Detection of nucleic acids and other low abundance components in native bone and osteosarcoma extracellular matrix by isotope enrichment and DNP-enhanced NMR†

Ieva Goldberg,a Rui Li,a Wing Ying Chow,a,b David G. Reid,a Ulyana Bashtanova,a Rakesh Rajan,a Anna Puszkarska,a Hartmut Oschkinat* and Melinda J. Duer*a

Sensitivity enhancement by isotope enrichment and DNP NMR enables detection of minor but biologically relevant species in native intact bone, including nucleic acids, choline from phospholipid headgroups, and histidinyl and hydroxylysyl groups. Labelled matrix from the aggressive osteosarcoma K7M2 cell line confirms the assignments of nucleic acid signals arising from purine, pyrimidine, ribose, and deoxyribose species. Detection of these species is an important and necessary step in elucidating the atomic level structural basis of their functions in intact tissue.

The extracellular matrix (ECM) is a complex, heterogeneous, and active, environment, playing structural roles, but also strongly influencing cell behaviour in both homeostasis and pathology. The difficulty of understanding the composition and regulation of the ECM parallels the situation within cells, but with the added complication that many molecular components are insoluble. Solid-state NMR (ssNMR) can contribute to a better understanding of the ECM through direct detection of sparse species, made possible by a strategy of isotopic enrichment of native biomaterials1 combined with rapidly evolving dynamic nuclear polarization (DNP) enhancement techniques.2 Briefly these are based on the transfer of the much greater electron spin polarization, excited by a microwave laser at the electron resonance frequency of an organic free radical dopant, to the (NMR active) nuclei of the molecules of interest, at low temperatures to maximize polarization. The conversion of electron to nuclear polarization results in NMR sensitivity enhancements of up to several orders of magnitude. Recently we have reported direct observation of important low abundance collagen post-translational modifications (PTMs) in intact native, and model, ECM.3 While we focussed on species derived from one specific amino acid (Lys) component of the ECM, it is also of great potential importance to investigate whether other sparse ECM species become observable with DNP enhancement. Non-protein minor species, besides low abundance amino acids, are of particular interest since they represent a realm in which NMR can provide atomic length scale molecular structural information complementing other current biophysical techniques. Incorporating non-protein molecules into the overall structural biology picture will broaden conceptions of which ECM components to investigate in further detail towards an integrated model of the biological and material properties of tissues.

Here, we report DNP-enhanced NMR on mouse bone4 enriched in a wide range of 13C, 15N-amino acids. Where necessary to distinguish between low abundance protein, and non-protein, signals, we confirmed amino acid type assignments by comparison with ECM produced by cells cultured on media supplemented with one or a few specific amino acids, using methods already described.4,5 Our previous analysis of bone focussed on the relatively abundant species observable with conventional ssNMR. The enhancement provided by DNP now reveals non-protein biomolecules biosynthesized from isotopically enriched amino acids and/or sugars, and their incorporation into bone. The preparation of stable isotope enriched biomaterials, and our DNP NMR methodology, is described in detail in the ESI,† Materials and methods sections.

We begin our investigation on bone with a suite of ssNMR characterizations under DNP signal enhancement conditions. A full 1D 15N DNP NMR spectrum is shown in ESI Fig. S1.† Apart from the predictable major signals (from backbone and side-chain amides, and abundant Lys and Arg) the sensitivity enhancement of enrichment-DNP reveals a number of more minor resonances (Fig. 1A) from lower abundance species; the shifts of many of those at high spectral frequency coincide
Fig. 1 (A) High frequency region of the $^{15}$N DNP-NMR spectrum showing assignments of resolved base signals of nucleic acids, and ring His signals; (B) $^{13}$C spectrum of His* FSOB ECM highlighting correspondence with the signal from the imidazole form of the His ring; (C) DARR assisted $^{15}$N-$^{13}$C correlation spectrum of labelled bone, transmitting magnetization along the Hyl (and Lys) sidechain to emphasize correlations between Hyl (as well as the intense Lys) N$_1$ and C$_5$. Also highlighted is the signal from the $^{(CH_3)}_3$N$^+$=CH$_2$=CH$_2$=O$^-$ correlation from phosphatidylcholine headgroups. Atom numbering of the nucleic acid purine (top) and pyrimidine skeletons is also shown.

closely with the common (deoxy)nucleotide base ring nitrogens (Biological Magnetic Resonance Database, BMRB, as summarized in ESI Schemes S1† (DNA) and S2 (RNA)). Signals from the nitrogen atoms of protonated His imidazoloid (ca. 180 ppm) and unprotonated imidazole (ca. 250 ppm) ring forms are also observed. The His assignments are reinforced by $^{15}$N data from model osteoblast (bone precursor) ECM selectively enriched in [U-$^{13}$C$_6$, $^{15}$N$_3$]-His (His*) (Fig. 1B; see ESI† for preparation and comparison of $^{15}$N-$^{13}$C correlation spectra, Fig. S2†). The low frequency region of the bone DARR-assisted $^{15}$N-$^{13}$C correlation spectrum (Fig. 1C) displays a clear cross peak (65 ppm $^{13}$C, 52 ppm $^{15}$N) which we ascribe to the $^{(CH_3)}_3$N$^+$=CH$_2$=CH$_2$=O$^-$ correlation in the head groups of phosphatidylcholine (PC) lipids ($^{13}$C assignments based on Kroon et al.; see ESI Fig. S3† for overlays with model compound spectra). The PC ethanolamine substructure is biosynthesized from Glc and Glc* via Ser,2 will necessarily be concurrently labelled. (Deoxy)riboses are biosynthesized from Glc* and so all (deoxy)sugar rings will be concurrently $^{13}$C enriched. Fig. 2B details results of a through-bond (J-transmitted) INADEQUATE,$^{13}$ which clearly shows the predicted purine C$_4$-C$_5$-N$_7$ substructure (from intact Gly*), and to some extent pyrimidine C$_4$-C$_5$-C$_6$ from Glc* via Ser, will be concurrently labelled. Finally Fig. 1C also shows the distinctive Hyl N$_2$-C$_6$ fingerprint correlations; the close correspondence with (U-$^{13}$C$_6$, $^{15}$N$_3$)-Lys enriched ECM from vascular smooth muscle cells is shown in ESI Fig. S4, proving that our labelling – DNP NMR combination readily detects the Hyl PTM even in this complex organic–inorganic composite native biomaterial, as well as in simpler native (skin) and model ECMS. Chemical shifts are summarized in Table 1.

Detection of His, lipid components, and Hyl, in native biomaterials is not unexpected. In contrast nucleic acids in bone have been little discussed, although this tissue is a favourable source of forensic genetic material. Given this underrepresentation of bone in nucleic acid literature, we turned to ECM from the aggressively metastasizing K7M2 osteosarcoma cell line to substantiate these signals. A $^{13}$C DNP NMR spectrum of K7M2 ECM supplemented with (U-$^{13}$C$_6$, $^{15}$N$_3$)-Gly (Gly*) and (U-$^{13}$C$_6$)-glucose (Glc*) is shown in Fig. 2A; it shows a number of clear signals with shifts consistent with the common (deoxy)nucleotide bases (ESI Schemes S1 and S2†). Abundant nucleic acid in the K7M2 ECM is confirmed by conventional in situ propidium iodide staining and confocal fluorescence microscopy (ESI Fig. S5†). The biosynthetic routes to the bases (ESI Scheme S3†) dictates that all the base carbon atoms will be at least partially $^{13}$C enriched by Gly* and Glc*, although only the purine C$_4$-C$_5$-N$_7$ substructure (from intact Gly*), and to some extent pyrimidine C$_4$-C$_5$-C$_6$ from Glc* via Ser, will be concurrently labelled. (Deoxy)riboses are biosynthesized from Glc* and so all (deoxy)sugar rings will be concurrently $^{13}$C enriched. Fig. 2B details results of a through-bond (J-transmitted) INADEQUATE, which clearly shows the predicted purine C$_4$-C$_5$-N$_7$ substructure (from intact Gly*), and to some extent pyrimidine C$_4$-C$_5$-C$_6$ from Glc* via Ser, will be concurrently labelled. Finally Fig. 1C also shows the distinctive Hyl N$_2$-C$_6$ fingerprint correlations; the close correspondence with (U-$^{13}$C$_6$, $^{15}$N$_3$)-Lys enriched ECM from vascular smooth muscle cells is shown in ESI Fig. S4, proving that our labelling – DNP NMR combination readily detects the Hyl PTM even in this complex organic–inorganic composite native biomaterial, as well as in simpler native (skin) and model ECMS. Chemical shifts are summarized in Table 1.

Table 1 Chemical shifts of resolved signals from low abundance species detected in $^{13}$C, $^{15}$N enriched bone by DNP NMR

| Amino acids/PTMs | $^{15}$N ppm | $^{13}$C ppm |
|------------------|-------------|-------------|
| Hyl C$_6$        | 69          | 44          |
| C$_2$            | 129         | 117         |
| N$_2$            | 135         | 180         |
| N$_3$            | 250         |
| His C$_4$        | 224         |
| C$_{22}$         | 148         |
| N$_2$            | 235         |
| N$_3$            | 197         |
| N$_3$/$^*$       | 162         |
| Lipids           |             |             |
| PC $^{N([CH_3])_3}$ | 52        |
| $^{CH_2}=N^+$    | 65          |
| $^{15}$N ppm     |             |             |
| $^{13}$C ppm     |             |             |

$^a$ Protonated imidazoloid form. $^b$ Unprotonated imidazoloid form. $^c$ Single unresolved signal.
NMR shows a clear signal from purine N7’s, corresponding closely to that of the putative purine N7’s in bone (Fig. 2E). Also a 13C(15N) TEDOR\(^{15}\) (Fig. 2F) mapping short carbon-nitrogen distances, shows the expected purine N7–C5 and weaker N7–C4 correlations, but also purine N7–C6 cross peaks (the latter atom acquired from Glc\(^*\) via N-formyl tetrahydrofolate). All DNA and RNA shifts measured in K7M2 ECM are summarized in Table 2.

There are large number of catalogued nucleic acid shifts in the BMRB but as yet no robust equivalent of the chemical shift database shift values (ESI Schemes S1 and S2\(^\text{†}\)) are labelled; (B) high frequency region of INADEQUATE J-transmitted SQ-DQ correlation spectrum (265 ppm \(^{13}\)C DQ sweep width to avoid aliasing; \(\xi\) + \(\eta\) – correlations between lipid olefinic carbons); (C) low frequency region of the same spectrum tracing out correlations within ribose (Rib), and deoxyribose (dRib), rings; (D) high frequency region of the \(^{13}\)C–\(^{13}\)C DARR correlation spectrum base; (E) \(^{15}\)N DNP NMR of Gly\(^*\)Glc\(^*\) K7M2 ECM compared with that of labelled bone, supporting the assignment of the 235 ppm signal as purine N7; (F) \(^{13}\)C\(^{15}\)N TEDOR spectrum showing the expected purine N7–C5 and N7–C6, as well as the putative purine ring N7–C8, through-space correlations (15N sweep width 438 ppm to cover full 15N range).

**Table 2 Chemical shifts of resolved (deoxy)nucleotide signals in the K7M2 rat osteosarcoma cell line**

|       | C2  | C4  | C5  | C6  | C8  | N7  |
|-------|-----|-----|-----|-----|-----|-----|
| Ade   | 153 | 147 | 116 | 133 | 139 | 236 |
| Gua   | 153 | 150 | 114 | 167 | 136 | 239 |
| Cyt   | 153 | 165 | 96  | 140 |     |     |
| Thy   | 153 | 167 | 110 | 137 |     |     |
| Ura   | 105 |     |     |     | 145 |     |

|       | C1’ | C2’ | C3’ | C4’ | C5’ |
|-------|-----|-----|-----|-----|-----|
| dRib  | 82  | 37  | 78  | 64  | 61  |
| Rib   | 90  | 72  | 69  | 79  | 61  |

\(\xi\) Signals observed in 1D spectrum although no SQ-DQ or DARR correlations observable; assignments are inferred from average shifts reported in BMRB. \(\eta\) Assignments to Ade and Gua are based on rank order of BMRB average shifts, and may be reversed. \(\text{†}\) From J-transmitted INADEQUATE SQ-DQ correlation. \(\text{‡}\) From \(^{13}\)C–\(^{13}\)C DARR correlation. \(\text{§}\) From \(^{15}\)N–\(^{13}\)C TEDOR.

Fig. 2 13C DNP NMR spectra of Gly\(^*\), Glc\(^*\) supplemented K7M2 ECM. (A) \(^{13}\)C CP-MAS; nucleotide base signals assigned by 2D experiments and database shift values (ESI Schemes S1 and S2\(^\text{†}\)) are labelled; (B) high frequency region of INADEQUATE J-transmitted SQ-DQ correlation spectrum (265 ppm \(^{13}\)C DQ sweep width to avoid aliasing; \(\xi\) + \(\eta\) – correlations between lipid olefinic carbons); (C) low frequency region of the same spectrum tracing out correlations within ribose (Rib), and deoxyribose (dRib), rings; (D) high frequency region of the \(^{13}\)C–\(^{13}\)C DARR correlation spectrum base; (E) \(^{15}\)N DNP NMR of Gly\(^*\)Glc\(^*\) K7M2 ECM compared with that of labelled bone, supporting the assignment of the 235 ppm signal as purine N7; (F) \(^{13}\)C\(^{15}\)N TEDOR spectrum showing the expected purine N7–C5 and N7–C6, as well as the putative purine ring N7–C8, through-space correlations (15N sweep width 438 ppm to cover full 15N range).
calcium phosphate precipitation in normal and pathological biomineralization. The molecular mechanisms underlying the initial bioprecipitation events and processes remain to be elucidated, in particular the role of matrix vesicle phospholipids. Direct observation of the PC headgroup by NMR in healthy bone is a significant step in atomic level understanding of the relevant intermolecular and interionic interactions. In addition to matrix vesicles, of 40–100 nm sized exosomes are the basis of a pathway for cancer metastasis and therapy resistance through delivery of proteins, lipids, RNA and DNA to remote cells. The cargo carried by these vesicles may play unexpected roles if released directly in the extracellular space and allowed to interact with other components of the ECM. DNP-enhanced NMR is potential non-targeted method of better understanding the role played by these extracellular vesicles.

Hyl and its subsequent modifications are essential to the structure, properties, and functions, of collagen. The type I collagen heterotrimer contains 12 His residues. Both of those in each z1 (I) chain are two residues from those Lys residues which are sequence specifically β-hydroxylated to Hyl, in a strongly conserved Lys-Gly-His-Arg tetrad. These His residues have been proposed to play a part in targeting lysyl hydroxylases and thereby the formation of the specific Lys and Hyl crosslinks essential to maintaining collagen and tissue structure. Some of the same His residues are not only hydroxylated to Hyl, but of these some are subsequently also enzymatically glycosylated to [Glc]GalHyl, an essential event in directing the as-yet poorly understood process of collagen fibril assembly. It has recently been proposed that these self same residues are also preferentially glycated in non-enzymatic, usually harmful, reactions with metabolic sugars. Finally there is evidence for a trivalent crosslink, histidinohydroxylysinosonorleucine, between these same Lys/Hyl and His residues. Thus His and Lys/Hyl residues co-exist in crosslinking and glycosylation/glycation “hotspots” which may be key to better understanding the spatial relationships underlying orderly, and pathological, collagen assembly and structure. Detecting His, Lys and Hyl simultaneously, as enabled by isotopic enrichment and DNP-enhanced NMR, in native and model ECM, opens fertile new possibilities for studying these processes.

This communication demonstrates again the ability of ssNMR, combined with signal enhancement by DNP and targeted isotope labelling, to detect low abundance components in intact tissue that may have surprising biological roles, thus complementing other biophysical and molecular biology techniques in exploring diverse areas of matrix biology.

Conflicts of interest

There are no conflicts of interest. Animal procedures were unregulated in terms of the U. K. Animals (Scientific Procedures) Act 1986, but nevertheless were reviewed for compliance with institutional ethical guidelines by the Animal Welfare Ethical Review Board, University of Cambridge.

Acknowledgements

U. K. Medical Research Council (DGR, RR, RG75828), the Engineering and Physical Sciences Research Council PhD studentship (IG, AP), CSC Cambridge Scholarship for PhD studentship (RL), postdoctoral fellowship from the DAAD and Leibniz Association (WYC), Raymond and Beverly Sackler Fund for Physics of Medicine, University of Cambridge (AP), and support by iNEXT, grant number 653706, funded by the Horizon 2020 programme of the European Commission.

Notes and references

1 V. W. C. Wong, D. G. Reid, W. Y. Chow, R. Rajan, M. Green, R. A. Brooks and M. J. Duer, J. Biomol. NMR, 2015, 63, 119.
2 A. S. L. Thankamony, J. J. Wittmann, M. Kaushik and B. Corzilius, Prog. Nucl. Magn. Reson. Spectrosc., 2017, 102–3, 120.
3 W. Y. Chow, R. Li, I. Goldberga, D. G. Reid, R. Rajan, J. Clark, H. Oschkinat, M. J. Duer, R. Hayward and C. M. Shanahan, Chem. Commun., 2018, 54, 12570.
4 W. Y. Chow, R. Rajan, K. H. Muller, D. G. Reid, J. N. Skeeper, W. C. Wong, R. A. Brooks, M. Green, D. Bihan, R. W. Farndale, D. A. Slatter, C. M. Shanahan and M. J. Duer, Science, 2014, 344, 742.
5 R. Li, R. Rajan, W. C. V. Wong, D. G. Reid, M. J. Duer, V. J. Somovilla, N. Martinez-Saez, G. J. L. Bernardes, R. Hayward and C. M. Shanahan, Chem. Commun., 2017, 53, 13316.
6 S. H. Li and M. Hong, J. Am. Chem. Soc., 2011, 133, 1534.
7 F. Hu, K. Schmidt-Rohr and M. Hong, J. Am. Chem. Soc., 2012, 134, 3703.
8 P. A. Kroon, D. M. Quinn and E. H. Cordes, Biochemistry, 1982, 21, 2745.
9 T. Kitayama, Y. Ogawa, K. Fujii, H. Nakahara, N. Mizuno, K. Sekiguchi, K. Kasai, N. Yurino, T. Yokoi, Y. Pukuma, K. Yamamoto, T. Oki, H. Asamura and H. Fukushima, Leg. Med., 2010, 12, 84.
10 C. Khanna, J. Khan, P. Nguyen, J. Prehn, J. Caylor, C. Yeung, J. Trepel, P. Meltzer and L. Helman, Cancer Res., 2001, 61, 3750.
11 I. V. Sergeyev, L. A. Day, A. Goldbourt and A. E. McDermott, J. Am. Chem. Soc., 2011, 133, 20208.
12 D. L. Nelson and H. M. Cox, Lehninger: Biochemistry, W. H. Freeman & Co., New York, 6th edn, 2013.
13 A. Lesage, C. Auger, S. Caldarelli and L. Emsley, J. Am. Chem. Soc., 1997, 119, 7867.
14 K. Takegoshi, S. Nakamura and T. Terao, Chem. Phys. Lett., 2001, 344, 631.
15 C. P. Jaroniec, C. Filip and R. G. Griffin, J. Am. Chem. Soc., 2002, 124, 10728.
16 D. S. Wishart, Prog. Nucl. Magn. Reson. Spectrosc., 2011, 58, 62.
17 P. N. Borer, S. R. LaPlante, N. Zanatta and G. C. Levy, Nucleic Acids Res., 1988, 16, 2323.
18 S. R. LaPlante, E. A. Boudreau, N. Zanatta, G. C. Levy, P. N. Borer, J. Ashcroft and D. Cowburn, *Biochemistry*, 1988, 27, 7902.
19 B. L. Gaffney, B. Goswami and R. A. Jones, *J. Am. Chem. Soc.*, 1993, 115, 12607.
20 X. P. Xu and S. C. F. Au-Yeung, *J. Phys. Chem. B*, 2000, 104, 5641.
21 X. Gao and R. A. Jones, *J. Am. Chem. Soc.*, 1987, 109, 3169.
22 M. P. Yavropoulou and J. G. Yovos, *J. Musculoskeletal Neuronal Interact.*, 2018, 18, 18.
23 R. L. Jilka, B. Noble and R. S. Weinstein, *Bone*, 2013, 54, 264.
24 O. M. Loreille, T. M. Diegoli, J. A. Irwin, M. D. Coble and T. J. Parsons, *Forensic Sci. Int.: Genet.*, 2007, 1, 191.
25 R. Coscas, M. Bensussan, M. P. Jacob, L. Louedec, Z. Massy, J. Sadoine, M. Daudon, C. Chaussain, D. Bazin and J. B. Michel, *Atherosclerosis*, 2017, 259, 60.
26 A. L. Boskey and D. M. Timchak, *Metab. Bone Dis. Relat. Res.*, 1983, 5, 81.
27 R. A. Haraszti, M. C. Didiot, E. Sapp, J. Leszyk, S. A. Shaffer, H. E. Rockwell, F. Gao, N. R. Narain, M. DiFiglia, M. A. Kiebish, N. Aronin and A. Khvorova, *J. Extracell. Vesicles*, 2016, 5, 32570.