Etudes des assemblages biomoléculaires par RMN

Jérôme Boisbouvier

Institut de Biologie Structurale - Jean-Pierre Ebel

Jérôme Boisbouvier

Dossier de Titres et Travaux pour obtenir :

L'Habilitation à Diriger des Recherches

Université Joseph Fourier – Grenoble 1
UFR de Biologie

"Etudes des assemblages biomoléculaires par RMN"

Soutenue le 22 octobre 2008, devant le jury:

Frédéric Dardel,	rapporteur
Eric Guittet,	rapporteur
Michel Kochoyan,	rapporteur
Bernhard Brutscher,	examinateur
Eva Pebay-Peyroula	Présidente

SOMMAIRE

Article 1 : *Relaxation Optimized NMR Spectroscopy of Methylene Groups in Protein and Nucleic Acids.*

Miclet, E., Williams, D., Clore, M., Bryce, D., Boisbouvier, J. & Bax, A. **J. Am. Chem. Soc** 126, 10560-70 (2004).

Article 2 : *Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids.*

Boisbouvier, J., Delaglio, F. & Bax, A. **Proc. Natl. Acad. Sci.** USA, 100, 11333-8 (2003).

Article 3 : *Liquid Crystal NMR Structure of HIV TAR RNA Bound to its SELEX RNA Aptamer Reveals the Origins of the High Stability of the Complex.*

Van Melckebeke, H., Devany, H.H., Di Primo, C., Beaurain, F., Toulmé, J.J., Bryce, D.L. Boisbouvier, J. **Proc. Natl. Acad. Sci. USA** 105, 9210-5 (2008)**.**

Article 4 : *High-Accuracy Distance Measurement Between Remote Methyls in Specifically Protonated Proteins.*

Sounier, R., Blanchard, L., Wu, Z. & Boisbouvier J. **J. Am. Chem. Soc** 129, 472-3 (2007).

CURRICULUM VITAE

Jérôme BOISBOUVIER

E-mail : jerome.boisbouvier@ibs.fr
Tel : 04 76 38 95 60 - Fax : 04 76 88 54 94

Institut de Biologie Structurale
41, Rue Jules Horowitz
38027 Grenoble Cedex 1

Diplômes

2000	Doctorat de **Physique**, Université J. Fourier (Grenoble).
1997	Magistère de **Chimie**, Ecole Normale Supérieure (Paris).
1996	DEA de **Biophysique** Moléculaire, Université P. & M. Curie (Paris).
1995	Maîtrise de **Chimie**, Ecole Normale Supérieure (Paris).
1994	Admission à l'Ecole Normale Supérieure, spécialité **Biologie** (Paris).

Activités de Recherche

2008-	Chargé de Recherche de 1ère classe au CNRS **Institut de Biologie Structurale**
2004-2008	Chargé de Recherche de 2nde classe au CNRS **Institut de Biologie Structurale**
2001-04	Post-doc, **National Institutes of Health**, Directeur : A. Bax (Bethesda, USA).
2000-01	Scientifique du contingent, Centre de Recher. du Service de Santé des Armées.
1997-00	Doctorant, **Institut de Biologie Structurale**, Directeur: D. Marion (Grenoble).
1999	Stage à **University of Georgia**, Directeur: Pr. P Legault (Athens, USA).
1996-97	Stage de magistère, **Institut de Biologie Structurale**, Directeur: D. Marion.
1996	Stage de DEA, **Institut Pasteur**, Directeur: M. Delepierre (Paris).

Activités d'enseignement et d'encadrement

01/08-09/11	Responsable d'un **stage postdoctoral** (Bourse de l'ARC)
06/08-08/09	Responsable d'un **stage postdoctoral** (Financement sur contrat)
10-12/07	Responsable d'un stage de **Master Life Science & Tech.** (Univ. Leiden & Delft)
04-06/07	Responsable d'un stage de **M1 Chimie/Biologie** Univ. Grenoble 1
06/06-07/08	Responsable d'un **stage postdoctoral** (Bourse Ministère Ens. Sup. et Recherche)
03-06/06	Responsable d'un stage de de **M1 Physique** Univ. Grenoble 1.
09/05-03/06	Responsable d'un **stage postdoctoral** (Bourse ANRS)
04-06/05	Responsable du stage de **M1 Sc. du Vivant** Univ. Grenoble 1
04/05-09/06	Responsable d'un **stage postdoctoral** (poste rouge CNRS)
09/04-06/08	**Codirecteur de thèse** (Univ. Grenoble 1).
2001	Qualifié aux fonctions de Maître de Conférences en Chimie et Biochimie.
03-06/99	Responsable d'un stage de **M1 Physique** Univ. Grenoble 1.
09/98-06/00	**Enseignant de Chimie, niveau Licence & Agrégation**, Université de Grenoble.
09/95-06/01	Préparation des classes préparatoires aux épreuves orales des concours d'entrée aux Grandes Écoles (Paris, Grenoble).

Distinctions, Prix et Subventions

2008	**Prix Paoletti** du département des Sciences du Vivant du CNRS.
2008	**Subvention de l'ANRS** : *"Etude Fonctionnelle et Structurale en solution des complexes RNP (eIF3-IRES) engagés dans l'initiation de la traduction du VHC"*. Co-participant (112 k€ / 2 ans).
2008	**Médaille de Bronze du CNRS**, présenté par les départements Chimie et Sciences du Vivant.
2007	Subvention du **programme interdisciplinaire CNRS Interface Physique-chimie-biologie**, Soutien à la prise de risque : "*Nano Machines thermophiles en action : une approche par RMN en temps réel"*. **Coordinateur** (55 k€ / 1 an).
2006	**Grant Human Frontier Science Program** (rang 8/240 équipes internationales) : "*RNA Shape Recognition and Structure in MicroRNA Processing"*. **Coordinateur** (400 k$ / 3 ans).
2006	**ANR jeune chercheur** : "*RNA Shape Recognition and Structure in MicroRNA Processing"*. **Coordinateur** (50 k€ / 4 ans).
2006	**Subvention EcoS / SECYT** pour financer les échanges avec l'Institut de Biologie de Rosario (Argentine) : *"Etudes structurales par RMN de la biogenèse des microRNAs"*. Co-participant (16 k€ / 3 ans).
2005	Subvention du **Fond de Recherche France-Canada, accords de cooperation CNRS / Instituts de Recherche en Santé du Canada** pour financer les échanges avec l'université d'Ottawa : *"Etude structurale du complexe entre la boucle TAR de l'ARN du VIH et son aptamère TAR*GA par RMN en milieu cristallin liquide"*. **Coordinateur** (13 k€ / 2 ans).
2004	**Career Development Award Human Frontier Science Program** : "*Quaternary Structure Study of large complexes using Liquid Crystal and Cross-Correlated NMR* ". **Coordinateur** (180 k$ / 3 ans).
2003	Classé 1er au concours de recrutement des chargés de recherche CNRS (section 16)
2001	**Prix du Jeune Chercheur** de la Société Française de **Biophysique**.
2001	Bourses postdoctorales internationales décernées par l'INSERM (déclinée), EMBO (déclinée), and Human Frontier Science Program Organization (acceptée).
1998-00	Bourse du ministère de la Recherche : Allocataire Moniteur Normalien.
1994-98	Elève fonctionnaire-stagiaire de l'Ecole Normale Supérieure (Paris, Fr).

Autres Activités :

2007-	Membre du commité de rédaction de *IBS Actualités.*
2006-	Co-organisateur de *International School on High Field NMR Spectroscopy for Solids and Liquids*, Les Houches (France) 15-19 May 2006.
2001-	Reviewer pour Journal of Biomol. NMR, Journal of Am. Chem. Soc., Proteins.

Publications :

• 25 articles acceptés dans Biochemistry (1) ; Dev. Cell (1); Eur. J. Biochem. (1); FEBS letters(1); J. Am. Chem. Soc. (7); J. Biomol. NMR (10); J. Mol. Biol. (1); Proc. Natl. Acad. Sci. USA (2) ; Protein Science (1).
• 10 publications signées en 1er auteur et 6 en dernier auteur ou correspondant.
• h-index : 12 (Web of Science®).

LISTE DES PUBLICATIONS :

1. Rasia, R., Bologna, N. Imbert, L. Palatnik*, J. & Boisbouvier*, J. "*Specificity of the recognition of pre-miRNA target by Arabidopsis primary microRNA processing protein HYL1"* **(in preparation).**

2. Imbert, L., Blot, D., Favier, A. & Boisbouvier*, J. "*Automated optimization of protein solubility and stability"* **(in preparation).**

3. Sounier, R., Ayala, I., Dura, A, Franzetti, B., Gans, P., Retel, J., Tugarinov, V., Kay, L.E., Grishaev, A. & Boisbouvier*, J. "*Detection of long range NOE in macromolecular assemblies beyond 100 kDa: possibilities and limits"* **(in preparation).**

4. Rasia, R., Palatnik, J. Brutscher, B. & Boisbouvier*, J. "*Fast Fold Determination using chemical shift information combined with liquid crystal NMR"* **(in preparation).**

5. Sounier, R., Wu, Z., & Boisbouvier*, J. "*Long range distance and angular information between remote methyls measured by liquid crystal NMR"* **(in preparation).**

6. Gans, P., Hamelin, O., Ayala, I., Dura, A, Franzetti, B., Sounier, R., Boisbouvier*, J *"Stereospecific Labeling of Leucine and Valine Methyls and Application to the NMR Study of Biological NanoMachines "* **(in preparation).**

7. Farjon, J., Boisbouvier, J., Schanda, P., Simorre, J.-P., Pardi, A. & Brutscher*, B. *"New sensitive longitudinal relaxation optimized NMR experiments for the study of RNA in solution"* **(in preparation).**

8. Amero, C., Schanda, P., Dura, A., Ayala, I., Marion, D., Franzetti, B., Brutscher, B., Boisbouvier*, J. *"2D NMR Spectroscopy of High Molecular Weight Protein Assemblies in Second Time Range"* **(in preparation).**

9. Ayala, I., Sounier, R., Use, N., Gans, P., & Boisbouvier*, J. *"An efficient protocol for the complete incorporation of methyl specifically protonated alanine in perdeuterated protein"* **Journal of Biomolecular NMR** (in press).

10. Rasia, R., Noirclerc-Savoye, M., Gallet, B., Bologna, N., Plevin, M., Blanchard, L., Palatnik, J., Brutscher, B., Vernet*, T. & Boisbouvier*, J. *"Parallel screening and optimization of protein constructs for structural studies"* **Protein Science** (in press).

11. Perard, J., Rasia, R., Medenbach, J., Ayala, I., Boisbouvier, J., Drouet, E. & Baudin*, F. *"Initiation factor eIF3 subunit b interacts with HCV IRES RNA through its N-terminal RNA recognition motif"* **FEBS Letters** (in press).

12. Van Melckebeke, H. Devany, H.H., Di Primo, C., Beaurain, F., Toulmé, J.J., Bryce*, D.L. Boisbouvier*, J. "*Liquid Crystal NMR Structure of HIV TAR RNA Bound to its SELEX RNA Aptamer Reveals the Origins of the High Stability of the Complex*" **Proc. Natl. Acad. Sci. USA** 105, 9210-5 (2008).

13. Palatnik, J., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D., Carrington, J., Weigel, D. *"Sequence and expression differences underlie functional specialization of arabidopsis microRNAs miR159 and miR319"* **Dev. Cell** 13, 115-25 (2007).

14. Schanda, P., Lescop, E., Falge, M., Sounier, R Boisbouvier* J., Brutscher* B. "*Sensitivity-optimized experiment for the measurement of residual dipolar couplings between amide protons*" **J. Biomol NMR** 38, 47-55 (2007).

15. Sounier, R. Blanchard, L. Wu, Z. & Boisbouvier* J. "*High-Accuracy Distance Measurement Between Remote Methyls in Specifically Protonated Proteins"* **J. Am. Chem. Soc** 129, 472-3 (2007).

16. Van Melckebeke, H., Pardi, A., Boisbouvier, J., Simorre, J.-P. & Brutscher*, B. "*High resolution base-type selective HCN experiment for nucleic acids"* **J. Biomol NMR** 32, 265-71 (2005).

17. Jaroniec , C., Boisbouvier, J., Nikonowicz, E. & Bax A. *"Quantitative J correlation methods for the accurate measurement of 13C-15N dipolar couplings in nucleic bases"* **J. Biomol NMR**, 31, 231-41 (2005).

18. Miclet, E., Boisbouvier, J., & Bax* A. "*Measurement of eight coupling from a single 3D NMR multiplet in proteins and nucleic acids"* **J. Biomol NMR**, 31, 201-16 (2005).

19. Boisbouvier, J., Bryce, D., O'Neil-Cabello, E., Nikonowicz, E. & Bax* A. *"Resolutionoptimized NMR measurement of 1DCH, 1DCC and 2DCH residual dipolar couplings in nucleic acids bases."* **J. Biomol. NMR**, 30, 287-301 (2004).

20. Bryce, D., Boisbouvier, J. & Bax* A. *"Experimental and Theoretical Determination of Nucleic Acid Magnetic Susceptibility: Importance for the Study of Dynamics by FieldInduced Residual Dipolar Couplings"* **J. Am. Chem. Soc** 126, 10820-1 (2004).

21. Miclet, E., Williams, D., Clore, M., Bryce, D., Boisbouvier*, J. & Bax* A. "*Relaxation Optimized NMR Spectroscopy of Methylene Groups in Protein and Nucleic Acids***". J. Am. Chem. Soc** 126, 10560-70 (2004).

22. Boisbouvier, J. , Delaglio, F. & Bax* A. "*Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids"* **Proc. Natl. Acad. Sci.** USA, 100, 11333-8 (2003).

23. Boisbouvier, J. , Wu, Z., Ono, A, Kainosho, M. & Bax* A. *"Anisotropic Rotational Diffusion Tensor Determination of Nucleic Acids from 13C NMR Relaxation"* **J. Biomol. NMR**, 27, 133-42 (2003).

24. Boisbouvier, J. & Bax* A.*" Long-Range Magnetization Transfer Between Uncoupled Nuclei by Dipole-Dipole Cross-correlated Relaxation : A precise Probe of β-Sheet Geometry in Proteins"* **J. Am. Chem. Soc.** 124, 11038-45 (2002).

25. Brutscher, B., Boisbouvier, J., Kupce, E., Tisné, C., Dardel, F., Marion, D. & Simorre* J.-P. "*Base-type-selective high resolution ^{13}C edited NOESY for sequential assignment of large RNA*" **J. Biomol. NMR,** 19, 141-51 (2001).

26. Boisbouvier, J., Brutscher, B., Pardi, A., Marion, D Simorre*, J.-P. "*NMR determination of*

sugar puckers in nucleic acids from CSA-dipolar cross-correlated relaxation" **J. Am. Chem. Soc.**, 122, 6779-80 (2000).

27. Boisbouvier, J., Blackledge, J., Sollier, A. & Marion* D. "*Simultaneous determination of disulphide bridge topology and three-dimensional structure using ambiguous intersulphur distance restraints : Possibilities and limitations*" **J. Biomol. NMR.,** 16, 197-208 (2000).

28. Delepierre* M., Prochnicka-Chalufour, A., Boisbouvier, J. & Possani, L.D. "*Pi7, an Orphan Peptide from the Scorpion Pandinus Imperator : A ^{1}H NMR Analysis Using a Nano-NMR Probe*" **Biochemistry,** 38, 16756-65 (1999).

29. Boisbouvier, J., Gans, P., Blackledge, M., Brutscher B. & Marion*, D. "*Long-range structural information in NMR studies of paramagnetic molecules from electron spin - nuclear spin cross-correlated relaxation*" **J. Am. Chem. Soc**., 121, 7700-01 (1999).

30. Boisbouvier, J., Brutscher, B,. Simorre J.-P & Marion* D. "*^{13}C spin relaxation measurements in RNA : Sensitivity and resolution improvement using spin-state selective correlation experiments*" **J. Biomol. NMR**., 14, 241-52 (1999).

31. Brutscher, B., Boisbouvier, J., Pardi, A., Marion, D. & Simorre* J.-P. "*Improved Sensitivity and Resolution in ^{1}H-^{13}C NMR Experiments of RNA*" **J. Am. Chem. Soc.**, 120, 11845-51 (1998).

32. Boisbouvier, J., Albrand, J.P., Blackledge, M. Jaquinod, M., Schweitz, H., Lazdunski, M. & Marion* D. "*A Structural Homologue of Mammal Pancreatic Colipase in Black Mamba Venom Revealed by NMR Floating Disulphide Bridge Analysis*" **J. Mol. Biol**., 283, 205-19 (1998).

33. Boisbouvier, J., Prochnicka-Chalufour, A., Nieto, A.R., Torres, J.A., Nanard, N., Rodriguez, M.H., Possani, L.D. & Delepierre* M., "*Structural information on a cecropin-like synthetic peptide, Shiva-3 toxic to the development of Plamodium berghei*" **Eur. J. Biochem**., 257, 263-73 (1998).

ACTIVITES D'ENCADREMENT

Les différents sujets qui ont constitué mon travail de recherche ont été menés en collaboration avec d'autres chercheurs ou étudiants. Depuis mon installation à l'IBS en tant que chargé de recherche CNRS, j'ai eu la chance d'encadrer cinq post-doctorants. En 2005 et 2006, j'ai formé les Drs M. Devany (bourse CNRS) et H. Van Melckebeke (bourse ANRS) à la RMN en milieu cristallin liquide, des acides nucléiques et de leurs complexes. Ils occupent désormais respectivement, un poste fixe de *spectrometer manager* of City University of New York (USA), et un poste de chercheur post-doctoral en RMN du solide à l'ETH de Zurich. De juin 2006 à juillet 2008, j'ai encadré Dr Rasia (bourse ministère de la recherche) qui a eu en charge l'étude des complexes protéines/ARNs impliquées dans la biogenèse des microRNAs chez les plantes. Dr Rasia occupe désormais un poste de chercheur permanent en Argentine (SECyT). Ces travaux ont déjà donné lieu à deux articles (van Melckebeke et al. 2005 & 2008) et trois autres, avec R. Rasia en premier auteur, sont en préparation. Depuis janvier 2008, j'accueille Dr. M. PLEVIN, chercheur expérimenté qui a déjà effectué 4 ans de recherche post-doctorale au *Ontario Cancer Institute,* sous la direction du Prof. M Ikura. Dr Plevin a en charge l'étude des interactions protéines/RNA et protéines/protéines impliquées dans la maturation des microRNAs oncogènes chez l'homme (Bourse ARC – 3 ans). L'ensemble de ces thématiques requiert la préparation d'échantillons d'ARN pour les études structurales en solution. Pour ce faire, j'ai mis en place une plateforme de production et de marquage isotopique des ARNs et j'encadre L. Imbert (ingénieur d'étude CNRS recruté sur contrat) qui a implémenté cette activité et assure la production à grande échelle des ARNs.

L'étude des assemblages biomoléculaires de grande taille a été initiée au cours de la thèse de R. Sounier (bourse ministère de la recherche) dont j'ai assuré la formation et la co-direction depuis octobre 2004 avec Dr. Jean-Pierre Simorre (IBS, Grenoble). Deux premiers articles issus de ce travail ont déjà été publiés (Sounier et al. 2007; Schanda et al. 2007) et deux autres, avec R. Sounier en premier auteur, sont actuellement en préparation. Ces travaux ont débouché sur l'étude RMN hors d'équilibre de NanoMachines hyperthermophiles de plusieurs centaines de kDa. Ce dernier projet a été confié au Dr. C Amero qui a rejoint l'équipe en juin 2008, pour être formé sous ma direction à la RMN des assemblages supramoléculaires.

J'ai également encadré, plusieurs étudiants en Master de Biologie de l'UJF. Ils ont été impliqués dans la mise au point de protocoles de protonation spécifique des méthyles des isoleucines pour K. Treche (2005) et des alanines pour N. Usé (2007). Deux étudiants en Masters de Physique ont été impliqués dans les développements pour l'attribution des ponts disulfures pour A. Sollier (1999) et M Falges (2006) pour la mesure rapide de RDC ^{1}H-^{1}H. Plus récemment, j'ai également formé J. Retel, un étudiant hollandais effectuant son stage de *Master Life Science & Technology* (programme commun des universités de Leiden et de Delft, NL 2007). Son travail de stage consiste à généraliser les outils développés lors de la thèse de Remy Sounier. Cette implication dans la formation des étudiants de Master a été validée par quatre articles dans lesquels les étudiants de Master sont co-auteurs (Boisbouvier et al. 2000; Scanda et al. 2007; Ayala et al., in press; Sounier et al., en préparation) et vient compléter l'expérience d'enseignement en Licence et Agrégation que j'ai acquise pendant deux ans de monitorat.

ACTIVITES DE RECHERCHE

L'ensemble de ma formation et mon parcours en recherche se situe à l'interface entre la Physique, la Biologie et la Chimie. Je me suis spécialisé dans le développement et l'application de la spectroscopie par Résonance Magnétique Nucléaire pour l'étude structurale et dynamique des macromolécules biologiques. Au cours de ma thèse puis d'un stage post-doctoral réalisé aux *National Institutes of Health*, j'ai contribué à étendre les possibilités qu'offre la RMN pour l'étude structurale des acides nucléiques. Ces travaux ont notamment permis, de repousser considérablement les limites en taille des ARN et ADN étudiables par RMN[1-2], tout en augmentant la précision des modèles structuraux qui peuvent être obtenus par cette spectroscopie[3,4]. Depuis mon recrutement en 2004 à l'Institut de Biologie Structurale, l'ensemble de mes activités concerne l'étude des assemblages macromoléculaires par RMN. Ces travaux inclus le développement méthodologique d'outils innovants et leurs applications à des systèmes d'intérêt majeur en biologie.

A) Expériences RMN Adaptées des Acides Nucléiques de Grandes Tailles.

Période :1998-2008

Bourses :
Allocation de Moniteur Normalien
Human Frontier Science Program

Laboratoires :
IBS (Grenoble)
NIH (Bethesda, USA)

Collaboration :
A. Pardi (Colorado)
F. Dardel (Paris)
C. Tisné (Paris)
M. Kainosho (Tokyo)

Publications :

- "*Improved Sensitivity and Resolution in ^{1}H-^{13}C NMR Experiments of RNA*" Brutscher, B., Boisbouvier, J., Pardi, A., Marion, D. & Simorre J.-P. **J. Am. Chem. Soc**., 120, pp 11845-11851 (1998).
- "*^{13}C spin relaxation measurements in RNA : Sensitivity and resolution improvement using spin-state selective correlation experiments*" Boisbouvier, J., Brutscher, B,. Simorre J.-P & Marion, D. **J. Biomol. NMR**., 14, pp 241-252 (1999).
- "*Base-type-selective high resolution ^{13}C edited NOESY for sequential assignment of large RNA*" Brutscher, B., Boisbouvier, J., Kupce, E., Tisné, C., Dardel, F., Marion, D. & Simorre J.-P. **J. Biomol. NMR**, 19, pp 141-151 (2001).
- "*Anisotropic Rotational Diffusion Tensor Determination of Nucleic Acids from ^{13}C NMR Relaxation*" Boisbouvier, J. , Wu, Z., Ono, A, Kainosho, M. & Bax A. **J. Biomol. NMR**, 27, pp133-142 (2003).
- "*Relaxation Optimized NMR Spectroscopy of Methylene Groups in Protein and Nucleic Acids*" Miclet, Williams, Clore, Bryce, Boisbouvier & Bax **J. Am. Chem. Soc** 126, pp 10560 (2004).
- "Resolution-enhanced base-type-edited HCN experiment for RNA" Van Melckebeke, H., Pardi, A., Boisbouvier, J., Simorre, J.-P., Brutscher, B. **J. Biomol. NMR**, 32, pp265-271 (2005).
- *"New sensitive longitudinal relaxation optimized NMR experiments for the study of RNA in solution."* Farjon, J., Boisbouvier, J., Schanda, P., Simorre, J.-P., Pardi, A. & Brutscher*, B **(in preparation).**

L'introduction de méthodes d'enrichissement isotopique (^{13}C, ^{15}N) des acides nucléiques a permis le développement de nombreuses expériences RMN hétéronucléaires pour l'étude dynamique et structurale des ADN et ARN. Mais ces techniques sont limitées par la relaxation des spins (vitesse de retour à l'équilibre de l'aimantation) qui augmente avec la taille moléculaire.

Ces effets de relaxation sont responsables de la perte de sensibilité et de résolution des spectres RMN, restreignant l'étude structurale par RMN à des acides nucléiques de moins de 15kDa.

1) Amélioration de la sensibilité et de la résolution des spectres de corrélation ^{1}H-^{13}C des bases nucléiques.

En RMN biomoléculaire, la relaxation des signaux est principalement dûe aux interactions dipolaires entre spins et au CSA (*Chemical Shift Anisotropy*) de chaque noyau. Puisque les mêmes mouvements moléculaires modulent ces deux types d'interaction, il est possible d'observer un phénomène d'interférence entre deux interactions, également nommée Corrélation Croisée. Dans un système à deux spins couplés, la relaxation croisée dipole-CSA est responsable de l'élargissement d'une composante du doublet observé en RMN et du rétrécissement de la seconde composante. La sélection de la composante du doublet la plus fine par une séquence appropriée conduit à un gain en résolution et en sensibilité des expériences. Cette approche, dénommée TROSY (*Transverse Relaxation Optimized Spectroscopy*)[5-6] a été adaptée pour les acides nucléiques enrichis en ^{13}C. Cette nouvelle stratégie permet une augmentation de la résolution et de la sensibilité jusqu'à un facteur 5, aux champs magnétiques accessibles, des spectres correspondant aux bases nucléiques[7]. La sélection de la composante avec des propriétés de relaxation favorables permet l'utilisation de filtre ^{13}C-^{13}C et l'augmentation de la résolution avec une perte limitée de signal[8]. L'application à un ARN de 26 kDa est présentée sur la figure 1.

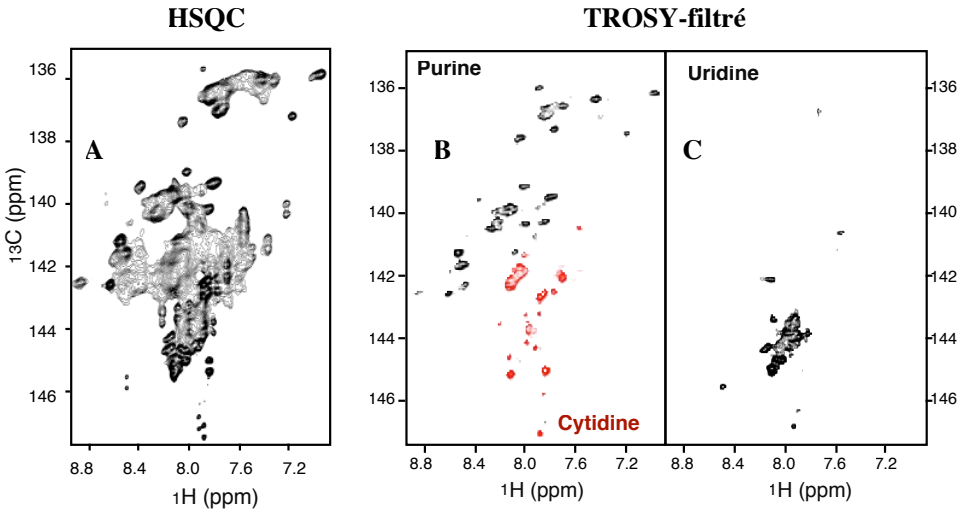

Figure 1. Spectres RMN ^{13}C-^{1}H correspondant aux bases de l'ARNt^{LYS3} (26 kDa). L'expérience a été enregistrée à 800MHz. **A**. La résolution des spectres RMN conventionnels est trop faible pour permettre une étude structurale par RMN. **B** & **C**. Le gain en résolution, obtenu avec l'expérience TROSY-filtré , permet d'envisager l'étude de la structure de plus gros ARN par RMN.

2) Amélioration de la résolution des spectres de corrélation des groupements CH_2.

Une large fraction des hydrogènes dans les acides nucléiques comme dans les protéines, correspond à des groupements méthylènes. L'étude précise de leur conformation ou de leur dynamique, est cependant limitée par leurs propriétés de relaxation très rapide, donnant lieu à une

faible sensibilité et résolution des spectres RMN. En utilisant simultanément six mécanismes de corrélation croisée, impliquant les interactions dipolaires ^{1}H-^{13}C et ^{1}H-^{1}H et le CSA ^{1}H et ^{13}C, il a été possible de réduire considérablement ces problèmes (annexe – article 1). Des gains en résolution et en sensibilité d'un facteur, 3.2 et 2 respectivement, sont obtenus à champ magnétique élevé[9], et ceux-ci restent substantiels à champ magnétique plus faible. Le gain en résolution obtenu, rend possible une analyse détaillée de régions spectrales réputées inaccessibles du fait de nombreuses superpositions spectrales.

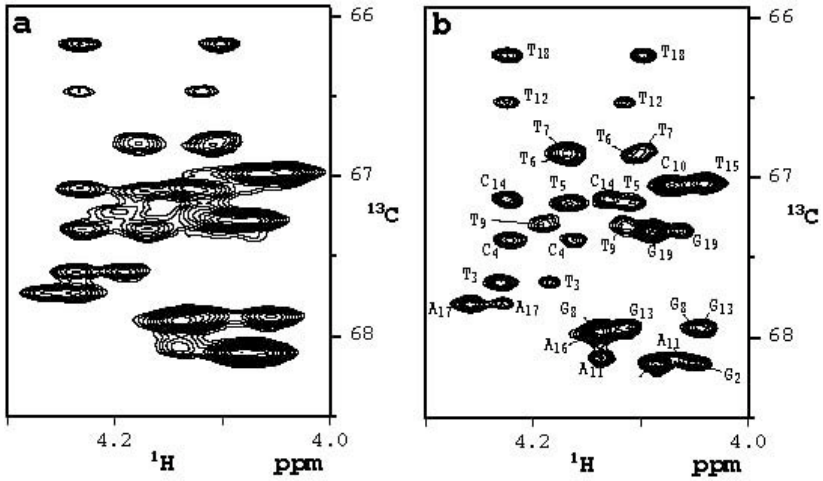

Figure 2. Spectres de corrélation ^{13}C-^{1}H des méthylènes $C_{5'}$ des acides nucléiques. Comparaison des expériences HSQC (a) optimisée pour les CH_2 et l'expérience CH_2-TROSY (b) enregistrée à 800 MHz. L'échantillon contient un ADN de 19 paires de bases dont un brin est enrichi en ^{13}C.

3) Application à l'étude de la structure et la dynamique des acides nucléiques.

Les séquences d'impulsions corrélant les fréquences des spins ^{1}H et ^{13}C sont des outils essentiels pour la plupart des expériences permettant l'étude des acides nucléiques. Des expériences tridimensionnelles de haute résolution ont été développées pour éditer les connectivités nOe (*nuclear Overhauser effect*) avec le gain en résolution de l'approche TROSY développée[8-9]. Les protocoles de calculs de structures standard sont basés sur ces contraintes de distance ^{1}H-^{1}H issues des nOe observés. La séquence TROSY a également été utilisée pour mesurer les couplages scalaires et dipolaires. Ces couplages apportent une information angulaire qui complète bien les contraintes de distances nOe et qui a récemment été montrée comme particulièrement utile pour affiner la structure des acides nucléiques (cf. § B.4 et § C.1).

Utilisant le gain en résolution et en sensibilité de l'approche TROSY, de nouvelles expériences ont été proposées pour sonder les caractéristiques dynamiques des biomolécules[10-11]. Avec les propriétés de relaxation favorable de la composante ^{13}C fine, une séquence d'impulsion très sensible a été développée afin de détecter les phénomènes d'échanges conformationnels. La contribution de cet échange à la relaxation de l'aimantation nous renseigne sur les mouvements de temps caractéristiques compris entre la milli et la microseconde. La détection de tels mouvements est importante car ceux-ci se produisent souvent pour des nucléotides impliqués dans la reconnaissance de partenaire(s) physiologique(s).

B) Développement de Nouvelles Sondes Structurales

Période : 1999-2004

Bourses :
Allocation Moniteur Normalien
Human Frontier Science Program

Laboratoires :
IBS (Grenoble),
NIH (Bethesda, USA)

Collaboration :
A. Pardi (Colorado)
E. Nikonowicz (Texas)

Publications :

- "*Long-range structural information in NMR studies of paramagnetic molecules from electron spin - nuclear spin cross-correlated relaxation*" Boisbouvier, J., Gans, P., Blackledge, M., Brutscher, B. & Marion, D. **J. Am. Chem. Soc**., 121, pp 7700-7701 (1999).
- "*NMR determination of sugar puckers in nucleic acids from CSA-dipolar cross-correlated relaxation*" Boisbouvier, J., Brutscher, B., Pardi, A., Marion, D Simorre, J.-P**. J. Am. Chem. Soc**., 122, pp 6779-6780 (2000).
- "*Long-Range Magnetization Transfer Between Uncoupled Nuclei by Dipole-Dipole Cross-correlated Relaxation : A precise Probe of β-Sheet Geometry in Proteins*" Boisbouvier, J. & Bax A. **J. Am. Chem. Soc** 124, pp 11038-11045 (2002).
- "*Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids*" Boisbouvier, Delaglio & Bax **Proc. Natl. Acad. Sci. USA** 100, pp11333-11338 (2003).
- "*Resolution-optimized NMR measurement of $^{1}D_{CH}$, $^{1}D_{CC}$ and $^{2}D_{CH}$ residual dipolar couplings in nucleic acids bases*." Boisbouvier, J., Bryce, D., O'Neil-Cabello, E., Nikonowicz, E. & Bax A. "**J. Biomol. NMR**, 30, pp287-301 (2004).
- "*Experimental and Theoretical Determination of Nucleic Acid Magnetic Susceptibility: Importance for the Study of Dynamics by Field-Induced Residual Dipolar Couplings*" Bryce, D., Boisbouvier, J. & Bax A. **J. Am. Chem. Soc** 126, pp 10820-10821 (2004).
- "*Measurement of eight coupling from a single 3D NMR multiplet in proteins and nucleic acids*" Miclet, E., Boisbouvier, J., & Bax A. **J. Biomol NMR** 31, pp201-216 (2005).
- "*Quantitative J correlation methods for the accurate measurement of ^{13}C-^{15}N dipolar couplings in nucleic bases*" Jaroniec , C., Boisbouvier, J., Nikonowicz, E. & Bax A. **J. Biomol NMR** 31, pp231-241 (2005).

Les approches RMN standard sont basées sur la détection de corrélations NOE observables entre deux protons séparés par moins de 6 Å. L'analyse de ces informations et la conversion en contraintes de distance sont des processus itératifs qui peuvent prendre jusqu'à plusieurs semaines de travail. Bien qu'il soit établi que ces approches standard permettent de résoudre la structure à résolution atomique des protéines globulaires, le manque de contraintes à longue portée reste un facteur limitant la précision des modèles structuraux pour les complexes ou les molécules allongées telles que les acides nucléiques. Mon activité de recherche s'est concentrée sur le développement et l'application de nouvelles méthodes RMN visant à s'attaquer à ces systèmes biologiques difficiles qui ne peuvent pas être étudiés précisément par les méthodes RMN standard. Pour mener à bien ce projet, j'ai exploité les possibilités offertes par la RMN en milieu liquide cristallin et les phénomènes de corrélation croisée.

1) Détermination de la conformation des sucres dans les acides nucléiques de grandes tailles.

Les ARN et ADN de tailles importantes (plus de 30 nucléotides) sont caractérisés par une relaxation rapide de l'aimantation qui rend particulièrement difficile la mesure des contraintes structurales standard (typiquement nOe ^{1}H-^{1}H et constantes de couplages ^{3}J). Le déplacement chimique, et plus particulièrement son anisotropie (CSA), étant très dépendant de la conformation locale des molécules[12], la quantification de la corrélation croisée CSA/dipole dans un système

(^{13}C-^{1}H) offre la possibilité de déterminer la conformation des sucres. Pour déterminer les modes de *puckering* de chaque nucléotide, j'ai proposé une méthode simple et fiable, basée sur la mesure de phénomènes d'interférences, pour extraire ce paramètre structural en l'absence d'information sur la dynamique des sucres[13] (Fig. 3). Contrairement aux méthodes RMN standard (nOe, ^{3}J), la quantification de ce nouveau paramètre est particulièrement utile pour les gros acides nucléiques car l'effet mesuré augmente avec la taille de la molécule étudiée.

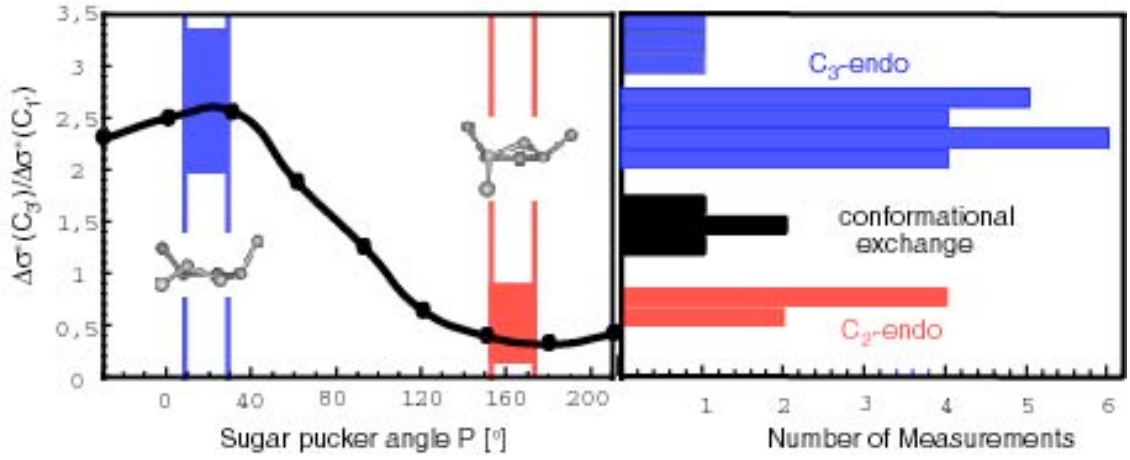

Figure 3: **A**. Dépendance du rapport des CSA des carbones $C_{3'}$ et $C_{1'}$ avec la conformation du sucre. **B**. Histogramme des rapports des vitesses de corrélation croisée CSA-Dipole mesurées pour les $C_{3'}$ et $C_{1'}$. Les conformations $C_{3'}$-*endo* (bleu) sont caractérisées par un rapport >2 Tandis que la conformation $C_{2'}$-*endo* (rouge) est caractérisée par des valeurs < 1.

2) Détermination de nouvelles informations à longue portée dans les biomolécules paramagnétiques.

La présence d'un centre paramagnétique, même distant de 20 Å des spins nucléaires, peut modifier notablement le CSA de ces noyaux. La quantification de cette variation du CSA, par suivi de la corrélation croisée «Curie-dipolaire», apporte un nouveau type d'information structurale à longue distance. J'ai élaboré et démontré cette stratégie sur un échantillon de cytochrome c' isotopiquement enrichi en ^{15}N. Un très bon accord a été observé entre les données expérimentales et la structure cristallographique de ce cytochrome (Fig. 4)[14]. La mesure de la corrélation croisée

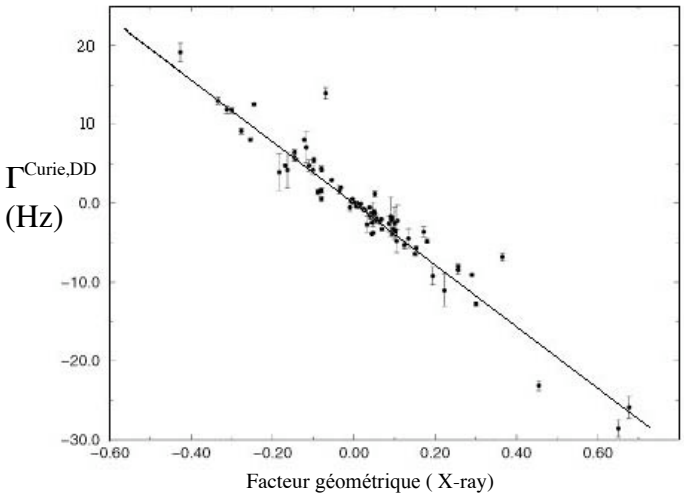

Figure 4. Comparaison entre la corrélation croisée Curie-Dipolaire ($\Gamma^{Curie,DD}$) mesurée sur le ferricytochrome c' et le facteur géométrique correspondant extrait de la structure cristallographique. Les données couvrent des distances, entre spins et centre paramagnétique, comprises entre 9.5 et 27.5 Å. Les déviations les plus conséquentes correspondent aux résidus S20 et G85 qui ont été montrés comme adoptant plusieurs conformations dans les structures cristallographiques résolues pour ce cytochrome.

Curie-Dipolaire, complètant les autres informations longues portées, a permis la détermination de la structure de ce cytochrome sans informations nOe[15]. Pour des molécules pouvant fixer des ions paramagnétiques, cette stratégie apparaît comme particulièrement utile quand les contraintes RMN standard ne permettent pas la détermination de la structure. Deux applications particulièrement prometteuses seront l'étude de molécules allongées telles que les acides nucléiques et le positionnement relatif des sous-unités d'un complexe.

3) Quantification précise des distances et des angles par transfert d'aimantation à longue portée entre spins non-couplés.

Il a été bien établi, au moins sur le principe, que la détection de phénomènes de corrélation croisée dans une macromolécule est possible pour toute paire d'interactions, quelle que soit la distance qui les sépare[16]. En pratique, les approches actuelles se sont limitées à l'étude de paires de spins liés covalemment, afin de générer des cohérences multispins, en utilisant le réseau de couplage scalaire. Arrivé dans le laboratoire du Dr. Ad BAX, j'ai mis au point un nouveau type d'expériences permettant le transfert d'aimantation, par corrélation croisée dipole-dipole, entre protons non-couplés entre eux[17]. L'efficacité de ce transfert est particulièrement sensible à la géométrie locale, en particulier à la distance interproton (r_{HH}) et à l'angle formé par ce vecteur et les liaisons C-H (θ_{CHH} - Fig. 5). L'application à l'étude des vecteurs $^{13}C_a$–$^{1}H_a$ qui se font face dans les protéines en feuillets β antiparallèles, a été démontrée pour le troisième domaine de la protéine G (GB3) et la protéase du VIH (22 kDa) compléxée à son inhibiteur DMP323.

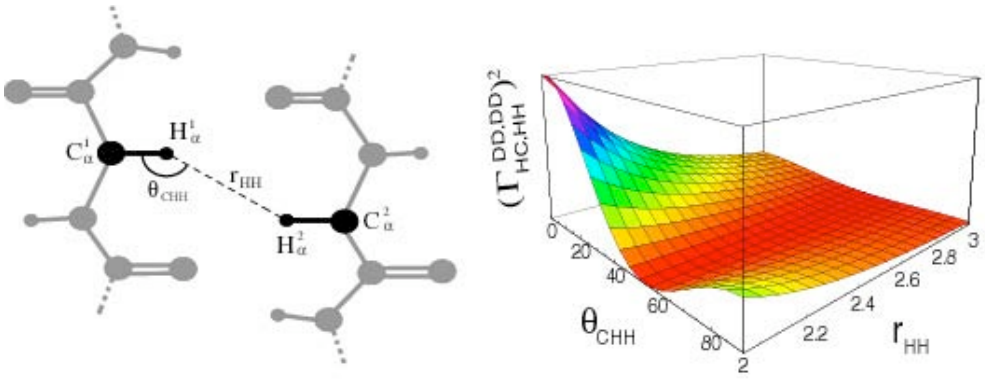

Figure 5: Dépendance structurale de la corrélation croisée Dipole-Dipole avec la géométrie des feuillets β antiparallèles.

Un bon accord est observé avec la structure cristallographique de GB3, déterminée à une résolution de 1.1Å. La mesure quantitative de cette corrélation croisée entre spins non-couplés nous apporte une information structurale très précise sur la géométrie des protéines riches en feuillet β. L'incertitude expérimentale se traduit par une erreur inférieure à 7 pm sur les distances interprotons et 3.5° pour les angles interspins. Puisque l'amplitude des phénomènes de corrélation croisée augmente avec la taille des molécules étudiées, l'application de ces nouvelles techniques n'est pas limitée aux systèmes de petites tailles. La corrélation croisée Dipole-Dipole entre

atomes non-couplés permettra de résoudre la structure des biomolécules à une précision encore inégalée.

4) Etudes structurales des acides nucléiques par RMN en milieu cristallin

Probablement, l'une des plus grandes avancées ces dernières années est le développement de milieu cristallin liquide permettant d'introduire en solution un alignement faible (~ 0,1- 1‰) des biomolécules par rapport au champ magnétique[18]. Cet alignement partiel permet de garder la haute résolution des spectres caractéristique de la RMN liquide, mais résulte en un moyennement incomplet de l'interaction dipolaire qui contient une information précise à la fois sur la distance et l'orientation caractérisant une paire de spins. La mesure de constantes de couplage dipolaire résiduel (RDCs) entre deux spins dont la distance est fixe, nous renseigne sur l'orientation des vecteurs internucléaires par rapport à un repère moléculaire fixe. Contrairement aux paramètres RMN standard, tels que les NOEs et les couplages scalaires, les RDCs apportent une information orientationnelle à longue portée. Pour les acides nucléiques, de formes généralement allongées, les RDCs sont particulièrement utiles pour définir l'ordre à longue portée qui ne peut être obtenu par les NOEs et couplages scalaires. Mais la plupart des études s'est focalisée sur un nombre modeste de RDC impliquant le proton et son hétéroatome (^{15}N ou ^{13}C)[19-20]. Le nombre de RDC ainsi mesurées (~ 6-7 par nucléotide) reste bien au-dessous du nombre de degrés de liberté caractérisant la conformation et l'orientation d'un nucléotide.

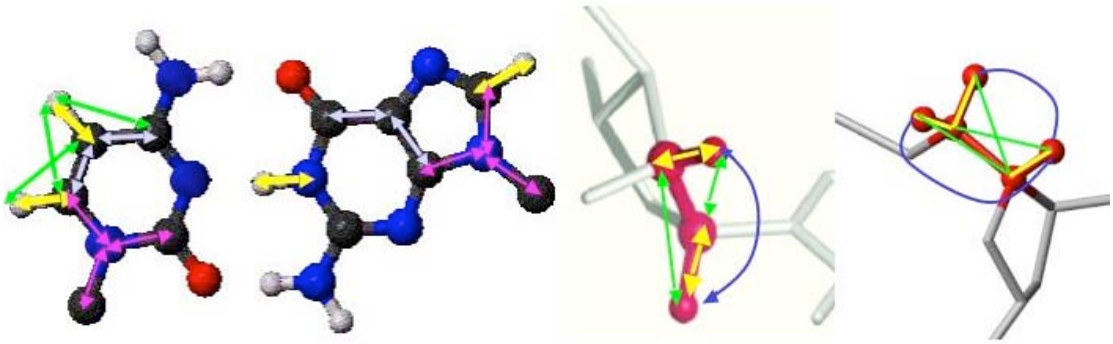

Figure 6. Exemples de RDCs mesurables dans les ARN. Les différents couplages dipolaires mesurables sont représentés par des flèches et codés par les couleurs suivantes : jaune $^{1}D_{CH}$, violet $^{1}D_{CC}$, vert $^{2}D_{CH}$, rouge $^{1}D_{CN}$ et bleu $^{3}D_{HH}$.

Au cours de mon séjour postdoctoral, j'ai entrepris de mettre au point une série d'expériences pour extraire un jeu important de RDC par nucléotide. Deux séries d'expériences ont été mises au point pour la mesure de 8 RDCs par groupement méthylène du squelette phosphodiester des acides nucléiques[21] et jusqu'à 7 RDCs par base nucléique[22] (Fig. 6). Une fois mes fonctions prises au CNRS, j'ai poursuivi ce projet en collaboration avec le laboratoire d'Ad BAX aux NIH. Nous avons entrepris de développer une expérience quantitative permettant la

mesure des couplages entre l'azote de la liaison glycosidique et chacun des trois carbones qui lui sont covalemment liés[23]. L'amplitude de l'effet à mesurer ne dépasse en général pas 3 Hz (25 fois plus faible que pour un couplage dipolaire CH) ; pour être utile en tant que contrainte structurale ce paramètre doit donc être quantifié très précisément. Une précision de l'ordre de 0.1 Hz a pu être atteinte grâce à une spectroscopie J-quantitative, où la valeur du couplage mesuré est codée dans l'intensité du signal détecté. Ce dernier jeu d'expérience porte à 11 le nombre de RDCs mesurables par base, au total jusqu'à 30 RDCs sont mesurables par nucléotide (Fig. 6), ce qui est bien supérieur au nombre de degré de liberté caractérisant la conformation et l'orientation de chaque nucléotide.

Les expériences présentées ci-dessus fournissent une information très précise sur l'orientation des nucléotides, mais ne nous renseignent nullement sur la position relative des nucléotides les uns par rapport aux autres. Pour réaliser ce projet, j'ai développé une méthode permettant la mesure spécifique de RDCs inter-nucléotides, entre protons séparés par une distance pouvant atteindre 12 Å (fig. 7)[24]. En comparaison avec l'analyse de NOE, ce type de paramètre donne lieu à des contraintes de distances précises et de portée beaucoup plus longue (annexe –article 2).

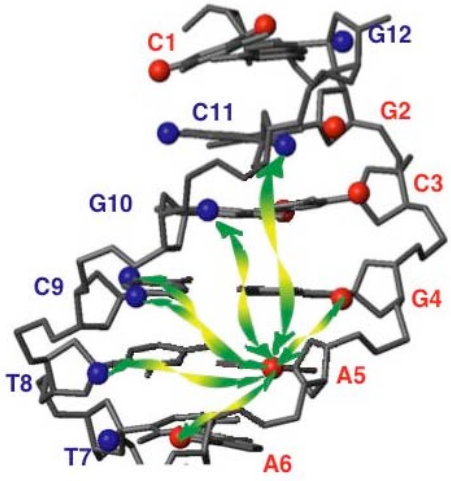

Figure 7. Exemple de RDCs inter-nucléotide mesurables à longue distance dans les acides nucléiques. Des transferts entre des protons séparés par 3 paires de bases (jusqu'à 12 Å) sont observables

C) Etudes structurales des ARNs viraux et de leurs complexes.

Période :2004-

Collaborations :
Pr. Bryce, U. Ottawa (Canada)
Dr. Toulmé (IECB, Bordeaux)
Pr Drouet (UVHCI-Grenoble)

Formation d'étudiants:
- H. Van Melckebeke, post doc 2005-2006
- M. Devany, post doc 2005-2006
- R. Rasia , post doc 2006-2007

Publications :

• "*Liquid Crystal NMR Structure of HIV TAR RNA Bound to its SELEX RNA Aptamer Reveals the Origins of the High Stability of the Complex*" Van Melckebeke, V., Devany, M., Di Primo, C., Beaurain, F.,. Toulmé, J.J., Bryce, D. & Boisbouvier J. **Proc. Natl. Acad. Sci. USA** 105, pp 9210-9215 (2008).

• *"Initiation factor eIF3 subunit b interacts with HCV IRES RNA through its N-terminal RNA recognition motif"* Perard, Rasia, Medenbach, Ayala, Boisbouvier, Drouet & Baudin **FEBS Letters** *(in press)* .

Contrats :
• Subventions du Fond de Recherche France-Canada Coordinateur. *"Etude structurale du complexe entre la boucle TAR de l'ARN du VIH et son aptamère TAR*GA par RMN en milieu cristallin liquide"*
• Accords de cooperation CNRS / Instituts de Recherche en Santé du Canada *"Etude structurale du complexe entre la boucle TAR de l'ARN du VIH et son aptamère TAR*GA par RMN en milieu cristallin liquide"*
• Subvention de l'ANRS : *"Etude Fonctionnelle et Structurale en solution des complexes RNP (eIF3-IRES) engagés dans l'initiation de la traduction du VHC".*

1) Bases structurales de la stabilité du complexe entre la boucle TAR du VIH et son aptamer d'ARN, déterminées par RMN en milieu cristallin.

La transactivation du génome du VIH requiert la liaison avec la protéine TAT sur un segment structuré d'une boucle de l'ARN virale appelée TAR (*Trans Activation-Responsiveelement*)[25]. Pour inhiber spécifiquement l'interaction entre TAT et TAR une approche *in vitro* (SELEX) a été utilisée pour sélectionner un aptamer de haute affinité capable de reconnaître sélectivement la boucle TAR[26]. Les aptamers, avec les plus hautes affinités pour TAR, adoptent une forme en épingle à cheveux et forme un *kissing complex* avec l'ARN cible, via une interaction boucle-boucle. La boucle de ces aptamers est parfaitement complémentaire des 6 nucléotides de la boucle de TAR. Une caractéristique intéressante de tous ces aptamers est que chacun possède une paire non-canonique G-A à la jonction entre leur tige et leur boucle[27].

Figure 8. Ensemble de structure du complexe TAR/TAR*GA déterminé par RMN en milieu cristallin liquide (rmsd=0.26 Å, les 17 meilleures structures sont superposées sur l'ensemble des atomes lourds du complexe).

Arrivé à l'IBS en tant qu'agent CNRS, j'ai appliqué les expériences de RMN en milieu cristallin précédemment développées pour déterminer la structure du complexe TAR-aptamer de plus haute affinité. Très peu de *kissing complexes* ont été étudiés au niveau structural par RMN ou cristallographie aux rayons X. La RMN en milieu cristallin nous a permis de déterminer la structure de ce complexe en solution avec une précision sans précédent (rmsd de 0.26 Å sur tous les atomes lourds du complexe)[28]. Nous avons pu montrer, grâce a cette structure haute résolution, que l'introduction d'une paire G-A à la jonction tige-boucle de l'aptamer permet la formation de 6 liaisons hydrogènes supplémentaires stabilisant le complexe (annexe – article 3). La comparaison avec d'autres études SELEX dirigées contre différentes tiges-boucles d'ARN, démontre que ces interactions stabilisantes sont conservées dans de nombreux *kissing complexes* de haute affinité. Nos résultats peuvent donc être utilisés comme base pour la conception de nouvelles drogues contre le VIH et pour la synthèse d'aptamers de haute affinité contre différentes boucles d'ARN d'importance biologique.

2) Etude de l'interaction d'eIF3 avec l'IRES du Virus de l'hépatite C.

L'IRES est un domaine de l'ARN du VHC reconnu par des protéines de la cellule hôte: facteurs d'initiation de la traduction (ou eIF) et protéines ribosomales. Ce complexe ribonucléoprotéique (RNP) est nécessaire pour l'initiation de la traduction virale et donc pour l'amplification du virus. En collaboration avec l'équipe du Pr. Drouet à Grenoble (UVHCI), nous focalisons nos efforts sur les complexes ribonucléoprotéiques issus de l'interaction entre le facteur cellulaire eIF3 et l'IRES du VHC et l'étude de leurs caractéristiques structurales. Le facteur eIF3 est un complexe multiprotéique de 750 kDa constitué de 13 sous-unités aujourd'hui clonées et séquencées[29]. Les études de prédiction de structures ont montré que seules les sous unités p44 (eIF3g) et p116 (eIF3b) présentent dans leur partie centrale un motif de reconnaissance de l'ARN (MRR). L'équipe du Pr. Drouet a pu démontrer que seule le domaine MRR de eIF3b reconnaissait spécifiquement l'IRES avec une affinité de 2.10^{-6} M.

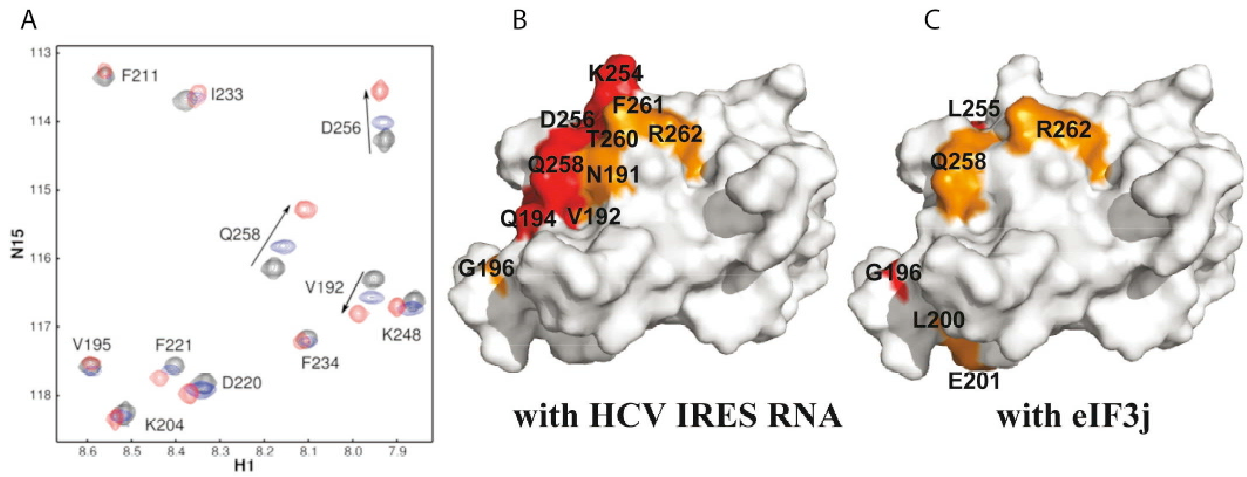

Figure 9. Titration du domaine MRR de eIF3b par l'IRES (100kDa). En **A** : Zoom du spectre [^{15}N,^{1}H]-TROSY pour un rapport IRES/MRR égale à 0 (Noir) 0.4 (bleu) et 1.25 (rouge). **B et C** : Représentation de la zone d'interaction de MRR-eIF3b avec l'IRES (B) et de MRR-eIF3b avec eIF3j (C) déterminée par titration RMN.

Dans le but d'identifier les résidus du MRR responsable de l'interaction spécifique avec l'IRES, nous avons effectué une étude des surfaces d'interaction par RMN. Nous avons mis au point un protocole de surexpression, marquage isotopique et de purification « RNAse-free » d'eIF3b-MRR (10 kDa) dans *E. Coli* en milieu minimum. Cet échantillon a été utilisé pour tester l'interaction avec l'IRES (100kDa). Malgré la taille importante du complexe (110 kDa) un spectre de très grande qualité a pu être obtenu sur un spectromètre RMN opérant à 800 MHz (Fig. 9A). Plusieurs signaux RMN sont déplacés par addition progressive d'IRES. La variation progressive des résonances de la protéine indique que le domaine MRR, l'IRES et le complexe sont dans une gamme d'échange rapide. Les résidus les plus affectés correspondent à deux régions dans la partie N-terminal (N191—G196) et C-terminal (K254-T260). De manière intéressante, il a été

récemment reporté que les résidus 254-260 sont directement impliqués dans l'interaction entre ce domaine de eIF3b avec la partie N-terminal d'eIF3j[30]. La similarité des surfaces d'interaction d'eIF3b avec eIF3j et l'IRES du VHC suggère fortement que ces interactions sont mutuellement exclusives. Ces travaux ont débouché sur la soumission d'un premier manuscrit[31] et l'obtention d'une subvention de l'ANRS pour la caractérisation structurale du complexe entre l'IRES du VHC et son domaine de fixation sur eIF3. Un chercheur post-doctoral a été recruté pour mener à bien ces travaux et débutera en janvier 2009.

PROJETS DE RECHERCHE

A) Projet: Etudes RMN des assemblages biologiques de tailles importantes.

Période : 2004-

Collaborations :
Dr. Wu (U. of Ohio Columbus, USA)
Pr. Chou (Harvard Univ. (Boston, USA)
Pr. Legault (Montreal, Canada)
Dr. Franzetti (IBS)
Dr. Vernet (IBS)

Formation d'étudiants:
R. Sounier, Doctorant 2004-2008
K. Treche, Master 2005
M. Falges ,Master 2006
N. Usé, Master 2007
J. Retel, Master 2007
C. Amero, Post-doc 2008-

Contrats :

• Career Development Award from Human Frontier Science Program : "*Quaternary Structure Study of large complexes using Liquid Crystal and Cross-Correlated NMR* ".

• Subvention du programme interdisciplinaire CNRS Interface physique-chimie-biologie, Soutien à la prise de risque : "*Nano Machines thermophiles en action : une approche par RMN en temps réel*".

Publications :

• "*High-Accuracy Distance Measurement Between Remote Methyls in Specifically Protonated Proteins*" Sounier, R. Blanchard, L. Wu, Z. & Boisbouvier **J. Am. Chem. Soc** 129, pp 472-473 (2007).

• "*Sensitivity-optimized experiment for the measurement of residual dipolar couplings between amide protons*" Schanda, P., Lescop, E., Falge, M., Sounier, R Boisbouvier J., Brutscher, B. **J. Biomol NMR** 38, pp47-55 (2007).

• *"An efficient protocol for the complete incorporation of methyl specifically protonated alanine in perdeuterated protein"* Ayala, Sounier, Use, Gans, & Boisbouvier* **J. Biomol NMR** (in press).

• *"2D NMR Spectroscopy of High Molecular Weight Protein Assemblies in Second Time Range"* Amero, C., Schanda, P., Dura, A., Ayala, I., Marion, D., Franzetti, B., Brutscher, B., Boisbouvier*, J. (in preparation).

• *"Stereospecific Labeling of Leucine and Valine Methyls and Application to the NMR Study of Biological NanoMachines "* Gans, P., Hamelin, O., Ayala, I., Dura, A, Franzetti, B., Sounier, R., Boisbouvier*, J (in preparation).

• "*Long range distance and angular information between remote methyls measured by liquid crystal NMR"* Sounier, Wu, & Boisbouvier* J. (en preparation).

• "*Detection of long range NOE in macromolecular assemblies beyond 100 kDa: possibilities and limits"* Sounier, Ayala, Dura, Franzetti, Gans, Retel, Tugarinov, Kay, Grishaev & Boisbouvier* (en préparation)

Le but de ce projet de recherche est de développer de nouvelles méthodes spectroscopiques permettant l'étude structurale et /ou dynamique d'assemblages biomoléculaires dépassant largement les limites actuelles de la RMN liquide. Ces recherches correspondent au projet que j'ai déposé pour mon recrutement au CNRS. Si l'on considère un complexe de deux partenaires, pour lesquels la structure tertiaire de chacun d'entre eux est connue, il n'y a que cinq degrés de liberté translationnelle et orientationnelle. Afin de situer les partenaires relativement les uns par rapport aux autres, il est donc nécessaire de disposer de seulement quelques contraintes de distance intermoléculaire précise. Pour les détecter dans des complexes macromoléculaires de plus de 100 kDa, il faut marquer spécifiquement les biomolécules étudiées. Ce projet est soutenu par un prix de *Human Frontier Science Program* qui m'a permis de mettre en place, dans mon laboratoire d'accueil, les stratégies de marquages spécifiques et des outils spectroscopiques adaptés présentés ci-dessous. Ces développements méthodologiques développés ces 3 dernières années sur des systèmes modèles sont actuellement appliqués à l'étude d'une machinerie biologique de 500 kDa.

1) Protonation spécifique des protéines.

La détection d'information structurale à longue portée nécessite le recours à une protéine perdeutérée sélectivement. Compte tenue de leur multiplicité et de leurs propriétés de relaxation favorables, les méthyles sont des sites intéressants pour introduire spécifiquement des protons. Le groupe du Pr. L Kay (Toronto) a développé une stratégie permettant de réintroduire les protons des méthyles δ_1 des isoleucines en utilisant des α-céto-butyrates spécifiquement protonés comme précurseurs[32]. Le protocole initialement publié a été implémenté et optimisé afin de permettre une réintroduction de protons à plus de 98% sur les méthyles δ_1 des isoleucines, tout en maintenant un niveau de deutération supérieur à 98 % sur les autres sites[33], sans surcoût du prix de revient part rapport à la perdeutération complète. Ce protocole optimisé peut également être appliqué pour le marquage spécifique des méthyles des Leucines et des Valines en utilisant l'α-cétoisovalerate comme précurseur spécifiquement protoné[34].

Les méthyles des alanines constituent un autre site de choix pour protoner spécifiquement une protéine. En effet compte tenu de leur abondance dans les protéines et de leur proximité avec le squelette, ils peuvent servir de sonde structurale et dynamique en de très nombreux points de la chaîne principale des protéines. Mais au contraire des isoleucines, leucines ou valines, l'alanine ne se trouve pas à la fin d'une voie de biosynthèse irréversible. L'alanine est en constant équilibre dans la cellule avec le pyruvate, un maillon central pour de très nombreuses voies métaboliques. Malgré la multiplicité des voies métaboliques impliquées, nous avons réussi à développer un protocole de marquage isotopique spécifique des alanines en milieu minimum[35]. Contrairement au protocole proposée au préalable[36], l'alanine spécifiquement protonée sur son méthyle peut être incorporée, à plus de 95%, dans des protéines surexprimées dans *E. coli,* avec des fuites isotopiques inférieures à 0.5% (Fig 10).

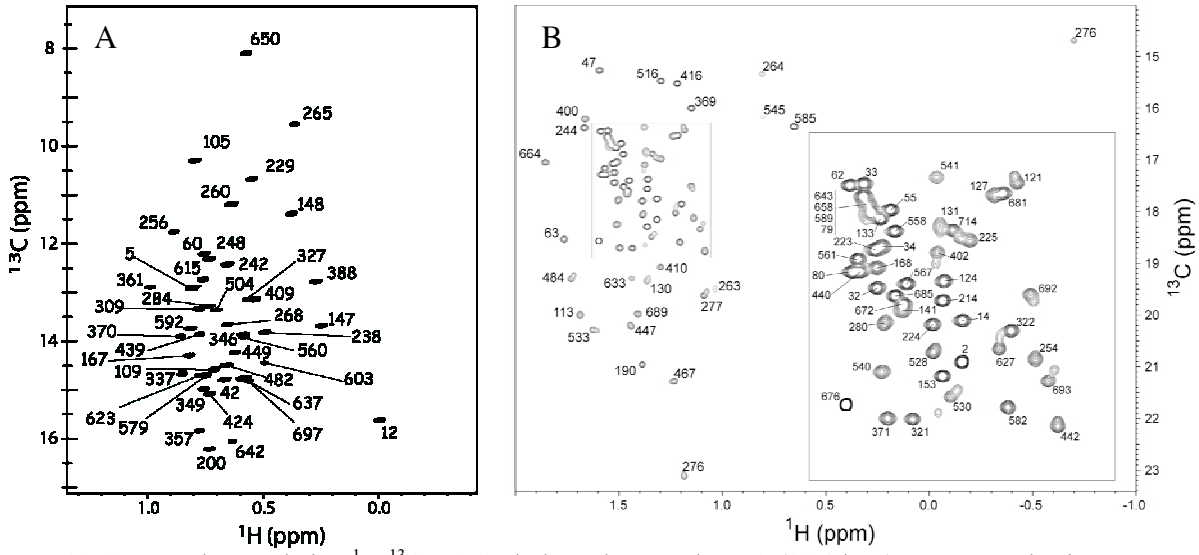

Figure 10. Spectre de corrélation ^{1}H-^{13}C HQC de la malate synthase G (81.4 kDa) protonée sélectivement sur les méthyles δ_1 des isoleucines (A) et des alanines (B).

2) Observation de contraintes de distances entre protons éloignés de plus de 10 Å dans les protéines.

Les méthodes RMN standard pour la détermination de la structure des biomacromolécules en solution reposent sur l'observation d'effets NOE à courte portée. Plus récemment, l'utilisation de spectromètre à haut champ et la perdeutération ont permis la détection de NOE entre protons séparés de 8 Å[37]. En utilisant la stratégie de marquage sélectif décrite précédemment, j'ai pu montrer que des contraintes de distances d'une précision inégalée pouvait être détectée entre protons séparés jusqu'à 12 Å dans les protéines globulaires de taille modérée (10-20 kDa). Cette approche a été publiée dans J. Am Chem Soc (annexe – article 4). L'intérêt de cette méthode est qu'elle est transposable à des systèmes de taille moléculaire bien au-delà des limites standard de la RMN. Appliquée à une protéine modulaire de 82 kDa et à un assemblage supramoléculaire de près de 500 kDa (Fig. 11), il est toujours possible de détecter des NOEs entre protons séparés par respectivement plus de 10 et 7 Å[38].

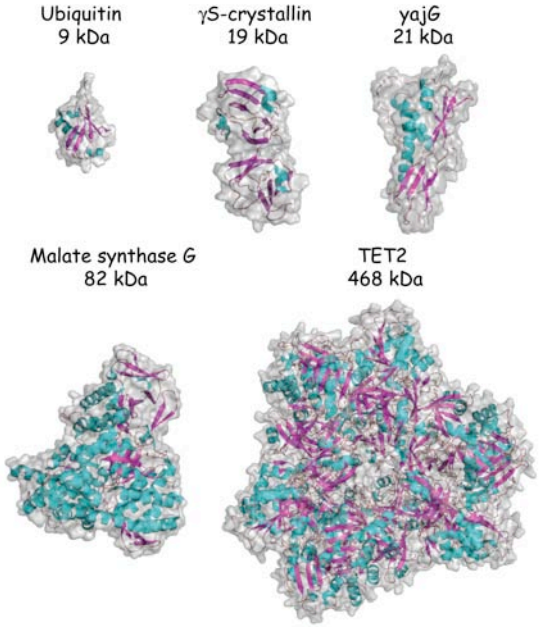

Figure 11. Tailles relatives et repliement globale des protéines utilisées pour la détection de NOE entre méthyles éloignés.

Le transfert d'aimantation par RDCs ^{1}H-^{1}H dans les protéines perdeutérées est limité à des protons séparés par 7 Å au plus[39-40], et ce, même dans des petites protéines de moins de 10 kDa. Cette limitation, par rapport à des acides nucléiques de taille comparable, est inhérente à la différence de densité de protons et à leurs fréquences de résonance caractéristique[41]. Concrètement, dans un acide nucléique de 8 kDa, il est possible de limiter le transfert qu'entre 25 protons correspondant aux H_1, qui sont distants les uns des autres d'au moins 5 Å. Dans le cas d'une protéine uniformément deutérée de taille comparable, on aura trois fois plus de protons. De plus les H_N peuvent s'approcher à moins de 3 Å les uns des autres dans les protéines. De ce fait, dans les protéines, il existe des interactions dipolaires très efficaces entre H_N proches qui tronquent complètement les transferts dipolaires entre protons plus éloignés. L'utilisation d'échantillons spécifiquement protonés sur les méthyles des ILE permet de limiter la densité de protons de 20 % (protons échangeables seulement) à 2-3 %, ce qui réduit considérablement le problème de la troncation dipolaire. L'efficacité de cette stratégie a été démontrée sur un échantillon modèle d'ubiquitine. Dans cette protéine de 9 kDa, il devient possible de détecter des

couplages dipolaires résiduels entre méthyles séparés de 14 Å. En utilisant simultanément les RDCs avec les NOE longue distance, il nous est désormais possible de déterminer précisément la longueur et l'orientation du vecteur reliant deux méthyles distant de plus de 10 Å dans une protéine[42].

3) Détection de corrélation croisée Curie-Dipolaire intermoléculaire.

La quantification de phénomènes d'interférence, entre atomes éloignés, apporte des informations angulaires et des distances d'une précision remarquable. J'ai démontré que ce principe est particulièrement intéressant à utiliser en présence d'un cation paramagnétique, puisqu'un phénomène d'interférence entre spins électronique et nucléaires (appelée interférence Curie-Dipolaire) peut être mesuré avec précision à plus de 20 Å du centre paramagnétique. De plus, l'effet mesuré croît avec la taille de la molécule étudiée, ce qui est non négligeable pour l'étude de la structure quaternaire des assemblages biomoléculaires de taille importante. La généralisation de cette approche aux complexes diamagnétiques, nécessite l'incorporation d'un site de fixation de haute affinité de fixation de haute affinité pour un centre paramagnétique, sur l'un des partenaires. En marquant isotopiquement en ^{15}N et/ou ^{13}C un second partenaire du complexe, il doit ainsi être possible de détecter uniquement des interférences Curie-Dipolaires intermoléculaires. Une comparaison des propriétés des différents centres paramagnétiques bien caractérisés, montre que les lanthanides au degré d'oxydation +III (excepté le Gadolinium) remplissent le mieux ces conditions requises à l'observation d'un effet à longue portée. La tendance des lanthanides à se fixer au squelette phosphodiester des acides nucléiques ou à se substituer au calcium dans les protéines, rend nécessaire l'insertion d'un site de fixation de très haute affinité.

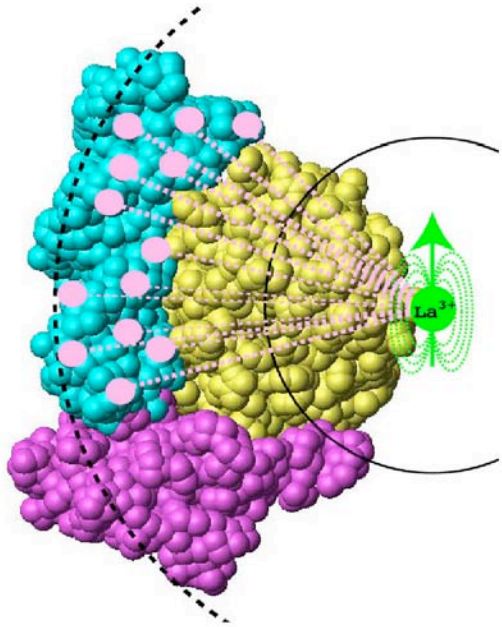

Figure 12. Stratégie développée pour mesurer spécifiquement des interférences intermoléculaires entre spin électronique (représenté en vert) et spins nucléaires (schématisé en rose) dans des assemblages biomoléculaires de tailles importantes.

Pour réaliser ce projet, j'ai choisi d'utiliser deux types de ligands. Le premier est un peptide de 15 AA optimisé afin de fixer les lanthanides avec une affinité nanomolaire[43]. Celui-ci peut être utilisé sous forme d'extension N ou C-terminale d'un des partenaires du complexe. Une seconde stratégie consiste à utiliser un dérivé de l'EDTA[44] avec une très forte affinité pour les lanthanides

au degré d'oxydation III ($K_d \sim 10^{-15}$). L'utilisation d'EDTA modifié avec une fonction thiol permet de le lier directement aux cystéines dans les protéines ou aux thiopyrimidines dans les acides nucléiques. En utilisant ces stratégies, deux complexes (ARN/protéine et protéine/protéine) pouvant spécifiquement fixer un lanthanide sont actuellement en cours d'étude.

4) Etude par RMN en temps réel d'une Nano Machine thermophile en action.

L'étude fonctionnelle et structurale des machines macromoléculaires est une tâche difficile compte tenu de la dimension des objets étudiés, de leur flexibilité et de la complexité des substrats manipulés (protéines, peptides, ADN, ARN …). Ces études nécessitent la combinaison de la cristallographie aux rayons X et de méthodes « basse» résolution : SAXS, SANS, cryomicroscopie électronique. Ces méthodes permettent difficilement d'obtenir des informations sur les parties mobiles des grands assemblages et n'autorisent pas des études cinétiques permettant de comprendre la dynamique fonctionnelle d'un système. Dans ce contexte, j'ai proposé d'utiliser les méthodes RMN développées pour les gros complexes dans le but d'obtenir des informations structurales et dynamiques sur les mécanismes de fonctionnement des machines macromoléculaires complexes. La protéase TET de l'Archaea hyperthermophile *Pyrococcus horikoshii* (homo-oligomèriques de 468 kDa) sera utilisée comme modèle d'étude[45]. De par sa solubilité, sa stabilité à haute température, ce système supramoléculaire biologique est particulièrement bien adapté au développement de ce nouvel outil d'étude structurale *in vitro* du fonctionnement des assemblages macromoléculaires en temps réel. L'ambition du projet est de fournir la première description du mode d'action des machines de dégradation TET par une approche cinétique et structurale originale, combinant la RMN et les propriétés de thermoactivation/stabilité des particules TET hyperthermophiles.

Le projet comprend 3 objectifs :

- Déterminer la structure des parties mobiles situées à l'intérieur des particules TET.
- Identifier les résidus responsables des différentes étapes du processus de digestion des polypeptides (reconnaissance, navigation à travers la particule, adressage aux sites actifs) ou de l'expulsion des acides aminés clivés.
- Réaliser une étude cinétique RMN en temps réel des différentes opérations et des mouvements structuraux associés au fonctionnement des machines TET

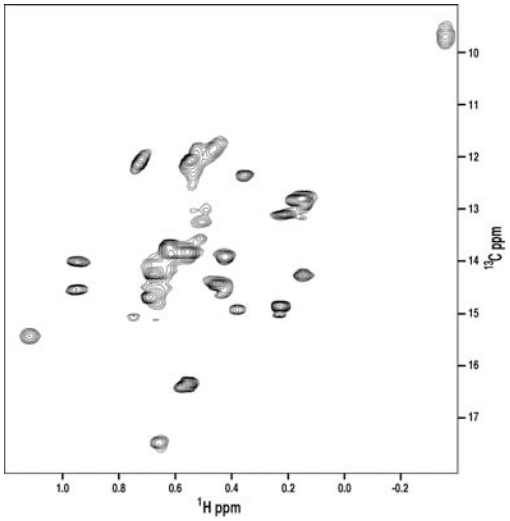

Figure 13 : Spectre de corrélation ^{13}C-^{1}H de la protéine TET2 (468 kDa) . L'échantillon contient 10 mg de protéine pure marquée spécifiquement ^{1}H,^{13}C sur les méthyles δ1 des isoleucines.

La protéine TET2 a pu déjà être produite sous forme deutérée en quantité compatible avec les études par RMN. L'équipe de Dr Franzetti a pu purifier la particule deutérée avec un haut degré de pureté (> 96 %), sous forme de complexes actifs, avec un rendement de plus de 20 mg par litre de culture. Cette protéine peut être concentrée à plus de 40 mg/mL pour les études RMN. Les études préliminaires RMN montrent que la plupart des méthyles est observables dans cet assemblage de près de 500 kDa (figure 13) et qu'un spectre RMN de corrélation ^{13}C-^{1}H de qualité peut être obtenu en une dizaine de secondes sur le spectromètre 800 MHz de l'IBS. La qualité des spectres fait de cet assemblage un système idéal pour extraire, par RMN, des informations quantitatives sur les distances interméthyles et pour caractériser la mobilité de ces résidus. Pour mener à bien ce projet ambitieux, j'ai initié une collaboration en interne avec un spécialiste des machineries hyperthermophiles (Dr. B. Franzetti) et le laboratoire d'Ingénierie des Macromolécules dirigé par Dr. Thierry Vernet. Ce projet a reçu en août 2007 un fonds d'amorçage du programme interdisciplinaire CNRS "Interface physique-chimie-biologie, soutien à la prise de risque ".

B) Projet: Etudes des bases structurales de la Biogenèse des microRNAs.

Période :2006-

Collaborations :
-Pr. Palatnik (IBR, Rosario, Argentine)
-Pr. Weigel (M.P.I. Tübingen, Allemagne)
- Dr. Vernet (IBS)
- Dr. Skoufias (IBS)

Formation d'étudiants:
-R. Rasia ,post doc 2006-2007
-M. Devany, post-doc 2005-2006
- N. Bologna, Master 2007
- M. Plevin, post-doc 2008-2011

Publications :

• *" Sequence and expression differences underlie functional specialization of arabidopsis microRNAs miR159 and miR319"* Palatnik, J., Wollmann, H., Schommer, C., Schwab, R., Boisbouvier, J., Rodriguez, R., Warthmann, N., Allen, E., Dezulian, T., Huson, D., Carrington, J., Weigel, D. **Dev. Cell** 13, pp 115-125 (2007).

• *"Parallel screening and optimization of protein constructs for structural studies"* Rasia, Noirclerc-Savoye, Gallet, Bologna, Plevin, Blanchard, Palatnik, Brutscher, Vernet* & Boisbouvier* Protein Science (in press)

• "*Fast Fold Determination using chemical shift information combined with liquid crystal NMR"* Rasia, Brutscher, & Boisbouvier* (in preparation).

• "*Automated optimization of protein solubility and stability"*Imbert, L., Blot, D., Favier, A. & Boisbouvier*, J. (in preparation).

• "*Specificity of the recognition of pre-miRNA target by Arabidopsis primary microRNA processing protein HYL1"* Rasia, Bologna, Imbert, Palatnik* & Boisbouvier* (in preparation).

Contrats :

• Grant de HFSP: "*RNA Shape Recognition And Structure In MicroRNA Processing in Plants"*.

• ANR jeune chercheur : "*RNA Shape Recognition And Structure In MicroRNA Processing"*.

• Subvention du programme EcoS / SECYT pour financer les échanges avec l'Institut de Biologie de Rosario (Argentine) : *"Etudes structurales par RMN de la biogenèse des microRNAs"*.

Ces dernières années, les petits ARN non codants, et en particulier les microRNAs (miRNAs) se sont placés très rapidement au premier plan de la recherche académique. Les premiers miRNAs ont été découverts au début des années 1990[46-47]. C'est seulement en 2000, qu'il fut reconnu que ces petits ARN (~21 nucléotides) étaient exprimés à partir du génome des

animaux comme des plantes. Il est maintenant largement admis que les miRNAs sont la clé d'un mécanisme biologique général permettant la régulation de l'expression des ARN messagers. Les miRNAs régulent très spécifiquement l'expression de leur ARNm cible, soit par dégradation de l'ARNm codant, soit par simple répression sans clivage de l'ARNm. Les miRNAs sont impliqués dans de nombreux processus biologiques aussi divers que la morphogenèse et l'organogenèse, la régulation du cycle cellulaire, l'initiation et la progression des tumeurs cancéreuses.

Bien qu'il y ait eu de rapides progrès dans la découverte de nouveaux miRNAs, les mécanismes moléculaires responsables de la sélectivité du clivage des miRNAs matures à partir de leur précurseur (pre-miRNA) restent à élucider. Cette étape est importante à comprendre car à l'heure actuelle, on peut prédire les cibles potentielles d'un miRNA du fait de la complémentarité partielle avec son ARNm cible, ainsi que la plupart des pre-miRNAs à partir de la séquence du génome d'un organisme, mais la séquence exacte d'un miRNA généré à partir d'un précurseur donné reste, elle, impossible à prédire. Le but de ce projet est l'étude, par Résonance Magnétique Nucléaire de la structure des complexes formés entre un pre-miRNA et les domaines de fixation à l'ARN des protéines impliquées dans la biogénèse des microRNA. Ces travaux nous permettrons de mieux comprendre cette voie de régulation du génome, d'identifier de nouveaux microRNAs et leurs cibles, et d'interférer spécifiquement avec leur biogenèse.

1) Du gène à la structure : mise au point d'un protocole haut débit de marquage isotopique et d'étude structurale par RMN.

La RMN en solution est une technique bien établie pour déterminer la structure des protéines de taille modérée. Le but de ce projet est de l'appliquer aux protéines impliquées dans la biogenèse des microARNs : il s'agit de protéines modulaires de grandes tailles (30 à 200 kDa) composées par un grand nombre de domaines indépendants de moins de 15 kDa (Fig 17). Une des difficultés de l'étude de tels domaines par RMN est la détermination des limites exactes des domaines structurés. Dans le but d'étudier ces systèmes modulaires complexes, nous avons mis au point un outil haut débit permettant d'accélérer et d'automatiser la détermination structurale par RMN de ces domaines. Cet outil comprend trois étapes distinctes :

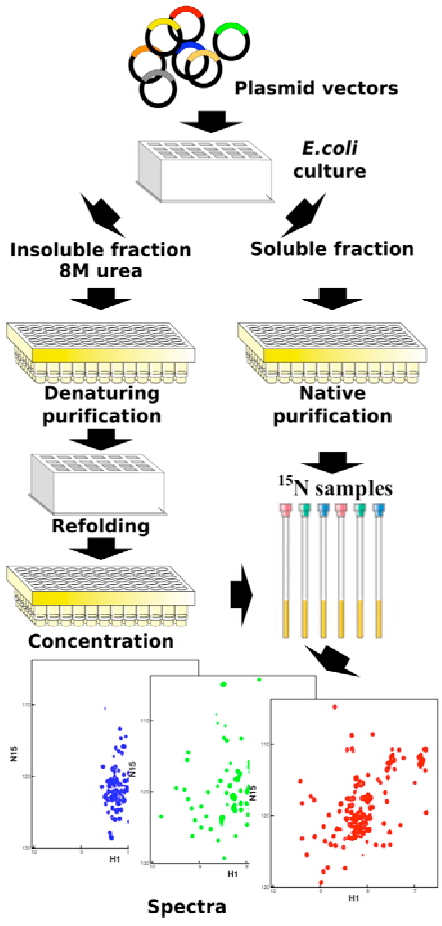

Figure 14 : Illustration schématique de la stratégie automatisée RoBiolMol-NMR

- Premièrement, en collaboration avec la plateforme RoBioMol de l'IBS, nous avons mis au point un protocole automatisé d'expression et de marquage isotopique des protéines. L'utilisation d'expériences RMN optimisées nous permet de détecter, en moins d'une heure, un spectre de corrélation 2D caractéristique, avec quelques nanomoles de protéines marquées ^{15}N. De tels spectres sont utilisés pour évaluer le nombre de résidus présents dans des régions structurées et non-structurées. Ces informations RMN nous permettent de déterminer la construction minimale formant un domaine structuré indépendant, une information cruciale pour optimiser les études structurales par RMN ou cristallographie aux rayons X. Le but de la plateforme RoBioMol-RMN est de pouvoir analyser 24 constructions différentes dans un délai de 3-5 jours (marquage isotopique et analyse RMN).

- Lorsqu'une construction, pour un domaine indépendant, a été optimisée, elle produite a grande échelle (Le protocole de marquage isotopique à petite échelle effectué sur la plateforme RoBioMol a l'avantage d'être facilement transposé à plus grande échelle). La solubilité et la stabilité sont optimisées grâce à des cribles mis en place sur les robots disponible à l'IBS. Une fois les conditions d'étude RMN optimisées, les fréquences spécifiques de chaque atome du squelette peptidique, peuvent être identifiées grâce à l'utilisation d'un set d'expérience RMN rapide. Le temps nécessaire pour l'attribution du squelette peptidique, pour des protéines de moins de 15 kDa (concentration ≥ 0.5 mM) est de l'ordre de 24 h, incluant l'acquisition de huit expériences 3D et leur analyse semi-automatique[48-49]. Cette étape d'attribution peut être utilisée pour identifier les résidus interagissant avec un partenaire physiologique.

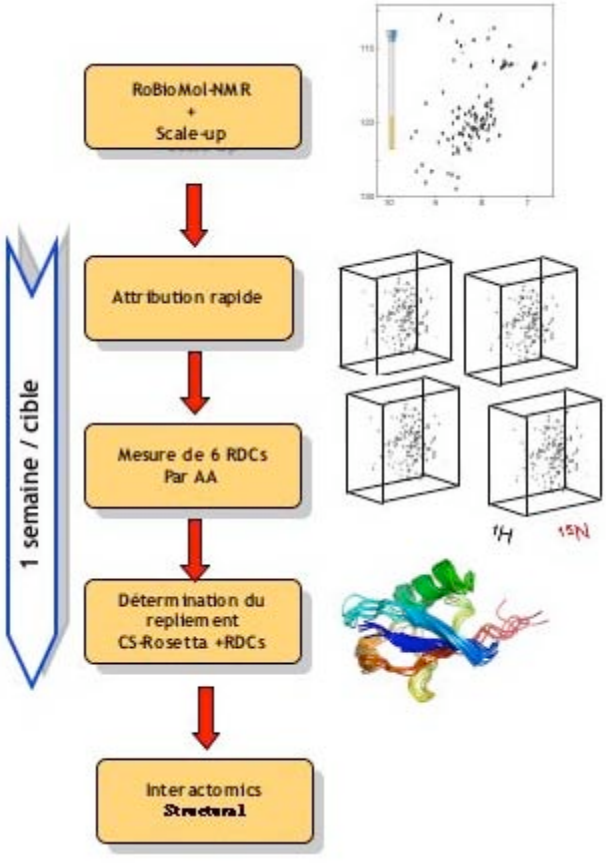

Figure 15 : Illustration schématique de la stratégie semi-automatisée CS-Rosetta-RDC

- La dernière étape consiste en la mise au point d'expériences de RMN rapides, permettant la détection de contraintes structurales. À ce jour, un jeu de 6 couplages résiduels dipolaires par plan peptidique peut être obtenu en deux jours. Basé sur la mesure rapide d'un jeu partiel de contraintes structurales, un protocole de calcul rapide a été optimisé pour déterminer le repliement

correspondant à la protéine à partir des seules informations de déplacement chimiques et des RDCs. La combinaison de ces deux approches permet de réduire considérablement le délai nécessaire pour déterminer la structure de protéine ou d'un domaine de taille modérée (< 15 kDa).

2) Base structurale de la spécificité de la bio-genèse des microARNs chez les plantes.

La biogenèse des miRNAs est un processus complexe dans lequel la structure du pre-miRNA joue un rôle clé dans l'identification de la position de la vingtaine de nucléotides qui constituera le miRNA mature. La biogenèse commence avec la transcription d'un transcrit primaire par une ARN polymérase II. Chez les plantes, les microRNAs sont transcrits comme de longs précurseurs de 90 à 300 nt. (pre-miRNA). Les pre-miRNAs sont reconnus et clivés par une RNAse III appelée DCL1 (*Dicer Like 1*) en association avec des protéines de reconnaissance de l'ARN (Serrate et HYL1) qui permettent de libérer le microRNA mature. Les miRNAs des plantes sont alors méthylés par Hen1 afin de les stabiliser et de prévenir leur uridinylation (Fig. 16).

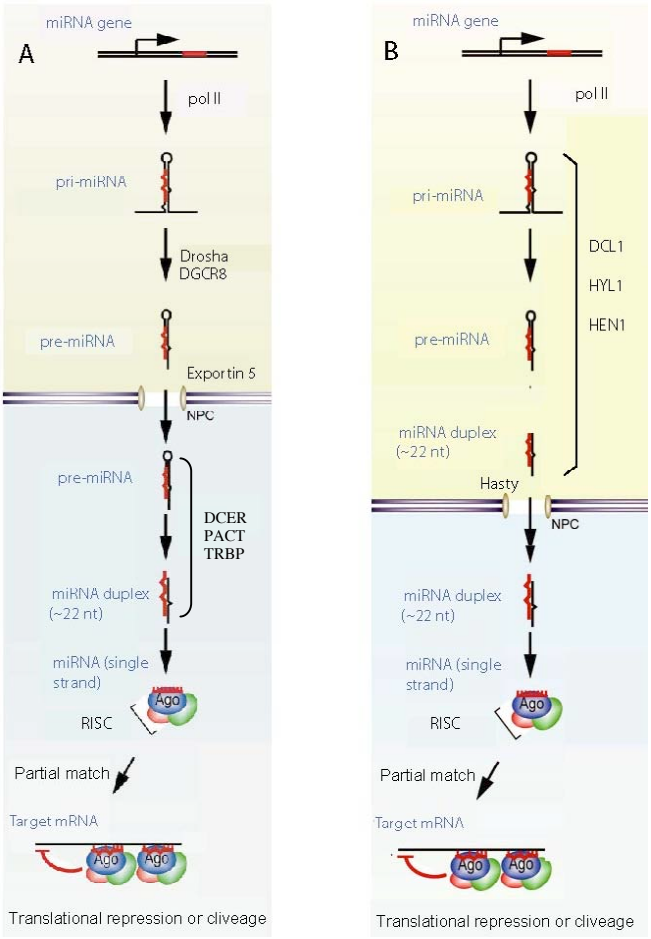

Figure 16 : Modèles du mécanisme de régulation de l'expression des gènes par les microRNAs chez l'homme (A) et chez les plantes (B).

La stratégie haut débit mise au point a été appliquée aux domaines de la protéine HYL1[50] et à trois autres domaines de la protéine DCL1 (Fig 17). Nous avons pu ainsi montrer qu'HYL1 était composée de deux domaines structurés et que le domaine C-terminal n'était pas structuré en solution que ce soit de façon isolée ou dans la protéine entière. Les deux premiers domaines d'HYL1 adoptent le repliement caractéristique des dsRBD et restent indépendants l'un de l'autre dans la protéine complète. Concernant DCL1, nous avons établi que parmi les domaines C-terminaux de liaison à l'ARN, seul le dernier formait un domaine dsRBD structuré et indépendant. Nous appliquons actuellement notre stratégie haut débit au domaine DUF283 situé au cœur de DCL1 pour lequel, il n'y a aucun homologue de séquence connue dont la structure ait

été déterminée. Suivant la même stratégie, de nombreux autres domaines de protéines impliquées dans la biogenèse des microRNAs de plante seront passés au crible à l'aide de la plateforme RoBioMol-NMR mise en place.

En janvier 2007, un ingénieur d'étude a été recruté sur contrat pour produire des ARN, afin d'identifier les surfaces d'interactions de ces protéines avec leur pre-miRNA. Un laboratoire a été équipé pour la production à grande échelle, le marquage isotopique et la purification d'ARN ; Plusieurs pre-miRNAs ont été produits et nous sommes en cours d'étude, par RMN et gel-retard, de l'interaction entre ces pre-miRNAs et les domaines indépendants de HYL1 et de DCL1 préalablement identifiés. Le but de ce projet est de résoudre la structure des complexes formés entre un pre-miRNA et les domaines de fixation à l'ARN des protéines impliquées dans la biogénèse des microRNA chez les plantes. À cette date, la plupart des ARN non-codant correspondant aux pre-miRNA est d'identifiée à partir de l'analyse du génome des organismes, mais la séquence du microRNA qui en dérive reste difficile à prédire. L'analyse des interactions impliquées à l'interface protéine/pre-miRNA servira de base à la compréhension de la sélectivité du clivage des pre-miRNA par DICER et à la découverte de nouveaux miRNAs et de leurs cibles. Ce projet, né d'une collaboration avec Dr. Palatnik, le biologiste qui a été le premier à pouvoir surexprimer des microRNAs dans les plantes[51], est supporté par un *grant* de *Human Frontier Science Program*, jusqu'en 2009.

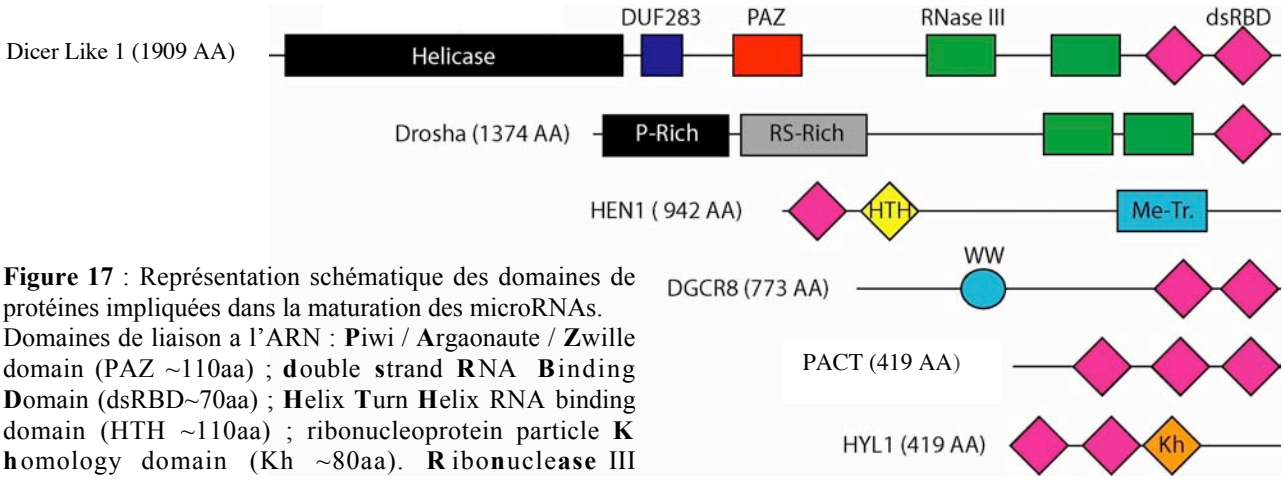

Figure 17 : Représentation schématique des domaines de protéines impliquées dans la maturation des microRNAs. Domaines de liaison a l'ARN : **P**iwi / **A**rgaonaute / **Z**wille domain (PAZ ~110aa) ; **d**ouble **s**trand **R**NA **B**inding **D**omain (dsRBD~70aa) ; **H**elix **T**urn **H**elix RNA binding domain (HTH ~110aa) ; ribonucleoprotein particle **K** **h**omology domain (Kh ~80aa). **R**ibo**n**ucl**ease** III catalytical domain (Rnase III ~150aa) ; **Me**- **Tr**ansferase domain (Me-Tr. ~150aa), **WW** Domain (30 AA).

3) Étude structurale de la de la biogenèse des microRNAs oncogènes humains.

Les microRNAs (miRNAs) sont à la clé du contrôle de nombreux processus chez l'homme. De récentes études, reliant le dérèglement de l'expression des miRNAs au cancer, ont placé ces petits ARNs au centre de l'oncologie moléculaire. Depuis la découverte en 2005 du premier miRNA oncogène[52-53], il est maintenant établi que la surexpression spécifique d'un miRNA peut

induire le développement d'une maladie néoplasique[54]. À ce jour, plus d'une vingtaine de miRNAs oncogènes a été identifiée chez l'homme. J'ai proposé d'étudier, au niveau atomique, les mécanismes à la base de la maturation des miRNAs oncogènes. Chez les animaux, les précurseurs des miRNAs sont constitués d'environ 60 à 80 nucléotides. Leur maturation implique un premier complexe de clivage appelé *microprocessor,* composé d'une RNAse de type III nommé DROSHA et d'une protéine d'ancrage sur l'ARN : DGCR8 (Fig. 16). Le *microprocessor* retire du transcrit primaire la structure tige-boucle constituant le pre-miRNA. Le récepteur nucléaire exportin-5 exporte le pre-miRNA vers le cytoplasme, où DICER accompagnée des protéines PACT et TRBP le clivera ensuite pour libérer le miRNA mature.

Pour mener à bien ce projet, j'ai établi une collaboration avec deux autres laboratoires de l'IBS afin de réunir les expertises nécessaires en Biochimie, Biophysique, Biologie Cellulaire, Moléculaire et Structurale. En fédérant nos compétences au sein d'un des 4 axes principaux de l'IBS, nous projetons d' :

1) Identifier et caractériser les domaines indépendants des protéines impliquées dans la maturation des miRNAs en utilisant la stratégie haut débit RoBioMol-NMR.

2) Identifier les surfaces d'interactions entre chacun de ces domaines et les précurseurs de miRNAs oncogènes.

3) De concevoir et tester, *in vitro* et *in vivo*, la capacité d'oligonucléotides d'interférer avec la biogenèse de cette nouvelle classe d'oncogènes.

Ce projet interdisciplinaire a pu être amorcé en 2007 grâce à un soutien de l'ANR, et est renforcé depuis janvier 2008 par le recrutement sur contrat d'un chercheur associé (financement ARC). Cependant des fonds supplémentaires seront nécessaires pour mettre en place les tests *in vitro* et *in vivo*, nécessaires pour le développement d'inhibiteurs spécifiques de la maturation des microRNAs oncogènes. À ce jour, nous avons pu optimiser l'expression et le marquage de deux domaines de DGCR8 (la protéine en charge d'ancrer le complexe de maturation nucléaire des microRNAs) d'un domaine inconnu de DICER (la protéine en charge de la maturation finale du microRNA dans le cytoplasme des cellules humaines) et de 4 domaines de PACT et TRBP (protéines nécessaires au bon fonctionnement de DICER chez l'homme). Les premiers précurseurs de microRNAs oncogènes sont en cours de production pour des études d'interaction avec les domaines de protéines impliquées dans leur maturation.

REFERENCES :

1. Lukavsky, P.J., Kim, I., Otto, G.A., Puglisi, J.D. *Structure of HCV IRES domain II determined by NMR.* **Nat. Struct. Biol.** 10, 1033-8 (2003).

2. D'Souza, V., Summers, M.F. *Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus.* **Nature**. 431, 586-90 (2004).

3. Bryce, D.L., Grishaev, A., Bax, A. *Measurement of ribose carbon chemical shift tensors for A-form RNA by liquid crystal NMR spectroscopy.* **J. Am. Chem. Soc**. 127, 7387-96 (2005).

4. Van Melckebeke, H., Devany, M., Di Primo, C., Beaurain, F., Toulmé, J.J., Bryce, D.L., Boisbouvier, J. *Liquid-crystal NMR structure of HIV TAR RNA bound to its SELEX RNA aptamer reveals the origins of the high stability of the complex*. **Proc. Natl. Acad. Sci. USA**. 105, 9210-5 (2008).

5. Pervushin, K., Riek, R., Wider, G., Wuthrich, K. *Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.* **Proc. Natl. Acad. Sci. USA.** 94, 12366-71 (1997).

6. Pervushin, K., Riek, R., Wider, G., Wuthrich, K. *Transverse Relaxation-Optimized Spectroscopy (TROSY) for NMR Studies of Aromatic Spin Systems in 13C-Labeled Proteins.* **J. Am. Chem. Soc**. 120, 6394-400 (1998).

7. Brutscher, B., Boisbouvier, J., Pardi, A., Marion, D., Simorre, J.-P. *Improved Sensitivity and Resolution in 1H-13C NMR Experiments of RNA.* **J. Am. Chem. Soc**. 120, 11845-51 (1998).

8. Brutscher, B., Boisbouvier, J., Kupce, E., Tisné, C., Dardel, F., Marion, D., Simorre, J.-P. *Base-type-selective high resolution 13C edited NOESY for sequential assignment of large RNA.* **J. Biomol. NMR**, 19,141-51 (2001).

9. Miclet, E., Williams, D., Clore, M., Bryce, D., Boisbouvier, J., Bax A. *Relaxation Optimized NMR Spectroscopy of Methylene Groups in Protein and Nucleic Acids.* **J. Am. Chem. Soc**. 126, 10560-70 (2004).

10. Boisbouvier, J., Brutscher, B,. Simorre J.-P., Marion, D. *13C spin relaxation measurements in RNA : Sensitivity and resolution improvement using spin-state selective correlation experiments.* **J. Biomol. NMR** 14, 241-52 (1999).

11. Boisbouvier, J. , Wu, Z., Ono, A, Kainosho, M., Bax, A. *Anisotropic Rotational Diffusion Tensor Determination of Nucleic Acids from 13C NMR Relaxation.* **J. Biomol. NMR** 27, 133-42 (2003).

12. Tjandra, N., Bax, A. *Large Variations in 13C Chemical Shift Anisotropy in Proteins Correlate with Secondary Structure.* **J. Am. Chem. Soc**. 119, 9576-7 (1997).

13. Boisbouvier, J., Brutscher, B., Pardi, A., Marion, D., Simorre, J.-P. *NMR determination of sugar puckers in nucleic acids from CSA-dipolar cross-correlated relaxation.* **J. Am. Chem. Soc**. 122, 6779-80 (2000).

14. Boisbouvier, J., Gans, P., Blackledge, M., Brutscher, B., Marion, D. *Long-range structural information in NMR studies of paramagnetic molecules from electron spin - nuclear spin cross-correlated relaxation.* **J. Am. Chem. Soc.** 121, 7700-1 (1999).

15. Hus, J.C., Marion, D., Blackledge, M. *De novo determination of protein structure by NMR using orientational and long-range order restraints*. **J. Mol. Biol.** 298, 927-36 (2001).

16. Wimperis, S., Bodenhausen, G. *Relaxation-Allowed Transfer of Coherence in NMR between Spins which are not Scalar Coupled.* **Chem. Phys. Lett.** 140, 41-45 (1987).

17. Boisbouvier, J., Bax, A. *Long-Range Magnetization Transfer Between Uncoupled Nuclei by Dipole-Dipole Cross-correlated Relaxation : A precise Probe of b-Sheet Geometry in Proteins*. **J. Am. Chem. Soc.** 124, 11038-45 (2002).

18. Tjandra, N., Bax, A. *Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium*. **Science**. 278,1111-4 (1997).

19. Hansen, M.R., Mueller, L., Pardi, A. *Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions*. **Nat. Struct. Biol.** 5, 1065-74 (1998).

20. Tjandra, N., Tate, S.I.; Ono, A., Kainosho, M., Bax, A. *The NMR Structure of a DNA Dodecamer in an Aqueous Dilute Liquid Crystalline Phase*. **J. Am. Chem. Soc.** 122, 6190-200 (2000).

21. Miclet, E., Boisbouvier, J., Bax. A. *Measurement of eight coupling from a single 3D NMR multiplet in proteins and nucleic acids*. **J. Biomol. NMR** 31, 201-16 (2005).

22. Boisbouvier, J., Bryce, D., O'Neil-Cabello, E., Nikonowicz, E., Bax, A. *Resolution optimized NMR measurement of 1DCH, 1DCC and 2DCH residual dipolar couplings in nucleic acids bases*. **J. Biomol. NMR** 30, 287-301 (2004).

23. Jaroniec , C., Boisbouvier, J., Nikonowicz, E., Bax, A. *Quantitative J correlation methods for the accurate measurement of 13C-15N dipolar couplings in nucleic bases.* **J. Biomol. NMR** 31, 231-41 (2005).

24. Boisbouvier, J. , Delaglio, F., Bax, A. *Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids.* **Proc. Natl. Acad. Sci. USA** 100, 11333-11338 (2003).

25. Weeks, K.M., Ampe, C., Schultz, S.C., Steitz, T.A., Crothers, D.M. *Fragments of the HIV-1 TAT protein specifically bind TAR RNA*. **Science** 249,1281-5 (1990).

26. Ducongé, F, Toulmé, J.J. *In vitro selection identifies key determinants for loop-loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1*. **RNA**. 5, 1605-14 (1999).

27. Ducongé, F., Di Primo, C., Toulmé, J.J. Is a closing *"GA pair" a rule for stable loop-loop RNA complexes?* **J. Biol. Chem.** 275,21287-94 (2000).

28. Van Melckebeke, H., Devany, M., Di Primo, C., Beaurain, F., Toulmé, J.J., Bryce, D.L., Boisbouvier, J. *Liquid-crystal NMR structure of HIV TAR RNA bound to its SELEX RNA aptamer reveals the origins of the high stability of the complex*. **Proc. Natl. Acad. Sci. USA.** 105, 9210-5 (2008).

29 Siridechadilok, B., Fraser, C.S., Hall, R.J., Doudna, J.A., Nogales, E. *Structural roles for human translation factor eIF3 in initiation of protein synthesis*. **Science**. 310, 1513-5 (2005).

30. ElAntak, L., Tzakos, A.G., Locker, N., Lukavsky, P.J. *Structure of eIF3b RNA recognition motif and its interaction with eIF3j: structural insights into the recruitment of eIF3b to the 40 S ribosomal subunit*. **J. Biol. Chem**. 282, 8165-74 (2007).

31. Perard, J., Rasia, R., Medenbach, J., Ayala, I., Boisbouvier, J., Drouet, E., Baudin, F. *Initiation factor eIF3 subunit b interacts with HCV IRES RNA through its N-terminal RNA recognition motif*. **(Submitted).**

32 Gardner, K. H., Kay, L. E. *Production and Incorporation of 15N, 13C, 2H (1H-1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies*. **J. Am. Chem. Soc**. 119,7599-600 (1997).

33. Sounier, R. Blanchard, L. Wu, Z., Boisbouvier, J. *High-Accuracy Distance Measurement Between Remote Methyls in Specifically Protonated Proteins*. **J. Am. Chem. Soc.** 129, 472-3 (2007).

34 Tugarinov V, Kay LE. *An isotope labeling strategy for methyl TROSY spectroscopy*. **J. Biomol. NMR** 28,165-72 (2004).

35 Ayala,I., Sounier, R., Usé, N., Gans, P., Boisbouvier, J. *An efficient protocol for the complete incorporation of methyl specifically protonated alanine in perdeuterated protein* **(in preparation).**

36. Isaacson, R.L., Simpson, P.J., Liu, M., Cota, E., Zhang, X., Freemont, P., Matthews, S. *A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues.* **J. Am. Chem. Soc.** 129,15428-9 (2007).

37. Koharudin, L.M., Bonvin, A.M., Kaptein, R., Boelens, R. *Use of very long-distance NOEs in a fully deuterated protein: an approach for rapid protein fold determination.* J. Magn. Reson. 163, 228-35 (2003).

38 Sounier, R., Ayala, I., Dura, A., Franzetti, B., Gans, P., Retel, J., Tugarinov, V., Kay, L., Grishaev, A., Boisbouvier, J. *Detection of long range NOE in macromolecular assemblies beyond 100 kDa: possibilities and limits.* **(in preparation)**.

39. Wu, Z., Bax, A. *Measurement of long-range 1H-1H dipolar couplings in weakly aligned proteins.* **J. Am. Chem. Soc.** 124, 9672-3 (2002).

40. Meier, S., Hussinger, D., Jensen, P., Rogowski, M., Grzesiek, S. *High-accuracy residual 1HN-13C and 1HN-1HN dipolar couplings in perdeuterated proteins*. **J. Am. Chem. Soc.** 125, 44-5 (2003).

41. Boisbouvier, J. , Delaglio, F., Bax, A. *Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids*. **Proc. Natl. Acad. Sci. USA**, 100, 11333-11338 (2003).

42. Sounier, R., Wu, Z., Boisbouvier, J. Long range distance and angular information between remote methyls measured by liquid crystal NMR. (**in preparation**).

43. Nitz, M., Sherawat, M., Franz, K.J., Peisach, E., Allen, K.N., Imperiali, B. *Structural origin of the high affinity of a chemically evolved lanthanide-binding peptide.* **Angew. Chem. Int. Ed. Engl**. 43, 3682-5 (2004).

44. Ikegami, T., Verdier, L., Sakhaii, P., Grimme, S., Pescatore, B., Saxena, K., Fiebig, K.M., Griesinger, C. *Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions.* **J. Biomol. NMR** 29, 339-49 (2004).

45. Franzetti, B., Schoehn, G., Hernandez, J.F., Jaquinod, M., Ruigrok, R.W., Zaccai, G. *Tetrahedral aminopeptidase: a novel large protease complex from archaea.* **EMBO J**. 21, 2132-8 (2002).

46a. Lee, R.C., Feinbaum, R.L., Ambros, V. *The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14*. **Cell**. 75, 843-54 (1993).

47. Wightman, B., Ha, I., Ruvkun, G. *Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans.* **Cell**. 75, 855-62 (1993).

48. Lescop, E., Rasia, R., Brutscher, B. *Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment.* **J. Am. Chem. Soc**. 130, 5014-5 (2008).

49. Lescop, E., Schanda, P., Rasia, R., Brutscher, B. *Automated spectral compression for fast multidimensional NMR and increased time resolution in real-time NMR spectroscopy.* **J. Am. Chem. Soc**. 129, 2756-7 (2007).

50. Han, M.H., Goud, S., Song, L., Fedoroff, N. *The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation.* **Proc. Natl. Acad. Sci. U S A**. 101, 1093-8 (2004).

51. Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., Weigel, D. *Control of leaf morphogenesis by microRNAs*. **Nature** 425, 257-63 (2003).

52. O'Donnell, K.A., Wentzel, E.A, Zeller, K.I., Dang, C.V., Mendell, J.T. *c-Myc-regulated microRNAs modulate E2F1 expression.* **Nature.** 435, 839-43 (2005).

53. He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J., Hammond, S.M. *A microRNA polycistron as a potential human oncogene*. **Nature**. 435, 828-33 (2005).

54. Costinean, S., Zanesi, N., Pekarsky, Y., Tili, E., Volinia, S., Heerema, N., Croce, C.M. *Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice.* **Proc. Natl. Acad. Sci. USA**. 103,7024-9 (2006).

ANNEXES :

J'ai choisi quatre articles parmi mes publications que je trouve les plus significatifs de mon activité scientifique.

Article 1 : *Relaxation Optimized NMR Spectroscopy of Methylene Groups in Protein and Nucleic Acids.*

Miclet, E., Williams, D., Clore, M., Bryce, D., Boisbouvier, J. & Bax, A. **J. Am. Chem. Soc** 126, 10560-70 (2004).

Article 2 : *Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids.*

Boisbouvier, J., Delaglio, F. & Bax, A. **Proc. Natl. Acad. Sci.** USA, 100, 11333-8 (2003).

Article 3 : *Liquid Crystal NMR Structure of HIV TAR RNA Bound to its SELEX RNA Aptamer Reveals the Origins of the High Stability of the Complex.*

Van Melckebeke, H., Devany, H.H., Di Primo, C., Beaurain, F., Toulmé, J.J., Bryce, D.L. Boisbouvier, J. **Proc. Natl. Acad. Sci. USA** 105, 9210-5 (2008)**.**

Article 4 : *High-Accuracy Distance Measurement Between Remote Methyls in Specifically Protonated Proteins.*

Sounier, R., Blanchard, L., Wu, Z. & Boisbouvier J. **J. Am. Chem. Soc** 129, 472-3 (2007).

J|A|C|S
ARTICLES
Published on Web 08/10/2004

Relaxation-Optimized NMR Spectroscopy of Methylene Groups in Proteins and Nucleic Acids

Emeric Miclet, David C. Williams Jr., G. Marius Clore, David L. Bryce, Jérôme Boisbouvier,*,† and Ad Bax*

Contribution from the Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0520

Received April 12, 2004; E-mail: jerome.boisbouvier@ibs.fr; bax@nih.gov

Abstract: A large fraction of hydrogens in proteins and nucleic acids is of the methylene type. Their detailed study, however, in terms of structure and dynamics by NMR spectroscopy is hampered by their fast relaxation properties, which give rise to low sensitivity and resolution. It is demonstrated that six different relaxation interference processes, involving $^1H-^{13}C$ and $^1H-^1H$ dipolar interactions and ^1H and ^{13}C chemical shift anisotropy, can be used simultaneously to mitigate these problems effectively. The approach is applicable to the majority of NMR experiments commonly used to study side chain and backbone conformation. For proteins, its efficiency is evaluated quantitatively for two samples: the third IgG-binding domain from Streptococcal Protein G and the protein calmodulin complexed with a 26-residue target peptide. Gains in both resolution and sensitivity by up to factors of 3.2 and 2.0, respectively, are observed for Gly residues at high magnetic field strengths, but even at much lower fields gains remain substantial. The resolution enhancement obtained for methylene groups makes possible a detailed analysis of spectral regions commonly considered inaccessible due to spectral crowding. For DNA, the high resolution now obtainable for C_5' sites permits an H_5'/H_5''-based sequential NOE assignment procedure, complementary to the conventional base-$H_1'/H_2'/H_2''$ pathway.

Introduction

In recent years, it has been shown that transverse-relaxation-optimized NMR spectroscopy (TROSY) methods can extend the application of NMR to considerably larger systems. To date, TROSY experiments have been described for optimizing observation of $^{15}N-^1H$ amide groups,[1,2] aromatic $^{13}C-^1H$ sites,[3−5] and most recently $^{13}C-^1H_3$ methyl groups.[6] No such methods have been available for methylene groups, even though they represent 46% of the observable protons in proteins, and the C_5' methylene group in nucleic acids represents a critical backbone element. Frequently, these methylene regions are difficult to access due to severe spectral crowding.

Here we present an approach, based on relaxation-optimized spectroscopy and multiplet simplification, that considerably improves both the resolution and sensitivity of methylene group NMR studies. The approach is applicable both to small molecules, where it is primarily the suppression of geminal $^1H-^1H$ splittings that enhances spectral appearance, and biological macromolecules, where the benefits of the improved relaxation properties selected in our experiments further enhance spectral quality. Conceptually, the new approach is an extension of earlier E.COSY based technology,[7] aimed at measurement of $^1H-^{13}C$ and $^1H-^1H$ dipolar couplings in protein[8] and nucleic acid[9] methylene groups. The element introduced here for optimized NMR observation of methylene groups can be incorporated in many of today's most common 2D and 3D NMR experiments. As an example, we show its incorporation in 3D ^{13}C-separated NOESY, enabling observation of the sequential $H_5'-H_1'$ NOEs in DNA and thereby providing sequential assignment information. The experiments also provide access to the measurement of the geminal $^1H-^1H$ scalar and dipolar coupling, as well as both corresponding one-bond $^1H-^{13}C$ interactions.

Methods

All NMR experiments were carried out on Bruker DRX spectrometers, operating at ^1H frequencies of 500 or 800 MHz and equipped respectively with a cryogenic (500 MHz) or a regular (800 MHz) triple resonance pulsed field gradient probehead, optimized for ^1H detection.

† Present address: Laboratoire de RMN, Institut de Biologie Structurale, Jean-Pierre Ebel, UMR 5075 CNRS-CEA-UJF, 41 rue Jules Horowitz, 38027 GRENOBLE Cedex 1, France.

(1) Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12366−12371.
(2) Riek, R.; Wider, G.; Pervushin, K.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4918−4923.
(3) Pervushin, K.; Riek, R.; Wider, G.; Wuthrich, K. *J. Am. Chem. Soc.* **1998**, *120*, 6394−6400.
(4) Brutscher, B.; Boisbouvier, J.; Pardi, A.; Marion, D.; Simorre, J. P. *J. Am. Chem. Soc.* **1998**, *120*, 11845−11851.
(5) Fiala, R.; Czernek, J.; Sklenar, V. *J. Biomol. NMR* **2000**, *16*, 291−302.
(6) Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 10420−10428.
(7) Griesinger, C.; Sørensen, O. W.; Ernst, R. R. *J. Chem. Phys.* **1986**, *85*, 6837−6852.
(8) Carlomagno, T.; Peti, W.; Griesinger, C. *J. Biomol. NMR* **2000**, *17*, 99−109.
(9) Miclet, E.; O'Neil-Cabello, E.; Nikonowicz, E. P.; Live, D.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 15740−15741.

All data sets have been processed with nmrPipe.[10] For resolution and sensitivity comparisons, no apodization of the time domain data was employed. Data were zero-filled by a factor of 2 in all dimensions prior to Fourier transformation.

Three different samples were employed. The first sample comprised 1.8 mM uniformly (>95%) ^{13}C,^{15}N-labeled third IgG-binding domain from Streptococcal Protein G, hereafter referred to as GB3, dissolved in D_2O, pH 5.6, 50 mM sodium phosphate (data were acquired at 10 °C). The second protein sample contained 1 mM uniformly (>95%) ^{13}C,^{15}N-enriched calmodulin, complexed with an unlabeled 26-residue peptide comprising the calmodulin-binding domain of skeletal muscle myosin light chain kinase, hereafter referred to as CaM/M13,[11] dissolved in D_2O, pH 6.8, 100 mM KCl (data acquired at 35 °C). The third sample corresponds to a 19-base-pair DNA binding site of the transcription factors Oct-1 and Sox-2 in the HoxB1 promoter.[12] One strand (purchased from Silantes GmbH, Munich, Germany) is uniformly ^{13}C,^{15}N-enriched (5′-TGTCTTTGTCATGCTAATG-3′) and forms a duplex with the unlabeled complementary DNA strand (5′-CATTAG-CATGACAAAGACA-3′). The DNA duplex was dissolved in D_2O, at a concentration of 0.8 mM, in 10 mM sodium phosphate, 10 mM dithiothreitol, and 0.02% NaN_3, pH 6.5 (data acquired at 35 °C).

Quantum chemical calculations of methylene proton chemical shielding tensors were carried out using the GIAO method as implemented in Gaussian03,[13] the hybrid B3LYP functional, and the 6-311++G(3df,3pd) basis set on all atoms. Results reported are for the methylene group of the model system $H_3C-N(H)-C(O)-CH_2-N(H)-CH_3$, which was constructed using standard bond lengths and angles, a N−C′−C^{α}−N dihedral angle (psi) of 135°, and a C′−C^{α}−N−C^{Me} dihedral angle of −139°, to mimic a glycine residue in an antiparallel beta strand geometry. Calculations of the absolute chemical shielding tensors for Gly-$^1H^{\alpha}$ methylene protons yield the following parameters: for the $H^{\alpha 2}$ proton (pro-R), $\sigma_{xx} = 23.0$ ppm, $\sigma_{yy} = 28.3$ ppm, $\sigma_{zz} = 30.4$ ppm; angles between the C−H vector and these three principal components are equal to 82.6°, 7.4°, 89.5°, and angles between the H−H vector and these three principal components are equal to 47.3°, 42.8°, 89.7°. For the $H^{\alpha 3}$ proton (pro-S), the principal components are $\sigma_{xx} = 23.9$ ppm, $\sigma_{yy} = 27.9$ ppm, $\sigma_{zz} = 33.6$ ppm; angles between the C−H vector and these three principal components are equal to 82.7°, 73.7°, 18.0°; and angles between the H−H vector and these three principal components are equal to 49.8°, 67.0°, 49.0°. Calculations at the same level of theory on tetramethylsilane (B3LYP/6-311+G* optimized structure) yield a proton isotropic chemical shielding of 31.6 ppm.

Theoretical Section

We consider a system of three coupled spin-$^1/_2$ nuclei (^{13}C, $^1H^1$, and $^1H^2$), corresponding to an isolated methylene group. Each spin is assumed to be characterized by an axially symmetric chemical shielding tensor. The single quantum NMR spectrum of each proton consists of a doublet of doublets, corresponding to the different spin states of the other ^1H and the ^{13}C. Each of the four components of proton H^1 (Figure 1a) can be described by a single-transition operator (H^{PQ}), which can be represented as a combination of product operator terms,[14] describing in-phase and antiphase transverse

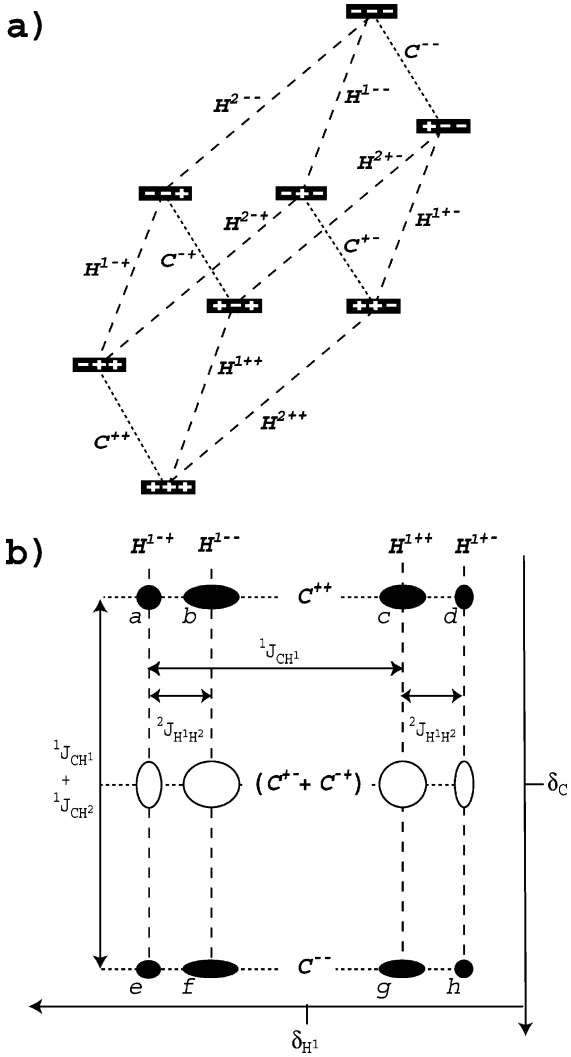

Figure 1. NMR transitions in a $^{13}CH_2$ group. (a) Energy level diagram of an isolated CH^1H^2 spin system. Each eigenstate is defined by three signs "+" or "−" (with "+" for $m_z = +^1/_2$ and "−" for $m_z = -^1/_2$), where the first sign corresponds to the ^{13}C spin state and the two latter signs to H^1 and H^2. Long-dashed lines depict ^1H transitions, and short-dashed lines are ^{13}C transitions, each labeled with its single transition operator, defined in eqs 1 and 3. (b) Schematic representation of the 2D multiplet pattern observed for an individual methylene ^1H in a fully coupled HSQC experiment. The diagram corresponds to $^1J_{CH} > 0$, $^2J_{HH} < 0$ and $^1J_{CH^1} = {}^1J_{CH^2} \gg |^2J_{H^1H^2}|$. For $^1J_{CH^1} = {}^1J_{CH^2}$, C^{-+} and C^{+-} transitions have vanishing intensity (open ellipsoids). The chemical shift frequencies are marked δ_C and δ_H. Horizontal short-dashed lines and vertical long-dashed lines mark ^{13}C and ^1H frequency transitions, respectively. The transverse relaxation rate for each transition is defined in eqs 2, 4, and 5. Italic characters *a*−*h* correspond to the panels of Figure 4 which display this multiplet component experimentally.

coherences H_x^1, $2H_x^1C_z$, $2H_x^1H_z^2$, $4H_x^1C_zH_z^2$:

$$H^{PQ} = H_x^1 + p \times 2H_x^1C_z + q \times 2H_x^1H_z^2 + pq \times 4H_x^1C_zH_z^2 \quad (1)$$

where P and Q characterize the eigenstates of C_z and H_z^2 respectively: $(P,Q) \in \{(+,+);(+,-);(-,+);(-,-)\}$, with "+" and "−" corresponding to $m_z = +^1/_2$ ($|\alpha\rangle$ spin state) and $m_z = -^1/_2$ ($|\beta\rangle$ spin state), respectively. Correspondingly, $p = 1$ for $P = +$; $p = -1$ for $P = -$; $q = 1$ for $Q = +$; $q = -1$ for $Q = -$. With the relative magnitude and the sign of couplings involved in a methylene spin system taken into account, the position of each multiplet component, corresponding to an

(10) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(11) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. *Science* **1992**, *256*, 632−638.
(12) Williams, D. C.; Cai, M. L.; Clore, G. M. *J. Biol. Chem.* **2004**, *279*, 1449−1457.
(13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Pople, J. A., et al. *Gaussian 03*, revision B.04; Gaussian Inc.: Pittsburgh, PA, 2003.
(14) Sørensen, O. W.; Eich, G. W.; Levitt, M. H.; Bodenhausen, G.; Ernst, R. R. *Prog. Nucl. Magn. Reson. Spectrosc.* **1983**, *16*, 163−192.

individual ^{1}H transition, is indicated in Figure 1b. As a result of interference between the different dipole−dipole and chemical shift anistropy (CSA) relaxation mechanisms operating on a given ^{1}H, different line widths are expected for each multiplet component. The relaxation rate of an individual transition is described by[15]

$$R(H^{PQ}) = \Gamma^{DD,DD}_{H^1C,H^1C} + \Gamma^{DD,DD}_{H^1H^2,H^1H^2} + \Gamma^{CSA,CSA}_{H^1,H^1} - p \times \Gamma^{CSA,DD}_{H^1,H^1C} - q \times \Gamma^{CSA,DD}_{H^1,H^1H^2} + pq \times \Gamma^{DD,DD}_{H^1C,H^1H^2} \quad (2)$$

where the first three terms correspond to dipolar ($\Gamma^{DD,DD}_{H^1C,H^1C}$, $\Gamma^{DD,DD}_{H^1H^2,H^1H^2}$) and CSA ($\Gamma^{CSA,CSA}_{H^1,H^1}$) autorelaxation rates, and the last three terms correspond to cross-correlated relaxation between H^1 CSA and H^1−C dipole−dipole interaction ($\Gamma^{CSA,DD}_{H^1,H^1C}$), H^1 CSA and H^1−H^2 dipole−dipole interaction ($\Gamma^{CSA,DD}_{H^1,H^1H^2}$), and dipole−dipole cross-correlated relaxation between H^1−C and H^1−H^2 ($\Gamma^{DD,DD}_{H^1C,H^1H^2}$). The four single-transition operators and corresponding relaxation rates for proton H^2 are obtained by interchanging indices 1 and 2 in eqs 1 and 2.

The four ^{13}C transitions (Figure 1a) are described by

$$C^{PQ} = C_x + p \times 2C_xH^1_z + q \times 2C_xH^2_z + pq \times 4C_xH^1_zH^2_z \quad (3)$$

where P and Q now represent the eigenstates of H^1_z and H^2_z, respectively. In a ^{1}H-coupled HSQC spectrum of a methylene group, the two center components of the doublet-of-doublets, corresponding to transitions C^{+-} and C^{-+}, are of opposite phase and cancel one another. The spectrum then appears as a doublet in the ^{13}C dimension, consisting of the two observable transitions C^{++} and C^{--}, separated by $^1J_{CH^1} + {}^1J_{CH^2}$. Analogous to eq 2, the relaxation rates of the ^{13}C multiplet components are described by

$$R(C^{PQ}) = \Gamma^{DD,DD}_{CH^1,CH^1} + \Gamma^{DD,DD}_{CH^2,CH^2} + \Gamma^{CSA,CSA}_{C,C} - p \times \Gamma^{CSA,DD}_{C,CH^1} - q \times \Gamma^{CSA,DD}_{C,CH^2} + pq \times \Gamma^{DD,DD}_{CH^1,CH^2} \quad (4)$$

Assuming isotropic rotational diffusion in the slow tumbling limit for a macromolecule with fast internal dynamics of moderate amplitude, the individual auto- and cross-correlated relaxation rates (eqs 2 and 4) are given by[15]

$$\Gamma^{DD,DD}_{HC,HC} = \Gamma^{DD,DD}_{CH,CH} = (\xi^{DD}_{HC})^2\tau_C\mathbf{S}^2/5 \quad (5a)$$

$$\Gamma^{DD,DD}_{H^1H^2,H^1H^2} = (\xi^{DD}_{H^1H^2})^2\tau_C\mathbf{S}^2/4 \quad (5b)$$

$$\Gamma^{CSA,CSA}_{A,A} = 4(\xi^{CSA}_{A})^2\tau_C\mathbf{S}^2/45 \quad (5c)$$

$$\Gamma^{CSA,DD}_{A,AM} = 4\xi^{CSA}_{A}\,\xi^{DD}_{AM}P_2(\cos\theta^{CSA,DD}_{A,AM})\tau_C\mathbf{S}^2/15 \quad (5d)$$

$$\Gamma^{DD,DD}_{AM,AX} = 2\xi^{DD}_{AM}\,\xi^{DD}_{AX}P_2(\cos\theta^{DD,DD}_{AM,AX})\tau_C\mathbf{S}^2/5 \quad (5e)$$

where A, M, and X denote the C, H^1, or H^2 spins, in arbitrary order, $\xi^{DD}_{AM} = (\mu_o h\gamma_A\gamma_M/8\pi^2)\langle r^{-3}_{AM}\rangle$, $\xi^{CSA}_{A} = \gamma_A B_0\Delta\sigma_A$, $\theta^{\mu,\mu'}$ is the angle between the principal axis of interactions μ and μ', τ_c is the global correlation time, r_{AM} is the distance between nuclei A and M, and $\mathbf{S}^2$ is the squared generalized order parameter.[16] Although, strictly speaking, a specific order parameter should be introduced for each pair of correlated interactions, in this study we assume a uniform value of $\mathbf{S}^2 = 0.8$, frequently a reasonable approximation.[17] The symbol $\Delta\sigma_A$ denotes the chemical shielding anisotropy of spin A, defined as $\Delta\sigma_A = (\sigma_{11} - (\sigma_{22} + \sigma_{33})/2)$, where σ_{11} is the chemical shielding tensor component furthest from the isotropic shielding value.

Results and Discussion

As described in the theoretical section, the fully coupled (in both ^{1}H and ^{13}C dimensions) HSQC spectrum will show eight multiplet components for each methylene proton of an isolated $^{13}CH_2$ group. The different coefficients applicable for each multiplet component when calculating the total relaxation rate from the sum of the autorelaxation and cross correlation terms of eqs 2−5, $\Gamma^{CSA,DD}_{H^1,H^1C}$, $\Gamma^{CSA,DD}_{H^1,H1H^2}$, $\Gamma^{DD,DD}_{H^1C,H^1H^2}$, $\Gamma^{CSA,DD}_{C,CH^1}$, $\Gamma^{CSA,DD}_{C,CH^2}$, and $\Gamma^{DD,DD}_{CH^1,CH^2}$, result in different line widths for each of these eight components, schematically indicated in Figure 1b. Analogous to the well-known ^{1}H−^{15}N TROSY experiment,[1] selection of the slowest relaxing component will enhance spectral resolution. As described below, such a selection can be made with high efficiency by a suitably optimized pulse scheme, yielding considerable resolution enhancement and even increases in sensitivity over conventional HSQC experiments. Below, the experiment is quantitatively analyzed for the three-spin system of Gly residues in uniformly ^{13}C-enriched proteins, dissolved in D_2O. Application to other methylene spin systems in biomacromolecules, outside of the isolated-three-spin approximation, will also be presented.

Methylene Spin-State-Selective Correlation. Figure 2a shows the implementation of a 2D experiment that specifically selects one of the eight methylene HSQC multiplet components. This 2D CH_2-TROSY pulse scheme is very similar to the standard inverse heteronuclear two-dimensional sensitivity-enhanced HSQC experiment[18] but lacks ^{1}H decoupling during t_1 and ^{13}C decoupling during t_2. It also uses modified timing durations in the final "Rance−Kay" element of the pulse sequence to accomplish $^{13}CH_2$ spin-state-selective coherence transfer or CH_2-S^3CT.

The pulse scheme starts with an initial INEPT transfer of ^{1}H magnetization to ^{13}C (time point *a*). An S^3E element,[19,20] between time points *a* and *b*, then selects the slowest relaxing, most downfield component of the ^{13}C multiplet, C^{--}. The total S^3E duration, normally $1/(4^1J_{CH})$ for methine ^{13}C−^{1}H groups, is reduced to $1/(8^1J_{CH}) \approx 880$ μs, to account for the 2^1J_{CH} separation between the outer multiplet components in the ^{13}C dimension. The C^{++} magnetization component is eliminated by the 90° ^{13}C pulse applied at time point *b*, which rotates it to the *z*-axis, whereas the C^{--} magnetization remains in the transverse plane and evolves during t_1 with frequency $\delta_C + {}^1J_{CH}$. This magnetization is then encoded by gradients G_4 and G_5. As pointed out above, C^{--} can be written as the sum of the four product operator terms: $C^{--} = C_x - 2C_xH^1_z - 2C_xH^2_z + 4C_xH^1_zH^2_z$. The subsequent methylene spin-state-selective coherence transfer (point *d* to *e* in Figure 2a) transforms C^{--} into the following four ^{1}H single quantum coherences: H^1_x,

(15) Kumar, A.; Grace, R. C. R.; Madhu, P. K. *Prog. Nucl. Magn. Reson. Spectrosc.* **2000**, *37*, 191−319.
(16) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(17) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.
(18) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663−10665.
(19) Meissner, A.; Duus, J. O.; Sorensen, O. W. *J. Magn. Reson.* **1997**, *128*, 92−97.
(20) Meissner, A.; Duus, J. O.; Sorensen, O. W. *J. Biomol. NMR* **1997**, *10*, 89−94.

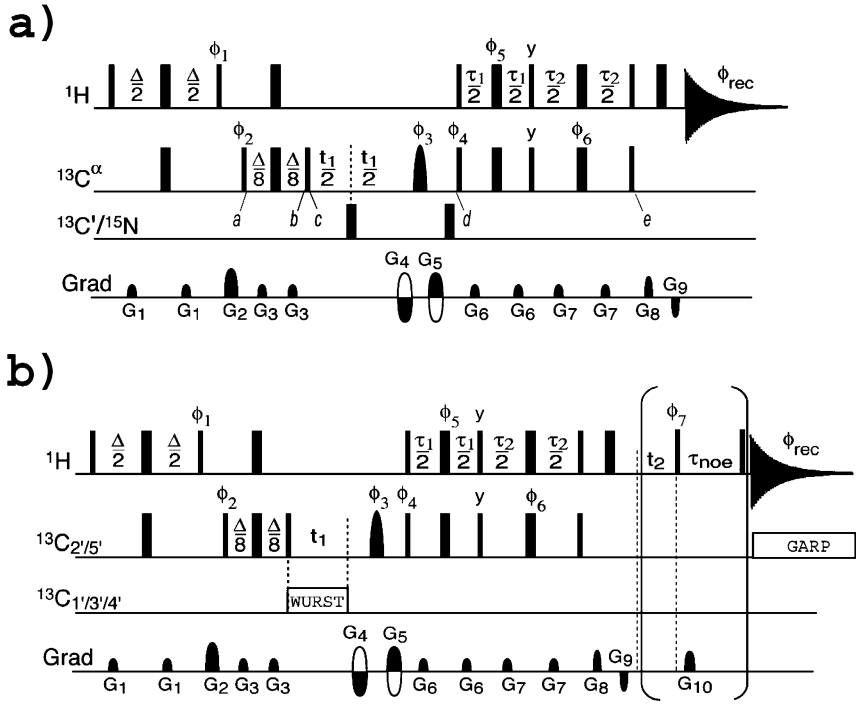

Figure 2. Pulse schemes of CH_2-TROSY experiments. Narrow and wide bars indicate nonselective 90° and 180° pulses, respectively. Unless specified, pulse phases are *x*. Delay durations: $\Delta = 1/(2J_{CH})$; $\tau_1 = 0.34/(J_{CH})$; $\tau_2 = 0.23/(J_{CH})$. For selection of the downfield, C^{--} components: $\phi_1 = -y$; $\phi_2 = (225°, 45°)$; $\phi_5 = x$; and $\phi_6 = x$ are used to select the H^{+-} component in the ^{1}H dimension (see Table 1); $\phi_3 = (x,x,y,y)$, $\phi_4 = x$ and $\phi_{rec} = (x,-x,-x,x)$. The experiment is recorded in the regular Rance−Kay manner:[18] for each t_1 increment, two FID's are acquired, one with G_4, G_5, and ϕ_4 inverted, and stored separately. Field gradients are sine-bell shaped with durations $G_{1,...,10}$ of 1, 2, 0.25, 1, 1, 0.2, 0.3, 0.35, 0.153, 0.6 ms; amplitudes of 10, 18, 12, 30, 30, 10, 12, 30, 30, 18 G/cm; and directions *x*, *xy*, *y*, −*z*, *z*, *x*, *y*, *z*, −*z*, *xyz*. (a) 2D CH_2-TROSY experiment, in the nonconstant-time mode (i.e., optimal for Gly). The shaped $^{13}C^{\alpha}$ pulse has a REBURP profile (3.1 ms duration at 800 MHz for a 6 ppm bandwidth inversion). $^{13}C'$ 180° pulses (carrier at 177 ppm) are sine-bell-shaped and have durations of 180 μs. (b) Pulse scheme of the 2D CH_2-TROSY and 3D CH_2-TROSY-NOESY adapted for nucleic acids. The shaped $^{13}C'_{2/5}$ pulse has been generated by coadding two REBURP profiles,[21] designed to refocus selectively two separate frequency regions: 65−70 ppm and 38−43 ppm (3.75 ms pulse duration at 800 MHz). Homonuclear band-selective $^{13}C'_{1,3,4}$ effective decoupling over the 76−89 ppm frequency range is achieved by means of a train of 180° adiabatic pulses (WURST-4 created with the Bruker shape tool, Q_0 factor = 4, for an inversion bandwidth of ±1300 Hz, maximum radio frequency field strength $\gamma B_1/2\pi = 804$ Hz, 5.25 ms duration at 800 MHz)[22] in a p5m4 composite decoupling sequence {2 × (0, 150, 60, 150, 0), 2 × (180, 330, 240, 330, 180)}.[23] ^{13}C decoupling during acquisition is only applied for the 3D experiments, using a GARP decoupling sequence with $\gamma B_1/2\pi = 4.2$ kHz. Quadrature detection in ^{13}C (t_1) and ^1H (t_2) dimension is achieved by recording four separate free induction decays or FIDs (E_1 to E_4) per pair of (t_1,t_2) values, with the following phases and gradient settings: $E_1 = \{\phi_4 = x, \phi_7 = x, G_4, G_5\}$; $E_2 = \{\phi_4 = -x, \phi_7 = x, -G_4, -G_5\}$; $E_3 = \{\phi_4 = x, \phi_7 = y, G_4, G_5\}$; and $E_4 = \{\phi_4 = -x, \phi_7 = y, -G_4, -G_5\}$. Prior to Fourier transformation, the FIDs are rearranged as follows: $E_1' = \{E_1 + E_2\}$; $E_2' = \{E_4 - E_3\}$; $E_3' = \{E_3 + E_4\}$; and $E_4' = \{E_1 - E_2\}$. Corresponding Bruker pulse programs and NMRpipe macros are available at http://spin.niddk.nih.gov/bax/.

$2H_x^1C_z$, $-2H_x^1H_z^2$, $-4H_x^1C_zH_z^2$. With identical values assumed for the two one-bond couplings, $^1J_{CH^1} = {}^1J_{CH^2} = J_{CH}$, the transfer efficiency in this CH_2-S^3CT element is the same for generation of $H_x^1(\tau_1, \tau_2)$ and $-4H_x^1C_zH_z^2(\tau_1, \tau_2)$ terms and described by a factor $f(\tau_1, \tau_2)$:

$$f(\tau_1, \tau_2) = \cos(\pi J_{CH}\tau_1)\sin(\pi J_{CH}\tau_2) + \sin(\pi J_{CH}\tau_1) \quad (6a)$$

Generation of $2H_x^1C_z(\tau_1, \tau_2)$ and $-2H_x^1H_z^2(\tau_1, \tau_2)$ also carries a common prefactor:

$$g(\tau_1, \tau_2) = \sin(\pi J_{CH}\tau_1)\cos(\pi J_{CH}\tau_1)\cos(\pi J_{CH}\tau_2) + \cos(\pi J_{CH}\tau_2)\sin(\pi J_{CH}\tau_2)[1 + \sin^2(\pi J_{CH}\tau_1)] \quad (6b)$$

If τ_1 and τ_2 delays are chosen such that $f(\tau_1, \tau_2) = g(\tau_1, \tau_2)$, cancellation occurs for signals corresponding to transitions H^{++}, H^{--}, and H^{-+} (Figure 3a). In contrast, the terms contributing to H^{+-} magnetization add constructively, yielding a ^{1}H multiplet in which only the most upfield component has nonzero intensity (Figure 3a). Although a continuous range of (τ_1, τ_2) values can satisfy $f(\tau_1, \tau_2) = g(\tau_1, \tau_2)$ (Figure 3b), the solution ($\tau_1 = 0.34/{}^1J_{CH}$; $\tau_2 = 0.23/{}^1J_{CH}$) corresponds to the highest transfer efficiency (Supporting Information Figure S1), and use of these values maximizes the sensitivity of correlating C^{--} and H^{+-} transitions.

Figure 4 demonstrates that each of the eight $^1H-^{13}C$ multiplet components for a given methylene proton can be selected separately with the pulse scheme of Figure 2a. The correlations correspond to the $C^{\alpha}H^{\alpha 3}$ spin pair of Gly-41 in GB3. Different correlations in the ^{1}H dimension are selected by changing phases ϕ_5 and ϕ_6 in the CH_2-S^3CT element in the manner described in Table 1. These phases control the signs of the (τ_1, τ_2)-dependent transverse ^{1}H magnetization terms, generated by the CH_2-S^3CT element: $H_x^1(\tau_1, \tau_2)$, $4H_x^1C_zH_z^2(\tau_1, \tau_2)$, $2H_x^1C_z(\tau_1, \tau_2)$, and $2H_x^1H_z^2(\tau_1, \tau_2)$. So, these terms can be individually inverted, without affecting their absolute magnitude. Next to selection of the desired proton transition by ϕ_5 and ϕ_6, the phase ϕ_2 of the ^{13}C 90° pulse preceding the S^3E element controls selection of the desired ^{13}C multiplet component, evolving during t_1. Analogous to previously proposed TROSY experiments for $^{15}N-^1H$ and aromatic $^{13}C-^1H$ spin systems,[4,24] the

(21) Geen, H.; Freeman, R. *J. Magn. Reson.* **1991**, *93*, 93−141.
(22) Kupce, E.; Freeman, R. *J. Magn. Reson., Ser. A* **1996**, *118*, 299−303.
(23) Fujiwara, T.; Nagayama, K. *J. Magn. Reson.* **1988**, *77*, 53−63.

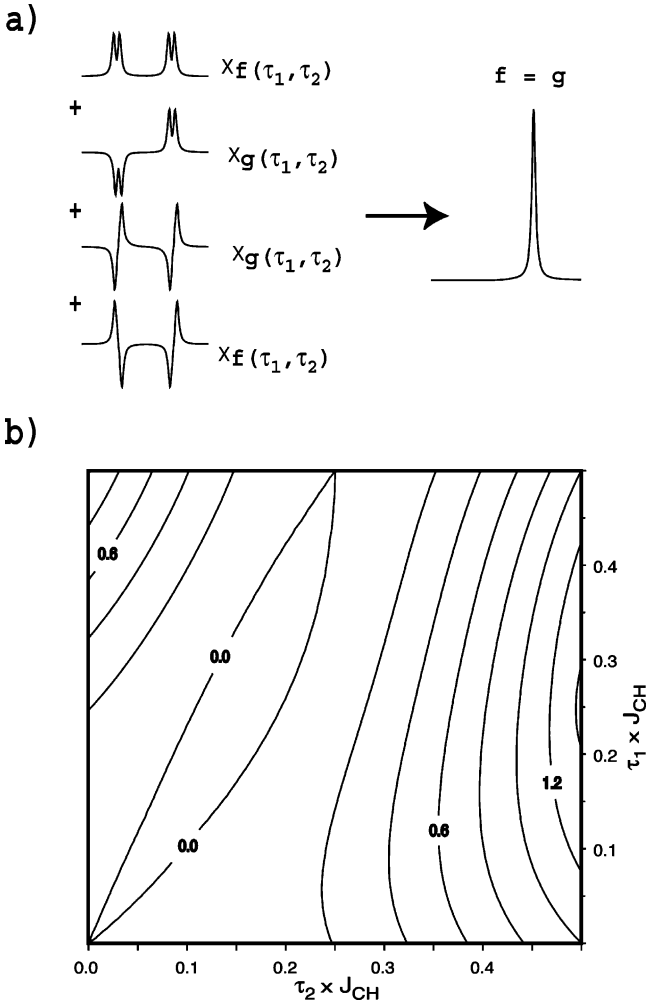

Figure 3. Selection of individual multiplet components in CH_2-TROSY. (a) Schematic representation of ^1H resonances corresponding to in-phase and antiphase terms (i) H_x, (ii) $2H_xC_z$, (iii) $-2H_xH_z$, and (iv) $-4H_xC_zH_z$, generated after the CH_2 spin-state selective coherence transfer. The magnitudes of the (i) and (iv) terms are proportionate to the function $f(\tau_1, \tau_2)$; (ii) and (iii) are proportionate to $g(\tau_1, \tau_2)$, with $f(\tau_1, \tau_2) = \cos(\pi J_{CH}\tau_1)\sin(\pi J_{CH}\tau_2) + \sin(\pi J_{CH}\tau_1)$ and $g(\tau_1, \tau_2) = \sin(\pi J_{CH}\tau_1)\cos(\pi J_{CH}\tau_1)\cos(\pi J_{CH}\tau_2) + \cos(\pi J_{CH}\tau_2)\sin(\pi J_{CH}\tau_2)(1 + \sin^2(\pi J_{CH}\tau_1))$. For $f(\tau_1, \tau_2) = g(\tau_1, \tau_2)$, only one ^1H multiplet component remains when coadding the spectra. (b) Contour plot of $f(\tau_1, \tau_2) - g(\tau_1, \tau_2)$.

Table 1. Phase Settings in Figure 2 for Selection of Each of the Eight Components of a CH_2 Group[a]

transition selected		ϕ_1[a]	ϕ_2[a]	ϕ_5[a]	ϕ_6[a]	component[b]
C^{++}	H^{+-}	y	135°/315°	y	x	d
	H^{++}	y	135°/315°	x	y	c
	H^{--}	y	135°/315°	y	y	b
	H^{-+}	y	135°/315°	x	x	a
C^{--}	$\boldsymbol{H^{+-}}$	$\boldsymbol{-y}$	**225°/45°**	$\boldsymbol{x}$	$\boldsymbol{x}$	**h**
	H^{++}	$-y$	225°/45°	y	y	g
	H^{--}	$-y$	225°/45°	x	y	f
	H^{-+}	$-y$	225°/45°	y	x	e

[a] Bold entries correspond to selection of the narrowest component, i.e., to the CH_2-TROSY experiment. [b] The character corresponds to the Figure 4 panel containing this multiplet component.

steady-state Boltzmann ^{13}C polarization can be added to the magnetization transferred from ^1H by appropriate selection of the phase ϕ_1 in Figure 2, thereby enhancing sensitivity. For each of the eight multiplet components, the optimized combinations of radio frequency phases are given in Table 1.

Incomplete suppression of multiplet components not selected by a given combination of phases in Table 1 can arise from deviations from the $\tau_1 = 0.34/^1J_{CH}$ and $\tau_2 = 0.23/^1J_{CH}$ condition, which cannot be met simultaneously for the experimental range of $^1J_{CH}$ values (referred to as $^1J_{CH}$ mismatching).[25,26] Other causes of incomplete suppression include differential relaxation,[27,28] the finite lifetime of the passive spin state,[29] and, in particular for the methylene case, evolution resulting from the relatively large $^2J_{HH}$ coupling. The selection efficiency of the S^3E element in the presence of $^1J_{CH}$ mismatching has already been well documented.[26,30] Simulation[31] of effects induced by the neglected $^2J_{HH}$ evolution or due to $^1J_{CH}$ mismatch during the CH_2-S^3CT element demonstrate that the intensity of multiplet components not selected by any given phase cycle combination of Table 1 will be ≤0.6% (for $^2J_{HH} = -18$ Hz) and ≤1% (for $\Delta^1J_{CH}/^1J_{CH} \leq 0.1$). Because the transfer delays τ_1 and τ_2 are relatively short, the effect of passive spin state flips or differential relaxation during the final CH_2-S^3CT element is relatively small and does not result in significant breakthrough of unselected components. Perhaps surprisingly, initial experiments using 500-μL sample volumes in regular NMR tubes revealed that B_1-field inhomogeneity can result in detectable spurious peaks (up to 4% of the selected component). However, in all tested cases (using 500 and 600 MHz cryogenic probeheads and 600 and 800 MHz room-temperature probes), this artifact is reduced below 1% by restricting the sample height, using a 280-μL volume in a thin-wall Shigemi microcell. One-dimensional cross sections, shown in Figure 4, confirm that no selection artifacts are observable at the noise threshold. However, accurate pulse calibrations remain critical to achieve clean selection: simulated and experimental results indicate that ^1H or ^{13}C pulse miscalibration by as little as 5% can result in selection artifacts of up to 7%.

Whereas, in standard ^1H-coupled HSQC experiments, correlations involving the C^{-+} and C^{+-} transitions are not observed, they can potentially lead to undesirable correlations in the CH_2-TROSY experiment. It can be shown that such artifacts arise from the ^{13}C Boltzmann polarization (vide infra, eq 8). In this case, the CH_2-S^3CT element converts the C^{-+} and C^{+-} coherences into the $H^{++} - H^{--}$ combination of single-transition operators, with an efficiency that depends on delays τ_1 and τ_2 (data not shown). To study this artifact more precisely, CH_2-TROSY spectra with enhanced intensity of this artifact were recorded for the GB3 sample by starting from in-phase ^{13}C magnetization. The undesired signal at the ^{13}C chemical shift frequency, δ_C, then consists of a doublet, antiphase in the ^1H dimension and separated by $^1J_{CH} + {}^2J_{HH}$, in agreement with theoretical analysis (data not shown). Calculations for $\tau_1 = 0.34/^1J_{CH}$ and $\tau_2 = 0.23/^1J_{CH}$ indicate that the intensity of this artifact

(24) Pervushin, K. V.; Wider, G.; Wuthrich, K. *J. Biomol. NMR* **1998**, *12*, 345−348.
(25) Meissner, A.; Schulte-Herbruggen, T.; Briand, J.; Sorensen, O. W. *Mol. Phys.* **1998**, *95*, 1137−1142.
(26) Sorensen, M. D.; Meissner, A.; Sorensen, O. W. *J. Biomol. NMR* **1997**, *10*, 181−186.
(27) Rance, M.; Loria, J. P.; Palmer, A. G. *J. Magn. Reson.* **1999**, *136*, 92−101.
(28) Kojima, C.; Kainosho, M. *J. Magn. Reson.* **2000**, *143*, 417−422.
(29) Meissner, A.; Schulte-Herbruggen, T.; Sorensen, O. W. *J. Am. Chem. Soc.* **1998**, *120*, 7989−7990.
(30) Meissner, A.; Schulte-Herbruggen, T.; Sorensen, O. W. *J. Am. Chem. Soc.* **1998**, *120*, 3803−3804.
(31) Nicholas, P.; Fushman, D.; Ruchinsky, V.; Cowburn, D. *J. Magn. Reson.* **2000**, *145*, 262−275.

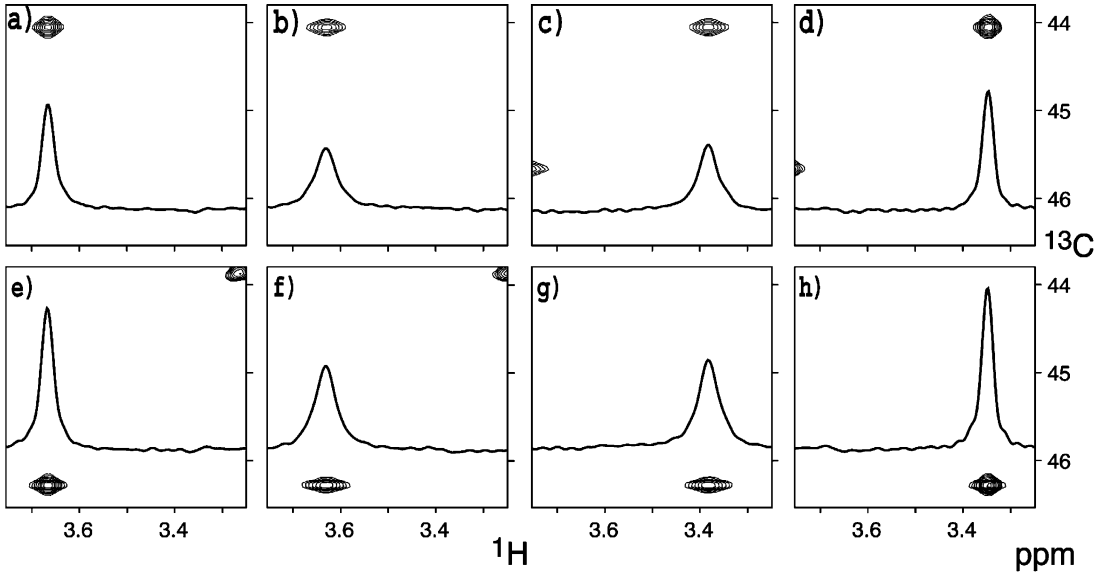

Figure 4. Experimental spectra for the eight multiplet components of Figure 1b, recorded with the pulse scheme of Figure 2a, for the Gly-41 $H^{\alpha 3}$ in GB3. The eight spectra were recorded at 10 °C, 500 MHz ^{1}H frequency, with the pulse phases listed in Table 1. Spectra a−d correspond to the C^{++} transition; e−h, to C^{--}; a and e, to H^{-+}; b and f, to H^{--}; c and g, to H^{++}, d and h, to H^{+-}. A cross section through each of the selected components is shown to highlight differences in ^{1}H line widths.

represents only ca. 2% of the selected correlation in the regular CH_2-TROSY spectrum. Use of longer delays, which still satisfy the equality $f(\tau_1, \tau_2) = g(\tau_1, \tau_2)$ (Figure 3b), can decrease the intensity of this artifact even further (Supporting Information, Figure S2). Total cancellation can be obtained for $\tau_1 = 0.50/{}^1J_{CH}$ and $\tau_2 = 0.25/{}^1J_{CH}$ but would occur at the expense of a 20% drop in sensitivity (Supporting Information, Figure S1) and an increase of other sources of artifacts. Therefore, as a compromise between sensitivity and artifact intensity, values of $\tau_1 = 0.4/{}^1J_{CH}$ and $\tau_2 = 0.24/{}^1J_{CH}$ are used. With these delays, no artifact larger than 1% is observed and the experiment still retains ≥95% of its maximum sensitivity.

High Resolution ^{1}H−^{13}C Spectra. As expected, intensities and line widths of the eight multiplet components of methylene ^{1}H−^{13}C correlation spectra vary considerably (Figure 4). The correlation between transitions C^{--} and H^{+-} exhibits the most favorable relaxation properties (Figure 4h). The longer relaxation time of H^{+-} compared to H^{-+} indicates that at the position of proton H^1, the sum of the dipolar fields corresponding to $^{13}C = |\alpha\rangle$ and $H^2 = |\beta\rangle$ decreases the "local field" of the H^1 CSA term. Similarly, the longer relaxation time of C^{--} compared to C^{++} indicates that, at the C position, the sum of the dipolar fields decreases the "local field" of the ^{13}C CSA term when both protons are in the $|\beta\rangle$ spin state. Therefore, optimal resolution is obtained when the CH_2-TROSY experiment selects this correlation.

As an example, Figure 5 compares the Gly region of such a CH_2-TROSY spectrum, recorded for the 20 kDa CaM/M13 complex, with a regular sensitivity-enhanced HSQC experiment, optimized for methylene correlations.[18,32] Comparison of ^{1}H line widths in these two spectra highlights the increased resolution.

We define a resolution enhancement factor λ_R as the ratio of the NMR signal line width measured in the HSQC and in the CH_2-TROSY spectra. The resolution enhancement factor in the ^{1}H dimension (λ_H) results from removal of the $^2J_{HH}$ splitting ($^J\lambda_H$) and selection of the slowest relaxing multiplet component ($^\Gamma\lambda_H$). To a first approximation, it is described by

$$\lambda_H = \frac{R_H + \pi \times |^2J_{HH}|}{R_H + \Gamma^{CSA,DD}_{H^1,H^1H^2} - \Gamma^{CSA,DD}_{H^1,H^1C} - \Gamma^{DD,DD}_{H^1C,H^1H^2}}$$

$$= {}^J\lambda_H \times {}^\Gamma\lambda_H = \left(1 + \frac{\pi \times |^2J_{HH}|}{R_H}\right) \times \left(1 + \frac{\Gamma^{CSA,DD}_{H^1,H^1H^2} - \Gamma^{CSA,DD}_{H^1,H^1C} - \Gamma^{DD,DD}_{H^1C,H^1H^2}}{R_H}\right)^{-1} \quad (7a)$$

where R_H is the proton autorelaxation rate: $R_H = \Gamma^{DD,DD}_{H^1C,H^1C} + \Gamma^{DD,DD}_{H^1H^2,H^1H^2} + \Gamma^{CSA,CSA}_{H^1,H^1}$. Each of the three pertinent cross-correlated rates, $\Gamma^{CSA,DD}_{H^1,H^1H^2}$, $\Gamma^{CSA,DD}_{H^1,H^1C}$, and $\Gamma^{DD,DD}_{H^1C,H^1H^2}$ can be extracted from the difference of ^{1}H line widths for suitably chosen pairs of resonances (eq 2). Analysis of spectra recorded for the CaM/M13 complex, at 35 °C and at a ^{1}H frequency of 800 MHz (data not shown), yields average values of $\Gamma^{DD,DD}_{H^1C,H^1H^2} = 27.4 \pm 3.2$ Hz, $\Gamma^{CSA,DD}_{H^1,H^1C} = 6.2 \pm 3.9$ Hz, and $\Gamma^{CSA,DD}_{H^1,H^1H^2} = -0.6 \pm 3.9$ Hz. Therefore, for moderate size proteins, the two most important factors for the line narrowing effect in the ^{1}H dimension are the dipole−dipole cross-correlated rate $\Gamma^{DD,DD}_{H^1C,H^1H^2}$ and the suppression of the $^2J_{HH}$ splitting (17−18 Hz for Gly). Whereas the absolute line narrowing caused by removal of the $^2J_{HH}$ splitting remains constant, the contribution from the cross-correlated relaxation increases with the correlation time and thereby the size of the protein. For the CaM/M13 complex, the experimentally determined resolution enhancement factor is $\lambda_H = 2.30$, with $^J\lambda_H = 1.54$ and $^\Gamma\lambda_H = 1.49$. For the tetrahedral geometry of the Gly methylene groups, substitution of our experimental value of $\Gamma^{DD,DD}_{H^1C,H^1H^2}$ into eq 5e yields $\tau_c \mathbf{S}^2 = 6.74$ ns. With a generalized order parameter $\mathbf{S}^2 = 0.8$ for Gly-C^α assumed, this corresponds to $\tau_c = 8.4$ ns, in good agreement with previous measurements.[33] No prior experimental data for the Gly-$^1H^\alpha$ chemical shielding tensor magnitude or orientation

(32) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sørensen, O. W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 301−306.

are available to confirm the measured cross-correlated relaxation rates, and instead we use the results from DFT calculations (see Methods). For $\tau_c S^2 = 6.74$ ns and DFT-derived ^{1}H shielding tensors, the predicted cross-correlated relaxation rates averaged over the two methylene protons ($\Gamma^{CSA,DD}_{H^1,CH^1} = +6.56$ Hz and $\Gamma^{CSA,DD}_{H^1,H^1H^2} = -0.17$ Hz) are in very good agreement with experimental results ($\Gamma^{CSA,DD}_{H^1,CH^1} = +6.2$ Hz and $\Gamma^{CSA,DD}_{H^1,H^1H^2} = -0.6$ Hz).

In the ^{13}C dimension, the resolution enhancement factor λ_C is defined as the ratio of the HSQC line width over that in CH_2-TROSY:

$$\lambda_C = \frac{R_C + \Gamma^{DD,DD}_{CH^1,CH^2}}{R_C + \Gamma^{DD,DD}_{CH^1,CH^2} + \Gamma^{CSA,DD}_{C,CH^1} + \Gamma^{CSA,DD}_{C,CH^2}} \qquad (7b)$$

where R_C is the carbon autorelaxation rate: $R_C = \Gamma^{DD,DD}_{CH^1,CH^1} + \Gamma^{DD,DD}_{CH^2,CH^2} + \Gamma^{CSA,CSA}_{C,C}$. Again assuming tetrahedral geometry and $\tau_c S^2 = 6.74$ ns, calculation of the dipole−dipole cross-correlation rate yields $\Gamma^{DD,DD}_{CH^1,CH^2} \approx -20.5$ Hz. Although this corresponds to a ca. 25% decrease in ^{13}C transverse relaxation rate compared to R_C, $\Gamma^{DD,DD}_{CH^1,CH^2}$ affects the two ^{13}C multiplet components, C^{--} and C^{++}, observed in HSQC and CH_2-TROSY identically. In contrast, the two CSA-dipole cross-correlated rates $\Gamma^{CSA,DD}_{C,CH^1}$ and $\Gamma^{CSA,DD}_{C,CH^2}$ broaden the upfield ^{13}C multiplet component and narrow the downfield, C^{--} transition, observed in the CH_2-TROSY experiment. For calmodulin Gly residues at 800 MHz, measurement of the decay rates of C^{--} and C^{++} yields $\Gamma^{CSA,DD}_{C,CH^1} + \Gamma^{CSA,DD}_{C,CH^2} = -17 \pm 5$ Hz. This corresponds to a ^{13}C resolution enhancement factor $\lambda_C = 1.4$.

Assuming $\theta^{CSA,DD}_{C,CH^1} = \theta^{CSA,DD}_{C,CH^2}$, the value of $\Delta\sigma_C \times P_2(\cos\theta^{CSA,DD}_{C,CH})$ is determined from eq 5d and the experimental $\Gamma^{CSA,DD}_{C,CH^1} + \Gamma^{CSA,DD}_{C,CH^2}$ rate. This yields $\Delta\sigma_C \times P_2(\cos\theta^{CSA,DD}_{C,CH}) = -25$ ppm, in good agreement with the experimental CSA values measured by solid-state NMR for Gly-C^α.[34]

Finally, the total resolution enhancement factor of the CH_2-TROSY over the HSQC experiment is $\lambda_R = \lambda_H \times \lambda_C = 3.2$. With the relative sign of each cross-correlated relaxation rate involved taken into account, the CH_2-TROSY experiment makes it possible to benefit from all six relaxation interference effects to reduce the decay rate of the selected multiplet component. Cross-correlated relaxation rates increase with the molecular correlation tumbling time τ_C. For the CSA values reported above, the resolution enhancement resulting from interference effects, $^\Gamma\lambda_H \times \lambda_C$, becomes larger than $^J\lambda_H$ for molecules with $S^2\tau_C$ above 5.3 ns. CSA-dipole cross-correlated relaxation increases linearly with magnetic field strength, whereas $^2J_{HH}$ coupling and, to a good approximation, dipole−dipole cross-correlated relaxation rates are field-independent. However, considering that the magnitudes of ^{1}H and ^{13}C CSA are moderate for aliphatic groups, suppression of the $^2J_{HH}$ splitting and dipole−dipole cross-correlation are the dominant factors behind the increased resolution. Therefore, the CH_2-TROSY experiment remains quite efficient at low field: a resolution enhancement factor of $\lambda_R = 2.8$ is observed for the CaM/M13 complex at 500 MHz, compared to the above-mentioned $\lambda_R = 3.2$ at 800 MHz (Table 2).

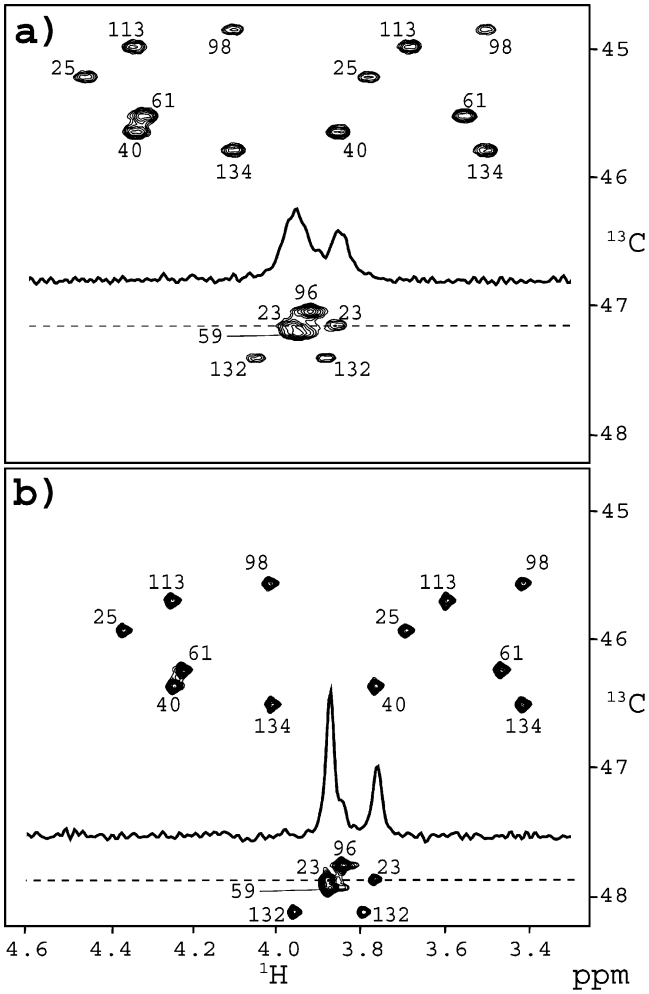

Figure 5. Comparison of (a) HSQC and (b) CH_2-TROSY spectra of Gly $C^\alpha H_2$ groups in ^{13}C-enriched calmodulin, complexed with unlabeled M13 peptide. Correlations are labeled by residue number. Both experiments were optimized for methylene detection, with the τ_1 delay of the Rance−Kay element in the HSQC experiment adjusted to $1/(8^1J_{CH})$.[32] Spectra were recorded at 800 MHz, 35 °C, with a relaxation delay of 2.5 s and total measuring times of 2.3 h each. Identical acquisition and processing parameters were used: time domain matrices of 192* × 640* data points, with acquisition times of 96 ms (t_1) and 80 ms (t_2), no apodization of the time domain data. Spectra are plotted at identical contour levels, and correlations are labeled by residue number. Inset cross sections, taken at the position of the dashed lines, correspond to Gly-23. The average resolution enhancement obtained with CH_2-TROSY is $\lambda_R = 3.2$, while sensitivity is increased by the factor $\lambda_S = 2$.

Table 2. Average Observed Gly $C^\alpha H_2$ Resolution and Sensitivity Enhancement Factors in a Calmodulin/Peptide Complex[a]

B_0 (T)[b]	λ_H[c]	λ_C[c]	λ_R[c]	λ_S[c]
11.75	2.3	1.2	2.8	1.6
18.8	2.3	1.4	3.2	2.0

[a] Data at 35 °C for a (^{13}C, ^{15}N)-labeled sample of the 20 kDa CaM/M13 complex in D_2O. [b] The recycle delay, optimized for maximum sensitivity, is 1.8 s at 11.75 T and 2.5 s at 18.8 T. [c] The resolution enhancement factors λ_H and λ_C are defined as the ratios of the ^{1}H (λ_H) and the ^{13}C line widths in the regular HSQC and CH_2-TROSY spectra. The total resolution enhancement factor, λ_R, equals the product of λ_H and λ_C. The sensitivity enhancement factor, λ_S, is the ratio of the intensity in the CH_2-TROSY and regular HSQC spectra, recorded with identical acquisition and processing parameters.

Sensitivity Enhancement. Compared to the regular HSQC experiment only one-half of the initial ^{1}H polarization is retained after the S^3E filter ($\lambda_{S^3E} = 0.5$). This loss is partially compensated

(33) Lee, A. L.; Sharp, K. A.; Kranz, J. K.; Song, X. J.; Wand, A. J. *Biochemistry* **2002**, *41*, 13814−13825.
(34) Yao, X. L.; Hong, M. *J. Am. Chem. Soc.* **2002**, *124*, 2730−2738.

by the addition of the ^{13}C Boltzmann polarization, which contributes to the detected signal in the CH_2-TROSY experiment, as discussed previously. At the end of the INEPT element (time point *a*, Figure 2a) the signal intensity is described by the following: $I(\delta) = M_H(\delta) \times [2C_xH_z^1 + 2C_xH_z^2] + M_C(\delta) \times [C_x]$, where $M_X(\delta)$ is the initial Boltzmann magnetization for nucleus X, using a recycling delay, δ. $I(\delta)$ can be decomposed in terms of single-transition operators C^{PQ} (eq 3):

$$I(\delta) = C^{--}\left(\frac{M_C(\delta)}{4} + \frac{M_H(\delta)}{2}\right) + C^{++}\left(\frac{M_C(\delta)}{4} - \frac{M_H(\delta)}{2}\right) + \frac{M_C(\delta)}{4}(C^{+-}+C^{-+}) \quad (8)$$

Only C^{--} is selected in the CH_2-TROSY experiment. Because M_X is directly proportional to the gyromagnetic ratio, γ_X, of X ($\gamma_H/\gamma_C \approx 4$; *i.e.*, $M_H/M_C \approx 4$), eq 8 indicates that a 12.5% signal increase is expected when the ^{13}C Boltzmann magnetization is added. A larger sensitivity gain is expected under nonequilibrium, steady-state conditions for the case where $T_1(^1H) > T_1(^{13}C)$.[4,24]

The dependence of signal contributions resulting from ^1H and ^{13}C steady-state polarization as a function of the recycle delay, δ, has been measured for the calmodulin Gly residues at a 18.8-T (800 MHz ^1H frequency) magnetic field strength (Supporting Information, Figure S3). Optimal signal-to-noise per unit of measuring time is obtained for $\delta \approx 2.5$ s. At this interscan delay, the ^{13}C Boltzmann magnetization corresponds to a signal increase of 33% over the case where only the magnetization transferred from ^1H is selected. In practice, shorter interscan delay times are frequently used to complete all required phase cycling steps in a finite amount of available measurement time. For a typical $\delta \approx 1$ s duration, the ^{13}C Boltzmann component increases the signal by 50%; i.e., the total amount of observed ^1H magnetization becomes 75% of that in a regular HSQC experiment.

The line narrowing effect resulting from cross-correlated relaxation in both ^1H and ^{13}C dimensions also contributes to increased intensity of the detected signal in the CH_2-TROSY experiment. However, if the acquisition time in the ^{13}C dimension were adjusted to a fixed multiple of the effective ^{13}C magnetization decay constant, the slower decay of ^{13}C magnetization has a negligible effect on the sensitivity of the experiment per unit of measuring time. In contrast, in the directly detected ^1H dimension, where longer sampling does not increase the total time of the experiment, sensitivity per unit of measuring time increases with the square root of the obtained resolution enhancement factor, again assuming that the acquisition time equals a fixed multiple of the decay constant in the two experiments. In practice, sensitivity comparisons of regular and TROSY type experiments are commonly carried out under conditions where identical sampling times are used for the regular and the TROSY-type experiments,[2,35] a comparison which increases the apparent sensitivity in the TROSY-type experiment, but which obviates calculation of the precise, residue-dependent timing parameters needed for optimizing each experiment. For simplicity, we therefore adhere to this common practice and compare data sets that have identical sampling durations in both ^1H and ^{13}C dimensions and process the spectra without any apodization of the time domain data.

Under these conditions, the expected sensitivity enhancement factor coming from favorable relaxation properties of the CH_2-TROSY experiment equals $2 \times \lambda_C \times {}^\Gamma\lambda_H$, where λ_C and ${}^\Gamma\lambda_H$ are defined in eq 7, and where the factor 2 is due to the suppression of the $^2J_{HH}$ splitting, assuming $^2J_{HH}$ is resolved in the HSQC spectrum. As discussed above, experimentally we find $\lambda_C = 1.4$ and ${}^\Gamma\lambda_H = 1.49$ for the Gly-$C^\alpha H_2$ spin system at $B_0 = 18.8$ T. However, an additional factor, λ_T, needs to be defined to account for the reduced transfer efficiency of the CH_2-S^3CT transfer relative to a methylene-optimized, sensitivity-enhanced Rance−Kay element. This factor depends on the delays τ_1 and τ_2 (Supporting Information Figure S1) and has a maximum for $\tau_1 = 0.34/^1J_{CH}$ and $\tau_2 = 0.23/^1J_{CH}$ ($\lambda_T = 0.85$). With all of the described effects taken into account (^{13}C spin state selection, use of ^{13}C Boltzmann polarization, cross-correlated relaxation, proton−proton decoupling enhancements, and transfer efficiency), the total sensitivity enhancement is then given by

$$\lambda_S = \lambda_{S^3E} \times \left(\frac{M_C(\delta) + M_H(\delta)}{M_H(\delta)}\right) \times 2 \times \lambda_C \times {}^\Gamma\lambda_H \times \lambda_T \quad (9)$$

yielding a value of $\lambda_S = 2.2$, in good agreement with a value of 2.0, observed experimentally at 800 MHz (Table 2). The 10% discrepancy between predicted and experimentally determined sensitivity enhancements is likely to result from two factors neglected in the above calculation: the effect of relaxation during the very short S^3E filtering delay and partial overlap of the two ^1H doublet components in the reference HSQC spectrum. This partial overlap increases the apparent HSQC intensity and, therefore, reduces the observed gain of the CH_2-TROSY experiment. When it is considered that much of the gain in resolution and sensitivity results from dipole−dipole cross correlation effects, which are independent of magnetic field strength, the CH_2-TROSY experiment remains advantageous across the entire range of commonly used magnetic field strengths. Indeed, a sensitivity enhancement factor of $\lambda_S = 1.6$ was observed for the CaM/M13 sample at 500 MHz, only 20% less compared to the gain observed at 800 MHz.

Effect of Nongeminal $^1H-^1H$ Interactions. Above, discussions focused on an isolated three-spin system, exemplified through the study of Gly methylene groups. However, the experiment is directly applicable to all methylene sites in macromolecules. For such nonisolated methylenes, remote $^1H-^1H$ interactions need to be taken into account when evaluating the merits of the CH_2-TROSY experiment. Both passive scalar interaction, $^3J_{HH}$, and dipolar $^1H-^1H$ relaxation with the additional protons will broaden the ^1H signal, whereas the effect of the additional protons on the ^{13}C decay will be negligible compared to that of its directly bonded protons. For the methylene protons, the difference in ^1H line width between the CH_2-TROSY and regular HSQC experiments is not affected by the additional presence of neighboring protons, but the relative resolution and sensitivity gain decrease as the remote protons broaden the ^1H line width in both types of experiments by the same amount.

The effect of protons in the vicinity of a methylene group can be described by adding the following terms to both the

(35) Tugarinov, V.; Kay, L. E. *J. Mol. Biol.* **2003**, *327*, 1121−1133.

numerator and denominator of λ_H (eq 7a): $\Sigma_{j\neq1,2}\,\pi \times {}^3J_{H^iH^j} + \Sigma_{j\neq1,2}\,\Gamma^{DD,DD}_{H^iHj,H^iHj}$, where H^i ($i = 1,2$) denotes one of the methylene protons. For macromolecules, such line broadening often will cause the geminal $^2J_{HH}$ coupling to be unresolved in the reference HSQC spectrum. Removal of the $^2J_{HH}$ coupling contribution to the ^{1}H line shape will narrow the resonance in the CH_2-TROSY spectrum but will appear less pronounced than for the case of isolated methylenes. Also, the additional ^{1}H−^{1}H dipolar interactions increase the effective ^{1}H longitudinal relaxation rate of methylene protons, without significantly affecting the relaxation rate of ^{13}C, thereby reducing the fractional contribution of the Boltzmann ^{13}C polarization.

To evaluate the enhancement offered by the CH_2-TROSY experiment for nonisolated methylene groups experimentally, we recorded CH_2-TROSY and regular ^{1}H−^{13}C HSQC spectra for GB3 at 800 MHz and 10 °C. The GB3 protein (56 residues) is much smaller than calmodulin and yields resolved correlations for most of its methylene resonances, in both the CH_2-TROSY and HSQC spectra (Supporting Information, Figure S4). The spectra can be recorded either in the constant-time (CT) mode (using a CT ^{13}C evolution period of $1/{}^1J_{CC} \approx 28$ ms)[36] or in the regular manner. If the CH_2-TROSY spectrum is recorded in the CT mode, ${}^nJ_{CH}$ couplings ($n \geq 2$) are not decoupled and cause some attenuation at the end of the CT evolution period: for example, ${}^nJ_{CH} = 7$ Hz results in a 20% loss of magnetization. On the other hand, as is the case for 3D ^{15}N TROSY-based experiments,[37] the favorable ^{13}C relaxation properties apply for the full duration of the CT period, T, whereas in the regular CT-HSQC favorable relaxation applies to the two multiplet components for durations of $T - t_1/2$ and $t_1/2$. This changes the intensity of the CH_2-TROSY signal by a factor of $2(\Gamma^{CSA,DD}_{C,CH^1} + \Gamma^{CSA,DD}_{C,CH^2}) \times T/(1 - \exp -2(\Gamma^{CSA,DD}_{C,CH^1} + \Gamma^{CSA,DD}_{C,CH^2})T)$ relative to that in the HSQC experiment. In the CT mode, this factor should substitute for λ_C in eq 9 when calculating the sensitivity enhancement factor, λ_S. This λ_S factor increases with the size of the molecule, and its magnitude becomes larger than the enhancement obtained for regular, non-CT acquisition mode (eq 7b) when $\mathbf{S}^2\tau_C \geq 5.7$ ns, assuming a regular CT duration of $T = 28$ ms. For side chain methylene carbons, we find that the intensity ratio observed in CH_2-TROSY versus HSQC spectra is also strongly influenced by the amplitude of internal dynamics of the methylene group, with more internal motion resulting in a smaller value of λ_S.[38] In constant-time experiments, improved spectral appearance therefore mainly results from the line narrowing obtained in the ^{1}H dimension. This latter effect depends strongly on the number of protons geminal to the methylene carbon in question. C^β methylenes of Cys, Ser, Asp, Asn, Phe, Tyr, Trp, and His each have only a single proton (H^α) that is J-coupled to H^β. For these residues, the gain in ^{1}H resolution is, on average, $\lambda_H = 1.7$ (Table 3). Smaller gains are obtained for C^β methylenes of Met, Glu, Gln, Arg, and Lys residues (Table 3), where three vicinal protons have $^3J_{HH}$ and dipolar interactions with H^β.

When comparing C_2' to C_5' methylene groups in DNA, we also find smaller enhancements in resolution and sensitivity: $\lambda_H = 1.6$ vs 1.9 and $\lambda_S = 1.0$ vs 1.3 (Table 3). The disappearance of a sensitivity gain is again consistent with the topology of the ^{1}H−^{1}H network: H_5'/H_5'' are only coupled to H_4', whereas H_2'/H_2'' protons are J-coupled to both H_1' and H_3'. Nevertheless, as will be demonstrated below, even this moderate (60−90%) gain in resolution can considerably enhance spectral appearance.

Table 3. Resolution and Sensitivity Enhancement Factors for Methylene Sites in Proteins and DNA[a]

CH_2 type	G^b	C, S, D, N, F, Y, W, H[b]	M, Q, E, K, R[b]	C_2'[c]	C_5'[c]
λ_H	2.3	1.7	1.4	1.6	1.9
λ_S	1.7	1.0	0.8	1.0	1.3

[a] Data recorded at 800 MHz, using the pulse sequence of Figure 2a for samples of (^{13}C, ^{15}N)-enriched GB3 (T = 10 °C), and using the pulse sequence of Figure 2b for a 19-basepair DNA oligomer (T = 35 °C), both in D_2O. The resolution enhancement factor, λ_H, corresponds to the average ^{1}H line width in the regular HSQC experiment divided by that in the CH_2-TROSY experiment. The sensitivity enhancement factor, λ_S, is the ratio of the intensity in the CH_2-TROSY spectrum divided by that in the regular HSQC spectrum, recorded with identical acquisition and processing parameters. [b] One-letter amino acid residue type. Numbers listed apply for $C^\beta H_2$. On average, the optimal CH_2-TROSY recycle delay was found to be ca. 1 s for Gly-C^α and ca. 0.5 s for side chain CH_2. [c] Methylene groups in the DNA oligomer. The optimal CH_2-TROSY recycle delay for CH_2 groups in DNA was 0.8 s.

The smaller enhancements observed in the CH_2-TROSY experiment when applied to groups embedded in an environment rich in protons is similar to that observed in other TROSY methods, such as those employed for enhancing signals of ^{15}N−^{1}H pairs[1] or ^{13}C−$^1H_3$ groups.[6] In these latter experiments, signal enhancement is considerably enhanced when neighboring protons are replaced by deuterons. For the CH_2-TROSY experiment described here, in the absence of deuteration, a dense proton environment may lead to a small loss in sensitivity compared to a standard, methylene-optimized HSQC experiment. However, the gain in ^{1}H resolution remains quite pronounced, making the experiment useful for unravelling the typically crowded methylene region of protein and nucleic acid spectra.

Application to Sequential Assignment of Nucleic Acids. The CH_2-TROSY enhancement can readily be incorporated in the wide range of commonly used 3D experiments. Here, its benefit is illustrated for the 3D ^{13}C-separated NOESY experiment, applied to DNA.

In nucleic acids, NOE interactions between adjacent nucleotides constitute the mainstay of sequential assignment. For ^{13}C-enriched RNA, the conventional TROSY approach has been shown to improve substantially the ^{13}C resolution of base carbons, benefiting the 3D ^{13}C-separated NOESY experiment for observation of sequential base $H_{6/8}$ to H_1' NOEs.[4] However, the shortest distances between sequential nucleotides in a standard duplex helix involve protons in the C_2' and C_5' positions.[39] These NOEs are often difficult to analyze because of spectral overlap, especially in the C_5' region. Below, we demonstrate that incorporation of the CH_2-TROSY element considerably improves spectral resolution, thereby permitting the study of sequential NOEs involving H_2'/H_2'' and H_5'/H_5'' protons.

Table 3 reports the resolution and sensitivity enhancement factors for the C_2' and C_5' methylene groups in a 19-base-pair DNA fragment,[12] when comparing the 2D version of the CH_2-TROSY experiment of Figure 2b with that of the analogous methylene-optimized HSQC experiment. Comparison of the C_5'

(36) Vuister, G. W.; Bax, A. *J. Magn. Reson.* **1992**, *98*, 428−435.
(37) Salzmann, M.; Pervushin, K.; Wider, G.; Senn, H.; Wuthrich, K. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13585−13590.
(38) Yang, D. W.; Mittermaier, A.; Mok, Y. K.; Kay, L. E. *J. Mol. Biol.* **1998**, *276*, 939−954.
(39) Wijmenga, S. S.; van Buuren, B. N. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287−387.

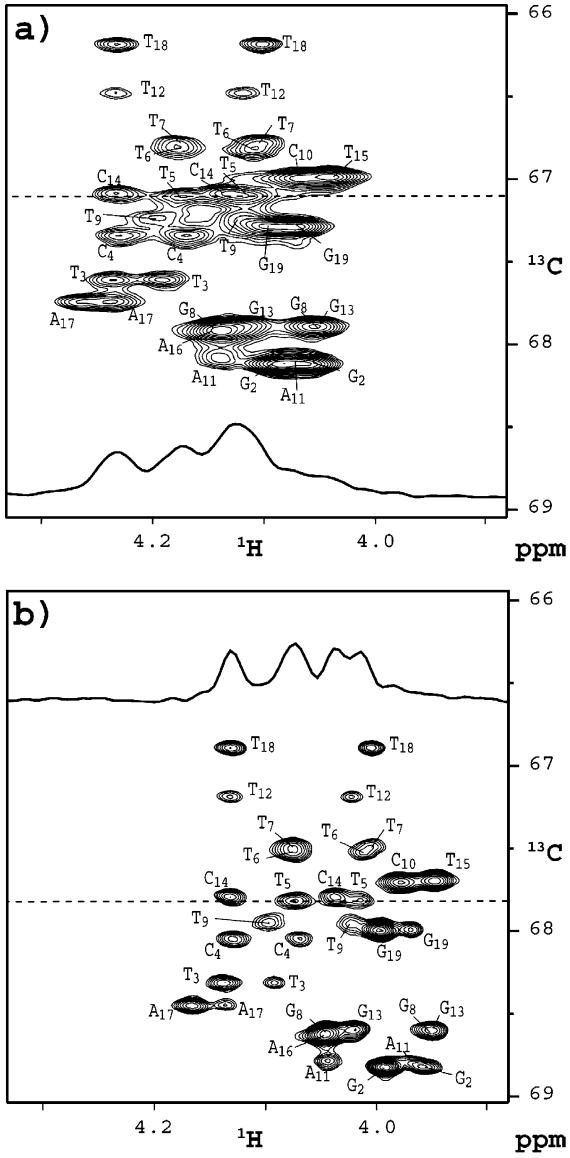

Figure 6. Comparison of (a) sensitivity-enhanced HSQC and (b) CH_2-TROSY spectra of the C_5' region of a 19-base-pair DNA oligomer for which one strand was uniformly enriched in ^{13}C. Spectra were recorded at 800 MHz, 35 °C, with measuring times of 1.2 h per spectrum. Spectra result from identical time domain matrices, consisting of 200* × 512* data points, recorded with acquisition times of 96 ms (t_1) and 64 ms (t_2). Data were processed identically, without apodization of the time domain data. Spectra are plotted at the same contour levels. Cross sections show the C_5' correlations of T5 and C14.

regions of the 2D CH_2-TROSY and HSQC spectra (Figure 6) confirms a considerable enhancement in spectral resolution, resulting in a nearly 2-fold increase in the number of resolved resonances.

In B-form DNA, the two shortest distances between sequential nucleotides i and $i + 1$ involve pairs $H_2''(i)/H_{6/8}(i + 1)$ and $H_1'(i)/H_5''(i + 1)$. We demonstrate that the latter interaction can be readily observed when the CH_2-TROSY element is incorporated in the 3D ^{13}C-separated NOE experiment. In our implementation (Figure 2b), the TROSY element is placed before the NOE mixing period, such that the enhancement resulting from the ^{13}C Boltzmann magnetization is retained and the relatively well dispersed H_1' and $H_{8/6}$ resonances are observed during the detection period, t_3. When it is considered that $^{13}C_2'$ and $^{13}C_5'$ chemical shifts are well separated from $^{13}C_1'$, $^{13}C_3'$, and $^{13}C_4'$, band-selective adiabatic homonuclear decoupling was used to eliminate the large one-bond $^1J_{CC}$ couplings ($^1J_{CC} \approx 40$ Hz).[40]

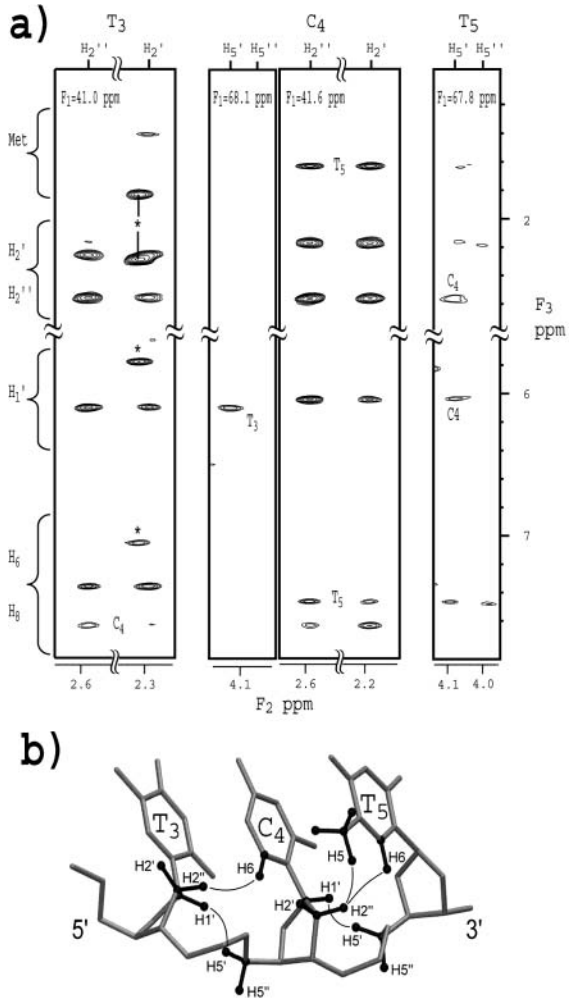

Figure 7. Sequential connectivity involving methylene protons in a DNA oligomer. (a) Strips taken for nucleotides T3, C4, and T5 from the 3D ^{13}C-separated CH_2-TROSY-NOESY spectrum, recorded on the 19-base-pair DNA duplex with the pulse scheme of Figure 2b. The spectrum was recorded at 800 MHz, 35 °C, 250 ms NOE mixing time, and a total measuring time of 48 h. The time domain matrix consisted of 122* × 55* × 512* data points, with acquisition times of 55 ms (t_1), 50 ms (t_2), and 64 ms (t_3). Only sequential NOEs are labeled in the figure. Due to partial overlap, the T3-C_2' strip also contains correlations for T18, marked by asterisks. (b) Summary of the observed sequential connectivities for T3-C4-T5, modeled in B-form helical geometry.

Figure 7 shows small regions of strips taken through the 3D CH_2-TROSY-NOESY spectrum of the 19-base-pair DNA oligomer, highlighting some of the NOE connectivities observable for C_2' and C_5' methylene protons. Because of the absence of homonuclear $^{13}C-^{13}C$ and geminal $^1H-^1H$ J splittings, as well as the slower transverse relaxation resulting from the CH_2-TROSY element, high spectral resolution is obtained in both the indirect ^{13}C (F_1) and ^{1}H (F_2) dimensions. This, for example, yields resolved correlations for the C_2' protons of nucleotide T3, which overlap with T18 in the regular HSQC-NOESY spectrum (data not shown). As expected, sequential NOEs among $H_2''(T3)/H_6(C4)$, $H_2''(C4)/H_6(T5)$, $H_5''(C4)/H_1'(T3)$, and $H_5''(T5)/H_1'(C4)$ are clearly visible and readily permit the sequential assignment to be made. Relatively intense intranucleotide NOE cross-peaks are also visible between H_2' and the

base H_6 proton, with a weaker correlation between H_2'' and H_6 resulting from spin diffusion via H_2' during the relatively long NOE mixing time (250 ms). For this mixing time, intranucleotide correlations between H_5''/H_5' and H_1' are absent and only the H_5''/H_1' sequential connectivity is observed. Therefore, the 3D ^{13}C-separated CH_2-TROSY-NOESY experiment offers two separate pathways for sequential assignment, which complement each other in cases of ambiguities resulting from spectral overlap.

Concluding Remarks

Detailed analysis of NMR signals involving the abundant methylene protons in macromolecules is frequently hampered by severe spectral crowding and rapid transverse relaxation. This problem provided the impetus behind a powerful but labor-intensive new isotope labeling procedure, known as SAIL,[41] where one of each pair of methylene hydrogens is substituted stereospecifically by deuterium. Instead, the CH_2-TROSY approach is aimed at the more general case where uniform ^{13}C enrichment, but no selective deuteration, is used. The CH_2-TROSY experiment yields considerable enhancement in resolution in both the ^1H and ^{13}C dimensions and can offer a modest increase in sensitivity too. As most of the gain in resolution results from dipole−dipole cross-correlation and from removal of the $^2J_{HH}$ splitting, which are both independent of magnetic field strength, the approach is beneficial over the entire range of commonly used magnetic field strengths.

The CH_2-TROSY element can be incorporated in a wide variety of NMR experiments, thereby potentially facilitating both resonance assignment and the extraction of structural and dynamic parameters involving methylene groups. The spectral enhancement resulting from the CH_2-TROSY element is largest when applied to isolated methylene groups, such as those found in Gly residues. These are often located in turn regions in proteins, whose precise geometry is frequently difficult to identify from the low number of restraints that typically can be extracted for these residues from conventional spectra. When the molecule is weakly aligned relative to the magnetic field, the ability to separately observe any of the methylene $^{13}C-^1H$ multiplet components also permits measurement of all three dipolar couplings for such a group. Besides the important structural information this offers, we currently are also investigating the application of the measurement of three independent dipolar couplings in the plane of the methylene moiety to characterize its dynamic properties.

Acknowledgment. We thank F. Delaglio for writing the special processing macro required for the Fourier transform of the 3D TROSY-NOESY and H. Kovacs for help in implementing homonuclear ^{13}C adiabatic decoupling. J.B. is supported by an HFSP fellowship; E.M., by an ARC fellowship; D.C.W., by a PRAT fellowship; and D.L.B., by an NSERC fellowship. This work was supported in part by the Intramural AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health.

Supporting Information Available: Four figures showing the following: CH_2-TROSY transfer efficiency as a function of transfer delays τ_1 and τ_2; the relative intensity of spurious peak corresponding to transition C^{-+} and C^{+-}; experimental signal-to-noise as a function of interscan delay in CH_2-TROSY; comparison of side chain methylene regions of GB3, recorded with CH_2-TROSY and regular HSQC. This material is available free of charge via the Internet at http://pubs.acs.org.

JA047904V

(40) Brutscher, B.; Boisbouvier, J.; Kupce, E.; Tisne, C.; Dardel, F.; Marion, D.; Simorre, J. P. *J. Biomol. NMR* **2001**, *19*, 141−151.
(41) Torizawa, T.; Terauchi, T.; Kainosho, M. *Seikagaku* **2002**, *74*, 1279−1284.

Direct observation of dipolar couplings between distant protons in weakly aligned nucleic acids

Jérôme Boisbouvier, Frank Delaglio, and Ad Bax†

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892-0520

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected on April 30, 2002.

Contributed by Ad Bax, July 23, 2003

Under conditions where macromolecules are aligned very weakly with respect to an external magnetic field, Brownian diffusion no longer averages internuclear dipole–dipole interactions to zero. The resulting residual dipolar coupling, although typically 3 orders of magnitude weaker than in a fully aligned sample, can readily be measured by solution NMR methods. To date, application of this idea has focused primarily on pairs of nuclei separated by one or two covalent bonds, where the internuclear separation is known and the measured dipolar coupling provides direct information on the orientation of the internuclear vector. A method is described that allows observation of dipolar interactions over much larger distances. By decoupling nearest-neighbor interactions, it is readily possible to observe direct dipolar interactions between protons separated by up to 12 Å. The approach is demonstrated for the DNA dodecamer d(CGCGAATTCGCG)$_2$, where direct interactions are observed between protons up to three base pairs apart.

In general, molecules diffusing in an environment with asymmetric boundary conditions sample orientational space in a nonuniform manner. When placed in a magnetic field, magnetic dipole–dipole interactions in such molecules no longer average to zero. As a result, NMR spectra of small organic solute molecules diffusing in a magnetically oriented liquid crystalline medium contain a wealth of such coupling information, which can reveal very precise information on the average geometry of the solute (1–3). The high degree of order of the solute molecule in such systems results in very large dipole–dipole couplings for proximate pairs of protons, up to 50% of what would be observed for a perfectly oriented molecule, and numerous smaller couplings to more remote protons. The resulting spectral complexity is such that complete analysis is not feasible for systems containing more than about a dozen magnetically active nuclei.

The spectral complexity resulting from the multitude of couplings that are larger than the resonance width will be reduced when the degree of alignment is scaled down. Very weak alignment (A_a), on the order of $A_a = 10^{-4}$, can result from a molecule's intrinsic magnetic susceptibility anisotropy when placed in a sufficiently strong magnetic field. Use of this method of alignment frequently exploits the presence of a paramagnetic ion, either endogenous (4–6) or tagged to the protein by a synthetic chelating group (7, 8). Despite the weakness of such alignment, and the proportionately small-scale factor for the dipolar interaction, the effects of the largest dipolar interactions remain detectable. This was first demonstrated for the one-bond backbone ^{15}N–^{1}H interactions in paramagnetic myoglobin (4), where dipolar couplings with magnitudes up to several hertz were observed, smaller than the natural resonance width but nevertheless easily detectable from a change in the one-bond $^{1}J_{NH}$ splitting.

Measurement of these couplings, also in diamagnetic systems, became much easier when it was shown to be feasible to induce a tunable, weak degree of macromolecular alignment by using very dilute, aqueous, lyotropic liquid crystalline solutions (9). The first such medium used for this purpose contained so-called bicelles (10), which (over a limited temperature and concentration range) are self-assembling, highly porous α-lamellar bilayers (11, 12) consisting of dimyristoyl phosphatidyl choline and detergent. When placed in a magnetic field, the bilayers of this liquid crystalline medium align with their bilayer normal orthogonal to the field. Originally, this medium was developed to study, mainly by solid-state NMR technology, the structure and membrane binding properties of small lipophilic molecules, anchored to the highly ordered bilayers (13, 14). The application of the bicelle medium to water-soluble proteins is rather different and takes advantage of the very small deviation from a purely random distribution of orientations that is sterically imposed on asymmetrically shaped proteins when immersed in such an ordered environment. A host of other liquid crystalline systems, including nematic suspensions of filamentous bacteriophages (15, 16) or crystalline cellulose fibers (17), and polyethylene glycol (18) or cetylpyridinium-based lamellar phases (19), are available to induce the required weak degree of macromolecular alignment. More recently, the use of anisotropically compressed aqueous acrylamide gels has been added to this arsenal (20, 21). Such media make it easily possible to increase the degree of ordering to $A_a \approx 10^{-3}$, where spectral complexity has not yet increased much compared with the regular solution NMR spectrum, but smaller interactions such as the one-bond ^{13}C–^{13}C and ^{13}C–^{15}N, and two-bond ^{13}C–^{1}H and ^{15}N–^{1}H also become accessible in proteins (9, 22) and nucleic acids (23, 24).

So far, most measurements of residual dipolar couplings have focused on heteronuclear one-bond or two-bond interactions, where the internuclear distance is known accurately from covalent geometry. The size of the observed coupling then depends uniquely on the orientation of the internuclear vector with respect to the molecule's alignment frame. For structure calculation purposes, however, such simplicity offers little advantage, and homonuclear dipolar coupling between protons far apart in the covalent network can readily be measured (9) and incorporated in the structure calculation protocol (25). We recently demonstrated the measurement of nearly 200 ^{1}H-^{1}H dipolar couplings for the DNA dodecamer d(CGCGAATTCGCG)$_2$ (26), a palyndromic oligomer that has been studied extensively by x-ray crystallography (27–29), NMR (30–32), and computational methods (33, 34). The structure of the center six base pairs of the oligomer in the crystal structure agrees remarkably well with the solution NMR structure (26), whereas the conformation of the flanking base pairs is affected by crystal packing and Mg^{2+} coordination, absent in solution. NMR structures calculated from ^{1}H–^{1}H dipolar couplings agreed well with those derived from ^{1}H–^{13}C dipolar couplings, but no interactions between protons separated by >4 Å were observable (34).

The problem with observing ^{1}H–^{1}H dipolar couplings for the longest distances is not so much that the absolute value of the coupling becomes too small, but primarily that there are too many much larger ^{1}H–^{1}H interactions, which obscure the small magnetic dephasing effects from the more remote interactions. In proteins, we have demonstrated that perdeuteration can

†To whom correspondence should be addressed. E-mail: bax@nih.gov.

overcome this problem, essentially chemically removing the majority of ^{1}H–^{1}H interactions and permitting the study of the remaining amide protons, for distances of up to 7 Å (35). We now demonstrate that interactions between remote protons, separated by distances of up to 12 Å, become accessible in fully protonated biomolecules by appropriate manipulation of the nuclear spin Hamiltonian, which allows for effective decoupling of all protons outside a spectral region of interest.

Experimental Section

All NMR experiments were carried out at 35°C on a Bruker DRX spectrometer, operating at 600 MHz ^{1}H frequency equipped with a triple resonance, three-axes pulsed-field gradient probe head, optimized for ^{1}H detection. Samples contained 3 mM of unlabeled DNA duplex d(CGCGAATTCGCG)$_2$, 50 mM KCl, 1 mM EDTA, and 10 mM potassium phosphate buffer (pH 7.0) in 99.9% D_2O, in a 300-μl Shigemi microcell. For the aligned sample 324 μg of cetyltrimethylammonium bromide and 360 μl of a 10% (wt/vol) nonhydrolyzable ether bicelle stock solution (composed of 1,2-*O*-ditetradecyl-*sn*-glycero-3-phosphocholine and 1,2-*O*-dihexyl-*sn*-glycero-3-phosphocholine in a molar ratio 3:1) were mixed with 300 μl of the DNA solution. The sample was then lyophilized and redissolved in 300 μl of D_2O. The final solution of DNA, containing 12% (wt/vol) bicelles, was characterized by a deuterium quadrupole splitting of 35.7 Hz. All data sets were processed and analyzed with the NMRPIPE software system (36). The pulse sequence to detect long-range ^{1}H–^{1}H dipolar couplings, and the NMRPIPE macro used to process the data are available on our laboratory web site: http://spin.niddk.nih.gov/bax.

Results

Band-Selective Removal of Coupling Interactions. In a weakly aligned molecule, the truncated spin Hamiltonian in the absence of external radiofrequency fields is given by

$$H = \sum_{\mathrm{j}} \gamma_H B_0 (1 - \sigma_j) I_Z^j + \sum_{j \neq k} 2\pi J_{jk} I_Z^j I_Z^k + \sum_{j \neq k} 2\pi D_{jk} I_Z^j I_Z^k, \qquad [1a]$$

where the first and second term are the Zeeman (H_Z) and the scalar coupling part (H_J) of the Hamiltonian, and the last one is the residual dipolar coupling term (H_{DD}). The residual dipolar coupling between a pair of protons, j and k, is given by

$$D_{jk} = -\frac{\mu_0 \hbar \gamma_{\mathrm{H}}^2}{8\pi^2 r_{jk}^3} A_a \left[(3\cos^2\theta - 1) + \frac{3}{2} R \sin^2\theta \cos 2\phi \right], \qquad [1b]$$

where θ and ϕ describe the orientation of the j-k vector relative to the molecular alignment tensor, r_{jk} is the interproton distance, A_a is the magnitude of the alignment, and R is its rhombicity (9).

In the absence of alignment, $H_{DD} = 0$, and spectral dispersion in the ^{1}H NMR spectrum is dominated by differences in chemical shift, σ, of the individual protons, with a multiplet structure superimposed on each resonance, caused by H_J, the through-bond coupling to the magnetically active nuclei three or fewer bonds removed from it. In the absence of isotopic enrichment, only interactions to geminal and vicinal hydrogens need to be considered, and multiplet structures are simple. For example, the anomeric H1′ proton in a deoxyribose ring appears as a doublet of doublets, resulting from its interaction with the H2′ and H2″ protons. Frequently this interaction is comparable to the natural line width of the proton resonance, resulting in unresolved splittings and an apparent broadening of the resonance (Fig. 1*B*). In the presence of weak alignment, a multitude of ^{1}H–^{1}H dipolar interactions are also present, typically in the range from 0 to ±30 Hz. Because of the large number of such interactions and the spread in their magnitudes, these couplings usually do not result in resolved multiplets and manifest themselves as severe apparent broadening of the ^{1}H resonance, and a concomitant loss in resolution (Fig. 1*A*).

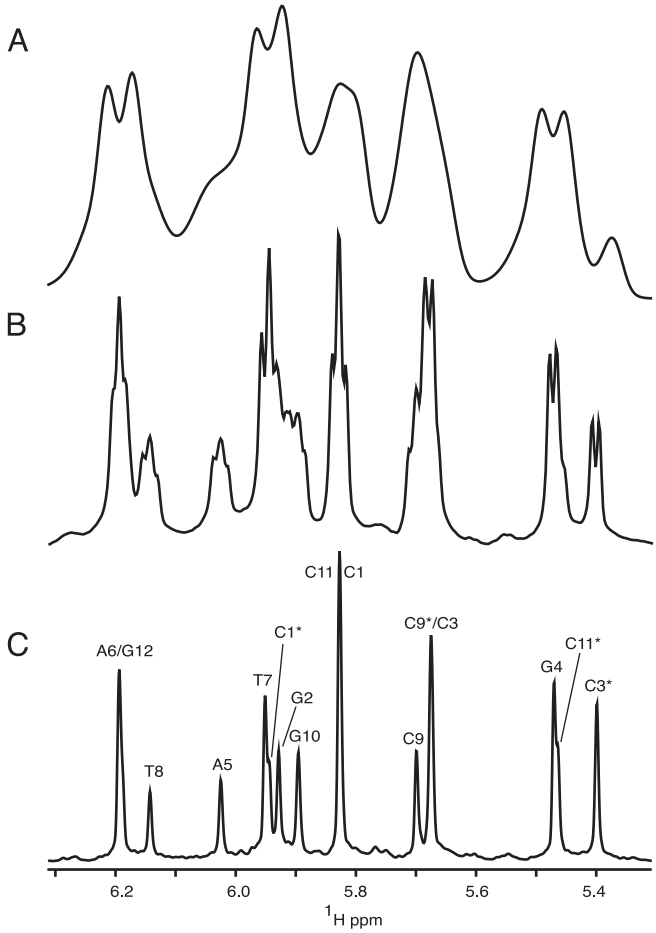

Fig. 1. Comparison of the H1′/H5 region of 1D spectra of d(CGCGAATTCGCG)$_2$, recorded without (*A* and *B*) or with (*C*) homonuclear decoupling. Spectrum *A* corresponds to the partially aligned state [12% (wt/vol) ether bicelle], and spectra *B* and *C* have been recorded on the corresponding isotropic sample of DNA. The duration of the sampled free induction decay equals 307 ms for spectra *A* and *B*. Apodization using a cosine-squared function, followed by zero-filling, was used before Fourier transformation. Spectrum *C* has been acquired and processed as described in the legend to Fig. 3. In *C*, $H_{1'}$ and H_5 protons are labeled by their nucleotide numbers, including * for H_5 resonances.

In principle, the loss in resolution caused by the homonuclear scalar and dipolar ^{1}H–^{1}H interactions can be eliminated by using the original homonuclear J spectroscopy experiment (37). In practice, however, the difficulty in obtaining absorption mode spectra limits the use of this approach. Instead, we describe a simple alternative method that removes the couplings between protons resonating in a selected band of the spectrum to all protons outside of this band. The pulse scheme is sketched in Fig. 2*A*. During a so-called constant-time evolution period (38) of duration 2 T, only the transverse magnetization of a narrow band of protons is selected by the application of appropriate gradients and a so-called REBURP 180° pulse (39), applied at time $T + t_1/2$ after initial excitation. This pulse eliminates evolution of homonuclear coupling between protons inside and outside the selected band from the last 2 $T - t_1$ fraction of the constant-time

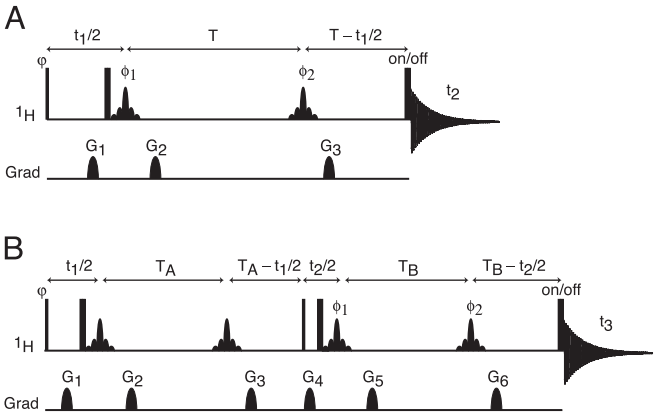

Fig. 2. Pulse schemes for homonuclear decoupled spectroscopy. (*A*) Basic pulse element for band-selective ^{1}H homonuclear decoupling. (*B*) Selective constant-time COSY pulse scheme, in which band-selective decoupling is implemented to specifically detect internucleotide residual dipolar coupling. Narrow and wide pulses correspond to flip angles of 90° and 180°, respectively. Unless specified, pulses are applied along the *x* axis. Shaped ^{1}H pulses are of the 180° REBURP type (39) (8-ms duration at 600 MHz, for a 500-Hz inversion bandwidth, centered at 5.8 ppm). Phase cycling: $\varphi = x,-x$; $\phi_1 = 2(x),2(-x)$; $\phi_2 = 4(x),4(-x)$; and receiver = $y,-y,-y,y,-y,y,y,-y$. (*A*) Quadrature detection in the indirect ^{1}H dimension is achieved by applying an extra ^{1}H 180° pulse just before the acquisition for half of the scans, storing signals separately, and using the usual ''echo-antiecho'' Fourier transform processing method (40, 41). All gradients are sine bell-shaped, with durations of 200 μs each, and (*x*,*y*,*z*) peak amplitudes in G/cm of (15,−33,20) for G_1, (9,9,12) for G_2, and (24,−24,32) for G_3. (*B*) Quadrature detection in F_1 is achieved by incrementing φ in the usual States-TPPI manner, and in F_2 by insertion on alternate scans of the extra ^{1}H 180° pulse just before acquisition (as in *A*). All gradients are sine bell-shaped, with durations of 200 μs, and (*x*,*y*,*z*) peak amplitudes in G/cm of (7,7,24) for G_1, (5,5,−7) for G_2, (12,12,17) for G_3, (15,−33,20) for G_4, (9,9,12) for G_5, and (24,−24,32) for G_6.

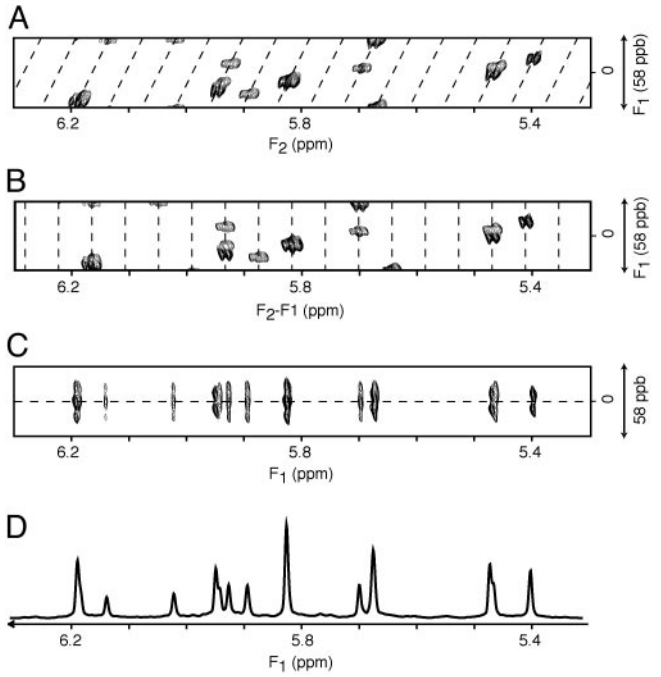

Fig. 3. Processing scheme for diagonal unfolding. (*A*) The original spectrum, processed in the standard manner, in which the diagonal signal (dashed line) is repeatedly folded in the F_1 dimension. Spectra have been recorded with the pulse scheme of Fig. 2*A*. The data matrix consists of 6*(t_1) × 1,024*(t_2) data points, with acquisition times of 144 ms (t_1) and 307 ms (t_2) ($2T$ = 164 ms). Linear prediction was used to double the number of time-domain points in the F_1 dimension, and both dimensions were apodized by cosine-squared functions and zero-filled by a factor of two before Fourier transformation. (*B*) An F_1-dependent frequency shift is applied to the F_2 dimension, such that the diagonal signals are now positioned vertically. (*C*) Vertical bands from the shifted spectrum are transposed and reassembled horizontally to construct an unfolded spectrum in the F_1 dimension. (*D*) A 20-Hz band, corresponding to the center of the F_2 dimension in *C*, is extracted, apodized by a sine-bell function, and then projected onto the F_1 dimension, to yield a homonuclear decoupled 1D spectrum. The spectral widths for the horizontal and vertical dimensions are, respectively, 600 and 34.9 Hz. For display purposes, the vertical axis is expanded 2-fold relative to the horizontal axis.

evolution period. Homonuclear coupling is eliminated from the remaining, first t_1 fraction of this period by a nonselective 180°, band-selective 180° pulse pair, applied at time $t_1/2$ after initial excitation. Temporarily ignoring couplings between protons within the selected band of resonances, chemical shift evolution during the constant-time evolution period has transformed an initial transverse spin operator, I_y, into $I_y\cos(2\pi\delta_i t_1) - I_x\sin(2\pi\delta_i t_1)$, where δ_i denotes the chemical shift of spin *i*. Signal detected during the subsequent detection period, t_2, is described by

$$S_i(t_1, t_2) = C_i\exp(-i2\pi\delta_i t_1) \times \exp(-i2\pi\delta_i t_2)\Pi_k\cos[2\pi(J_{i,k} + D_{i,k})T], \quad [2]$$

where C_i is an experimental constant, depending on sample concentration and instrumental factors, and the product extends over all spins *k*, coupled to *i* and resonating within the selected band. A second experiment, which inserts a 180° pulse between the end of the evolution period and the start of the detection changes the $\exp(-i2\pi\delta_i t_1)$ term to $\exp(i2\pi\delta_i t_1)$, and the paired data sets can be used to generate purely absorptive 2D NMR spectra in the standard manner (40, 41).

In the final spectrum, the F_1 dimension yields the homonuclear decoupled resonances at frequencies δ_i, whereas in the F_2 dimension the regular multiplet, centered around δ_i, is present (Fig. 3*A*). As demonstrated in Fig. 3*A*, extensive folding can be used in the F_1 dimension without introducing resonance overlap. The constant-time evolution mode in the t_1 dimension allows the use of mirror-image linear prediction (42) for further enhancement of the spectral resolution. A shearing transformation of the data matrix coordinates according to $(F_1', F_2') = (F_1, F_2 - F_1)$ yields a spectrum where the F_1' coordinate remains δ_i, but the F_2' dimension now represents the multiplet shape (Fig. 3*B*). Rearrangement of this data matrix (Fig. 3*C*) allows one to ''unfold'' such a spectrum, and a projection on its F_1' axis yields homonuclear decoupled spectra (Figs. 1*C* and 3*D*), which exhibit much improved spectral resolution.

Detection of Remote Dipolar Interactions. The above described homonuclear decoupling procedure only removes the couplings between protons resonating inside the spectral region selected by the band-selective pulse and those outside. When selecting, for example, the H1′ region in a DNA spectrum, which also contains the cytidine H5 resonances, dephasing resulting from couplings between different H1′ protons, or between H1′ and H5 protons remains active during the t_1 evolution period. Considering that the distance between such pairs of protons almost always exceeds 4 Å, the corresponding dipolar couplings are very small. However, with all other large intranucleotide couplings effectively removed, these small couplings can readily be detected.

The pulse scheme used for this purpose (Fig. 2*B*) essentially duplicates the evolution period of Fig. 2*A* and separates the first and second evolution period by a 90° mixing pulse, which serves the same function as the 90° mixing pulse in the classic COSY experiment (43). During the t_1 and t_2 evolution periods, and

focusing on the spectral region encompassing the H1′ and H5 protons, $H_J = 0$, but H_z and H_{DD} terms for H1′/H1′ and H1′/H5 interactions remain intact. As a result, at the end of the second constant time evolution period, of duration 2 T_B, the signal originating on proton i and remaining on proton i is described by

$$S_{i,j}(t_1, t_2) = C_i \sin(2\pi\delta_i t_1)\Pi_k \cos(2\pi D_{i,k} T_A)\cos(2\pi D_{i,k} T_B) \times \exp(i2\pi\delta_i t_2). \quad [3a]$$

Magnetization transferred from i to in-phase magnetization of spin j is described by

$$S_{i,j}(t_1, t_2) = C_i \sin(2\pi\delta_i t_1)\sin(2\pi D_{i,j} T_A)\sin(2\pi D_{i,j} T_B) \times \Pi_{k\neq j}\cos(2\pi D_{i,k} T_A)\cos(2\pi D_{i,k} T_B)\exp(i2\pi\delta_i t_2). \quad [3b]$$

2D Fourier transformation with respect to t_1 and t_2 then yields a doubly constant-time, in-phase COSY spectrum, with an $F_{i,j}(\delta_i, \delta_j)$ to $F_{i,i}(\delta_i, \delta_i)$ intensity ratio given by $\tan(2\pi D_{i,j} T_A)\tan(2\pi D_{i,j} T_B)$, neglecting the difference in relaxation of protons i and j. Similar $j \rightarrow i$ terms are obtained by interchanging the indices i and j. In practice, the experiment is recorded in a 3D mode, but as very extensive folding can be used (Fig. 3) relatively few complex points are required for the F_2 dimension.

For each duration of t_1, the (t_2, t_3) plane is processed as described above for generating the 1D homonuclear decoupled spectrum, resulting in a $S(t_1, F_2)$ spectrum, which is then Fourier-transformed with respect to t_1, yielding the final 2D spectrum. For extracting the homonuclear dipolar coupling directly from this projected 2D spectrum, an additional scale factor applies that depends on the mode of projection (skyline or integration) and apodization used in the projection process of each (t_2,t_3) plane. Its effect can be eliminated by calculating $D_{i,j}$ from:

$$\tan^2(2\pi D_{i,j} T_A)\tan^2(2\pi D_{i,j} T_B) = [F_{i,j}(\delta_i, \delta_j)F_{j,i}(\delta_j, \delta_i)]/[F_{i,i}(\delta_i, \delta_i)F_{j,j}(\delta_j, \delta_j)], \quad [4]$$

where $F_{i,j}(\delta_i, \delta_j)$ represents the cross peak intensity between protons i (F_1 dimension) and j (F_2 dimension).

Application to a B-Form DNA Oligomer. The experiments described above are demonstrated for a double-stranded B-form DNA oligomer: d(CGCGAATTCGCG)$_2$. In the absence of ^{13}C enrichment, transverse relaxation times of the H1′ resonances are quite favorable and are ~150 ms. This allows the use of reasonably long durations of the constant time evolution periods. For example, for generating the spectrum of Fig. 1*C*, a t_1 acquisition time of 144 ms was used, resulting in excellent resolution. For the constant-time COSY experiment of Fig. 2*B*, two such evolution periods are present, and for sensitivity considerations shorter transfer delays of 2 T = 81 ms are used for the constant-time evolution periods. The alignment of the dodecamer in the ether-bicelle medium (44) is described by a nearly axially symmetric alignment tensor (rhombicity, $R = 0.25$) of magnitude $D_a^{CH} = 60.8$ Hz ($A_a = 2.77 \times 10^{-3}$). For a pair of protons, separated by 7 Å on a vector parallel to the unique axis of the tensor, this corresponds to a D_{HH} dipolar coupling of 1.83 Hz, or a cross peak to diagonal peak ratio of 0.25. The smallest ratios observed in the spectrum are on the order of 0.006, corresponding to an internuclear separation of 12.7 Å if the vector were parallel to the unique axis of the alignment tensor.

A cross section through the spectrum of Fig. 4, taken at the diagonal resonance of A5-H1′ reveals several of the weaker dipolar interactions, annotated with the corresponding pairs of protons involved in the structure. At 10.4 Å, the longest distance observed is that between A5-H1′ and C9-H5 on the opposing strand (Fig. 5), giving rise to a 0.3-Hz coupling. The orientation of this vector is nearly orthogonal to the helix axis ($\theta \approx \pi/2$ in Eq. **1b**), and therefore corresponds to this relatively small coupling. The A5-H1′ to C11- H1′ vector is nearly parallel to the helix axis and at a comparable internuclear separation of 9.7 Å yields a >2-fold larger coupling and a correspondingly more intense cross peak (Fig. 4*B*).

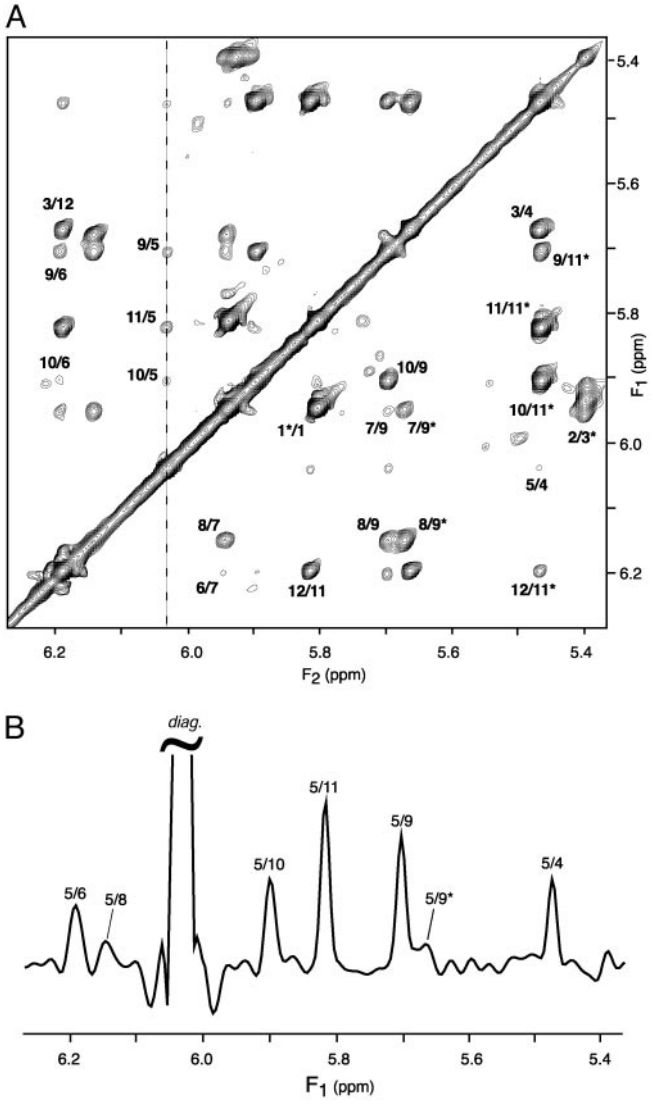

Fig. 4. ^{1}H-^{1}H dipolar correlation spectra of d(CGCGAATTCGCG)$_2$, homonuclear band-selectively decoupled, recorded at 600 MHz in 12% ether bicelle medium. (*A*) Projection of the 3D spectrum along the F_3 axis onto the 2D (F_1,F_2) plane, displaying the dipolar connectivities between $H_{1'}$ and H_5 protons. (*B*) F_1 cross section, taken at the frequency of A5-$H_{1'}$ (6.03 ppm, marked by a dashed line in *A*). Spectra have been recorded with the pulse scheme of Fig. 2*B*. Total measurement time is 36 h, with the data matrix consisting of 40*(t_1) × 20*(t_2) × 256*(t_3) data points, with acquisition times of 62.4 ms (t_1), 60.8 ms (t_2) and 51 ms (t_3) (2 T_A = 2 T_B = 81 ms). Linear prediction was used to double the number of time-domain points in both the t_1 and t_2 dimensions, and all three dimensions were apodized by cosine-squared functions and zero-filled before Fourier transformation. Each (F_2,F_3) plane has been reduced to its homonuclear decoupled 1D equivalent, using the protocol of Fig. 3 (using a 60-Hz width for the projected central band). Intrastrand connectivities (intranucleotide, sequential, and long range) are labeled by their corresponding residue numbers on the lower right half of the spectrum, whereas the interstrand cross peaks are labeled in the upper left half of the spectrum. Only positive contour levels are displayed.

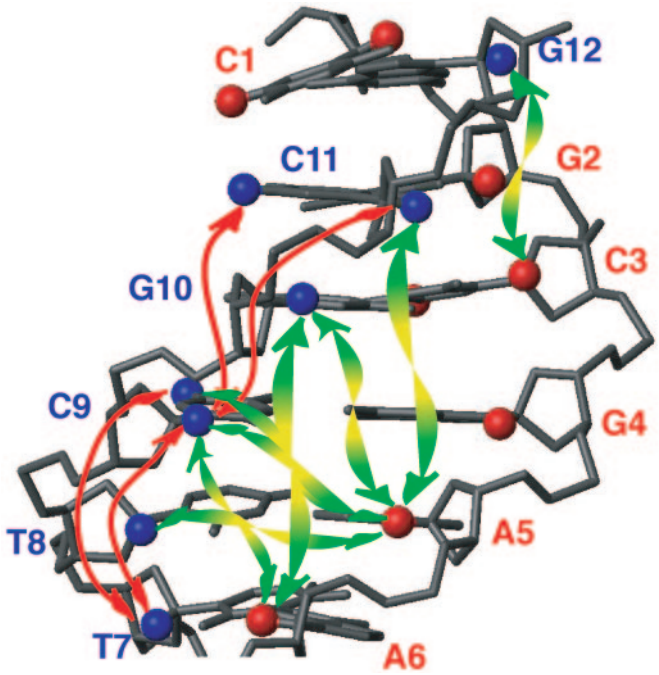

Fig. 5. Long-range ^{1}H-^{1}H connectivities observed for the d(CGCGAATTCGCG)$_2$ dodecamer, with only half of the palyndromic oligomer shown. H1′ and H5 are displayed as spheres. Nucleotides are labeled by their nucleotide numbers, with red and blue marking the opposing strands. Intrastrand and interstrand long-range connectivities are indicated by red and green arrows, respectively.

Discussion

Weak alignment offers numerous new opportunities in structural studies of macromolecules. They not only improve local structural characteristics, such as the Ramachandran map distribution when applied to proteins (45), but also can define global features at high accuracy, including DNA bending (32, 46, 47) and relative orientations of proteins in complexes (48). In addition, they potentially can reveal details about internal dynamics, including their amplitude and directions, provided the motional amplitudes are sufficiently large (49).

De novo structure determination on the basis of dipolar couplings has also been demonstrated, but remains challenging if only one-bond heteronuclear dipolar couplings are available (50–52). In this case, the lack of distance restraints makes it difficult to correctly translate elements of well-defined local structure relative to one another. ^{1}H–^{1}H dipolar couplings may prove to be a useful complement in this respect. Their measurement is not afflicted by spin diffusion effects and therefore can provide more accurate distance information, albeit with increased error bounds when the dipolar interactions are very weak. Our results demonstrate that ^{1}H–^{1}H dipolar interactions are detectable up to distances of at least 12 Å, well beyond the reach of direct nuclear Overhauser effect interactions ($\leq$6 Å). These remote ^{1}H–^{1}H dipolar couplings may prove particularly useful in the assembly of parts of well-defined local structure, such as helical regions in complex nucleic acid structures, or elements of secondary structure in proteins.

The concept of band-selective homonuclear ^{1}H decoupling is likely to be beneficial to a wide range of experiments, both in proteins and nucleic acids, and is not restricted to the measurement of dipolar couplings. Particularly in experiments that use direct observation of the ^{13}C spectrum, where homonuclear couplings severely affect attainable resolution, significant gains may be achievable.

Although we demonstrated the application of homonuclear decoupling and remote dipolar coupling measurement for an unlabeled nucleic acid, where ^{1}H transverse relaxation properties are favorable, comparably favorable relaxation characteristics are encountered when considering ^{1}H–^{15}N or aliphatic ^{1}H–^{13}C two-spin coherence terms. Therefore, extension of this approach to isotopically enriched proteins and nucleic acids also appears feasible.

We thank Dennis Torchia for useful suggestions. This work was supported by a fellowship from the Human Frontier Science Program (to J.B.).

1. Saupe, A. & Englert, G. (1963) *Phys. Rev. Lett.* **11,** 462–464.
2. Emsley, J. W. (1996) in *Encyclopedia of Nuclear Magnetic Resonance*, eds. Grant, D. M. & Harris, R. K. (Wiley, Chichester, U.K.), Vol. 4, pp. 2788–2799.
3. Burnell, E. E. & DeLange, C. A. (2003) *NMR or Ordered Liquids* (Kluwer, Boston).
4. Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 9279–9283.
5. Banci, L., Bertini, I., Huber, J. G., Luchinat, C. & Rosato, A. (1998) *J. Am. Chem. Soc.* **120,** 12903–12909.
6. Demene, H., Tsan, P., Gans, P. & Marion, D. (2000) *J. Phys. Chem. B* **104,** 2559–2569.
7. Ma, C. & Opella, S. J. (2000) *J. Magn. Reson.* **146,** 381–384.
8. Gaponenko, V., Dvoretsky, A., Walsby, C., Hoffman, B. M. & Rosevear, P. R. (2000) *Biochemistry* **39,** 15217–15224.
9. Tjandra, N. & Bax, A. (1997) *Science* **278,** 1111–1114.
10. Bax, A. & Tjandra, N. (1997) *J. Biomol. NMR* **10,** 289–292.
11. Gaemers, S. & Bax, A. (2001) *J. Am. Chem. Soc.* **123,** 12343–12352.
12. Nieh, M. P., Glinka, C. J., Krueger, S., Prosser, R. S. & Katsaras, J. (2001) *Langmuir* **17,** 2629–2638.
13. Sanders, C. R. & Schwonek, J. P. (1992) *Biochemistry* **31,** 8898–8905.
14. Sanders, C. R., Hare, B. J., Howard, K. P. & Prestegard, J. H. (1994) *Prog. Nucl. Magn. Reson. Spectrosc.* **26,** 421–444.
15. Clore, G. M., Starich, M. R. & Gronenborn, A. M. (1998) *J. Am. Chem. Soc.* **120,** 10571–10572.
16. Hansen, M. R., Mueller, L. & Pardi, A. (1998) *Nat. Struct. Biol.* **5,** 1065–1074.
17. Fleming, K., Gray, D., Prasannan, S. & Matthews, S. (2000) *J. Am. Chem. Soc.* **122,** 5224–5225.
18. Ruckert, M. & Otting, G. (2000) *J. Am. Chem. Soc.* **122,** 7793–7797.
19. Prosser, R. S., Losonczi, J. A. & Shiyanovskaya, I. V. (1998) *J. Am. Chem. Soc.* **120,** 11010–11011.
20. Tycko, R., Blanco, F. J. & Ishii, Y. (2000) *J. Am. Chem. Soc.* **122,** 9340–9341.
21. Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T. & Grzesiek, S. (2000) *J. Biomol. NMR* **18,** 303–309.
22. Yang, D. W., Venters, R. A., Mueller, G. A., Choy, W. Y. & Kay, L. E. (1999) *J. Biomol. NMR* **14,** 333–343.
23. Zidek, L., Wu, H. H., Feigon, J. & Sklenar, V. (2001) *J. Biomol. NMR* **21,** 153–160.
24. Yan, J. L., Corpora, T., Pradhan, P. & Bushweller, J. H. (2002) *J. Biomol. NMR* **22,** 9–20.
25. Tjandra, N., Marquardt, J. & Clore, G. M. (2000) *J. Magn. Reson.* **142,** 393–396.
26. Wu, Z., Delaglio, F., Tjandra, N., Zhurkin, V. B. & Bax, A. (2003) *J. Biomol. NMR* **26,** 297–315.
27. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980) *Nature* **287,** 755–758.
28. Tereshko, V., Minasov, G. & Egli, M. (1999) *J. Am. Chem. Soc.* **121,** 470–471.
29. Shui, X. Q., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998) *Biochemistry* **37,** 8341–8355.
30. Nerdal, W., Hare, D. R. & Reid, B. R. (1989) *Biochemistry* **28,** 10008–10021.
31. Denisov, A. Y., Zamaratski, E. V., Maltseva, T. V., Sandstrom, A., Bekiroglu, S., Altmann, K. H., Egli, M. & Chattopadhyaya, J. (1998) *J. Biomol. Struct. Dyn.* **16,** 547–568.
32. Tjandra, N., Tate, S., Ono, A., Kainosho, M. & Bax, A. (2000) *J. Am. Chem. Soc.* **122,** 6190–6200.
33. Young, M. A., Ravishanker, G. & Beveridge, D. L. (1997) *Biophys. J.* **73,** 2313–2336.
34. McConnell, K. J. & Beveridge, D. L. (2000) *J. Mol. Biol.* **304,** 803–820.
35. Wu, Z. R. & Bax, A. (2002) *J. Am. Chem. Soc.* **124,** 9672–9673.
36. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995) *J. Biomol. NMR* **6,** 277–293.
37. Aue, W. P., Karhan, J. & Ernst, R. R. (1976) *J. Chem. Phys.* **64,** 4226–4227.
38. Bax, A., Mehlkopf, A. F. & Smidt, J. (1979) *J. Magn. Reson.* **35,** 167–169.

39. Geen, H. & Freeman, R. (1991) *J. Magn. Reson.* **93,** 93–141.
40. Bachmann, P., Aue, W. P., Müller, L. & Ernst, R. R. (1977) *J. Magn. Reson.* **28,** 29–39.
41. Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. (1991) *J. Magn. Reson.* **93,** 151–170.
42. Zhu, G. & Bax, A. (1990) *J. Magn. Reson.* **90,** 405–410.
43. Aue, W. P., Bartholdi, E. & Ernst, R. R. (1976) *J. Chem. Phys.* **64,** 2229–2246.
44. Ottiger, M. & Bax, A. (1999) *J. Biomol. NMR* **13,** 187–191.
45. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. (1997) *Nat. Struct. Biol.* **4,** 732–738.
46. Vermeulen, A., Zhou, H. J. & Pardi, A. (2000) *J. Am. Chem. Soc.* **122,** 9638–9647.
47. Barbic, A., Zimmer, D. P. & Crothers, D. M. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 2369–2373.
48. Clore, G. M. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 9021–9025.
49. Meiler, J., Prompers, J. J., Peti, W., Griesinger, C. & Bruschweiler, R. (2001) *J. Am. Chem. Soc.* **123,** 6098–6107.
50. Delaglio, F., Kontaxis, G. & Bax, A. (2000) *J. Am. Chem. Soc.* **122,** 2142–2143.
51. Hus, J. C., Marion, D. & Blackledge, M. (2001) *J. Am. Chem. Soc.* **123,** 1541–1542.
52. Rohl, C. A. & Baker, D. (2002) *J. Am. Chem. Soc.* **124,** 2723–2729.

Liquid-crystal NMR structure of HIV TAR RNA bound to its SELEX RNA aptamer reveals the origins of the high stability of the complex

Hélène Van Melckebeke[†], Matthew Devany[†], Carmelo Di Primo[‡§], François Beaurain[‡§], Jean-Jacques Toulmé[‡§], David L. Bryce[¶‖], and Jérôme Boisbouvier[†‖]

[†]Institut de Biologie Structurale (IBS) Jean-Pierre Ebel, 41 Rue Jules Horowitz, Commissariat à l'Energie Atomique (CEA), Centre National de la Recherche Scientifique (CNRS), Université Joseph-Fourier, 41 Rue Jules Horowitz, F-38027 Grenoble, France; [‡]INSERM U869, Institut Européen de Chimie et Biologie, 2 Rue Robert Escarpit, 33607 Pessac Cedex, France; [§]Université Victor Segalen, 146 Rue Léo Saignat, 33076 Bordeaux Cedex, France; and [¶]Department of Chemistry, University of Ottawa, 10 Marie Curie Private, Ottawa, ON K1N 6N5, Canada

Edited by Adriaan Bax, National Institutes of Health, Bethesda, MD, and approved April 14, 2008 (received for review January 3, 2008)

Transactivation-response element (TAR) is a stable stem–loop structure of HIV RNA, which plays a crucial role during the life cycle of the virus. The apical loop of TAR acts as a binding site for essential cellular cofactors required for the replication of HIV. High-affinity aptamers directed against the apical loop of TAR have been identified by the SELEX approach. The RNA aptamers with the highest affinity for TAR fold as hairpins and form kissing complexes with the targeted RNA through loop–loop interactions. The aptamers with the strongest binding properties all possess a GA base pair combination at the loop-closing position. Using liquid-crystal NMR methodology, we have obtained a structural model in solution of a TAR–aptamer kissing complex with an unprecedented accuracy. This high-resolution structure reveals that the GA base pair is unilaterally shifted toward the 5′ strand and is stabilized by a network of intersugar hydrogen bonds. This specific conformation of the GA base pair allows for the formation of two supplementary stable base-pair interactions. By systematic permutations of the loop-closing base pair, we establish that the identified atomic interactions, which form the basis for the high stability of the complex, are maintained in several other kissing complexes. This study rationalizes the stabilizing role of the loop-closing GA base pairs in kissing complexes and may help the development or improvement of drugs against RNA loops of viruses or pathogens as well as the conception of biochemical tools targeting RNA hairpins involved in important biological functions.

residual dipolar coupling | RNA–RNA complex | kissing complex | GA base pair

The transactivation of the genome transcription of HIV-1 requires the binding of the Tat protein to a structured segment of the viral RNA called transactivation-response element (TAR). TAR is a 57-nt-long RNA that folds into a stable stem–loop structure (1, 2). In the absence of Tat, the transcription of the virus is initiated, but the RNA polymerase II disengages prematurely from the template (3). Mutations in the apical loop of TAR that do not interfere with Tat binding do indeed modulate the transactivation, suggesting that the loop region acts as a binding site for essential cellular cofactors, such as CycT1, P-TEFb, and CDK9 (4–7). The apical loop of TAR RNA is therefore a good target to inhibit artificially the replication of HIV-1.

Several strategies have allowed for the identification of various ligands of TAR RNA (8–11). A large field of investigation concerns RNA aptamers that specifically bind TAR, which potentially offer a good selectivity and affinity for the RNA target. Antisense oligonucleotides complementary to the top loop of the TAR element were initially tested (12). Subsequently, a first class of stem–loop rationally designed RNA ligands has been proposed, whose loops can interact with the six bases of the TAR apical loop (13) to form a so-called "kissing complex" (14). An *in vitro* selection (SELEX) approach has been used in an attempt to identify oligonucleotides that display high affinity and specificity for the HIV-1 TAR RNA element (15). The RNA aptamers with highest affinity for TAR fold as imperfect hairpins and form kissing complexes with the targeted RNA through loop–loop interactions. The apical loop of these aptamers contains six central bases, which are complementary to the 6-nt loop of TAR. Interestingly, all aptamers with the strongest binding properties also possess a GA combination at the end of the stem (loop-closing position). The role of loop-closing G and A nucleotides was investigated by systematic permutation (16) and by molecular dynamics studies (17). It was shown that the nature of the residues closing the loop of the aptamer was crucial for kissing complex stability, and that the selected G-A combination leads to the most stable complex.

Recently, chemically modified oligonucleotides directed toward the TAR stem loop, derived from the selected RNA kissing aptamer have been prepared to increase resistance to nucleases and affinity for the target (18–20). Several of these ligands have been shown to regulate TAR-dependent transcription *in vitro* or in cultured cells (18, 21). Chemically modified RNA aptamers, which strongly interact with the apical loop of the TAR hairpin could then be potential candidates for the development of anti-HIV drugs able to disrupt the ternary TAR–Tat–CycT1 complex, leading to abortive RNA synthesis.

In the present work, we used high-field NMR spectroscopy and liquid-crystalline NMR methodology to obtain a high-resolution solution structure of the TAR–aptamer complex of highest stability. NMR data recorded on weakly aligned samples allowed us to obtain a structural model in solution of an RNA–RNA kissing complex with higher accuracy and precision than previously reported kissing complex solution structures (14, 20, 22–26). This structural study revealed the key atomic interactions responsible for the stabilizing role of the GA base pair at the loop-closing position of kissing complexes.

Results

Minimal Functional System Suitable for High-Resolution NMR Study. To investigate the stabilizing role of the GA closing base pair of the aptamer, we performed NMR experiments on a complex formed by

Author contributions: J.-J.T., D.L.B., and J.B. designed research; H.V.M., M.D., C.D.P., and F.B. performed research; H.V.M., M.D., C.D.P., J.-J.T., D.L.B., and J.B. analyzed data; and H.V.M., D.L.B., and J.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The structures have been deposited in Protein Data Bank (PDB ID code 2rn1), and assignment and restraints have been deposited in the Biological Magnetic Resonance Bank (BMRB entry 11014).

‖To whom correspondence may be addressed. E-mail: dbryce@uottawa.ca or jerome.boisbouvier@ibs.fr.

This article contains supporting information online at www.pnas.org/cgi/content/full/0712121105/DCSupplemental.

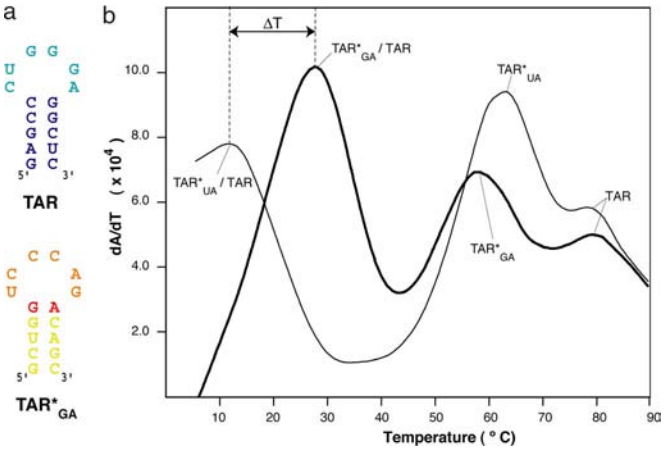

Fig. 1. Stability of TAR–aptamer kissing complex suitable for NMR studies. (*a*) Sequences of the RNAs used in the study, denoted TAR and TAR*$_{GA}$. The TAR*$_{UA}$ aptamer is obtained by replacing G*5 of TAR*$_{GA}$ by a U (highlighted in red). (*b*) First derivatives of the UV absorbance (A) at 260 nm of TAR-TAR*$_{GA}$ and TAR-TAR*$_{UA}$ complexes as a function of temperature. The experiments were performed for 1 μM each oligomer RNA in a 10-mM sodium phosphate buffer at pH 6.6, 50 mM NaCl, and 0.01 mM EDTA.

the stem–loop corresponding to the apical part of TAR and a truncated version of the aptamer identified by SELEX for its high affinity for the TAR target (15). These two molecules are each 16 nt long (Fig. 1*a*) and will be denoted TAR and TAR*$_{GA}$ in the following sections (the last two letters indicate the nature of the loop-closing bases). To determine whether the GA bases closing the aptamer loop maintain their stabilizing role in this simplified complex, we performed thermal denaturation experiments monitored by UV spectroscopy (Fig. 1*b*). A drop in the melting temperature is observed between the TAR*$_{GA}$-TAR (27.2 $\pm$ 0.5°C) and TAR*$_{UA}$-TAR complexes (11.9 $\pm$ 0.5°C). These results are consistent with the melting temperatures measured for complexes between TAR and RNA aptamers with longer stems (16, 17). This validates the choice of shorter RNAs to investigate the role of the GA loop-closing bases of the aptamer. Despite the importance of Mg^{2+} ions for the stability of kissing complexes (27), the interaction is still strong enough to enable a high-resolution NMR study in the absence of magnesium ($K_d \approx 0.5$ μM). The addition of five equivalents of magnesium ions does not significantly change the NMR spectra, as described for a similar kissing complex (14). Furthermore, the 15°C drop in the melting temperature of the complex when replacing the GA loop-closing bases by a UA base pair is conserved in the absence (Fig. 1*b*) or presence of 0.3 mM magnesium chloride (32.5 $\pm$ 0.5°C for TAR-TAR*$_{GA}$ and 18.4 $\pm$ 0.5°C for TAR-TAR*$_{UA}$). Because the magnesium ions do not play a specific role in the stabilization of the complex by the GA loop-closing bases of the aptamer, all NMR experiments were carried out in the absence of magnesium to compare our structure with a previously determined solution model of the kissing complex lacking the loop-closing GA base pair (14).

The Aptamer Loop-Closing GA Base Pair Is of the N_1-N_1 Carbonyl-Amino Type. The imino region of the 1D NMR spectra is composed of 16 signals, which have been assigned based on the observable sequential imino/imino NOE connectivities. The observation of the signals for the 16 different imino protons indicated they are all protected from solvent exchange by base-pairing interactions. To further identify nucleotides that are base-paired, experiments correlating ^{15}N and ^{1}H nuclei linked by indirect $^{2h}J_{NN}$ couplings across the hydrogen bonds were recorded (28). The 5 bp of the TAR stem, the first 4 bp of the TAR*$_{GA}$ stem, and the 12 bases of the interacting loops exhibit chemical shifts, $^{2h}J_{NN}$ couplings, and ^{1}H-^{1}H NOEs characteristic of canonical Watson–Crick base pairing, confirming the secondary structure proposed for the TAR/TAR*$_{GA}$ kissing complex (Fig. 2*a*). Analysis of $^{2h}J_{NN}$ connectivity of the TAR*$_{GA}$ G*5 nucleotide revealed that the imino proton H_1 of G*5 forms a hydrogen bond with N_1 of A*12 [supporting information (SI) Fig. S1*f*], confirming that the loop-closing G and A nucleotides are base-paired together. An additional NOE correlation peak observed between H_1 of G*5 and H_2 of A*12 indicates the close proximity of these two atoms (Fig. S1*b*). The GA base-pairing geometry has thus been identified as the N_1-N_1 carbonyl-amino type (Fig. S1*a*), because all other base-pairing modes are incompatible with the NMR data.

Liquid-Crystal NMR Structure of the TAR-TAR*$_{GA}$ Complex. The hydrogen-bonding network of Watson–Crick and N1-N1 carbonyl-amino type base pairs determined experimentally for the TAR-TAR*$_{GA}$ complex served as a basis for the use of a set of 62 distance restraints. Three hundred ninety-two proton–proton distance constraints were extracted from NOESY experiments. To supplement the local information gained from NOE data, we measured 143 ^{13}C-^{1}H residual dipolar couplings (RDCs) to provide long-range orientational information for both base planes and sugar rings. These data were completed by 171 dihedral angle restraints applied to the glycosidic angles and phosphodiester angles (*SI Text*). An average of 24 constraints per nucleotide (Table 1) were used to determine a structure ensemble of 17 conformers, which superimposed with an average pairwise rmsd of 0.26 $\pm$ 0.07 Å (Fig. 2*b*). Liquid-crystal NMR parameters are particularly sensitive to bond orientation and provide a unique opportunity to define precisely molecular conformation (29). The absence of significant violation of RDCs in the calculated structures indicates that the dipolar coupling data can be fit by a single structure. Analysis of the structure ensemble revealed that the puckering phases of the

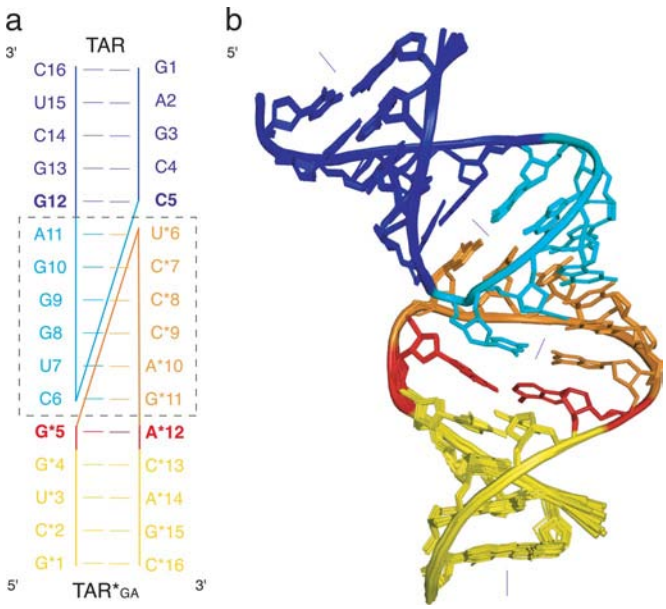

Fig. 2. Structure of the TAR-TAR*$_{GA}$ complex. (*a*) Schematic representation of the secondary structure of the kissing complex formed by the HIV-1 TAR RNA element and the TAR*$_{GA}$ aptamer used in this study (the aptamer nucleotides are labeled with a star). Both loop-closing base pairs are in boldface, in blue for TAR and in red for TAR*$_{GA}$. The intermolecular helix is highlighted with a dashed square. (*b*) Side view of the NMR ensemble of the TAR-TAR*$_{GA}$ complex. The superposition of 17 structures of the NMR ensemble has been performed on the lowest-energy structure for all heavy atoms of A2-U15, C*2-G*15 (all bases except the first base pairs of the stems) with a rmsd equal to 0.26 Å. The directions of helices of the mean structure are depicted by thin lines.

BIOPHYSICS

Table 1. Structural statistics

Distance constraints (intra/seq./long dist., $j > i + 1$)		392 (149/186/57)
Hydrogen bond restraints		62
Dihedral restraints		171
^{13}C-^{1}H RDCs (class 1 / class 2 / class 3)		143 (90/31/22)
rmsd from experimental restraints†	RDCs class 1 (Hz)	1.41 ± 0.03
	dihedral angles (°)	0.33 ± 0.02
	NOEs (Å)	0.0487 ± 0.001
Deviations from idealized covalent geometry†	bonds (Å)	0.0038 ± 0.0001
	angles (°)	0.58 ± 0.005
	impropers (°)	1.45 ± 0.03
Cross validation of a subset of the class 1 RDCs (Q^{free})		13.5%
Heavy atoms pairwise average root mean square deviation†, Å	TAR stem helix (2–5,12–15)	0.13 ± 0.08
	TAR*$_{GA}$ helix (2*-5*,12*-15*)	0.26 ± 0.11
	Intermol. helix (6–11, 6*-11*)	0.12 ± 0.06
	Overall (2–15, 2*-15*)	0.26 ± 0.07

†The statistics are given as an average for the 17 lowest-energy structures.

different ribose rings can be determined with an average precision lower than 2°, whereas the buckle, propeller twist, opening, tilt, roll, and twist angles can be derived with an average precision on the order of 1° (Tables S1–S3). To further investigate the accuracy of the calculated structures, we backcalculated a set of structures with a small fraction of RDCs omitted to derive the quality factor Q^{free} (30). The computed average Q^{free} factor is low (Q^{free} = 13.5%), attesting to the accuracy of the structure of the TAR–TAR*$_{GA}$ complex (31).

The overall shape of the kissing complex can be decomposed into three parts: two intramolecular helices containing 5 bp each, linked by a central intermolecular helix consisting of the six bases of each loop paired together. These three elements stack on top of each other to form a quasicontinuous helix. A strong curvature of 45° toward the major groove of the loop–loop helix is observed between the directions of the intramolecular stem helices of the two RNA molecules (Fig. 2*b*). Analysis of the geometry of the complex (Tables S1–S3) shows that the three helices exhibit a regular A form shape, except at the interhelical junction sites. The unusual G*5-A*12 base pair of TAR*$_{GA}$ is characterized by large values of opening (−26.4 ± 1.8°) and buckle (24.7 ± 0.7°), compared with standard Watson–Crick base pair (Table S2) but remains well stacked with the neighboring aptamer base pair G*4-C*13 and the intermolecular base pair C6-G*11 (Fig. 3 and Table S3). All of the sugars adopt a $C_{3'}$-*endo* conformation, except three riboses in the vicinity of the aptamer GA loop-closing base pair, which adopt a $C_{4'}$-*exo* conformation (G*5 and A*11) or a $C_{2'}$-*endo* conformation (G*4). At the junctions between the stem helices and the intermolecular helix, a different behavior of the two strands is observed. Although the 3′ strand smoothly deviates from a canonical A form RNA helix, larger deviations of the phosphodiester backbone are observed for the 5′ side. This deviation from ideal helix geometry is different for the two junctions of the complex: a straight kink of the phosphodiester backbone occurs for TAR at the C6 phosphorus position (Fig. S2*a*), whereas a continuous bending is observed between the aptamer phosphorus atoms of G*5 and U*6 (Fig. S2*b*).

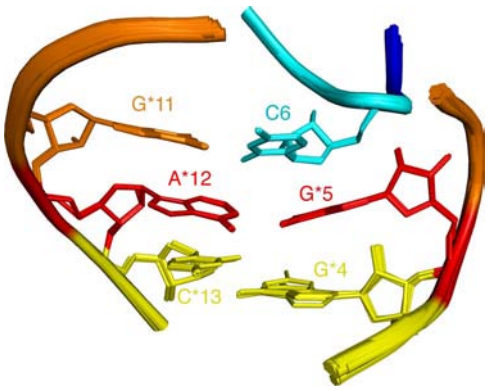

Fig. 3. Detailed view of the base pair stacking at the TAR*$_{GA}$ loop-closing position. The superposition of the 17 structures of the NMR ensemble has been performed on the lowest energy structure for all base and sugar heavy atoms of nucleotides C6, G*4, G*5, G*11, A*12, and C*13 with a rmsd equal to 0.13 Å.

Discussion

The TAR*$_{GA}$ Loop-Closing Base Pair Conformation Is Stabilized by a Network of Intersugar Hydrogen Bonds and Induces the Formation of Two Additional Base Pairs. One major consequence of the introduction of a purine–purine base pair is an increase by ≈2 Å of the $C_{1'}$-$C_{1'}$ distance at the loop-closing position in the TAR-TAR*$_{GA}$ complex compared to the TAR-TAR*$_{UA}$ structure (14). Very interestingly, superimposition of the TAR*$_{GA}$ ensemble on the first 4 bp of the TAR*$_{UA}$ stem reveals a very good overlap between the two loop-closing A*12 bases, whereas the U*5 of TAR*$_{UA}$ superimposes on the six-atom cycle of TAR*$_{GA}$ G*5 (Fig. 4). The resulting unilateral shift of G*5 is stabilized by a network of new hydrogen bonds. The $H_{2'}$ proton of G*5 is placed in the center of a cluster of three hydrogen bond acceptors located on the TAR phosphodiester backbone between C5 and C6 (namely $O_{3'}$, O_{1P}, and $O_{5'}$). Analysis of the different possible rotameric states around the $C_{2'}$—$O_{2'}$ bond indicates that $H_{2'}$ of G*5 can form a stable intermolecular hydrogen bond with the $O_{5'}$ or O_{1P} acceptors of C6 (Fig. 5). Because this proton was not detectable by NMR (32, 33), the exact nature of the TAR accepting group cannot be determined directly by NMR. To add experimental evidence that the $H_{2'}$ proton of G*5 participates in an intermolecular hydrogen bond, −2′OMe modified versions of TAR*$_{GA}$ were synthesized to perform thermal denaturation experiments. The replacement of the $H_{2'}$ of the aptamer G*5 by a methyl group induces a drop of the melting temperature by 10°C (Table S4). In a control experiment, no effect was detected on the stability of the kissing complex when a similar modification to the sugar of A*12 was introduced. These results strongly support the notion that this intermolecular hydrogen bond involving $H_{2'}$ of G*5 plays a substantial role in the stabilization of the kissing complex, as suggested earlier by molecular dynamics simulations (17). Similar analysis of the liquid-crystal NMR structure ensemble indicates that the $H_{2'}$ of C5 may form a supplementary weak intermolecular hydrogen bond with U*6-O_{2P}. The G*5 stretched conformation may be also stabilized by the $C_{2'}$-*endo* conformation of G*4, which places its $H_{2'}$ proton in position to form a strong hydrogen bond with the $O_{4'}$ acceptor of G*5. The sugar ring of G*4 is itself stabilized by a hydrogen bond between its $O_{5'}$

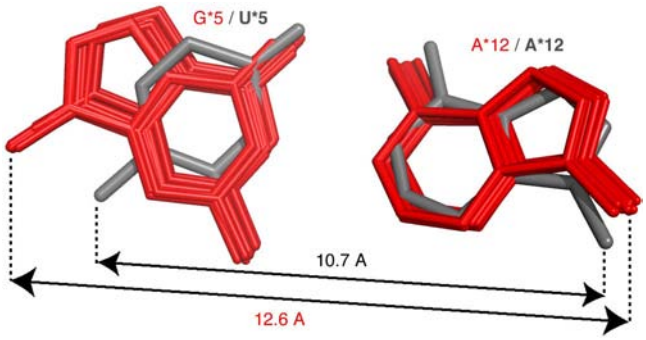

Fig. 4. Superimposition of TAR*$_{GA}$ and TAR*$_{UA}$ loop-closing base pairs. The stem of TAR*$_{UA}$ from the previously determined structure of the TAR-TAR*$_{UA}$ complex (ref. 14; PDB ID code 1KIS) was superimposed on the stem of TAR*$_{GA}$. All the structures (TAR*$_{UA}$ and TAR*$_{GA}$ structure ensemble) have been superimposed on base heavy atoms of nucleotides 2, 3, 4, 13, 14, and 15 of the TAR*$_{GA}$ lowest-energy structure. The G*5-A*12 base pairs are represented in red for the TAR*$_{GA}$ structure ensemble, whereas U*5-A*12 of TAR*$_{UA}$ is represented in gray.

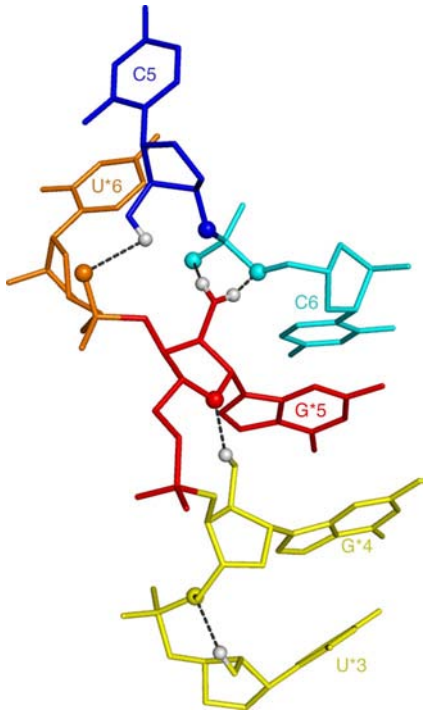

Fig. 5. Stabilization of the conformation of G*5 by an intersugar hydrogen bond network. The proposed hydrogen bonds are displayed as dashed lines. Heavy atoms of nucleotides U*3, G*4, and G*5 of TAR*$_{GA}$ and C5 and C6 of TAR of the lowest-energy structure are displayed. The different possible rotameric states around the $C_{2'}$—$O_{2'}$ bonds have been analyzed to define the most favorable location of $H_{2'}$ protons represented [these rotameric states are characterized by the dihedral angle (θ) defined between $H_{2'}$-$C_{2'}$-$O_{2'}$-$H_{2'}$]. The $H_{2'}$ of G*5 can form a hydrogen bond with either $O_{5'}$ of TAR cytidine 6 (with $\theta \approx 315°$ and the $O_{5'}$-$H_{2'}$-$O_{2'}$ angle denoted $\alpha \approx 155°$) or O_{1P} of the same nucleotide (with $\theta \approx 45°$ and $\alpha \approx 145°$). Because solution NMR experimental data do not allow discrimination between these two conformations, the two favorable rotameric states have been represented . $H_{2'}$ of G*4 forms a strong hydrogen bond with $O_{4'}$ of G*5 ($d_{O.H} < 2$ Å, $\alpha \approx 135°$, for $O_{2'}$-$C_{2'}$ rotameric angle $\theta \approx 135°$). The $H_{2'}$ proton of U*3 forms a hydrogen bond with $O_{5'}$ of G*4 ($d_{O.H} = 2.1$ Å, $\alpha \approx 130°$, for $O_{2'}$-$C_{2'}$ rotameric angle $\theta \approx 60°$). The $H_{2'}$ of TAR-C5 can form a supplementary weak intermolecular hydrogen bond with U*6-O_{2P} ($d_{O.H} = 2.4$ Å, $\alpha \approx 95°$).

and the $H_{2'}$ proton of U*3. However, no destabilization effects on the kissing complex were detected when the $H_{2'}$ of G*4 and U*3 were replaced by methyl groups (Table S4). These results indicate that these two intramolecular hydrogen bonds are likely not critical for the stabilization of the kissing complex but may potentially stabilize the aptamer both in the TAR/TAR*$_{GA}$ complex and in its free form.

The stabilization of the G*5 stretched conformation also allows the concomitant stabilization of two supplementary base pairs: the N1-N1 carbonyl-amino loop-closing G*5-A*12 itself, and the adjacent intermolecular base pair connecting the 5′ end nucleotide of the aptamer loop (U*6) with the 3′ end nucleotide of the TAR loop (A11). These two imino protons are not NMR-detectable in a complex formed between TAR and TAR*$_{UA}$ (13), indicating these hydrogens are not protected from chemical exchange with water by a stable base-pairing interaction. Together, the presence of these four supplementary base–base hydrogen bonds, along with the formation of the intersugar hydrogen bond between TAR*$_{GA}$ and TAR described above (Fig. 5), are key factors responsible for the 50-fold decrease in the dissociation constant of the aptamer–TAR kissing complex induced by the introduction of GA as a loop-closing base pair (16). The atoms involved in these intermolecular hydrogen bonds or stabilizing interactions should be preserved in any chemically modified aptamers to maintain the high affinity of the complex.

A GA Base Pair Closing the Loop: A General Rule to Stabilize Kissing Complexes? Using thermal denaturation, it was demonstrated that the combination of two purines (and especially the GA base pair) located at the stem–loop junction of the aptamer is the key factor enabling the stability of the TAR–aptamer kissing complex. A ranking of the stabilizing effects of the different base pair at the junction between the aptamer stem and loop has been obtained from these data: GA > GG > GU > AA > GC > UA > CA,CU (Table 2; refs. 16 and 17). To test whether these conclusions are restricted to this aptamer or whether they are more general, we investigated the effect of the change of the loop-closing base pair of TAR hairpin. This study reveals that the GA base pair located at the stem–loop junction of the TAR sequence also confers the highest stability for this kissing complex (Table 2). An increase of the melting temperature by 14°C is obtained on average, when replacing a Watson–Crick base pair by a GA base pair in the TAR loop-closing position.

In vitro selection of aptamers against the anticodon hairpin of yeast $tRNA^{Phe}$ also converged to the formation of kissing complexes (34). Interestingly, the authors showed that a GA base pair is preferred at the stem–loop junction of the $tRNA^{Phe}$ aptamer. To complement these results, we have performed a denaturation analysis of the $tRNA^{Phe}$–aptamer complex. The results show that the replacement of the GA base pair closing the aptamer loop by a GC base pair results in a drop of T_m by >10°C (Table S5 and Fig. S3). Similar studies have been performed on the RNAI–RNAII kissing complex involved in the control of the *colE1* plasmid replication. This kissing complex has been shown to possess no linking residues between the 7-nt loop–loop helix and the intramolecular helices (23, 35). Again, a mutation of the GC closing base pair to a GA base pair leads to a stabilization of the complex (16).

Despite the fact that all these previous examples indicate that the addition of a GA base pair to close the aptamer loop seems to be a general rule to stabilize kissing complexes, some slightly different results have been observed for complexes that possess some unpaired nucleotides. In the DIS-DIS HIV-1 kissing complex, switching from a purine–purine to a purine–pyrimidine base pair at the loop-closing position still induces destabilization of the DIS kissing complex (36). However, an AA combination is preferred at this position compared with the GA pair in this kissing complex containing a 6-nt loop–loop helix and three unpaired purines.

Table 2. Effects of the loop closing base pair on the stability of the kissing complex

TAR/aptamer	UA	GA	CU	GC	GU	CA
UA	13.0 ± 0.8	33.2 ± 1.6	<10	18.7 ± 1.9	23.4 ± 1.0	<10
GA	24.5 ± 0.5	41.8 ± 2.1	17.3 ± 0.2	31.5 ± 0.6	33.8 ± 2.6	23.2 ± 0.8
CU	<10	<10	<10	<10	<10	<10
CG	<10	33.1 ± 1.2	<10	14.9 ± 0.3	20.0 ± 0.8	<10

Melting temperatures (T_m) (°C) of complexes formed by different mutants of TAR (vertically) and SELEX identified aptamer (horizontally). The mutations concern the stem–loop closing base pair for both RNAs and are indicated in the different columns. A short version of TAR without the stem-stabilizing GC base pair (G3-C14, 14 nt), and a nontruncated apatamer ($R06_{24}$, refs. 16 and 17) has been used in this study. The given values are the mean and standard deviation of two or three measurements performed for 1 μM of each oligomer RNA in a 20 mM sodium cacodylate buffer (pH 7.3 at 20°C) containing 20 mM sodium chloride and 140 mM potassium chloride, and in the presence of 0.3 mM Mg^{2+}.

Curvature of Kissing Complexes Lacking Unpaired Residues. It must be noted that standard NMR data, based on the detection of NOEs and indirect scalar couplings, mainly define local geometries and do not provide direct evidence for global bending of the overall helical axis of the complex (14). The introduction of restraints based on RDCs (37–39) greatly improves this situation, allowing a precise determination of the bending of nucleic acids (40–46). The solution structure of a kissing complex determined here by liquid-crystal NMR technology is a particularly high resolution, allowing us to assess precisely the global bending of the kissing complex in solution. For these reasons, the global bending of the TAR*$_{GA}$-TAR structures should preferentially be compared with x-ray structures of RNA–RNA complexes.

As for the TAR-TAR*$_{GA}$ complex, the structures of these kissing complexes share a common quasicontinuous helix fold formed by two intramolecular and one intermolecular helices but differ in several respects. First, the number of paired loop residues can vary from two up to seven (22, 23, 35). In addition, some of the kissing complexes possess nonpaired bases (22, 24, 25, 47–49). For example, the crystal structures of the HIV-1 MAL and LAI kissing complexes (49), the anticodon–anticodon interaction in $tRNA^{Asp}$ (47), and a loop–loop interaction in 23S rRNA (48) show coaxial alignment of the two opposite stems due to flanking residues on the 5′ side. The stabilization observed in the UV-melting data for this type of 5′ configuration is likely to be caused by coaxial stacking (36). The unpaired bases provide a linker between the loop and the stem to cross the major groove of the loop–loop helix and give enough flexibility at the junction to maximize the stacking between each helical segment.

In contrast, the absence of a linker implies the existence of severe structural constraints at the stem–loop junctions for the HIV TAR aptamer complex. Using liquid-crystal NMR technology, we have shown that a unilateral stretching of the loop-closing GA base pair located at the stem–loop junction alleviates the tensions created by the lack of unpaired linking residues in the TAR-TAR*$_{GA}$ complex, consequently providing a higher stability. This results in a 45° bending of the TAR-TAR*$_{GA}$ kissing complex toward the major groove (Fig. 2*b*). Similar bending has been documented for a kissing complex formed between ColEl plasmid specific RNA I and RNA II transcripts, comprising seven intermolecular base pairs and no unpaired nucleotides (23, 35).

Conclusion

The high-resolution structure of HIV-1 TAR complexed with a high-affinity RNA aptamer has allowed us to elucidate the molecular origins of the gain in binding affinity associated with the aptamer GA loop-closing base pair, compared with rationally designed hairpin ligands containing only Watson–Crick base pairs. The larger interglycosidic distance imposed by this purine–purine base pair induces a unilateral shift of the G base, allowing for the formation of two supplementary stable base-pair interactions. This specific conformation is furthermore stabilized by a network of intersugar hydrogen bonds. Systematic permutation of the closing base pair and comparisons with other kissing complexes indicate that the stabilization role of the GA base pair seems to be a general rule for kissing complexes lacking unpaired nucleotides between the interacting loop and the intramolecular stems. Thus, the precise atomic interactions crucial for the stability of RNA–RNA kissing complexes, which are revealed by this liquid-crystal NMR study, may serve as a basis to produce modified aptamers with improved nuclease resistance and that maintain the highest affinity for the HIV TAR sequence. This knowledge will also be useful for the design of high-affinity RNA aptamers against RNA loops of viruses or pathogens as well as in the conception of biochemical tools targeting RNA hairpins involved in important biological functions.

Materials and Methods

Sample Preparation. Unlabeled RNA molecules used for this study were synthesized on a solid phase. ^{15}N-^{13}C-labeled (98%) 16-nt TAR and TAR*$_{GA}$ were prepared *in vitro* by using T7 RNA polymerase (Silantes). RNA samples were dialyzed against 50 mM NaCl at pH 6.6 in a 0.01 mM EDTA, 10 mM sodium phosphate buffer containing 0.4 g·liter^{-1} sodium azide. Only one of the partners was labeled with ^{15}N and ^{13}C in TAR/TAR*$_{GA}$ samples used for NMR studies. The final sample concentrations of the labeled species were between 0.4 and 0.8 mM, and the relative concentration of the unlabeled species in the mixed samples was 2:1. All of the NMR data were acquired at 25°C in D_2O buffer, except experiments involving exchangeable protons, which were recorded at 10 and 5°C in H_2O buffer. Further details on assignment, dihedral, and distance restraint determination are given in *SI Text* and Table S6.

Liquid-Crystal NMR Spectroscopy. For liquid-crystal NMR studies, the ^{15}N-^{13}C-labeled RNA complexes were aligned by adding to an isotropic sample a Pf1 filamentous phage solution (ref. 39; ASLA Biotech) to give a final concentration of ≈13 mg/ml. The ^2H NMR splitting observed for ^2H_2O at 25°C in these samples was 14.0 ± 0.1 Hz for the complex containing ^{15}N-^{13}C-labeled TAR RNA and 12.4 ± 0.1 Hz for the sample containing ^{15}N-^{13}C-labeled TAR*$_{GA}$ RNA. Spin-state selective experiments (50–52) were used to extract one-bond ^{13}C-^1H couplings in isotropic and Pf1-aligned ^{13}C-^{15}N-labeled RNA complex samples. Extracted RDCs were scaled linearly with respect to the observed D_2O splitting, to take into account the small difference in the magnitude of the alignment of different samples. The linearity of measured dipolar couplings with the D_2O splitting was verified experimentally by varying the amount of phages in the same sample of TAR/TAR*$_{GA}$ complex (data not shown). The RDCs were separated into three classes to account for overlap of the peaks in the spectra, with error bars estimated at 1 Hz (well resolved), 5 Hz (partial overlap), and 10 Hz (overlap).

Structure Calculation. A first estimate of the magnitude of the axial and rhombic components of the alignment tensor was obtained by using histogram methods (53–55). The values have been further refined by a grid search over the D_a and R values of the tensor to determine the minimal value of the dipolar energy for the set of structures calculated using Xplor-NIH (56) with a parameter file specifically optimized for nucleic acids (ref. 57; details in *SI Text*). Furthermore, the grid search procedure was repeated using only either RDCs for TAR or for TAR*GA molecules, independently. Both calculations converged on similar values of D_a and R, suggesting that the different parts of the complex are not subject to internal flexibility. This is corroborated by the absence of unpaired or highly mobile nucleotides at the junction between the three different helical segments. A single alignment tensor was then used to calculate a final set of 800 structures from

which the 17 structures with lowest energies were selected to represent the structure ensemble, and these were deposited in the Protein Data Bank (PDB ID code 2rn1).

Cross-validation of the structures was carried out by selecting a subset of high-precision one-bond ^{13}C-^{1}H RDCs (class 1), excluding them from the list of restraints enforced during the structure calculation, and then evaluating how well the excluded RDCs are predicted on the basis of the resulting structures. Specifically, 12 precise one-bond ^{13}C-^{1}H RDCs for TAR*$_{GA}$ were divided into four groups, and each of the four groups was used for cross-validation in succession. A similar procedure was performed independently for TAR.

Thermal denaturation of experiments was monitored on a Uvikon XL (Bio-Tek Instruments) spectrophotometer interfaced with a Peltier effect device that controls temperature within $\pm$0.1°C. The two RNAs were mixed at room temperature and allowed to interact for 20 min before they were cooled to 4°C. The experiment started after 20 additional minutes at this temperature. Denaturation of the samples was achieved by increasing the temperature at a rate of 0.4°C per min from 4 to 90°C and was followed at 260 nm.

ACKNOWLEDGMENTS. We thank Drs. Brutscher, Simorre, and Sounier for helpful discussions. This work was supported by the Commissariat à l'Energie Atomique, the Centre National de la Recherche Scientifique (CNRS), University of Grenoble, University of Ottawa Faculty of Science, Canadian Institutes of Health Research (CIHR), and Human Frontier Science Program Organization (HFSP). H.V.M. and M.D. acknowledge the receipt of fellowships from the Agence Nationale de Recherche sur le Sida et les Hépatites Virales and CNRS. J.B. is a recipient of a Career Development Award from HFSP (Grant CDA0029/2004-C); J.B and D.L.B. acknowledge funding from the France-Canada Research Fund, the CIHR-CNRS program, and the Jacques Cartier Center.

1. Muesing MA, Smith DH, Capon DJ (1987) Regulation of messenger-RNA accumulation by a human-immunodeficiency-virus transactivator protein. *Cell* 48:691–701.
2. Weeks KM, Ampe C, Schultz SC, Steitz TA, Crothers DM (1990) Fragments of the HIV-1 TAT protein specifically bind TAR RNA. *Science* 249:1281–1285.
3. Kao SY, Calman AF, Luciw PA, Peterlin BM (1987) Anti-termination of transcription within the long terminal repeat of HIV-1 by TAT gene-product. *Nature* 330:489–493.
4. Feng S, Holland EC (1988) HIV-1 TAT trans-activation requires the loop sequence within TAR. *Nature* 334:165–167.
5. Wei P, Garber ME, Fang SM, Fischer WH, Jones KA (1998) A novel CDK9-associated C-type cyclin interacts directly with HIV-1 Tat and mediates its high-affinity, loop-specific binding to TAR RNA. *Cell* 92:451–462.
6. Ivanov D, *et al.* (1999) Cyclin T1 domains involved in complex formation with Tat and TAR RNA are critical for tat-activation. *J Mol Biol* 288:41–56.
7. Bieniasz PD, Grdina TA, Bogerd HP, Cullen BR (1999) Recruitment of cyclin T1/P-TEFb to an HIV type I long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. *Proc Natl Acad Sci USA* 96:7791–7796.
8. Hamy F, *et al.* (1997) An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. *Proc Natl Acad Sci USA* 94:3548–3553.
9. Dassonneville L, Hamy F, Colson P, Houssier C, Bailly C (1997) Binding of Hoechst 33258 to the TAR RNA of HIV-1. Recognition of a pyrimidine bulge-dependent structure. *Nucleic Acids Res* 25:4487–4492.
10. Sullenger BA, Gallardo HF, Ungers GE, Gilboa E (1990) Overexpression of TAR sequences renders cells resistant to Human-Immunodeficiency-Virus replication. *Cell* 63:601–608.
11. Boiziau C, Dausse E, Yurchenko L, Toulme JJ (1999) DNA aptamers selected against the HIV-1 trans-activation-responsive RNA element form RNA-DNA kissing complexes. *J Biol Chem* 274:12730–12737.
12. Ecker DJ, *et al.* (1992) Pseudo half-knot formation with RNA. *Science* 257:958–961.
13. Chang KY, Tinoco I (1994) Characterization of a kissing hairpin complex derived from the human-immunodeficiency-virus genome. *Proc Natl Acad Sci USA* 91:8705–8709.
14. Chang KY, Tinoco I (1997) The structure of an RNA ''kissing'' hairpin complex of the HIV TAR hairpin loop and its complement. *J Mol Biol* 269:52–66.
15. Duconge F, Toulme JJ (1999) *In vitro* selection identifies key determinants for loop-loop interactions: RNA aptamers selective for the TAR RNA element of HIV-1. *RNA* 5:1605–1614.
16. Duconge F, Di Primo C, Toulme JJ (2000) Is a closing ''GA pair'' a rule for stable loop-loop RNA complexes? *J Biol Chem* 275:21287–21294.
17. Beaurain F, Di Primo C, Toulme JJ, Laguerre M (2003) Molecular dynamics reveals the stabilizing role of loop-closing residues in kissing interactions: comparison between TAR-TAR* and TAR-aptamer. *Nucleic Acids Res* 31:4275–4284.
18. Darfeuille F, *et al.* (2002) Loop-loop interaction of HIV-1 TAR RNA with N3′-> P5′ deoxyphosphoramidate aptamers inhibits in vitro Tat-mediated transcription. *Proc Natl Acad Sci USA* 99:9709–9714.
19. Kolb G, *et al.* (2005) Hexitol nucleic acid-containing aptamers are efficient ligands of HIV-1 TAR RNA. *Biochemistry* 44:2926–2933.
20. Lebars I, Richard T, Di Primo C, Toulme JJ (2007) NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer. *Nucleic Acids Res* 35:6103–6114.
21. Darfeuille F, *et al.* (2006) Aptamers targeted to an RNA hairpin show improved specificity compared to that of complementary oligonucleotides. *Biochemistry* 45:12076–12082.
22. Kim CH, Tinoco I (2000) A retroviral RNA kissing complex containing only two G center dot C base pairs. *Proc Natl Acad Sci USA* 97:9396–9401.
23. Lee AJ, Crothers DM (1998) The solution structure of an RNA loop-loop complex: the ColE1 inverted loop sequence. *Structure* 6:993–1005.
24. Dardel F, Marquet R, Ehresmann C, Ehresmann B, Blanquet S (1998) Solution studies of the dimerization initiation site of HIV-1 genomic RNA. *Nucleic Acids Res* 26:3567–3571.
25. Mujeeb A, Clever JL, Billeci TM, James TL, Parslow TG (1998) Structure of the dimer initiation complex of HIV-1 genomic RNA. *Nat Struct Biol* 5:432–436.
26. Baba S, *et al.* (2005) Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *J Biochem* 138:583–592.
27. Gregorian RS, Crothers DM (1995) Determinants of RNA hairpin loop-loop complex stability. *J Mol Biol* 248:968–984.
28. Dingley AJ, Grzesiek S (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide (2)J(NN) couplings. *J Am Chem Soc* 120:8293–8297.
29. Wu Z, Delaglio F, Tjandra N, Zhurkin VB, Bax A (2003) Overall structure and sugar dynamics of a DNA dodecamer from homo and heteronuclear dipolar couplings and P-31 chemical shift anisotropy. *J Biomol Nmr* 26:297–315.
30. Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J Am Chem Soc* 120:6836–6837.
31. Bax A, Grishaev A (2005) Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr Opin Struct Biol* 15:563–570.
32. Ying J, Bax A (2006) 2′-hydroxyl proton positions in helical RNA from simultaneously measured heteronuclear scalar couplings and NOEs. *J Am Chem Soc* 128:8372–8373.
33. Hennig M, Fohrer J, Carlomagno T (2005) Assignment and NOE analysis of 2′-hydroxyl protons in RNA: implications for stabilization of RNA A-form duplexes. *J Am Chem Soc* 127:2028–2029.
34. Scarabino D, Crisari A, Lorenzini S, Williams K, Tocchini-Valentini GP (1999) tRNA prefers to kiss. *EMBO J* 18:4571–4578.
35. Marino JP, Gregorian RS, Csankovszki G, Crothers DM (1995) Bent helix formation between RNA hairpins with complementary loops. *Science* 268:1448–1454.
36. Lorenz C, Piganeau N, Schroeder R (2006) Stabilities of HIV-1 DIS type RNA loop-loop interactions *in vitro* and *in vivo*. *Nucleic Acids Res* 34:334–342.
37. Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278:1111–1114.
38. Tjandra N, Garrett DS, Gronenborn AM, Bax A, Clore GM (1997) Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nat Struct Biol* 4:443–449.
39. Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* 5:1065–1074.
40. Bayer P, Varani L, Varani G (1999) Refinement of the structure of protein-RNA complexes by residual dipolar coupling analysis. *J Biomol Nmr* 14:149–155.
41. Vermeulen A, Zhou HJ, Pardi A (2000) Determining DNA global structure and DNA bending by application of NMR residual dipolar couplings. *J Am Chem Soc* 122:9638–9647.
42. Mollova ET, Hansen MR, Pardi A (2000) Global structure of RNA determined with residual dipolar couplings. *J Am Chem Soc* 122:11561–11562.
43. Warren JJ, Moore PB (2001) Application of dipolar coupling data to the refinement of the solution structure of the Sarcin-Ricin loop RNA. *J Biomol Nmr* 20:311–323.
44. Sibille N, Pardi A, Simorre JP, Blackledge M (2001) Refinement of local and long-range structural order in theophylline-binding RNA using C-13-H-1 residual dipolar couplings and restrained molecular dynamics. *J Am Chem Soc* 123:12135–12146.
45. Padrta P, Stefl R, Kralik L, Zidek L, Sklenar V (2002) Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings. *J Biomol Nmr* 24:1–14.
46. Wu Z, Maderia M, Barchi JJ, Marquez VE, Bax A (2005) Changes in DNA bending induced by restricting nucleotide ring pucker studied by weak alignment NMR spectroscopy. *Proc Natl Acad Sci USA* 102:24–28.
47. Westhof E, Dumas P, Moras D (1988) Restrained refinement of 2 crytalline forms of yeast aspartic-acid and phenylalanine transfer-RNA crystals. *Acta Crystallogr A* 44:112–123.
48. Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 angstrom resolution. *Science* 289:905–920.
49. Ennifar E, Walter P, Ehresmann B, Ehresmann C, Dumas P (2001) Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol* 8:1064–1068.
50. Brutscher B, Boisbouvier J, Pardi A, Marion D, Simorre JP (1998) Improved sensitivity and resolution in H-1-C-13 NMR experiments of RNA. *J Am Chem Soc* 120:11845–11851.
51. Miclet E, Boisbouvier J, Bax A (2005) Measurement of eight scalar and dipolar couplings for methine-methylene pairs in proteins and nucleic acids. *J Biomol Nmr* 31:201–216.
52. Boisbouvier J, Bryce DL, O'Neil-Cabello E, Nikonowicz EP, Bax A (2004) Resolution-optimized NMR measurement of D-1(CH), D-1(CH) and D-2(CH) residual dipolar couplings in nucleic acid bases. *J Biomol Nmr* 30:287–301.
53. Clore GM, Gronenborn AM, Bax A (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J Magn Reson* 133:216–221.
54. McCallum SA, Pardi A (2003) Refined solution structure of the iron-responsive element RNA using residual dipolar couplings. *J Mol Biol* 326:1037–1050.
55. Bryce DL, Bax A (2004) Application of correlated residual dipolar couplings to the determination of the molecular alignment tensor magnitude of oriented proteins and nucleic acids *J Biomol NMR* 28:273–287.
56. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Mag Reson* 160:65–73.
57. Bryce DL, Grishaev A, Bax A (2005) Measurement of ribose carbon chemical shift tensors for A-form RNA by liquid crystal NMR spectroscopy. *J Am Chem Soc* 127:7387–7396.

J|A|C|S
COMMUNICATIONS

Published on Web 12/23/2006

High-Accuracy Distance Measurement between Remote Methyls in Specifically Protonated Proteins

Remy Sounier,§ Laurence Blanchard,§ Zhengrong Wu,‡ and Jérôme Boisbouvier*,§

IBS, Institut de Biologie Structurale Jean-Pierre Ebel, 41 rue Jules Horowitz, F-38027 Grenoble, CEA, CNRS, Université Joseph Fourier, and Biochemistry Department, The Ohio State University, Columbus, Ohio 43210

Received October 10, 2006; E-mail: jerome.boisbouvier@ibs.fr

NMR structure determination of biological macromolecules in solution has relied on the use of semiquantitative NOE distance restraints and dihedral angle restraints. The introduction of orientational information derived from residual dipolar couplings has allowed a significant improvement in the precision of structural models of globular proteins.[1] Nevertheless the lack of long-range distances remains a major limiting factor for the accuracy of structure determination of modular and elongated biomolecules as well as for biomolecular complexes. The use of perdeuterated samples and improvement in spectrometer hardware have allowed in the last years the measurement of distance restraints between protons separated by up to 8 Å.[2] Detection of ^{1}H−^{1}H NOEs between more remote protons has been hampered by the small magnitude of such effects due to the r^{-6} dependence. Here we demonstrate that observation of ^{1}H−^{1}H NOEs over distances of up to 12 Å is feasible using a highly deuterated protein in which protons have been selectively incorporated into isoleucine δ_1-methyl sites. By reducing the proton density to 2−3% compared to fully protonated samples, the measurement of accurate long-range distance restraints becomes possible. The detection of NOEs between such remote protons benefits from three major effects. First, because of the proton multiplicity, the choice of methyls for selective protonation enhances signal transferred between two sites by a factor of 9, in the limit of long intermethyl distances. Second, in highly deuterated samples, methyl protons are characterized by favorable transverse and longitudinal relaxation rates providing high resolution and signal-to-noise ratio of the NMR spectra.[3] And last, by reducing the proton density, spin diffusion is considerably reduced, a prerequisite for accurate distance measurements. Long mixing times can then be used to transfer substantial amounts of magnetization through very weak NOEs.

The efficiency of the method for observing the long distance NOE is demonstrated on two proteins for which high resolution solution structures are available: ubiquitin and γS-crystallin, a 19 kDa protein comprising two domains.[4] Highly deuterated samples with selective incorporation of protonated methyls on Ile-δ_1 methyl sites of ubiquitin and γS-crystallin, have been prepared using 2-keto-3,3-d_2-butyrate as the unique proton source during protein overexpression in *E. coli* as previously proposed by Gardner and Kay.[5] A protocol has been implemented to incorporate selectively protonated methyls on Ile-δ_1 sites with an efficiency larger than 98% while the average deuteration level at all other sites in the overexpressed proteins remains larger than 98% (Supporting Information S1). As all the ^{1}H resonances of Ile-δ_1 methyls (7 for ubiquitin and 9 for γS-crystallin) are well resolved, all the NOESY spectra have been recorded using ^{1}H edited 2D-experiments (Supporting Information S2).

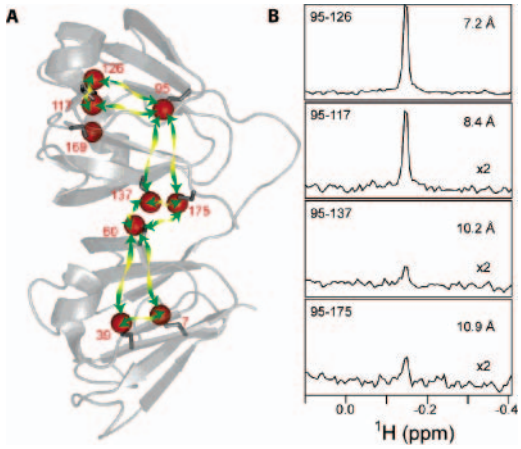

Figure 1. (A) Ile-δ_1 intermethyl NOEs observed in γS-crystallin are indicated by green arrows. Ile-169 does not give rise to detectable intermethyl NOEs owing to extensive line broadening. (B) Intermethyl NOEs detected for Ile-95-δ_1. F_1 cross-sections at δ_1 methyl frequency of Ile 126, 117, 175, and 137 are displayed. The 800 MHz NOESY spectrum has been acquired in 14 h on a 1 mM sample of U-[^{2}H],[CH_3]-Ile-δ_1 γS-crystallin.

The optimal NOE mixing time for transfer between protonated Ile-δ_1 methyls was determined from the build-up of cross-peak intensities in a series of short 2D NOESY spectra. Maximum cross-peak intensities were obtained for mixing times of 1.25 and 0.8 s for ubiquitin and γS-crystallin, respectively (data not shown). 2D NOESY spectra with optimized mixing times have been acquired at 25 °C in 14 h for each protein on a 800 MHz spectrometer equipped with a cryogenic probe. A total of ten pairs of intermethyl cross-peaks for ubiquitin and eleven for γS-crystallin has been detected. All the intermethyl NOEs observed for γS-crystallin are displayed in Figure 1A. 1D traces showing the four observed NOEs involving methyl δ_1 of Ile-95 are displayed in Figure 1B. Comparison of the distance distribution of Ile-δ_1 methyls predicted from the ubiquitin and γS-crystallin structures[4] with cross-peaks detected in the NOESY spectra, revealed that all observed NOEs correspond to intermethyl distances ranging from 4 to 12 Å (Figure 2). All methyl pairs distant by less than 10.5 Å give rise to observable NOEs. For intermethyl distances ranging from 10.5 to 12.5 Å, more than 60% of the expected NOEs are still detected on both sides of the NOESY diagonal with an average signal-to-noise ratio of 5:1.

To evaluate the accuracy of the method, a full-relaxation-matrix analysis was employed to extract distance restraints from NOESY spectra.[6] Given the high resolution of the spectra and the simplicity of the spin system in Ile-δ_1 methyl specific protonated samples, the relaxation matrix σ can be directly determined from diagonal- and cross-peak volumes in NOESY spectra (Supporting Information S3). For the simplified case of fast rotating methyl groups in an isotropically tumbling macromolecule,[7] the average distance be-

§ Institut de Biologie Structurale
‡ The Ohio State University.

10.1021/ja067260m CCC: $37.00

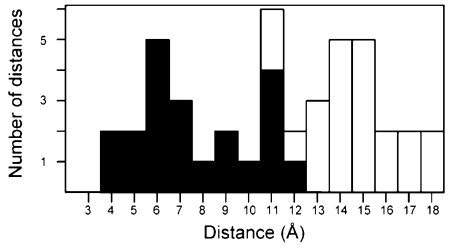

Figure 2. Distribution of Ile-δ_1 intermethyl distances *versus* average interproton distance calculated from structures of ubiquitin and γS-crystallin (except for Ile-169).[4] The experimentally detected NOEs are displayed by black bars.

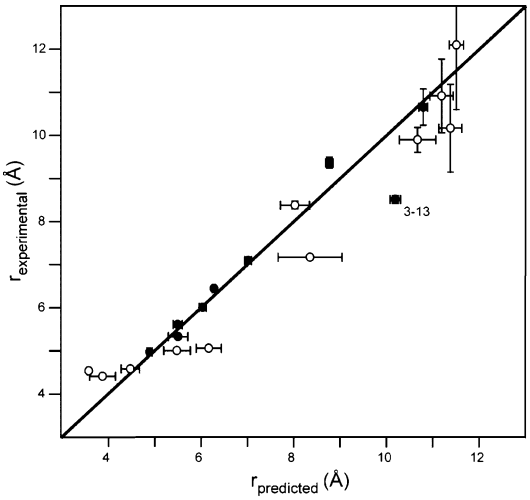

Figure 3. Comparison of the experimental distances calculated from NOE peak volumes using full-relaxation-matrix analysis with distances predicted from ubiquitin (filled circles) and γS-crystallin (open circles) structures.[4] The standard deviation of the distances predicted from the structure ensemble are represented by a horizontal error bar, with $r = (\langle r^{-3}\rangle)^{-1/3}$. A thousand sets of noise-corrupted NOESY peak volumes have been generated (using experimental noise) and used to extract the standard deviation on $r_{\text{experimental}}$ (displayed as vertical error bars). The largest discrepancy occurs between residues 3 and 13 of ubiquitin, which have been shown to populate different side-chain rotamers.[11]

tween methyls i and j can be estimated directly from the relaxation matrix elements σ_{ij}. Comparison of experimental distances with distances extracted from the ubiquitin and γS-crystallin structures shows a very good agreement (Figure 3), with a Pearson's correlation coefficient $R_P = 0.96$. The distances exceeding 5 Å can be determined with an average accuracy of 9% without the use of any internal distance reference. Despite the simplicity of the dynamic model, the NOE-derived distances remain accurate. Because of the $\langle r^{-3}\rangle^2$ dependence of the NOE, the extracted distances are relatively insensitive to error on τ_C. In the slow tumbling limit, an error of 30% in τ_C will result in an error of 5% in the experimentally determined distances.

Full-relaxation-matrix analysis also allows us to evaluate the contribution of spin diffusion to the NOE cross-peaks observed between Ile-δ_1 methyls. Although relatively long mixing times were used (0.8 s for a 17 ns tumbling molecule), spin diffusion only accounts for 12% of the cross-peak volumes, for methyls separated by more than 10 Å in U-[^{2}H], [CH_3]-Ile-δ_1-γS-crystallin (proton density 2%). Protonation of the methyls of Val and Leu residues (proton density 9%) or of all methyl sites in γS-crystallin (proton density 15%) would increase the contribution of spin diffusion to respectively 76% and 92% of the NOEs observed between methyls distant by more than 10 Å (Supporting Information S3), making accurate distance determination difficult. The method proposed here for measurement of accurate long-distance restraints between methyls relies on the reduction of spin diffusion effects by the protonation of only few specific sites in a protein. This specific protonation strategy can be generalized to other methyl-containing residues. For example, specific protonation of only one of the two methyls of Leu and Val residues can be achieved using 2-keto-3-methyl-d_3-3-d_1-butyrate as a precursor during *E. coli* overexpression.[8] The selective protonation of other methyl-containing residues can be obtained alternatively using a cell-free approach.[9] The efficiency of the method was demonstrated on the small model protein ubiquitin as well as on the 19 kDa protein γS-crystallin. The proposed approach can be transposed to larger proteins using ^{13}C,^{1}H-methyl specific labeling. Simulation indicates that in a 45 ns tumbling methyl-specific protonated protein, NOE cross-peaks will still be detectable between methyls separated by up to 10 Å using the recently proposed 4D methyl-TROSY−NOESY experiment.[10] In such large proteins, spin diffusion will contribute for less than 10% of the NOE cross-peak volumes, provided that the total proton density remains lower than 2−3%. The possibility to measure accurate translational restraints between distant methyls offers new opportunities for the study of large proteins and protein complexes by solution NMR.

Acknowledgment. This work was supported by the Commissariat à l'Energie Atomique, the Centre National de la Recherche Scientifique, and Human Frontier Science Program Organization. The authors thank Drs. Simorre, Schanda, Devany, Bersch, Bax, and Brutscher for helpful discussions, C. Jaroniec for acquisition of the initial NOESY data set on γS-crystallin, and I. Ayala and K. Treche for help in ubiquitin preparation.

Supporting Information Available: Protocol for specific protonation of Ile-δ_1 methyls, experimental details for detection of long-range methyl−methyl NOEs, and equations used to calculate experimental distances from NOESY spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111−1114.
(2) (a) Mueller, G. A.; Choy, W. Y.; Daiwen, Y.; Forman-Kay, J. D.; Venters, R. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *300*, 197−212. (b) Wu, Z.; Bax, A. *J. Am. Chem. Soc.* **2002**, *124*, 9672−9673. (c) Koharudin, L. M. I.; Bonvin, A. M. J. J.; Kaptein, R.; Boelens, R. *J. Magn. Reson.* **2003**, *163*, 228−235. (d) Meier, S.; Häussinger, D.; Jensen, P.; Rogowski, M.; Grzesiek, S. *J. Am. Chem. Soc.* **2003**, *125*, 44−45.
(3) Tugarinov, V.; Hwang, P. M.; Ollerenshaw, J. E.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 10420−10428.
(4) (a) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837. (b) Grishaev, A.; Wu, J.; Trewhella, J.; Bax, A. *J. Am. Chem. Soc.* **2005**, *127*, 16621−16628.
(5) Gardner, K. H.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119*, 7599−7600.
(6) Macura, S.; Ernst, R. R. *Mol. Phys.* **1980**, *41*, 95−107.
(7) Tropp, J. *J. Chem. Phys.* **1980**, *72*, 6035−6043.
(8) Tugarinov, V.; Kay, L. E. *J. Biomol. NMR* **2004**, *28*, 165−172.
(9) (a) Kigawa, T.; Muto, Y.; Yokoyama, S. *J. Biomol. NMR* **1995**, *6*, 129−134. (b) Torizawa, T.; Shimizu, M.; Taoka, M.; Miyano, H.; Kainosho, M. *J. Biomol. NMR* **2004**, *30*, 311−325. (c) Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Ono, A. M.; Güntert, P. *Nature* **2006**, *440*, 52−57.
(10) (a) Tugarinov, V.; Ollerenshaw, J. E.; Kay, L. E. *J. Am. Chem. Soc.* **2005**, *127*, 8214−8225. (b) Frueh, D.; Vosburg, D. A.; Walsh, C. T.; Wagner, G. *J. Biomol. NMR* **2006**, *34*, 3−172.
(11) Chou, J. J.; Case, D. A.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 8959−8966.

JA067260M