[24] Of note, the δ_1-methyl group of Ile-44 has a lower than average order parameter, yet participates in at least two $C_{Me}H\cdots\pi_{CO}$ interactions. All methyl group donors, and the majority of carbonyl group acceptors, for which $C_{Me}H\cdots\pi_{CO}$ correlations were observed are inaccessible to the solvent (Supplementary Figure S8). The most interesting example in this context is the $C_{Me}H\cdots\pi_{CO}$ interaction involving the side-chain peptide group of Gln-41 (Figure 3). This residue is completely buried in an apolar environment and the polar side chain peptide group forms no canonical hydrogen bonds. Gln-41 is the only example of a buried asparagine or gluatamine in ubiquitin.

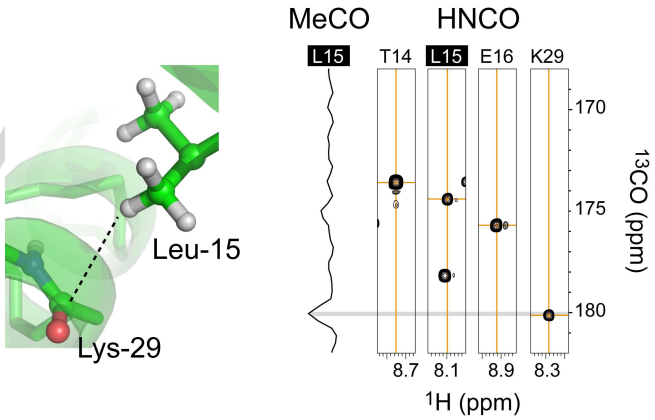


Figure 4. Assignment of Me/π_{CO} couplings. A 1D (^{13}C) trace of an 'MeCO' spectrum of ubiquitin corresponding to a ^{1}H resonance frequency of the pro*S* methyl group of Leu-15 (0.72 ppm) is shown next to (^{1}H,^{13}C) strips extracted from a 3D (^{1}H,^{13}C,^{15}N) HNCO spectrum. HNCO strips were selected based on sequence proximity to the donor residue (i-1, i and i+1) or Me/CO distance (*d* < 6 Å) and matching CO chemical shift.

The detection of small through-space $^{n\pi}J_{HMeCO}$ couplings provides strong experimental support for the presence of C−H···π interactions involving peptide bond acceptor groups in proteins. The labeling scheme employed here only permitted the

detection of $C_{Me}H\cdots\pi_{CO}$ interactions involving Ile-δ_1 and Leu-pro*S* methyl groups. However, it is important to note that C−H$\cdots\pi_{CO}$ interactions in proteins are by no means limited to these two methyl groups. $C_{Me}H\cdots\pi_{CO}$ interactions involving other methyl donor groups could be similarly investigated by NMR spectroscopy by using alternative methyl-labeling approaches,[25] while non-methyl aliphatic C−H donor groups could be characterized using approaches such as Stereo-Array Isotopic Labeling.[26] Moreover, other chemical groups (e.g. hydroxyl groups) in proteins could also form X−H$\cdots\pi_{CO}$ interactions and be investigated by NMR spectroscopy using an appropriately isotopically-labeled protein sample.

Even though the strengths of these non-covalent interactions are weaker than canonical hydrogen bonds, C−H$\cdots\pi_{CO}$ interactions are present in large numbers in proteins[3] and are therefore likely to make an important cumulative contribution to the structure, dynamics and function of a protein. Exploitation of the small and conformationally-sensitive *J* couplings associated with weak non-covalent interactions involving π-acceptor groups promises to be a valuable tool for interrogating the contribution of these hydrogen-bond like interactions to the structure and function of proteins.

Acknowledgements

F. P. thanks NSERC for graduate scholarship and a Banting Postdoctoral Fellowship. F. P. is also supported through a Spedding Fellowship funded by the Laboratory Directed Research and Development (LDRD) program at the Ames Laboratory. The Ames Laboratory is operated for the U.S. Department of Energy by Iowa State University under Contract No. DE-AC02-07CH11358. This work used the high-field NMR and isotopic-labeling facilities at the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). Research leading to these results was funded from the European Research Council under the EU Seventh Framework Program FP7/2007-2013 Grant Agreement no. 260887. D.L.B. thanks NSERC for funding. We thank Dr Remy Sounier for assistance with sample preparation, and Dr Gail Bartlett for critical reading of the manuscript.

Keywords: NMR • C−H$\cdots\pi$ interaction • protein • *J* coupling • DFT

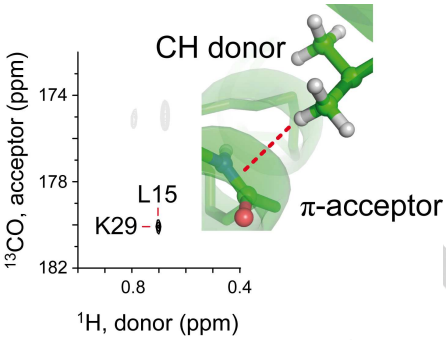
[1] M. S. Weiss, M. Brandl, J. Suhnel, D. Pal, R. Hilgenfeld, *Trends Biochem. Sci.* **2001**, *26*, 521–523.
[2] M. Nishio, *Phys. Chem. Chem. Phys.* **2011**, *13*, 13873–13900.
[3] M. Nishio, Y. Umezawa, J. Fantini, M. S. Weiss, P. Chakrabarti, *Phys. Chem. Chem. Phys.* **2014**, *16*, 12648–12683.
[4] G. R. Desiraju, *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 52–59.
[5] Z. S. Derewenda, L. Lee, U. Derewenda, *J. Mol. Biol.* **1995**, *252*, 248–262.
[6] F. Cordier, M. Barfield, S. Grzesiek, *J. Am. Chem. Soc.* **2003**, *125*, 15750–15751.
[7] J. D. Yesselman, S. Horowitz, C. L. Brooks, R. C. Trievel, *Proteins* **2015**, *83*, 403–410.
[8] G. J. Bartlett, D. N. Woolfson, *Protein Sci.* **2016**, *25*, 887–897.
[9] T. Steiner, G. Koellner, *J. Mol. Biol.* **2001**, *305*, 535–557.
[10] M. Brandl, M. S. Weiss, A. Jabs, J. Suhnel, R. Hilgenfeld, *J. Mol. Biol.* **2001**, *307*, 357–377.
[11] M. J. Plevin, D. L. Bryce, J. Boisbouvier, *Nat. Chem.* **2010**, *2*, 466–471.
[12] G. J. Bartlett, A. Choudhary, R. T. Raines, D. N. Woolfson, *Nat. Chem. Biol.* **2010**, *6*, 615–620.
[13] A. Choudhary, R. T. Raines, *Protein Sci.* **2011**, *20*, 1077–1081.
[14] J.-C. Hierso, *Chem. Rev.* **2014**, *114*, 4838–4867.
[15] G. Cornilescu, J.-S. Hu, A. Bax, *J. Am. Chem. Soc.* **1999**, *121*, 2949–2950.
[16] M. P. Ledbetter, G. Saielli, A. Bagno, N. Tran, M. V. Romalis, *PNAS* **2012**, *109*, 12393–12397.
[17] J. P. Foster, F. Weinhold, *J. Am. Chem. Soc.* **1980**, *102*, 7211–7218.
[18] A. E. Reed, F. Weinhold, *J. Chem. Phys.* **1985**, *83*, 1736–1740.
[19] J. Autschbach, B. Le Guennic, *J. Chem. Educ.* **2007**, *84*, 156.
[20] V. Tugarinov, P. M. Hwang, J. E. Ollerenshaw, L. E. Kay, *J. Am. Chem. Soc.* **2003**, *125*, 10420–10428.
[21] S. Vijay-Kumar, C. E. Bugg, W. J. Cook, *J. Mol. Biol.* **1987**, *194*, 531–544.
[22] S. Grzesiek, F. Cordier, V. Jaravine, M. Barfield, *Prog. Nuc. Mag. Res. Spect.* **2004**, *45*, 275–300.
[23] N. Tjandra, S. E. Feller, R. W. Pastor, A. Bax, *J. Am. Chem. Soc.* **1995**, *117*, 12562–12566.
[24] C. Farès, N.-A. Lakomek, K. F. A. Walter, B. T. C. Frank, J. Meiler, S. Becker, C. Griesinger, *J. Biomol. NMR* **2009**, *45*, 23–44.
[25] R. Kerfah, M. J. Plevin, R. Sounier, P. Gans, J. Boisbouvier, *Curr. Opin. Struct. Biol.* **2015**, *32*, 113–122.
[26] M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono, P. Guntert, *Nature* **2006**, *440*, 52–57.

WILEY-VCH

Entry for the Table of Contents

COMMUNICATION

Solution NMR spectroscopy and density functional theory calculations demonstrate the existence of weak, hydrogen-bond-like C−H··· π interactions in proteins involving methyl donor groups and peptide bond acceptor groups.



Frédéric A Perras, Dominique Marion, Jérôme Boisbouvier, David L Bryce,* and Michael J Plevin**

Page No. – Page No.

Observation of CH···π interactions between methyl and carbonyl groups in proteins

WILEY-VCH

SUPPLEMENTARY MATERIAL

Experimental methods

Supplementary Table S1. NBO interaction energies for an ethane-glycinamide dimer.

Supplementary Table S2. Donor groups, acceptor groups, distances and theoretical magnitude of $^{n\pi}J_{HMeCO}$ couplings.

Supplementary Figure S1. NBO analysis of Me/π_{CO} interactions.

Supplementary Figure S2. NMR pulse sequence and data processing method used to measure the long-range $^{n\pi}J_{HMeCO}$ couplings.

Supplementary Figure S3. Isotopic enrichment patterns in the protein samples used in this study.

Supplementary Figure S4. Geometric distribution of methyl donor and carbonyl acceptor groups of $C_{Me}H\cdots\pi_{CO}$ interactions detected by NMR spectroscopy.

Supplementary Figure S5. NBO and NMR analysis of CH/O hydrogen bond interaction involving Ile-61 and Leu-56.

Supplementary Figure S7. Order parameters for methyl and backbone amide groups in ubiquitin.

Supplementary Figure S8. Solvent accessibilities of the methyl and carbonyl groups in ubiquitin.

SUPPLEMENTARY METHODS

Calculations

All DFT calculations were performed with the use of the Amsterdam Density Functional program (ADF).[1] *J* coupling calculations were performed at the PBE0/jcpl[2] level of theory due to the method's good performance in the calculation of *J* coupling constants involving carbon and protons.[3] The structural models included the two, methyl-terminated, amino acids involved in the interaction. NBO/NLMO analyses[4] were performed with the use of the slightly less computationally demanding PBE method[5] and were analyzed using the NBO ver. 5.0 program[6] that is distributed with ADF. The energies of the NBO interactions for the glycinamide-ethane heterodimer are listed in Supplementary Table S1 below for various conformations.

Data analysis

After initial reconstruction of NMR data sets (see legend to Supplementary Figure S2), spectra were processed with nmrPipe[7] and analyzed using CCPN Analysis v2.[8] Protein structure images were made with Pymol v 0.99 (http://pymol.org). Superposition of molecular coordinates was performed using ProFit (http://www.bioinf.org.uk/programs/profit/). Solvent accessible surface areas were calculated using NACCESS using default settings (http://www.bioinf.manchester.ac.uk/naccess/).

Sample preparation

All protein samples used for measuring Me-π_{CO} couplings were expressed from *E. coli* BL21(DE3) cultures grown in standard M9 minimal media containing [*U*-^{15}N]-labelled ammonium chloride (Cambridge Isotope Ltd), [*U*-^{2}H,^{13}C]-d_6-glucose (Cambridge Isotopes Ltd) and 99.85 % D_2O (Cambridge Isotopes Ltd). Specific protonation of leucine δ_2- and valine γ_2-methyl groups was achieved by supplementing the growth medium with 300 mg/L of [*2*-H_3]methyl, [*4*-D_3]-acetolactate.[9] Acetolactate was prepared in house. Pro*R*/*S* nomenclature of Leu and Val methyl groups follows IUPAC/IUBMB/IUPAB guidelines.[10] Specific protonation of isoleucine δ_1-methyl groups was achieved by supplementing the growth medium with 70 mg/L of *2*-keto-[*3*-$^{2}H_3$]-butyrate.[11,12] Ubiquitin samples were purified as previously described.[13] Final NMR samples were prepared in 50 mM tris-HCl, pH 8.0 in 100 % D_2O, supplemented with sodium azide and protease inhibitors. The final concentration of both protein samples was approximately 2 mM.

The 1D projections in Figures 3b,c were recorded using [*U*-^{2}H,^{13}C,^{15}N] labeled ubiquitin with [^{1}H,^{13}C]-labelling of either isoleucine δ_1-, or leucine δ_2- and valine γ_2-methyl groups. This isotopic labeling pattern was achieved using the same approach as above but with the addition of *2*-keto-[*3*-$^{2}H_3$],*4*-[^{13}C]-butyrate or [*2*-^{13}C,H_3]methyl, [*4*-D_3]-acetolactate, respectively.

Supplementary Table S1. NBO interaction energies for the glycinamide-ethane model.

$d_{CMe-\pi}$ / Å	θ / °	$\pi_{CO} \rightarrow \sigma^*_{CH}$ / kcal mol^{-1}	$\sigma_{CO} \rightarrow \sigma^*_{CH}$ / kcal mol^{-1}	$\sigma_{CH} \rightarrow \pi^*_{CO}$ / kcal mol^{-1}
3.3	0	1.99	0.13	0.31
3.4	0	1.26	0.18	0.23
3.5	0	0.76	0.22	0.17
3.6	0	0.42	0.24	0.12
3.6	4	0.42	0.24	0.06
3.6	8	0.48	0.23	<0.05
3.6	12	0.55	0.21	<0.05
3.6	16	0.62	0.19	<0.05
3.6	20	0.69	0.17	<0.05
3.6	24	0.77	0.14	0.07
3.7	0	0.21	0.24	0.08
3.8	0	0.09	0.21	<0.05
3.9	0	<0.05	0.18	<0.05
4.0	0	<0.05	0.14	<0.05

Supplementary Table S2. Donor groups, acceptor groups, distances and DFT-calculated $^{n\pi}J_{HMeCO}$ couplings.

Methyl residue	Methyl proton chemical shift (ppm)	Carbonyl residue	Carbonyl carbon chemical shift (ppm)	Internuclear distance (Å)	DFT-predicted *J* (mHz)
Ile-3	0.56	Phe-4	175.1	4.7	24
Ile-23	0.56	Thr-55	176.3	4.0	40
		Thr-22	176.3	3.6	14
Ile-30	0.86	Val-26	177.9	4.0	25
		Gln-41 (s/c)	174.7	3.8	67
Ile-44	0.65	Arg-42	173.5	4.4	31
		Gly-47	173.5	5.1	18
		Leu-43	175.0	2.8	31 (42)[a]
		Gln-49	175.0	4.1	32
Ile-61	0.38	Leu-56	180.6	3.9	-51
Leu-15	0.73	Lys-29	180.1	4.8	16
Leu-50	-0.19	Lys-48	174.5	4.2	48
		Gln-49	175.4	3.2	46

(a) Couplings involving the Ile-44/Leu-43 $C_{Me}H\cdots\pi_{CO}$ interaction were calculated in two ways. The main value corresponds to a coupling calculated using the approach applied to other $C_{Me}H\cdots CO$ pairs. The value in parentheses corresponds to a model in which the Ile-44 has been replaced by methane to represent the δ_1-methyl group. This adapted model rules out the possibility that this coupling is due to a long range intramolecular 6-bond *J* coupling between the δ_1-methyl group of Ile-41 and the carbonyl group of Leu-43.

(s/c) denotes side-chain carbonyl group acceptor

Supplementary Figure S1. NBO analysis of $C_{Me}H\cdots\pi_{CO}$ interactions. (a) Interaction energy as a function of the distance, *d*, between the mid-point of the C=O bond and the methyl carbon atom; (b) interaction energy as a function the angle θ, described in Figure 1(a). Values for $\pi_{CO}\rightarrow\sigma^*_{CH}$ (black), $\sigma_{CO}\rightarrow\sigma^*_{CH}$ (red) and $\sigma_{CH}\rightarrow\pi^*_{CO}$ (green) interactions are plotted

Supplementary Figure S2. NMR pulse sequence (A) and data processing method (B) used to measure long-range $^{n\pi}J_{HMeCO}$ couplings. The spectra were acquired using a 600 MHz Varian-Agilent NMR spectrometer equipped with a cryogenically cooled probe. The standard SOFAST-HMQC experiment[14] was modified as follows: the transfer delay, Δ, was set to 0.4 s, which was chosen as a compromise between relaxation and transfer efficiency. In the absence of transverse relaxation, a larger value corresponding to $^1/_{2J}$ would have been optimal. This long Δ delay makes it possible to enhance the experimental sensitivity (by a factor of $\sqrt{2}$) by detecting the signal on both sides of the echo, i.e. for negative (depicted in grey) and positive (in black) t_2 values. The timing of the acquisition launch has to be optimized to ensure that the sampling grid entails one point sampled at exactly $t_2 = 0$. The ^{13}C carrier frequency is set at 175 ppm with a ^{13}C spectral width of 14 kHz (unless otherwise stated). During the periods Δ, evolution due to other J_{CH} couplings is eliminated by two WURST pulses[15] applied at 42 ppm with a 55 ppm bandwidth. Evolution due to J_{HD} couplings is eliminated by ^2H composite decoupling. The ^1H excitation pulse is a 4 ms PC9 selective pulse[16] and the refocusing pulse is a 1 ms r-SNOB pulse.[17] All pulses were applied along x with phase cycling of hard carbon pulses (φ_1 = x, -x; φ_2 = 2(x), 2(-x)) and the receiver (φ_{rec} = x. -x, -x, x). The 2D Me/CO spectrum of Ile-δ_1 labeled ubiquitin was recorded with matrix size of 192 × 2 (t_1) by 2048 × 2 (t_2) points, and 400 scans per increment for a total acquisition time of 80 hours. The 2D Me/CO spectrum of Leu/Val pro*S* labeled ubiquitin was recorded with matrix size of 128 × 2 (t_1) by 2048 × 2 (t_2) points, and 448 scans per increment for a total acquisition time of 60 hours. In both spectra, an equal number of points were collected at negative t_2 and at positive t_2. All data was collected at 310 K.

(B) Processing flow for the SOFAST-HMQC spectra. The spectra were processed using the Rowland NMR toolkit (http://rnmrtk.uchc.edu/rnmrtk/RNMRTK.html). The raw data are apodized (multiplication by a non-shifted sine-bell) and then zero-filled (2×). Because the implementation of standard Fast Fourier Transform (FFT) algorithms assumes that the first data point correspond to $t_2 = 0$, a circular permutation of the data is required. At this stage, the data points for the positive t_2 value (in black) are at the very beginning and that for the negative t_2 value (in grey) at the very end, in agreement with the circular periodic property of the discrete FT. The frequency domain in F_2 is thus easily obtained after FT and proper zero-order phase correction. Note that no first order phase correction is required as the sampling grid contains the $t_2 = 0$ data point. In the t_1 dimension, a standard processing approach is used, with the exception of the missing first data point (due to finite pulse length as compared to the large F_1 spectral width) that is restored by backward linear prediction.

Supplementary Figure S3. Isotopic enrichment patterns in the protein samples used in this study. Two samples of [^2H,^{13}C,^{15}N] labeled ubiquitin were prepared with

specific protonation of either: (a) Ile-δ_1; or (b) Leu/Val pro*S* methyl groups. Leucine and valine labeling was achieved using [2-H_3]methyl, [*4*-D_3]-acetolactate,[9] which results with [^{1}H,^{13}C]-labeling of the pro*S* prochiral methyl groups. For valine, this labeling scheme results in a [^{12}C]-carbonyl carbon, which means that $C_{Me}H\cdots\pi_{CO}$ interactions involving the valine carbonyl group cannot be detected using this sample. Isoleucine labeling was achieved using [*3*-D_3], [*4*-H_3]-α-ketobutyrate.[11,12] Again, the isotopic labeling pattern of this precursor results in a [^{12}C]-labeled carbonyl carbon, which means that Me/π_{CO} interactions involving the isoleucine carbonyl group cannot be detected using this sample.

Supplementary Figure S4. Geometric distribution of methyl donor and carbonyl acceptor groups of Me/π_{CO} interactions detected by NMR spectroscopy. PDB co-ordinates (1UBQ) for methyl group (Ile-C_γ-$C_{\delta1}H_3$ or Leu-C_γ-$C_{\delta2}H_3$) and peptide bond (CO-NH) components of each experimentally detected $C_{Me}H\cdots\pi_{CO}$ interaction were superposed on the CO-NH group.

Supplementary Figure S5. NBO and NMR analysis of CH/O hydrogen bond interaction involving Ile-61 and Leu-56. (A) Assignment of $C_{Me}H\cdots\pi_{CO}$ couplings. A 1D (^{13}C) trace of a $C_{Me}H\cdots\pi_{CO}$ spectrum of ubiquitin corresponding to a ^{1}H resonance frequency of the Ile-δ_1 methyl group of Ile-61 (0.38 ppm) is shown next to (^{1}H,^{13}C) strips extracted from a 3D (^{1}H,^{13}C,^{15}N) HNCO spectrum. HNCO strips were selected based on sequence proximity to the donor residue (i-1, i and i+1) or $C_{Me}H\cdots CO$ distance (d < 6 Å) and matching CO chemical shift. Two labeling schemes were used in this study (Leu/Val-specific or Ile-specific), which result in [^{13}C] labeling for all backbone carbonyl groups with the exception of those of Val or Ile, respectively. The HNCO strip corresponding to the CO group of Ile-61 (colored grey) was taken from data recorded using a sample with Leu/Val-specific labeling. These data also emphasize that the Me/π_{CO} cross-peaks detected do not result from long-range intra-residue $^5J_{HMeCO}$ couplings for the 1-5 % of ^{13}C carbonyl sites that are not derived from the methyl labeling precursor. (B) The NBOs participating in the $\sigma_{CO}\rightarrow\sigma^*_{CH}$ interaction, which is responsible for the through-space *J* coupling in the Leu-56/Ile-61 pair, are depicted. As can be seen, contrarily to the orbitals shown in Figure 2 in the main text, the vacant σ^*_{CH} NBO now interacts with the lobe of opposite sign of the σ_{CO} NBO. This change in sign would lead to an opposite sign in the Fermi contact contribution to the total *J* coupling and is the reason why this spin pair has a negative *J* coupling constant (see Supplementary Table S2).

Supplementary Figure S6. Comparison of DFT-calculated $^{n\pi}J$ couplings with experimentally-detected correlations. (A) a 1D trace (^{1}H = -0.2 ppm) from a 2D (^{1}H,^{13}C) HMQC spectrum of Leu/Val pro*S* labeled ubiquitin with the carbon offset at 140 ppm and spectral width of 200 ppm shows signals resulting from $^{n\pi}J$ couplings involving both carbonyl and aromatic acceptor groups. The magnitudes of the $^{n\pi}J$ couplings for $C_{Me}H\cdots\pi_{aromatic}$ interactions have been previously reported.[13] The peak at 175 ppm shows the unresolved resonances relating to $C_{Me}H\cdots\pi_{CO}$ interactions involving L50-δ_2 methyl donor group and K48 and Q49 carbonyl acceptor groups (see also Figure 3c). The intensity of this peak would correspond to a combined $^{n\pi}J$ coupling of 55 mHz, which is consistent with the values calculated from DFT

(Supplementary Table S2). (B) Graph comparing the signal intensities of correlations arising from $C_{Me}H\cdots\pi_{CO}$ interactions involving Ile-δ_1 methyl groups plotted as a function of the square of the DFT-calculated scalar coupling (J^2, mHz^2).

Supplementary Figure S7. Order parameters for methyl and backbone amide groups in ubiquitin. Order parameters (S^2) for methyl (top panel,[18]) and backbone amide (bottom panel; BioMagResBank accession code: bmrb6470, [19]) groups as a function of residue number.

The top panel shows order parameters of Leu-pro*S* or Ile-δ_1 methyl groups as a function of residue number. The positions of Leu and Ile residues are indicated. Coloring indicates those methyl groups for which $C_{Me}H\cdots\pi_{CO}$ interactions were detected by NMR spectroscopy (Red, Ile; Blue, Leu). The line at S^2 = 0.31 indicates the mean methyl order parameters.[18]

The bottom panel shows order parameters (S^2) of backbone amide groups as a function of peptide group. The peptide group number refers to residue of the carbonyl group. The ^{15}N S^2 value reported for residue i refers to the nitrogen of the same peptide group, i.e. residue i + 1. Coloring indicates residues with carbonyl groups that participate in $C_{Me}H\cdots\pi_{CO}$ interactions detected by NMR spectroscopy (Red, Ile; Blue, Leu). The line at S^2 = 0.84 indicates the mean amide order parameters.[19] The bottom panel also shows the location of secondary structure (grey) and the NH donor (blue) and CO acceptors (green) that participate in canonical hydrogen bonds.

The colored lines connecting the top and bottom panels indicate the $C_{Me}H\cdots\pi_{CO}$ interactions detected by NMR spectroscopy in this study.

Supplementary Figure S8. Solvent accessibilities of the methyl and carbonyl groups in ubiquitin. Per-atom surface accessible surface area (SASA) was calculated using NACCESS using the 3D structure of ubiquitin (1UBQ, [20]).

The top panel shows the SASA of Leu-pro*S* or Ile-δ_1 methyl groups as a function of residue number. The positions of Leu and Ile residues are indicated. Coloring indicates those methyl groups for which $C_{Me}H\cdots\pi_{CO}$ interactions were detected by NMR spectroscopy (Red, Ile; Blue, Leu). The grey line at 70 $Å^2$ indicates the approximate SASA of an isolated methyl group. The mean SASA for Ile-δ_1 or Leu-pro*S* methyl groups in ubiquitin was calculated to be 11 $Å^2$.

The bottom panel shows SASA for backbone carbonyl groups as a function of residue number. The grey line at 35 $Å^2$ indicates the approximate SASA of a carbonyl group in a Gly-Gly dipeptide, which can be taken to represent the maximum SASA for this group (note that the SASA C-terminal carbonyl group exceeds this as would be expected). The mean SASA for backbone carbonyl groups in ubiquitin was calculated to be 11 $Å^2$. The bottom panel also shows the location of secondary structure (grey) and the NH donor (blue) and CO acceptors (green) that participate in canonical hydrogen bonds. Hydrogen bond donor and acceptor groups correspond to those with detectable cross-hydrogen bond *J* couplings.[21]

The colored lines connecting the top and bottom panels indicate the $C_{Me}H\cdots\pi_{CO}$ interactions detected by NMR spectroscopy in this study.

Supplementary References

[1] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, *J Comp Chem* **2001**, *22*, 931–967.

[2] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **1996**, *105*, 9982–9985.

[3] T. Kupka, M. Nieradka, M. Stachów, T. Pluta, P. Nowak, H. Kjær, J. Kongsted, J. Kaminsky, *J. Phys. Chem. A* **2012**, *116*, 3728–3738.

[4] J. Autschbach, *J. Chem. Phys.* **2007**, *127*, 124106.

[5] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[6] E. D. Glendening, J. K. Badenhoop, A. E. Reed, J. E. Carpenter, J. A. Bohmann, C. M. Morales, C. R. Landis, F. Weinhold, Theoretical Chemistry Institute, University of Wisconsin, et al., **n.d.**

[7] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **1995**, *6*, 277–293.

[8] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins* **2005**, *59*, 687–696.

[9] P. Gans, O. Hamelin, R. Sounier, I. Ayala, M. A. Durá, C. D. Amero, M. Noirclerc-Savoye, B. Franzetti, M. J. Plevin, J. Boisbouvier, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 1958–1962.

[10] J. L. Markley, A. Bax, Y. Arata, C. W. Hilbers, R. Kaptein, B. D. Sykes, P. E. Wright, K. Wuthrich, *J. Biomol. NMR* **1998**, *12*, 1–23.

[11] K. H. Gardner, L. E. Kay, *J. Am. Chem. Soc.* **1997**, *119*, 7599–7600.

[12] V. Tugarinov, L. E. Kay, *J. Biomol. NMR* **2004**, *28*, 165–172.

[13] M. J. Plevin, D. L. Bryce, J. Boisbouvier, *Nat. Chem.* **2010**, *2*, 466–471.

[14] P. Schanda, E. Kupce, B. Brutscher, *J. Biomol. NMR* **2005**, *33*, 199–211.

[15] E. Kupce, R. Freeman, *J Magn Reson A* **1995**, *115*, 273–276.

[16] E. Kupce, R. Freeman, *J Magn Reson A* **1994**, *108*, 268–273.

[17] E. Kupce, J. Boyd, I. D. Campbell, *J. Magn. Reson. B.* **1995**, *106*, 300–303.

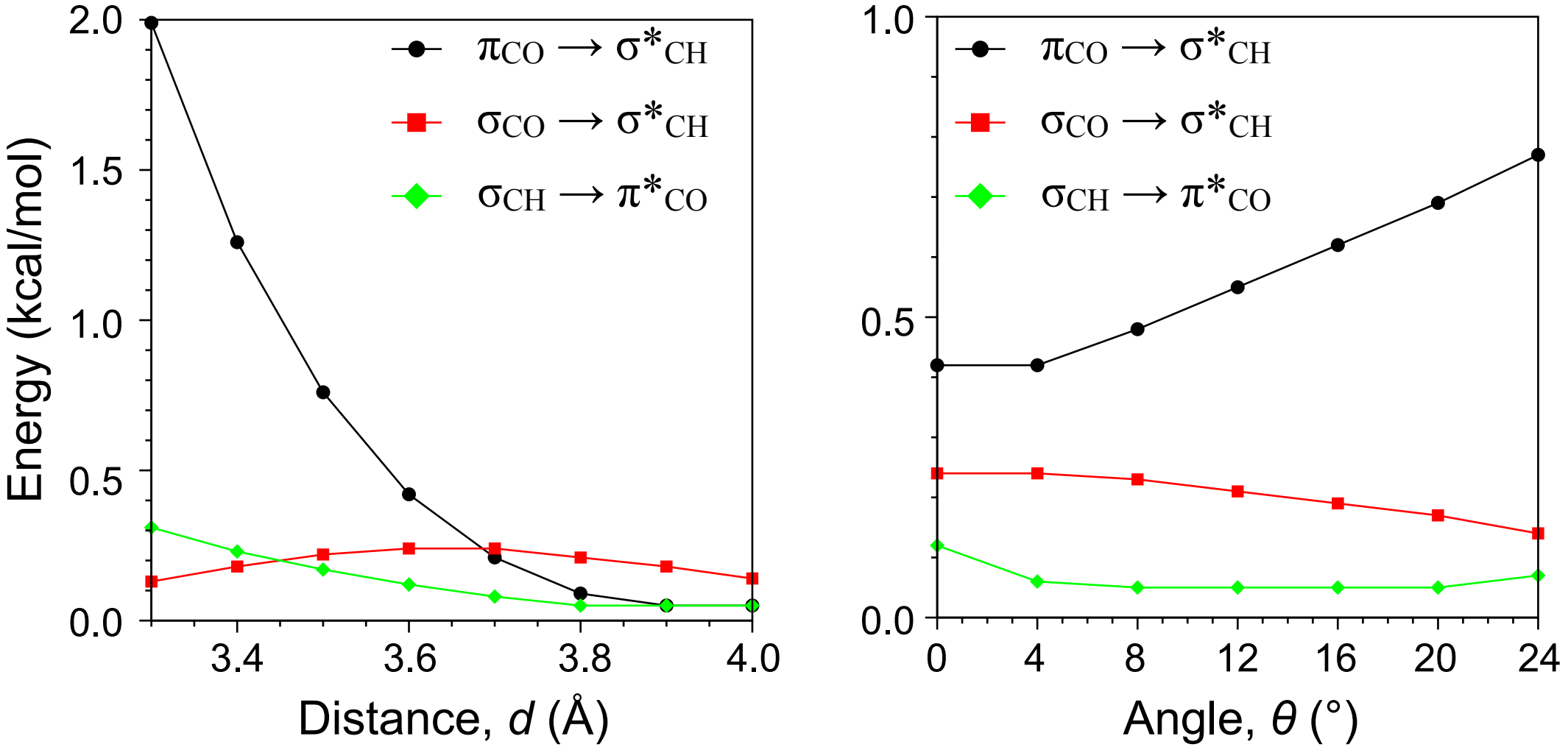
[18] C. Farès, N.-A. Lakomek, K. F. A. Walter, B. T. C. Frank, J. Meiler, S. Becker, C. Griesinger, *J. Biomol. NMR* **2009**, *45*, 23–44.

[19] N. Tjandra, S. E. Feller, R. W. Pastor, A. Bax, *J. Am. Chem. Soc.* **1995**, *117*, 12562–12566.

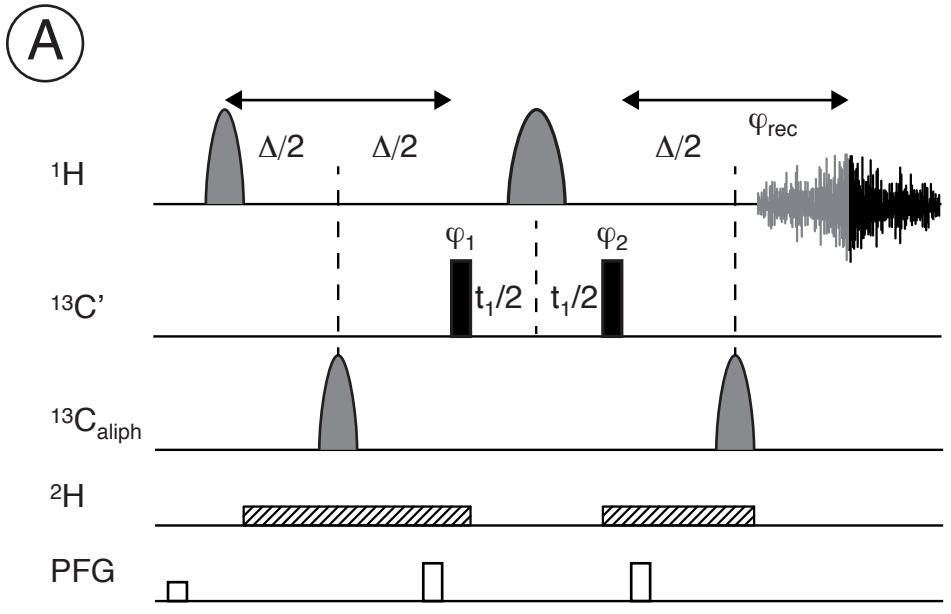
[20] S. Vijay-Kumar, C. E. Bugg, W. J. Cook, *J. Mol. Biol.* **1987**, *194*, 531–544.

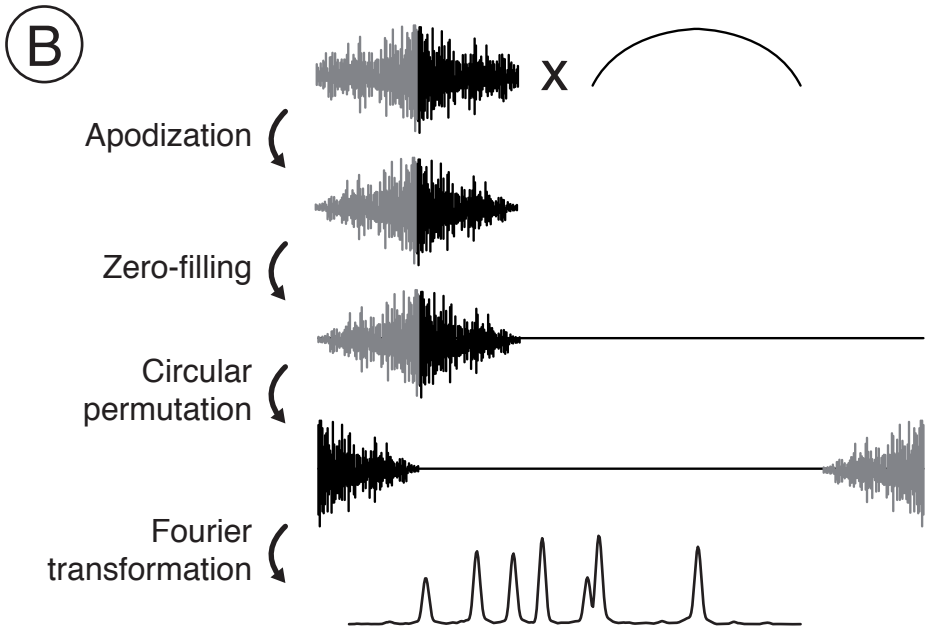
[21] S. Grzesiek, F. Cordier, V. Jaravine, M. Barfield, *Prog. Nuc. Mag. Res. Spect.* **2004**, *45*, 275–300.

Supplementary Figure S1

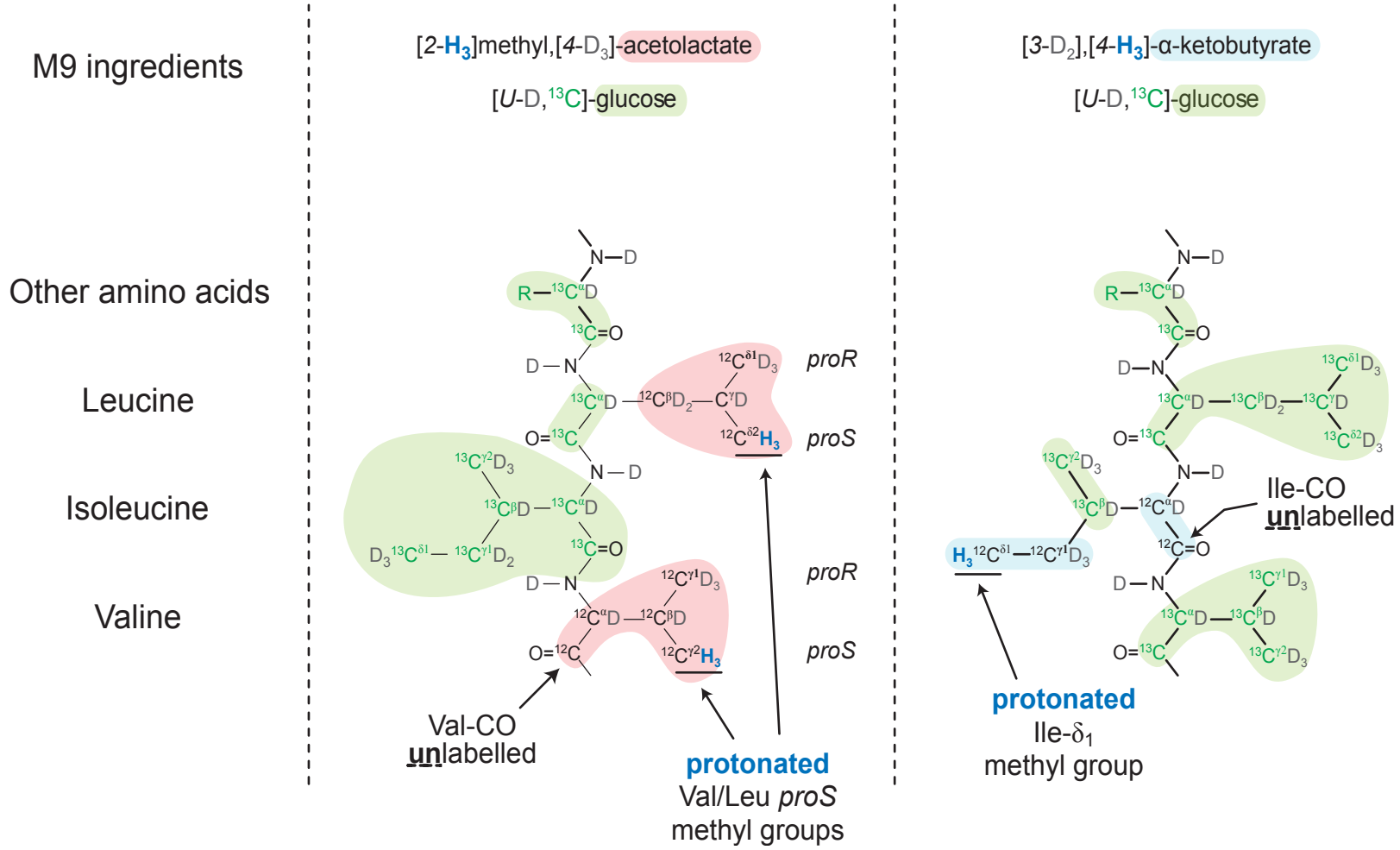


Supplementary Figure S2

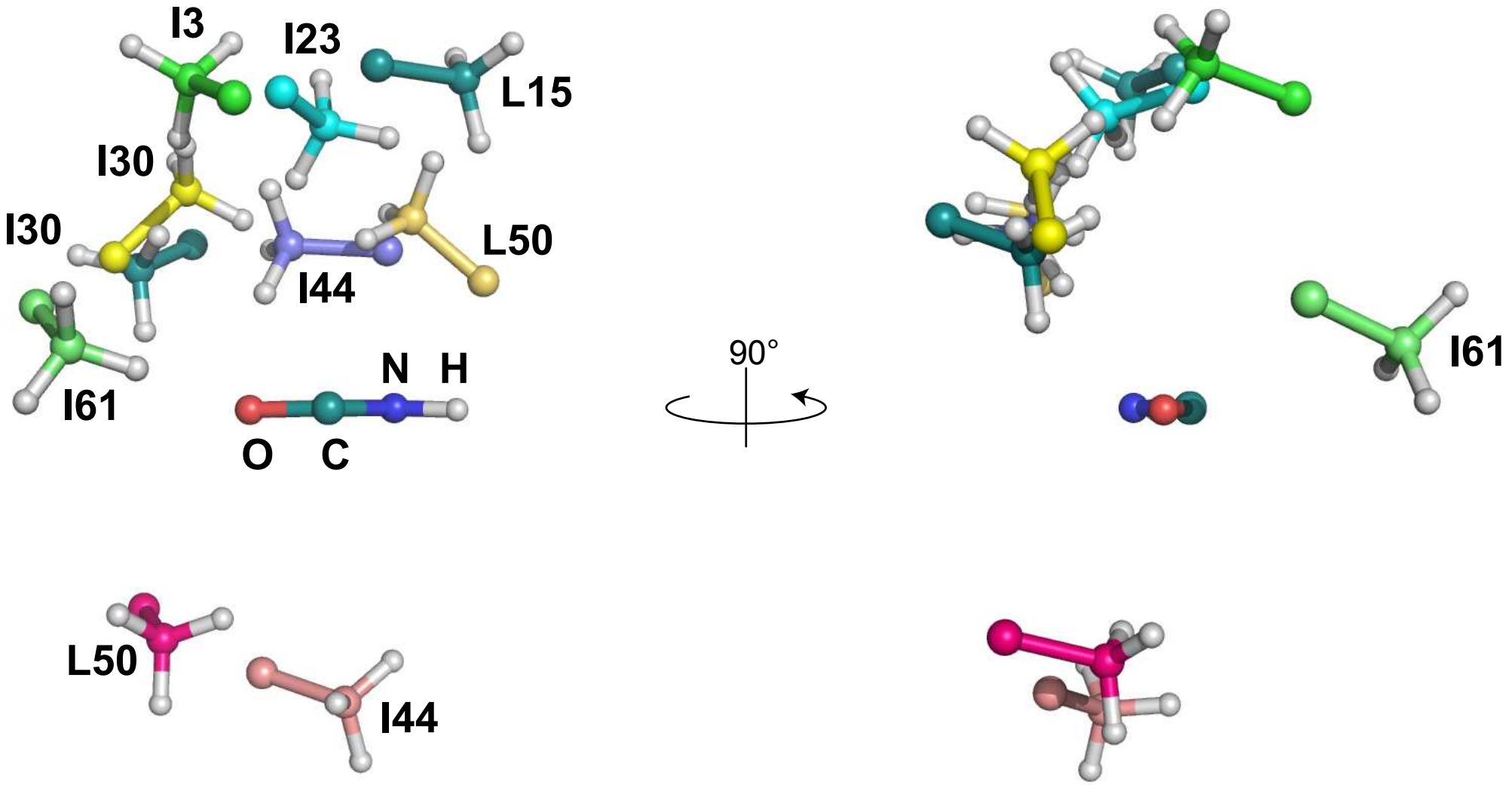




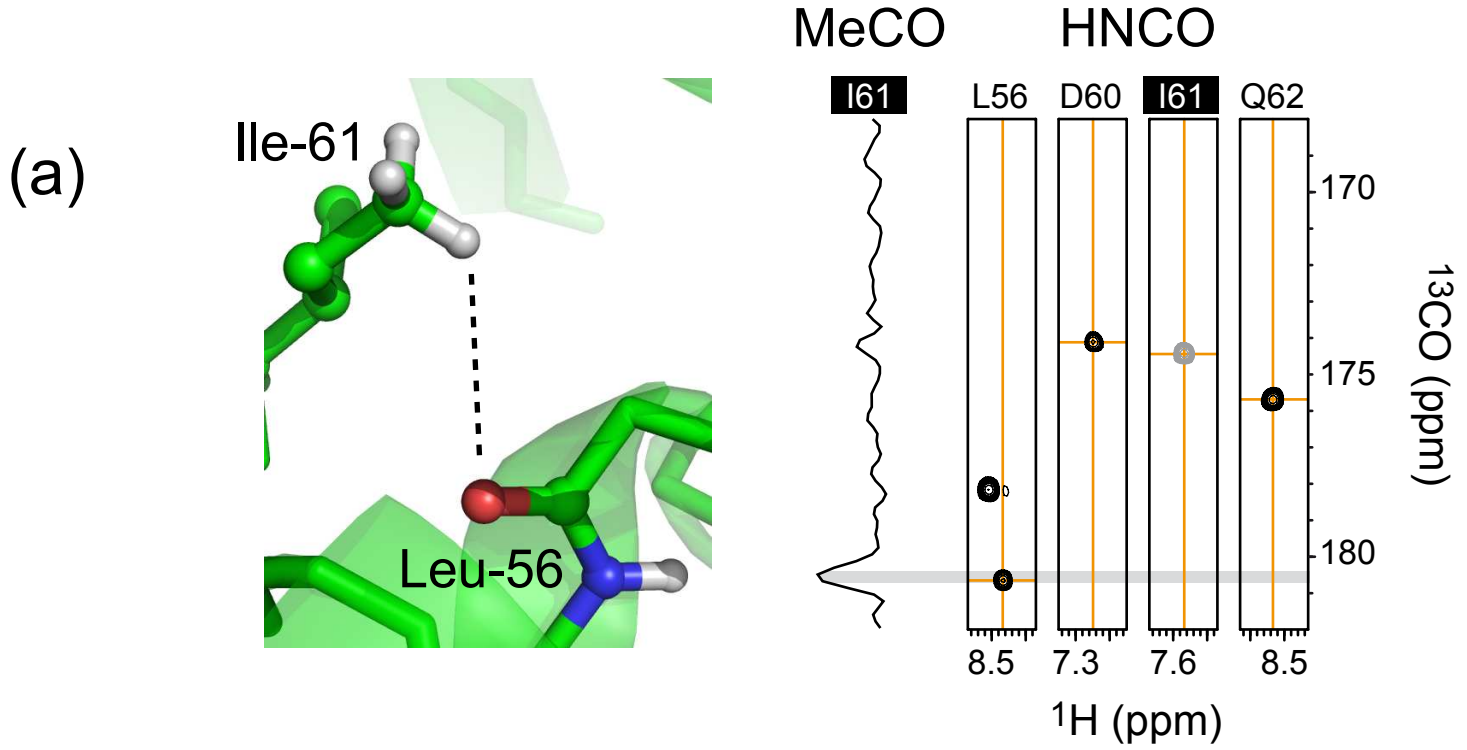
Supplementary Figure S3

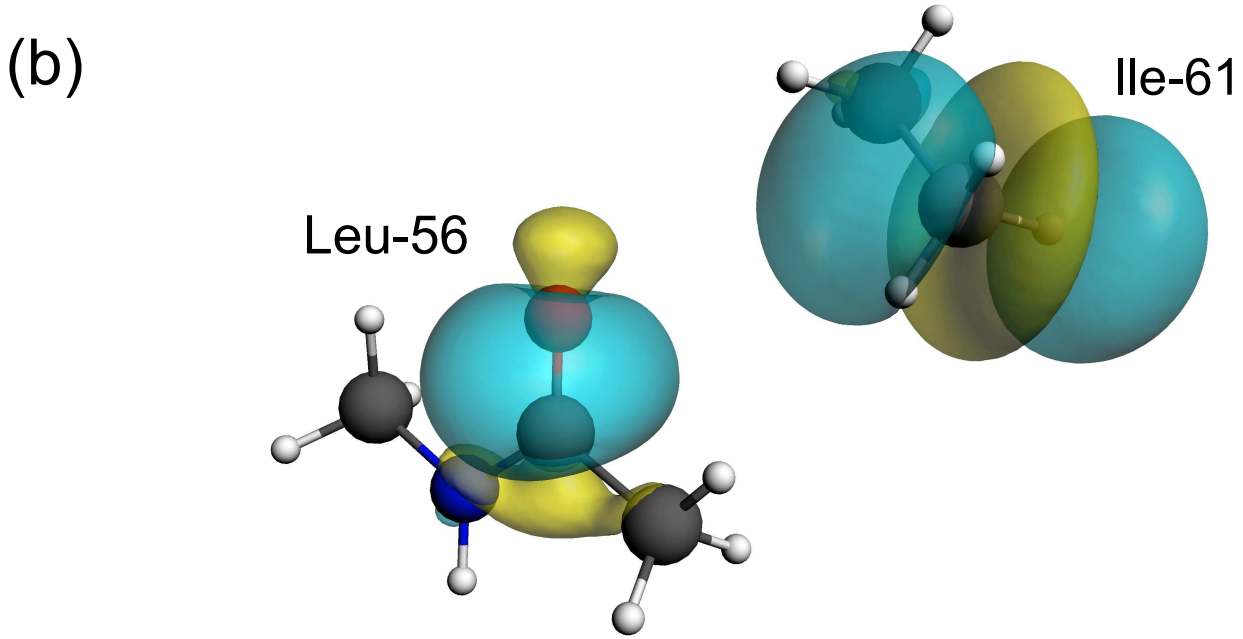


Supplementary Figure S4

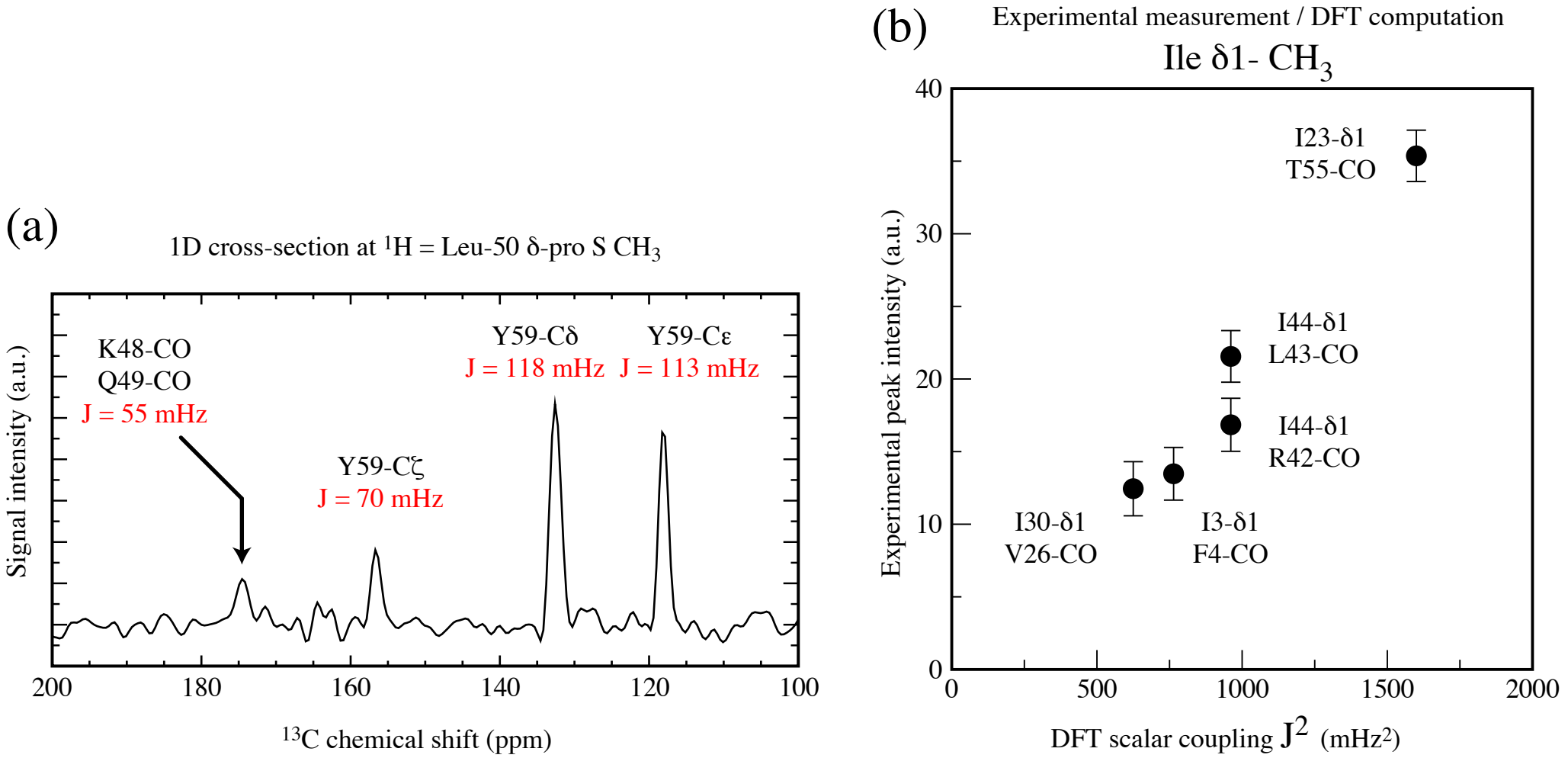


Supplementary Figure S5

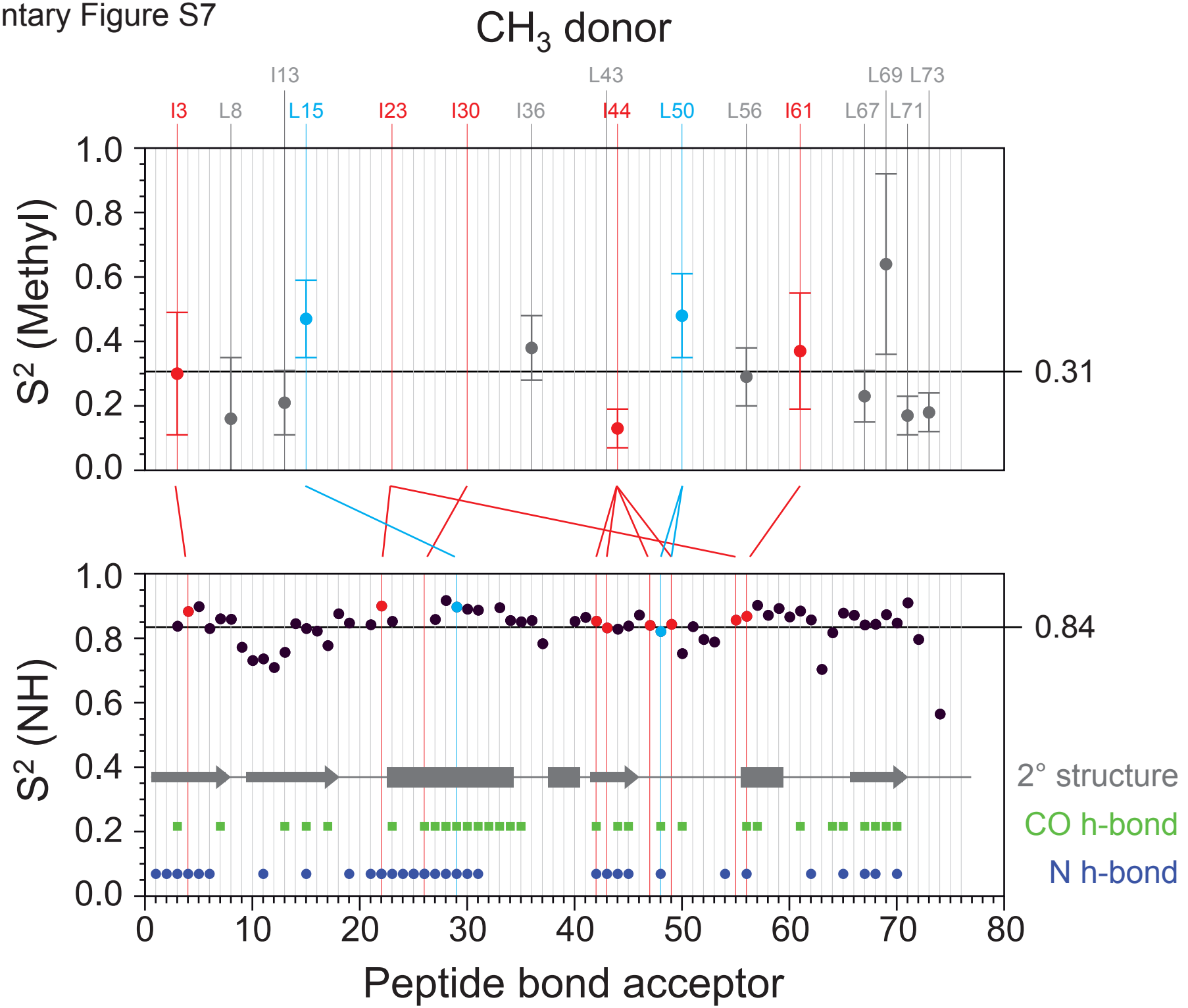
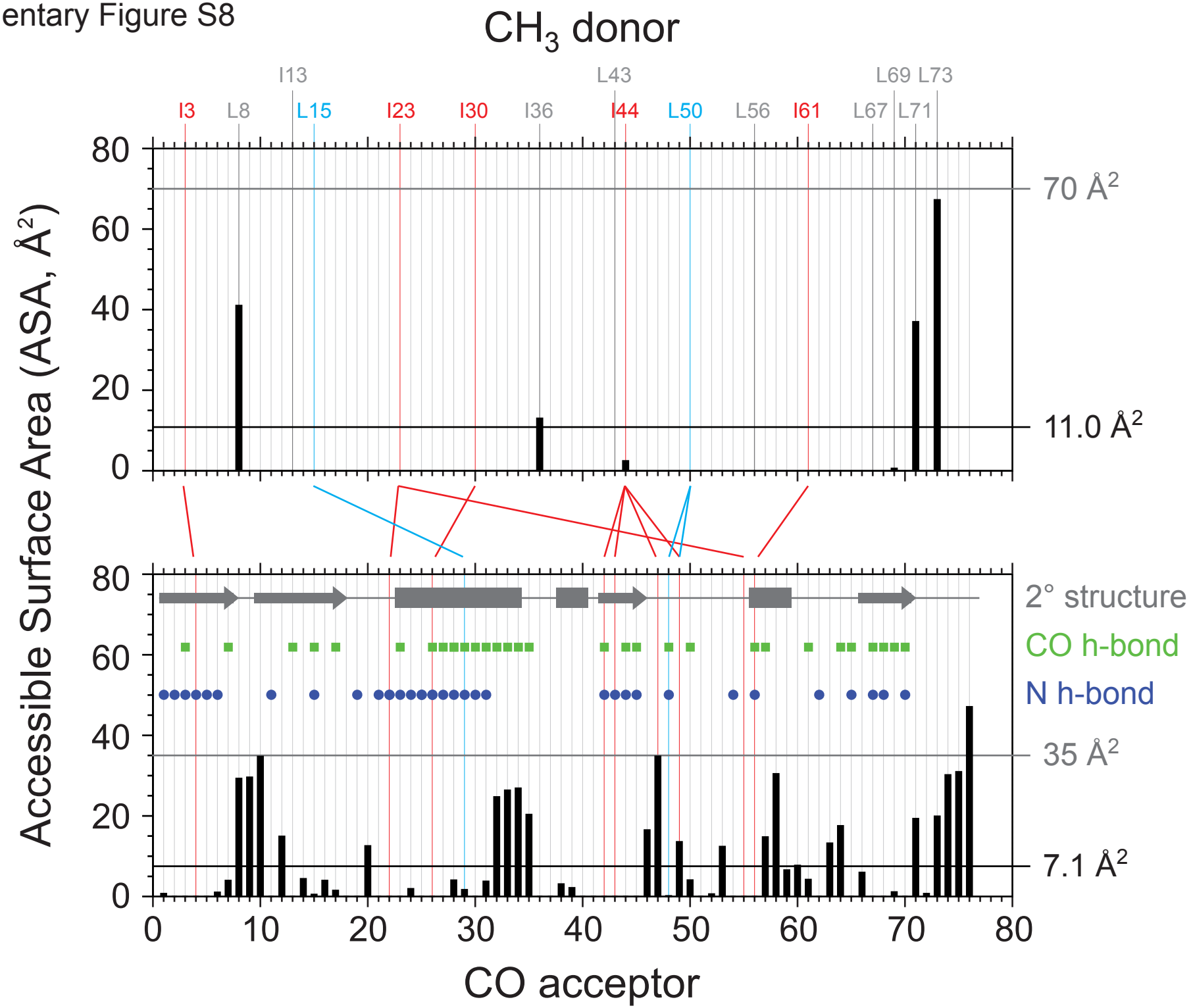




Supplementary Figure S6



Supplementary Figure S7



Supplementary Figure S8