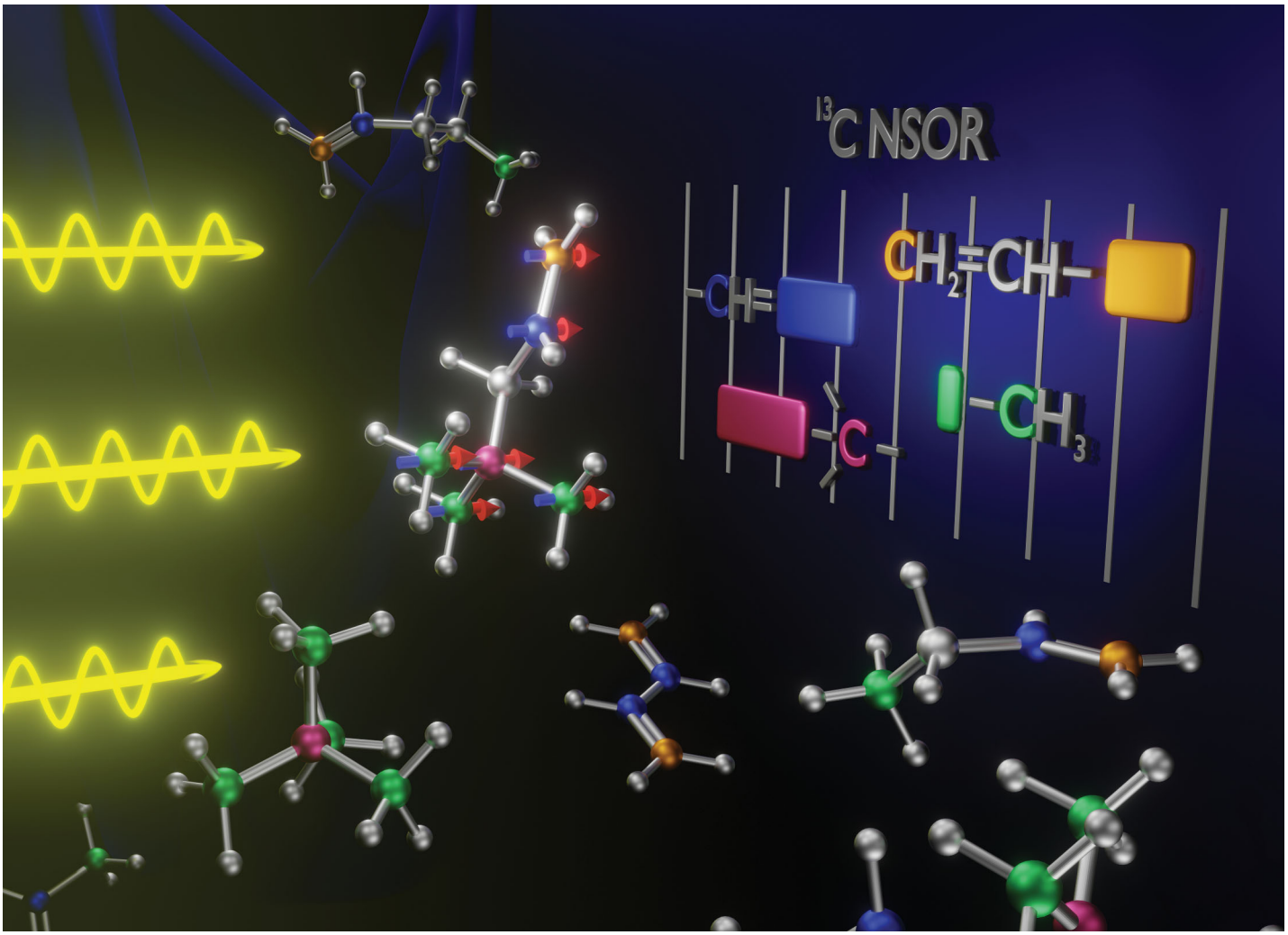

Showcasing research from the NMR Research Unit (University of Oulu, Finland).

Nuclear spin-induced optical rotation of functional groups in hydrocarbons

This article discusses the relationship between molecular structure and nuclear spin-induced optical rotation (NSOR) in hydrocarbons. NSOR is an optical rotation caused by oriented nuclear spins when a linearly polarized light beam passes through a substance. In this work, it is demonstrated that the NSOR effect is systematically related to the molecular structure of hydrocarbons, such that particular moieties exhibit characteristic angles of rotation. The distinct optical rotation by different chemical groups opens possibilities to develop a new form of spectroscopy sensitive to molecular structure, based on NSOR.

As featured in:

See Petr Štěpánek, *Phys. Chem. Chem. Phys.*, 2020, **22**, 22195.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

Registered charity number: 207890

PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2020, **22**, 22195

Nuclear spin-induced optical rotation of functional groups in hydrocarbons†

Petr Štěpánek *

Received 27th May 2020,
Accepted 14th July 2020

DOI: 10.1039/d0cp02856h

rsc.li/pccp

Nuclear spin-induced optical rotation (NSOR) is a nuclear magneto-optic effect manifesting as a rotation of the plane of polarization of linearly polarized light induced by nuclear magnetic moments within a molecule. NSOR probes molecular optical properties through localized nuclear interactions and has potential to be developed into a new spectroscopic tool. However, so far the connection between the molecular structure and NSOR response has not been systematically investigated. To obtain insight into this relation and to assess its viability as a foundation for a new spectroscopic method, NSOR of ^{1}H and ^{13}C nuclei in a set of hydrocarbon molecules with various structural motifs is theoretically investigated using density functional theory calculations. The results reveal that NSOR intensities are correlated with several structural features of the molecules, such as the position of the nucleus in the carbon chain, isomerism and presence of nearby unsaturated groups. Specific patterns connecting NSOR to the local chemical environment of the nucleus can be observed. It is also shown that this effect can be to a good approximation modelled as a sum of individual contributions from nearby chemical groups, allowing for a rapid estimation of its values. The demonstrated systematic dependence of the NSOR signal on the molecular structure is a desirable feature for theoretical and experimental development of new spectroscopic methods based on this phenomenon.

1 Introduction

Modern spectroscopic methods provide invaluable insight into the structure of molecules and materials. A prominent example of this is nuclear magnetic resonance (NMR), a technique known for its versatility in the investigation of molecular structure,[1] quantifying microscopic dynamic phenomena,[2] and non-invasive imaging.[3] The strength of NMR comes from its ability to distinguish between individual atoms in a molecule based on their local chemical surroundings.

Recently, a branch of spectroscopy related to NMR, known as nuclear magneto-optic spectroscopy (NMOS), has been explored both experimentally[4–11] and theoretically.[10–29] In contrast to classical NMR, which detects the signal through electromagnetic induction from the precessing nuclear magnetic moments, NMOS measures how these magnetic moments change the polarization state of the light passing through the sample. The interactions between the nuclei and the light, which are mediated by the electron cloud of the molecule, have shown promise for opening new types of spectroscopic measurements, such as observation of dipolar couplings in isotropic liquids[26] or mapping the properties of excited states.[23] So far, five NMOS effects have been described: nuclear spin-induced optical rotation (NSOR),[4–18] nuclear spin-induced circular dichroism (NSCD),[19–23] nuclear spin-induced Cotton–Mouton effect (NSCM),[24,25] nuclear spin-induced Cotton–Mouton effect in an external magnetic field (NSCM-B),[26,27] and nuclear quadrupole-induced Cotton–Mouton effect (NQCM).[28,29] Out of these, only NSOR, a circular birefringence caused by nuclear magnetic moments oriented along the light beam, has been experimentally measured so far.[4–11]

The NMOS field is still in its infancy with broad room for development both experimentally and theoretically. So far, several general observations have been reported, such as enhancement of NSOR due to a nearby electronic resonance[14] or conditions necessary for appearance of the NSCD signal[23] and its correlation with the global electronic structure.[22] A simple semi-empirical model for estimation of ratios of NSOR signals in a single molecule has also been proposed.[17] However, despite multiple studies on the subject the general connection between the molecular structure motifs and the corresponding NMOS response has not been established for any of these effects and the interpretation of experiments still largely relies on full rigorous quantum-chemical calculations.[10–13]

Motivated by the lack of such insight, this paper investigates the existence of systematic trends or characteristic features that

NMR Research Unit, Faculty of Science, University of Oulu, P.O. Box 3000, FI-90014, Oulu, Finland. E-mail: petr.stepanek@oulu.fi

† Electronic supplementary information (ESI) available: List of molecules, basis set benchmark, additional figures discussing the effects of molecular conformation, isomerism and solvent on the NSOR signal. See DOI: 10.1039/d0cp02856h

connect chemical moieties in the molecular structure with NMOS response. In particular, the study focuses on NSOR since it is the most experimentally developed NMOS effect and is likely to be one of the first to find an application in chemical structure determination in the near future.

In contrast to NMR, in which signals of nuclei of one isotope have the same strength, NSOR intensity, and even sign, depends on the chemical environment of the nucleus, an effect known as optical chemical shift.[10,14,17] The goal of the presented report is to quantify this variation with respect to molecular structure.

The study is based on computational evaluation of NSOR for a set of hydrocarbon molecules including alkanes, alkenes, alkynes and dienes. The hydrocarbons were chosen as the model set because they provide multiple isomers with recurring structural motifs and thus can produce a sufficiently large data set for assessing common features. They are also likely to be one of the first groups of molecules to be investigated in a systematic manner by high-resolution NSOR.

The analysis of the results reveals that NSOR signals of individual nuclei correlate with the local chemical structure, such as position of the nucleus in the chain, isomerism and presence of nearby unsaturated bonds. In addition, it is also found that NSOR signals can be approximately reconstructed from individual contributions of neighbouring atom groups. The validity of this approximation is tested by comparing the NSOR values obtained from the contribution model with full quantum-chemical calculations for several hydrocarbons not present in the original set. An overall good agreement is found, supporting the existence of common spectral markers of NSOR for nuclei in specific moieties, and their additivity.

2 Methods

The set of model molecules consisted of all structural isomers of linear and branched hydrocarbons which are possible to create from up to six carbon atoms. Because some structural features are less represented in this group than others, a number of larger hydrocarbons were added to the set, chosen so as to increase the number of data points corresponding to less abundant structural motifs. The list of all molecules is given in ESI† in Tables S1–S6. The studied systems included alkanes, alkenes, alkynes and dienes (isolated, conjugated and cumulated).

Each molecule was represented by a single conformer, built as an extended chain to minimize intramolecular interactions between different parts of the chain. The geometries were then optimized in Turbomole 7.4.1[30,31] at the RI-B3LYP/def-TZVP level.[32–34] The presence of a local geometry minimum was verified by calculation of harmonic vibrational frequencies at the same level of theory. For comparison, several highly branched molecules were also optimized with the addition of DFT-D3 dispersion correction by Grimme.[35]

The NSOR parameters were calculated in DALTON[36,37] using the quadratic response function formalism.[38] The angle of optical rotation per unit length, nuclear concentration and spin polarization can be expressed using the quadratic response function $\langle\langle \mu_\alpha;\mu_\beta,h^{\mathrm{hf}}_{K,\gamma}\rangle\rangle_{\omega,0}$ as[19]

$$\theta = -\frac{1}{12}\omega\mu_0 c_0 N_\mathrm{A} I_K \varepsilon_{\alpha\beta\gamma}\Im\left\langle\left\langle \mu_\alpha;\mu_\beta,h^{\mathrm{hf}}_{K,\gamma}\right\rangle\right\rangle_{\omega,0} \quad (1)$$

where ω, μ_0, c_0, N_A, I_K, and $\varepsilon_{\alpha\beta\gamma}$ are, in order, the angular frequency of the incident light, permeability of vacuum, speed of light in vacuum, Avogadro constant, spin quantum number of the nucleus K and Levi-Civita anti-symmetric tensor, and $\Im$ denotes the imaginary part. Implicit summation over repeated indices of Cartesian components α, β, γ is implied. The operators in the response function are electric dipole $\hat{\boldsymbol{\mu}}$ and hyperfine interaction $\hat{\boldsymbol{h}}^{\mathrm{hf}}_K$, the latter corresponding in non-relativistic theory of closed-shell systems to paramagnetic nuclear spin–electron orbit operator $\hat{\boldsymbol{h}}^{\mathrm{PSO}}_K$:

$$\hat{\boldsymbol{\mu}} = -e\sum_i \mathbf{r}_i \quad (2)$$

$$\hat{\boldsymbol{h}}^{\mathrm{PSO}}_K = \frac{e\hbar\mu_0}{4\pi m_\mathrm{e}}\gamma_K\sum_i \frac{\boldsymbol{l}_{K,i}}{r_{Ki}^3} \quad (3)$$

where e, m_e, $\hbar$, γ_K and $\boldsymbol{l}_{K,i}$ are the charge and mass of the electron, reduced Planck constant, gyromagnetic ratio, and angular momentum of electron i around the position of nucleus K, respectively.

Calculations of the response functions were performed using the BHandHLYP[33,39] density functional with the co2 basis set, in vacuum. The BHandHLYP functional has been shown to provide NSOR close to experimental and correlated *ab initio* coupled cluster singles and doubles (CCSD) results for organic molecules.[10,11] The basis set co2 is a completeness-optimized basis set[40] developed and tested for NSOR.[16] Additional benchmark tests against larger completeness-optimized basis sets[16,41] have shown that it provides a sufficient quality with modest computational effort (see Fig. S1 in ESI†). To assess the effect of solvent, several NSOR calculations were performed with the inclusion of an implicit solvent using the integral equation formalism of the polarizable continuum model (IEF-PCM) with solvent parameters for *n*-heptane.[42]

All calculations[43] were carried out for a laser wavelength of 405 nm, which is a commonly available wavelength in commercial laser diodes. The corresponding energy is also far enough from the lowest excitation energy for all studied molecules, avoiding accidental near-resonance effects from close lying excited states, which may significantly affect the intensity of NSOR signals.[14] Although such amplification of NSOR may be desirable on its own as a spectroscopic feature, accidental excitations would complicate assessment of the common trends in the present study. The NSOR values for chemically equivalent nuclei were averaged.

All NSOR signals throughout the text are reported normalized to unit spin polarization, concentration of nuclei and sample thickness in units of μrad mol^{-1} dm^3 cm^{-1}. The units are written in the text as μrad for short and normalization is always implied.

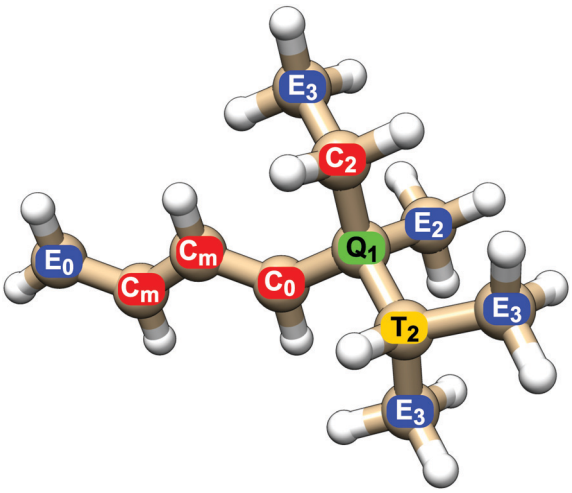

Fig. 1 Explanation of the labelling system of nuclei. Colours and capital letters distinguish carbon nuclei at the end of the chain (**E**), inside the chain bound to two carbon atoms (**C**), nuclei at a branching point sharing three carbon–carbon bonds (**T**) and quaternary carbon with four carbon neighbours (**Q**). The subscripts indicate the distance from the nearest unsaturated bond, counted as number of carbon–carbon bonds, with 0 zero being assigned to atoms sharing the bond and *m* to the atoms in the middle of a larger unsaturated system.

All NSOR signals were categorized based on the chemical identity of the nucleus according to two structural criteria: (a) atom type, as determined by its position in the chain, *i.e.*, how many other carbon atoms it is bound to, and (b) distance from the nearest unsaturated bond. An example of the labeling of nuclei is given in Fig. 1.

The first criterion for assigning the nuclei is their position in the hydrocarbon chain. For carbon nuclei we can distinguish four possibilities: nucleus in the group at the end of the chain (**E**, end), nucleus in the middle of a linear chain (**C**, chain), atom at a T-branch (**T**, tertiary), and quaternary carbon (**Q**). This notation reflects the number of attached carbon atoms, from 1 in primary carbons (**E**) to 4 in quaternary carbons (**Q**).

The second criterion for categorization of nuclei is their distance from the nearest unsaturated bond. The nuclei of carbons that share the unsaturated bond are labelled 0, their direct neighbours are labelled 1, the next ones in the chain 2, *etc.* (Fig. 1). In the case of molecules with several unsaturated bonds, the nucleus is numbered according to the closest one. This choice is based on the expected locality of NSOR due to the presence of the locally acting spin–orbit operator, as was already shown for NSOR[14] and closely related NSCD property.[23] The assumption that closer perturbation will have larger impact than more remote one is first verified on systems with only one unsaturated bond before applying it to more complex systems. In the case of alkanes only the first criterion is used to categorize the atoms as there are no unsaturated bonds.

A subtype of nuclei, labelled *m* instead of a number, is introduced for dienes. It describes nuclei sandwiched between two double bonds. The corresponding atom type $\mathbf{C}_m$ occurs in the following typical chains: (a) in isolated dienes it is a carbon nucleus neighboured on both sides by double bonds, in the chain of the form C═C–$\mathbf{C}_m$–C═C; (b) in conjugated dienes it is the nucleus in the middle of the conjugated chain, *i.e.* C═$\mathbf{C}_m$–$\mathbf{C}_m$═C; (c) in cumulated dienes it corresponds to the nucleus of the atom sharing both double bonds, in structure C═$\mathbf{C}_m$═C.

The above described labelling is used for carbon atoms. The hydrogen nuclei share for simplicity the atom type with the carbons they are attached to. For example, hydrogen atoms bound to carbon atom $\mathbf{E}_0$ in Fig. 1 are also labelled $\mathbf{E}_0$, even though they do not participate in the double bond themselves. In other words, all nuclei in a given CH_n group share the label of the carbon nucleus.

The molecular models were rendered using UCSF Chimera.[44]

3 Results

As noted above, the following results were obtained from a limited number of model molecules, each represented by a single conformer in vacuum, and with different number of data points for each type of nucleus. Therefore, the results should be understood as a qualitative description of the essential features of NSOR with respect to the general molecular framework. The quantitative prediction of experimental data for each molecule would have to include other effects, such as conformational flexibility and influence of the environment. So far, the available high-resolution experimental data for atom-specific NSOR signals[6] do not have a sufficiently high signal-to-noise ratio for quantitative comparison.

It should also be noted that since NSOR originates from the alignment of the light beam and nuclear magnetic moments, it is naturally modulated by the precession of the magnetization and is detected at the Larmor frequency corresponding to the NMR shift of the nucleus. This allows one to experimentally separate the NSOR signals according to their chemical shift in cases where NSOR intensities alone might lead to ambiguous results.

3.1 Simple hydrocarbons

3.1.1 Alkanes. The results of NSOR calculations for ^{1}H and ^{13}C in alkanes are shown in Fig. 2, panel I and II, respectively. On the *x*-axis is the NSOR in μrad mol^{-1} dm^3 cm^{-1}; the offsets on the *y*-axis, colours and symbols of data points differentiate between atom types, as described in the Methods section. At a glance it is clear that the signals tend to cluster for both elements according to their chemical nature into rather compact groups.

In the case of ^{1}H, there is an overlap between the **E** and **C** groups, while the hydrogen nuclei **T** bound to tertiary carbons are offset at higher NSOR values. The total spread of signals is rather limited, ranging between 1.5 and 2 μrad. The NSOR overall gets stronger with increasing branching at the site of the carbon it is attached to, as seen in increasing intensity from **E** to **T**.

Signals of ^{13}C nuclei form three distinct clusters corresponding to separate groups for atom types **E** and **C**, and the

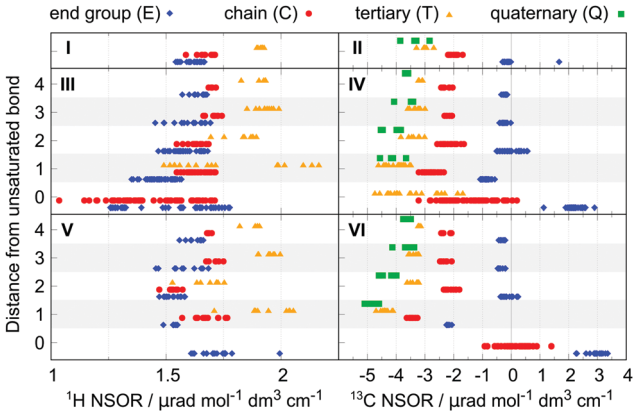

Fig. 2 Comparison of ^{1}H (left) and ^{13}C (right) NSOR signals for alkanes (top row, panels I and II), alkenes (middle row, panels III and IV) and alkynes (bottom row, panels V and VI). On the *x*-axis is the optical rotation normalized for unit path length, concentration of nuclei and unit spin polarization, and on the *y*-axis is the distance of the atom from the unsaturated bond. The chart distinguishes nuclei corresponding to atoms at the end of the chain (**E**, blue diamonds), inside the chain (**C**, red circles), carbons with three carbon neighbours (**T**, orange triangles) and quaternary carbons (**Q**, green squares).

overlapping group of **T** and **Q**. The values are distributed over a broader range compared to ^{1}H, from around 0 μrad for atom type **E** down to −4 μrad for **T** and **Q**. The data point at 1.67 μrad corresponds to methane and due to the unique structure of this molecule is not counted as a typical feature of alkanes. In general, a similar trend can be observed as in ^{1}H where the strength of NSOR signals increases with the number of carbon neighbours, although in this case with reversed sign and larger relative magnitude.

In summary, NSOR of each atom type in alkanes, as described by their position in the carbon skeleton, falls into a rather narrow range of values for both ^{1}H and ^{13}C nuclei and its magnitude is positively correlated with the number of neighbouring carbon atoms. The NSOR intensity can thus distinguish between primary, secondary, and to a limited degree tertiary and quaternary CH_x groups in alkanes.

3.1.2 Alkenes. The NSOR of alkenes is plotted in Fig. 2 in panels III and IV for ^{1}H and ^{13}C, respectively. In this and the following charts, the *y*-axis now also shows the distance from the unsaturated bond, as defined in the Methods section.

At first sight it is apparent that the signals now cover a much wider range. In the case of ^{1}H the NSOR values of $\mathbf{C}_0$ and $\mathbf{E}_0$-type nuclei at the double bond are not anymore clustered into separate groups. Instead, they are spread rather evenly, mostly towards negative with respect to alkanes, and overlap significantly. As we will discuss in detail below, this wide spread of values reflects *cis*/*trans* isomerism around the double bond. Next to the double bond the signals of atom types $\mathbf{E}_1$ and $\mathbf{C}_1$ start to cluster, with $\mathbf{E}_1$ having an overall weaker signal than either $\mathbf{C}_1$ or $\mathbf{T}_1$ and also being weaker than **E** in alkanes. NSOR of atom type $\mathbf{C}_1$ appears in roughly the same region as in alkanes and the signals of nuclei $\mathbf{T}_1$ are spread over a very wide range, providing the strongest signals at around 2.2 μrad, but also appearing at much lower values, overlapping with the whole range of $\mathbf{C}_1$ and part of $\mathbf{E}_1$. This spread is partially explained by *cis*/*trans* isomerism as NSOR of nuclei of this type in *cis*-alkenes appears around 1.6 μrad while *trans*-alkenes and alk-1-enes show higher NSOR values (see Fig. S2 in ESI†). However, this does not explain the wide separation fully, which is linked to other structural details not apparent from this set of data points. Further from the double bond, the NSOR strength of the $\mathbf{E}_2$ atom type slightly increases and $\mathbf{T}_2$ signals aggregate around 1.8 μrad. Interestingly, with increasing distance from the unsaturated bond the signals of all atom types drift towards roughly the same intervals as in alkanes. The double bond thus appears to affect the NSOR of nearby nuclei the most, and this effect falls off with the distance.

In the case of ^{13}C the atoms participating in the double bond show a wide distribution, but over quite characteristic regions. Similarly as in the case of ^{1}H, this wide spread can be linked to isomerism, and will be discussed below. The $\mathbf{E}_0$-type nuclei have strongly positive NSOR around 2.3 μrad, $\mathbf{C}_0$-type occupy a wide range between 0 and −3 μrad and $\mathbf{T}_0$ nuclei also show a large spread, but shifted to more negative values between −1.5 and −5 μrad. The outlier $\mathbf{E}_0$ signal at 1.13 μrad belongs to ethene. The nuclei neighbouring the double bond form more compact groups with $\mathbf{T}_1$ and $\mathbf{Q}_1$ around −4 μrad, $\mathbf{C}_1$ around −2.8 μrad, and $\mathbf{E}_1$ switching sign to slightly negative values around −0.9 μrad. In position 2 the signals of $\mathbf{E}_2$ nuclei drift towards zero, $\mathbf{C}_2$ signals towards −2 μrad, $\mathbf{T}_2$ signals move closer to −3 μrad, and $\mathbf{Q}_2$-type remain the most negative at around −4 μrad. At larger distances signals of all atom types monotonically shift towards the range of values characteristic for alkanes, similarly as in the case of ^{1}H.

We can see from these results that NSOR can distinguish, *e.g.*, the terminal CH_2= group, which gives a positive signal for the $\mathbf{E}_0$ nucleus, from the $\mathbf{C}_0$ in –CH=CH– inside the chain. The spectroscopic value of NSOR can be further enhanced when we take into account the Larmor frequency of the nuclei: the signals in position 0 will appear at chemical shifts characteristic for unsaturated bonds, while the NSOR of other nuclei will be modulated with frequencies typical for saturated hydrocarbons. This allows one, for example, to identify the ^{13}C $\mathbf{E}_1$ signal as it is the only NSOR signal in the range around −1 μrad with chemical shift outside of the alkene region.

Based on the observations we can argue that the double bond acts as a local perturbation of NSOR with respect to an alkane chain and affects most strongly nuclei up to roughly two to three bond-lengths away. Nuclei further from the double bond see rapidly diminishing influence. The local nature of NSOR has been observed before as signal enhancement of nuclei located in the chromophore when the energy of the incident light approaches its electronic resonance[14] and the present results suggest that the locality of NSOR manifests also at energies far from excitations. This behaviour will be observed in the forthcoming cases as well.

3.1.3 Alkynes. The ^{1}H and ^{13}C NSOR of alkynes is presented in Fig. 2, panels V and VI, respectively. The signals form in this case more compact groups than in alkenes.

For ^{1}H nuclei the $\mathbf{E}_0$ is located around 1.7 μrad, at slightly higher values than $\mathbf{E}$ in alkanes and covers the upper range of $\mathbf{E}_0$ of alkenes. The outlier signal at 1.99 μrad is ethyne. Next to the triple bond the NSOR of $\mathbf{C}_1$ and $\mathbf{T}_1$ is spread out at around the same mean value as in alkanes while $\mathbf{E}_1$ nuclei give slightly smaller NSOR. The intensities thus keep the general trend $\mathbf{T} > \mathbf{C} > \mathbf{E}$. Interestingly, in position 2 the NSOR of $\mathbf{C}_2$ and $\mathbf{T}_2$ appears to be slightly shifted towards smaller values by about 0.1 and 0.2 μrad, respectively, compared to alkanes and alkenes. At distances further from the triple bond the NSOR converges again towards values of alkanes for all three atom types.

The NSOR of ^{13}C qualitatively resembles that of alkenes, but with overall slightly stronger intensities. For atoms participating in the triple bond the $\mathbf{E}_0$ nuclei give NSOR around 3 μrad, compared to 2.3 μrad in alkenes. On the other hand, the signals of $\mathbf{C}_0$ are the exception and are overall weaker and spread quite evenly around zero. Next to the triple bond the NSOR of different types of nuclei forms distinct compact groups with all signals negative and following the progression $\mathbf{E}_1 > \mathbf{C}_1 > \mathbf{T}_1 > \mathbf{Q}_1$. The most notable increase in intensity compared to alkenes is for $\mathbf{E}_1$. As in previous cases the influence of the unsaturated bond can be observed to a lesser degree still in $\mathbf{Q}_2$ and $\mathbf{T}_2$, but drops off with distance.

From the point of view of spectroscopic application the carbon nuclei $\mathbf{E}_0$ and $\mathbf{C}_0$ at the triple bond are the most characteristic, being well separated in their magnitude. The other NSOR signals, which have chemical shifts in the region of saturated hydrocarbons, can be assigned for $\mathbf{E}_{\geq 2}$ nuclei with NSOR around zero, and for **T**/**Q** atom types at strongly negative NSOR values < -4 μrad. We note in passing that the **T** and **Q** could be further experimentally distinguished by their different NMR relaxation rates. The regions around -2 μrad and -3.5 μrad are ambiguous since NSOR values of $\mathbf{E}_1/\mathbf{C}_{n\geq 2}$ and $\mathbf{C}_1/\mathbf{T}_{n\geq 2}/\mathbf{Q}_{n\geq 3}$ overlap there, respectively.

Comparing the overall features of the alkynes, alkenes and alkanes, we can see that the NSOR of carbon nuclei $\mathbf{E}_0$, $\mathbf{E}_1$, $\mathbf{C}_1$, $\mathbf{T}_1$ and $\mathbf{Q}_1$ increases in intensity as the system becomes more unsaturated. On the other hand, the NSOR intensity of $\mathbf{C}_0$ atom type decreases towards zero. Additionally, the signals in alkenes are much more widely spread compared to the other two cases. This is partially due to the *cis*/*trans* isomerism and possibly also because of the sensitivity of NSOR of nuclei near the double bond to conformations and possible intramolecular interactions (see Fig. S3 in ESI†). Alkynes might be less sensitive to this effect, in part because the rest of the chain points straight away from the triple bond and thus decreases the possibility of close interactions, while alkenes can create more folded structures due to the bond angle of sp^2 carbon atoms.

3.1.4 Effect of isomerism on NSOR of alkenes. In the case of alkenes the ^{13}C NSOR of atom types $\mathbf{C}_0$ and $\mathbf{T}_0$ covers a wide range of values. These signals can be further divided into groups according to the isomerism of the molecules they originate from: one group corresponds to signals of alk-1-enes, which have the double bond at the end of the chain, and one stems from alk-*n*-enes ($n \in \{2,3\}$ in the studied set), where the double bond

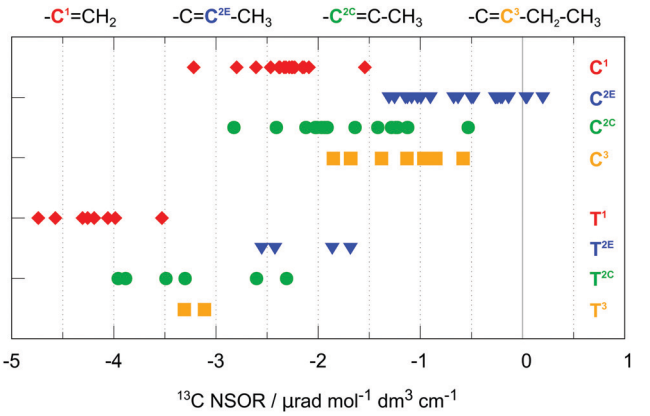

Fig. 3 Effect of position of the double bond in the alkene chain on ^{13}C NSOR of atom types **C** and **T**. The *x*-axis represents the normalized optical rotation, and the *y*-axis distinguishes different atom positions.

is inside the chain, as shown in Fig. 3. For both $\mathbf{C}_0$ and $\mathbf{T}_0$, the alk-1-enes tend to provide stronger, more negative signals compared to those of alk-2-enes, which are overall weaker. Moreover, in alk-2-enes the NSOR of nuclei $\mathbf{C}_0/\mathbf{T}_0$ closer to the end of the chain is weaker than the NSOR of those near the centre of the chain. Despite some overlap between the groups, their mean values are quite well separated (-2.3 μrad for alk-1-enes compared to -1.7 and -0.6 μrad for C^{2C} and C^{2E}, respectively; and -4.2 μrad compared to -3.2 and -2.1 μrad for T^{2C} and T^{2E}, respectively (Fig. 3)). The spread of the values inside each group remains, however, rather wide.

Thus, there are two possible general patterns: in the case of alk-1-enes, one of the two NSOR signals with chemical shifts corresponding to the unsaturated bond will be strongly positive ($\mathbf{E}_0$) and one strongly negative ($\mathbf{C}_0$, second carbon from the end, C^1 in Fig. 3). In the case of alk-2-enes, or in general, alk-*n*-enes, where the double bond is inside the chain, the positive $\mathbf{E}_0$ signal will be absent and instead two negative, less intense $\mathbf{C}_0$ signals will be observed (C^{2C}, C^{2E} or C^3 in Fig. 3). A similar argument can be made for the nuclei of type $\mathbf{T}_0$, depending on whether they are located next to the terminus of the molecule or inside the chain.

This observation is another demonstration of the influence of local electronic structure on NSOR. The atoms not only feel the presence of the nearby double bonds, but also the surrounding chain structure in their vicinity. The former effect appears to be the dominant one, though.

cis/*trans isomerism.* Another structural motif in alkenes is *cis*/*trans* isomerism. This has been briefly discussed in relation with ^{1}H $\mathbf{T}_1$, which shows some dependence on the isomerism of the molecules in which it is present. Here we look in more detail at the nuclei directly connected next to the double bond. For this purpose, the nuclei were categorized into four groups, according to the orientation of that particular nucleus with respect to the carbon chain(s) continuing on the opposite side of the double bond. When a carbon chain is attached in the *cis* or *trans* position, the nucleus is correspondingly labelled as *cis*

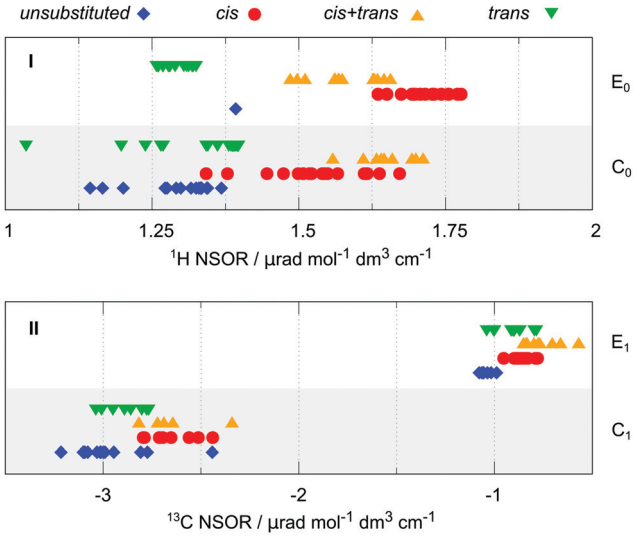

Fig. 4 Effect of *cis*/*trans* isomerism in alkenes for ^{1}H (top panel I) and ^{13}C (bottom panel II) NSOR signals. On the *x*-axis is the normalized optical rotation, and the *y*-axis distinguishes different atom positions.

and *trans*, respectively. If both *cis* and *trans* positions are occupied with carbon substituents, the nucleus is assigned to group *cis* + *trans*. Conversely, when there are no carbon chains attached at the other side of the double bond, the nucleus is labelled *unsubstituted*. We note that *unsubstituted* implies that the nucleus is located in a group neighbouring the end of the molecule.

The situation for ^{1}H and ^{13}C is summarized in Fig. 4 in panels I and II, respectively. The plot distinguishes between the atoms that are part of the end group of the molecule (type $\mathbf{E}_0$ for ^{1}H and $\mathbf{E}_1$ for ^{13}C) or inside the chain (type $\mathbf{C}_0$ or $\mathbf{C}_1$, respectively). The atom type **T** does not show any significant trend in this respect in the available data and is not reported. The *cis*/*trans* isomerism is marked in colour and shape of the symbols and by offset on the *y*-axis.

The ^{1}H NSOR shows a separation of different signals according to their local isomerism, with the signals of $\mathbf{E}_0$ nuclei decreasing in intensity in the order *cis* > *cis* + *trans* > *trans*. The single point in the *unsubstituted* group corresponds to ethene. Although there is an overlap between *cis* + *trans* and *cis* groups, the *trans* one is set apart at much lower intensity, thus being of a potential spectroscopic value.

The hydrogen nuclei $\mathbf{C}_0$ cover a slightly wider region of intensities. The groups of signals of nuclei in *cis* and *cis* + *trans* positions are close together at higher intensities than the *trans* and *unsubstituted* groups, similarly as in the case of $\mathbf{E}_0$. Interestingly, the *cis* + *trans* group provides in this case stronger average signal than *cis* and gives now the strongest NSOR, relative to the total spread of the signals. Because of the overall wider range of the values of the *cis* group, it now also overlaps with *trans*. Nevertheless, the region below 1.25 μrad is exclusively occupied by *trans* and *unsubstituted* groups, while *cis* and *cis* + *trans* are the only species present above 1.5 μrad.

In the case of ^{13}C the trend of intensities for different groups of atom type $\mathbf{E}_1$ follows roughly the pattern *cis* + *trans* > *cis* ≈ *trans* > *unsubstituted*. Interestingly, this pattern differs from that of the ^{1}H $\mathbf{E}_1$ nuclei, where *cis* is the group with the strongest signals. The $\mathbf{C}_1$ nuclei show the trend *cis* + *trans* ≈ *cis* > *trans* > *unsubstituted*. Although there is a slight progression of intensities between the groups, their separation is quite small, with significant overlap between them. Therefore, at least at this level of approximation, the ^{13}C signals might not be specific enough for unambiguous determination of the molecular structure. This issue is exacerbated by the presence of notable outliers in the case of $\mathbf{C}_1$.

In summary the ^{1}H signals are more specific than the ^{13}C signals in distinguishing *cis*/*trans* isomerism. Although the different groups are not completely separated, there is a rather strong distinction between *cis* and *cis* + *trans* on the one hand and *trans* and *unsubstituted* on the other hand, promising a potential spectroscopic feature of NSOR for structural elucidation. Given the constraints of this study it is not clear whether such distinction would be unambiguous to resolve all four cases under real experimental conditions. However, the results suggest at least a possibility to distinguish the limiting cases near the ends of this NSOR range.

3.2 Dienes

We will now discuss dienes as examples of common, more complicated hydrocarbon systems. We can distinguish three groups of dienes, depending on the mutual distance of the double bonds: isolated dienes, where the double bonds are separated by two or more single bonds; conjugated dienes, containing alternating single and double bonds; and cumulated dienes, with double bonds directly next to each other.

The NSOR for all three types is plotted in Fig. 5 together with the essential results for alkenes for comparison. The nuclei at

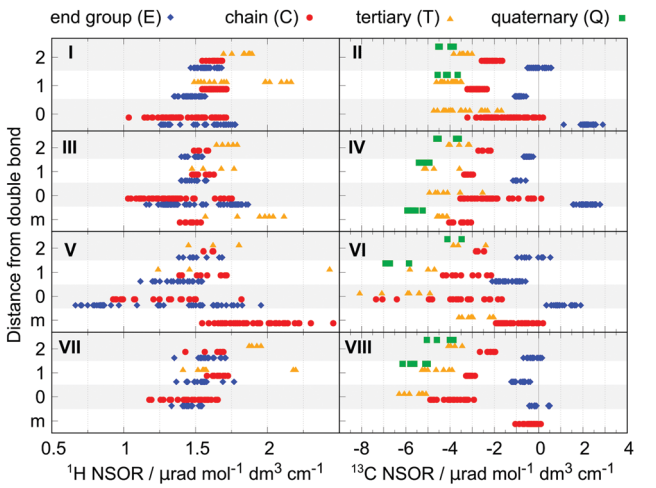

Fig. 5 Comparison of ^{1}H (left) and ^{13}C (right) NSOR signals for alkenes (top row, panels I and II), isolated dienes (second row, panels III and IV), conjugated dienes (third row, panels V and VI) and cumulated dienes (bottom row, panels VII and VIII). On the *x*-axis is the normalized optical rotation, and on the *y*-axis is the distance of the atom from the nearest unsaturated bond. The chart distinguishes nuclei corresponding to atoms at the end of the chain (**E**, blue diamonds), inside the chain (**C**, red circles), carbons with three carbon neighbours (**T**, orange triangles) and quaternary carbons (**Q**, green squares). Signals labelled *m* correspond to nuclei between two double bonds (see the Methods section for details).

distances up to 2 bonds from the nearest double bond are shown, as they are the most influenced and thus of the largest interest. (For extended figure showing also the further nuclei see ESI,† Fig. S4.)

3.2.1 Isolated dienes. As we have seen in the case of alkenes and alkynes, the unsaturated bond affects significantly nuclei participating in it, and the directly adjacent CH_x groups. Since double bonds are separated in isolated dienes, we can expect that they may locally behave like alkenes and their influence on their surroundings will be similar. Inspecting the results for ^{1}H and ^{13}C NSOR in isolated dienes shown in Fig. 5, panels III and IV, we can see that this is indeed the case and the ranges of intensities for both elements cover very similar regions as in alkenes.

For ^{1}H the signals of atom types $\mathbf{E}_0$ and $\mathbf{C}_0$ spread over the intervals of about 1.25–2.0 μrad and 1.0–1.75 μrad, respectively, and become more compact with increasing distance from the double bond. Both the spread of values and their mean position reflect those of alkenes. This similarity can be observed also in the signals of ^{13}C. A small deviation is present in $\mathbf{Q}_1$ and $\mathbf{T}_1$, which are pushed slightly towards negative. This can be explained by the influence of the second double bond, to which the **Q** is more sensitive than **C** at larger distances, as seen in $\mathbf{Q}_2$ compared to $\mathbf{C}_2$ in alkenes.

Of special interest are the signals of nuclei which neighbour simultaneously both double bonds. They can thus be considered a special case of atom types in position 1. In the case of ^{1}H the NSOR signals of $\mathbf{C}_m$ are slightly more negative compared to the $\mathbf{C}_1$ in alkenes, indicating some additional influence of the second double bond. On the other hand, NSOR of $\mathbf{T}_m$ is spread over a similar region as $\mathbf{T}_1$. However, we should note the wide distribution of the values, which makes this conclusion more tentative.

For ^{13}C nuclei the signals of $\mathbf{C}_m$ form quite a compact group and shift towards negative values compared to $\mathbf{C}_1$. This indicates that the presence of a second adjacent double bond to some degree strengthens the influence of the first one. Moreover, the shift from **C** in alkanes to $\mathbf{C}_m$ is roughly twice as large as the difference between **C** in alkanes and $\mathbf{C}_1$ in alkenes (−1.52 μrad compared to −0.83 μrad), suggesting that the effect of the two double bonds is approximately additive. The strengthening of the NSOR by the second double bond is also seen to a limited degree for $\mathbf{T}_m$, which is shifted by −1.33 μrad from the alkane value compared to the shift of −0.91 μrad for $\mathbf{T}_1$, and to an even larger extent in $\mathbf{Q}_m$ values compared to $\mathbf{Q}_1$ in alkenes (shift of −2.35 μrad and −0.79 μrad, respectively). This suggests that both double bonds contribute to the NSOR perturbation, a feature that will be explored later in the construction of an approximate model for this property.

Overall, the features of the isolated dienes reflect those of alkenes for their common atom types, suggesting the local nature of NSOR in molecules with non-interacting double bonds. In addition, the *cis*/*trans* isomerism also influences the NSOR in a similar manner as in alkenes (see Fig. S5 in ESI†).

3.2.2 Conjugated dienes. The results for ^{1}H and ^{13}C NSOR of conjugated dienes are shown in Fig. 5, panels V and VI, respectively. The first noticeable feature is that the range of values is very wide compared to other studied systems.

In the case of ^{1}H the spread is mostly towards weaker NSOR. The signals of atom types $\mathbf{E}_0$ completely overlap with the range of $\mathbf{C}_0$ and reach very low values around 0.7 μrad. Next to the double bond the nuclei $\mathbf{E}_1$ and $\mathbf{C}_1$ are spread less, but still over a wider interval than in alkenes. They also start to differ in their mean values with $\mathbf{C}_1$ being stronger than $\mathbf{E}_1$. In position 2 the signals approach the corresponding alkene values. The signals of $\mathbf{T}_1$ and $\mathbf{T}_2$ are found in a wider region than in alkenes, with no appreciable shift of the mean value, taking into account the limited number of points.

In ^{13}C the NSOR values for atom type $\mathbf{C}_0$ and $\mathbf{T}_0$ are shifted towards negative compared to alkenes, and cover a broader range. In particular, $\mathbf{T}_0$ shows the strongest negative NSOR among the studied molecules, up to −8 μrad. On the other hand, the signals of $\mathbf{E}_0$ are weaker than in alkenes and with comparable spread. Nuclei $\mathbf{E}_1$ and $\mathbf{C}_1$ neighboring the conjugated system both give signals distributed over wider regions than in alkenes, but with similar mean values while the NSOR of $\mathbf{Q}_1$ and $\mathbf{T}_1$ is quite significantly shifted towards negative. In position 2 all four atom types approach values similar to those in alkenes.

The nuclei of type *m*, corresponding to the atoms in the middle of the conjugated chain, do not have a directly comparable atom type in alkenes. In the case of ^{1}H the $\mathbf{C}_m$ nuclei exhibit significantly stronger NSOR compared to other types, reaching up to almost 2.5 μrad, which is one of the largest observed values for ^{1}H among the tested molecules. For ^{13}C the NSOR of $\mathbf{C}_m$ is at slightly negative values, approximately covering the upper range of $\mathbf{C}_0$ in alkenes. Likewise, tertiary carbons $\mathbf{T}_m$ also cover the upper range of $\mathbf{T}_0$ in alkenes. This might suggest some similarity between the NSOR of nuclei inside conjugated systems and those in double bonds of alk-*n*-enes, which occupy this region.

The overall feature of conjugated dienes is that both ^{1}H and ^{13}C signals cover a wide interval of values. In fact, the strongest and the weakest NSOR for ^{1}H and the most negative NSOR for ^{13}C can be found in this set of molecules. This suggests that NSOR of conjugated systems might be more sensitive to minor structural changes. Similarly as in the case of alkenes, some of the variability of the NSOR can be explained by the *cis*/*trans* isomerism (see Fig. S5 in ESI†). Further studies on extended conjugated systems, aromatic and anti-aromatic molecules might provide some additional insight.

3.2.3 Cumulated dienes. The results for NSOR of ^{1}H and ^{13}C in cumulated dienes are shown in Fig. 5, panels VII and VIII, respectively. Overall the different atom-type groups are rather compact for both ^{1}H and ^{13}C, in contrast to the above cases.

The ^{1}H NSOR is rather similar to that of alkenes for all atom types, having similar mean values and spread with slightly more compact width for the nuclei in position 0. In the case of ^{13}C the NSOR of nuclei of type $\mathbf{C}_0$ and $\mathbf{T}_0$ is shifted towards negative values compared to alkenes. On the other hand, the signals of $\mathbf{E}_0$-type nuclei are spread closely around zero. Interestingly, this can be seen as a trend for $\mathbf{E}_0$ signals which decrease in intensity from alkanes/isolated dienes through

conjugated dienes to cumulated ones as the double bond system gets more compact. In position 1 the values for $\mathbf{C}_1$ and $\mathbf{E}_1$ are already close to the NSOR of alkenes, while $\mathbf{T}_1$ and $\mathbf{Q}_1$ are slightly more negative. In position 2 all four atom types approach values close to alkenes.

Concerning the atom type $\mathbf{C}_m$, it shows a rather weak signal between 0 and −1 μrad. It is worth noting that similarly as for $\mathbf{E}_0$, the NSOR values of this atom type steadily decrease going from isolated dienes to conjugated to cumulated, as the double bonds approach one another.

3.3 Effect of solvent and geometry

As has been shown, NSOR is sensitive to the environment surrounding the molecule.[10–13] To gain insight into this effect for the present case, NSOR calculations including the implicit solvent model (PCM) were performed for a subset of molecules, with two representatives from each group of investigated hydrocarbons. The list of selected structures is given in Table S7 in ESI.† *n*-Heptane was used as solvent in the PCM, as it is a typical hydrocarbon and thus should provide a reasonable description if we consider the molecule to be surrounded by the bulk of its own neat liquid phase or other similar, nonpolar solvent.

The results are shown in Fig. 6. As we can see, the inclusion of the implicit solvent model rather systematically increases the NSOR values by about 25% for all atom types, with only small deviations. As the change is approximately proportional to the signal intensity, this effect enhances the distinction between the different atom types, but does not change the overall trends.

The increase of NSOR intensity can be compared to the changes in optical properties induced by the solvent. The excitation energies calculated in vacuum and with the PCM show very small differences (see ESI,† Fig. S6), and the increase in NSOR is thus not due to the presence of closer electronic resonances.[14] On the other hand, electric transition dipole moments, which formally appear in the sum-over-states expression of NSOR, are overall larger, correlating with the change in NSOR.

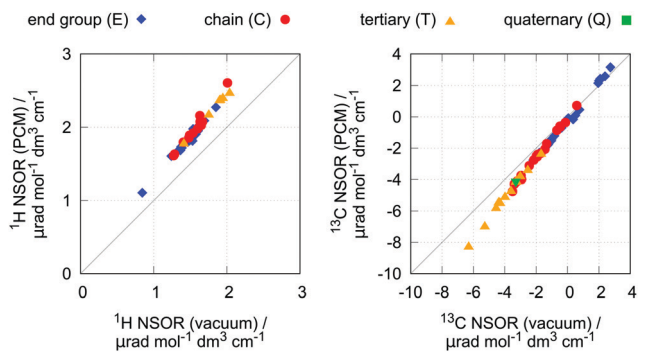

Fig. 6 Effect of the implicit solvent model on ^{1}H NSOR (left) and ^{13}C NSOR (right). The *x*-axis represents NSOR calculated for molecules in vacuum, and the *y*-axis shows NSOR calculated with the inclusion of the polarizable continuum model. Each point represents one chemically distinct nucleus.

We note that this description gives only a partial insight, since NSOR of different nuclei can also be influenced by specific localized solvent–solute intermolecular interactions, which are not present in the PCM. Thus, further changes in NSOR, likely more sensitive to the chemical nature of the nucleus, can be expected by including an explicit solvent. Such interactions might be prominent particularly in the case of more polar molecules than hydrocarbons.

Another aspect which might influence NSOR is geometry of the molecular structure. The effect of small variations was investigated by including the dispersion DFT-D3 term[35] in the geometry optimization of eight structures selected from the most branched molecules, where the influence of dispersion on the geometry is expected to be the most pronounced (see Table S7 in ESI†). It is found that this geometry perturbation does not change NSOR significantly (see Fig. S7 in ESI†). It should be noted, however, that the dispersion might have secondary effects on NSOR by affecting the relative populations of conformers.

The molecular conformation affects the NSOR intensity noticeably (see Fig. S3 in ESI†). The present results suggest that the saturated parts of molecules are less sensitive to such changes, and their NSOR reflects primarily the structural features determined by the molecular skeleton. The effect of conformation on the unsaturated parts of the molecule appears more pronounced. Quantitative assessment of the role of conformational freedom and its impact on NSOR and its trends is beyond the scope of the present work. Given that the differences in NSOR between conformers are comparable in magnitude to the effect of the solvent, even modelled implicitly, such evaluation would most likely require an accurate model including explicit solvent molecules.

3.4 Trends in NSOR of hydrocarbons

Let's summarize the main observations about NSOR in hydrocarbons so far:

(a) signals of alkanes are all present in narrow regions for each of the atom types **E**, **C**, **T** and **Q**

(b) an unsaturated bond locally perturbs the NSOR of nearby nuclei, shifting the values from that of alkanes; this change is largest for atoms participating in the bond and for those adjacent to it; NSOR of nuclei further apart approaches values characteristic for alkanes

(c) there is a monotonic progression in NSOR intensity for ^{13}C nuclei in position 0 and 1 from alkanes to alkenes and alkynes, and for $\mathbf{E}_0$ and $\mathbf{C}_m$ in dienes from isolated to cumulated, suggesting dependence of NSOR on the electron richness of the system

(d) for nuclei not involved in the unsaturated bond, the signal strength follows the pattern $\mathbf{Q} \approx \mathbf{T} > \mathbf{C} > \mathbf{E}$ for ^{13}C and $\mathbf{T} > \mathbf{C} \geq \mathbf{E}$ for ^{1}H

(e) generally, ^{13}C-NSOR is more sensitive to the backbone structure of the molecule, while ^{1}H is more sensitive to isomerism.

In addition, in systems with two isolated double bonds, their influence on NSOR of adjacent nuclei appears to be approximately additive as seen for $\mathbf{C}_m$, $\mathbf{T}_m$ and $\mathbf{Q}_m$ in isolated dienes.

This, however, is not observed in conjugated and cumulated dienes – if the unsaturated bonds form more extended systems, the new complex chromophore has its own specific NSOR properties.

Based on the results from the isolated dienes we may tentatively propose that NSOR can be approximated for non-interacting chromophores as an additive property. The central hypothesis of the model is that each chromophore contributes to the NSOR of nearby nuclei separately. Each contribution is specific for the atom type of the nucleus being influenced (*e.g.*, **E**), the type of the chromophore (*e.g.*, a double bond) and the distance between the chromophore and the nucleus. Because the NSOR values of nuclei far from the chromophore approach those of alkanes, we can take the alkane values as a baseline for each atom type and add the corrections, which are most significant near the chromophore and fall off towards zero with distance. Thus, the corrections can be estimated for each atom type **AT** at the distance N from the chromophore C as

$$\Delta(\mathrm{AT})_N^{\mathrm{C}} = \overline{\mathrm{NSOR}(\mathrm{AT})_N^{\mathrm{C}}} - \overline{\mathrm{NSOR}(\mathrm{AT})^{\mathrm{alkane}}} \qquad (4)$$

where the bar over the quantities represents their mean, evaluated from the test set. The NSOR contributions calculated according to eqn (4) for ^{13}C and ^{1}H are presented in Tables 1 and 2, respectively, for alkanes, alkenes, alkynes and cumulated dienes. Isolated dienes are not included due to their similarity to alkenes and the presence of two distinct chromophores, and conjugated dienes due to their wide range of values, which makes this simple model unsuitable for them. We note here that the values given in the tables are estimates influenced by the finite number of points in the training set of nuclei.

Using these data we can estimate the NSOR for a system composed of several chromophores as follows:

Table 1 Relative contributions of unsaturated functional groups to ^{13}C NSOR for different atom types depending on their distance from the chromophore. The values are reported in μrad mol^{-1} dm^3 cm^{-1} for carbon marked C^x

	Atom type of **C**^{x}			
Chromophore	E	C	T	Q
Alkane	−0.16	−2.00	−3.04	−3.34
CH_2=**C**^{x}– (alk-1-ene)	—	−0.34	−1.16	—
–C=**C**^{x}–CH_3 (alk-2-ene)	—	1.42	0.91	—
–**C**^{x}=C–CH_3 (alk-2-ene)	—	0.32	−0.17	—
–C=**C**^{x}– (general alkene)	2.42	0.58	−0.36	—
–C=C–**C**^{x}–	−0.72	−0.83	−0.91	−0.79
–C=C–C–**C**^{x}–	0.12	−0.15	−0.32	−0.72
–C=C–C–C–**C**^{x}–	−0.07	−0.16	−0.29	−0.24
–C≡**C**^{x}–	3.06	2.19	—	—
–C≡C–**C**^{x}–	−2.02	−1.44	−1.34	−1.51
–C≡C-C–**C**^{x}–	0.08	−0.11	−0.35	−0.94
–C≡C–C–C–**C**^{x}–	−0.20	−0.31	−0.38	−0.28
–C=**C**^{x}=C–	—	1.64	—	—
–C=C=**C**^{x}–	0.11	−1.88	−2.59	—
–C=C=C–**C**^{x}–	−0.67	−1.13	−1.70	−2.21
–C=C=C–C–**C**^{x}–	−0.13	−0.21	−0.78	−1.04
–C=C=C–C–C–**C**^{x}–	−0.11	−0.16	−0.22	−0.58

Table 2 Relative contributions of unsaturated functional groups to ^{1}H NSOR for different atom types depending on their distance from the chromophore and position of the nucleus around the double bond with respect to the carbon chain. The values are reported in μrad mol^{-1} dm^3 cm^{-1} for hydrogen nucleus marked H^x

	Atom type of **H**^{x}		
Chromophore	E	C	T
Alkanes	1.61	1.68	1.91
–C=C**H**^{x} *cis*	0.11	−0.15	—
–C=C**H**^{x} *trans*	−0.32	−0.35	—
–C=C**H**^{x} *cis* + *trans*	−0.03	−0.03	—
–C=C**H**^{x} *unsubstituted*	—	−0.39	—
–C=C–C**H**^{x}	−0.13	−0.05	−0.14
–C=C–C–C**H**^{x}	−0.04	−0.05	−0.07
–C=C–C–C–C**H**^{x}	−0.01	0.03	0.02
–C≡C**H**^{x}	0.09	0.00	—
–C≡C–C**H**^{x}	−0.08	−0.15	0.02
–C≡C–C–C**H**^{x}	−0.08	0.03	−0.23
–C≡C–C–C–C**H**^{x}	−0.03	0.00	0.04
–C=C=C**H**^{x}	−0.15	−0.21	—
–C=C=C–C**H**^{x}	−0.07	−0.03	−0.27
–C=C=C–C–C**H**^{x}	−0.03	−0.06	0.00
–C=C=C–C–C–C**H**^{x}	−0.01	−0.03	0.00

1. Identify the atom type (**E**, **C**, **T** or **Q**) for the nucleus of interest. Assign the nucleus the baseline NSOR value corresponding to its atom type in alkanes.

2. Find chromophores near the nucleus, at least up to two bonds away.

3. Add all relevant corrections corresponding to the atom type, chromophore, and the distance between them. Alkane substituents do not contribute.

3.4.1 Validation of the additive model. To assess the robustness of the model and its ability to estimate NSOR, several hydrocarbons not included in the training set have been chosen as test cases (Fig. 7). The molecules were selected rather arbitrarily, with the goal of having several structural features in combinations not present in the original set.

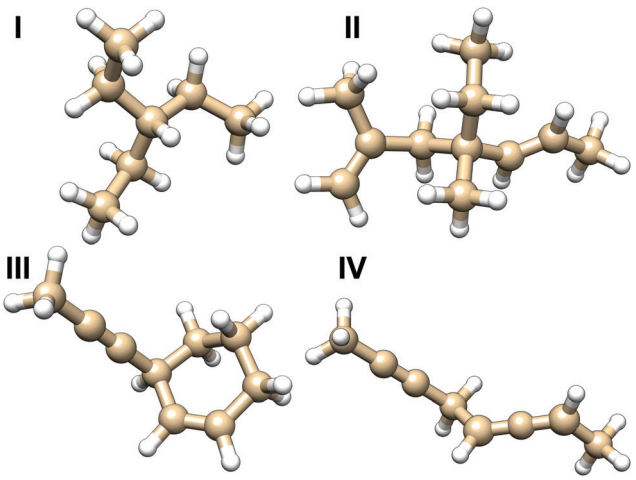

Fig. 7 Molecules used for testing the additive model of NSOR. The results for different variations of molecule **I** are in Table 3, and for molecules **II**, **III** and **IV** in Fig. 8.

Table 3 Comparison of ^{13}C NSOR values (in μrad mol^{-1} dm^3 cm^{-1}) of the central carbon atom in different variations of test system I obtained from full calculation of the molecule (Full) and estimated by adding contributions of the substituents to the baseline alkane value T = −3.04 μrad (Model). The relative error is calculated as Δ_r = |(Model − Full)/Full|

	Substituent			Contribution					
Molecule	S_1	S_2	S_3	S_1	S_2	S_3	Model	Full	Δ_r%
$\mathbf{I}_1$	–CH_2–CH_3	–CH_2–CH_3	–CH_2–CH_3	—	—	—	−3.04	−3.34	9
$\mathbf{I}_2$	–CH_2–CH_3	–CH_2–CH_3	–CH=CH_2	—	—	−0.91	−3.96	−3.95	0
$\mathbf{I}_3$	–CH_2–CH_3	–CH=CH_2	–CH=CH_2	—	−0.91	−0.91	−4.87	−4.80	2
$\mathbf{I}_4$	–CH=CH_2	–CH=CH_2	–CH=CH_2	−0.91	−0.91	−0.91	−5.79	−5.02	15
$\mathbf{I}_5$	–CH=CH_2	–CH=CH_2	–C≡CH	−0.91	−0.91	−1.34	−6.21	−5.12	21
$\mathbf{I}_6$	–CH_2–CH_3	–CH_2–CH_3	–C≡CH	—	—	−1.34	−4.38	−4.46	2
$\mathbf{I}_7$	–CH_2–CH_3	–CH=CH_2	–C≡CH	—	−0.91	−1.34	−5.29	−5.14	3
$\mathbf{I}_8$	–CH_2–CH_3	–C≡CH	–C≡CH	—	−1.34	−1.34	−5.72	−5.03	14
$\mathbf{I}_9$	–CH_2–CH_3	–CH=CH_2	–CH=C=CH_2	—	−0.91	−1.70	−5.66	−5.41	5
$\mathbf{I}_{10}$	–CH_2–CH_3	–CH=C=CH_2	–CH=C=CH_2	—	−1.70	−1.70	−6.45	−5.76	12
$\mathbf{I}_{11}$	–CH=C=CH_2	–CH=C=CH_2	–CH=C=CH_2	−1.70	−1.70	−1.70	−8.16	−7.23	13

The NSOR values for all nuclei were calculated using the additive model and compared with the results obtained from full quantum-chemical calculations on optimized structures.

The molecule **I** was chosen as the first test case to benchmark the model. Several structures were built by replacing some of the ethyl substituents attached to the central carbon with different side chains. The created structures are shown in Table 3 together with comparison of NSOR obtained from full calculation and from the additive model. Overall a reasonable agreement is found, with the largest relative difference between the model and the full calculation being about 21%. This is a rather good result considering the simplicity of the model with respect to full quantum-chemical computation. The error increases with the number of attached unsaturated chromophores, which is expected as the molecules in the training set contained only a single chromophore each.

The molecules **II**, **III** and **IV** feature a wider range of atom types at different distances from the chromophores. Their NSOR values obtained from full calculation and from the contribution model are shown in Fig. 8. The nuclei are numbered in this case with two indices, representing the distance from the first and the second chromophore, indicated for each molecule in the figure by red and orange, respectively. For each atom type the values obtained from full calculation are shown as black bars, and from the approximate model in green. The error bars show estimate of the precision of the values in the model, calculated for each nucleus as $\sigma_{\text{nucleus}} = \sqrt{\sigma_{c1}^2 + \sigma_{c2}^2}$, where σ_{c1}, σ_{c2} are the standard sample deviations of NSOR of the chromophores calculated for the corresponding atom type and distance. The coloured stacked bars show the baseline NSOR of alkane in blue and contributions from the two unsaturated groups in red and orange. Missing values are due to the particular carbons having no hydrogens attached, or more chemically inequivalent hydrogens being attached to the same carbon.

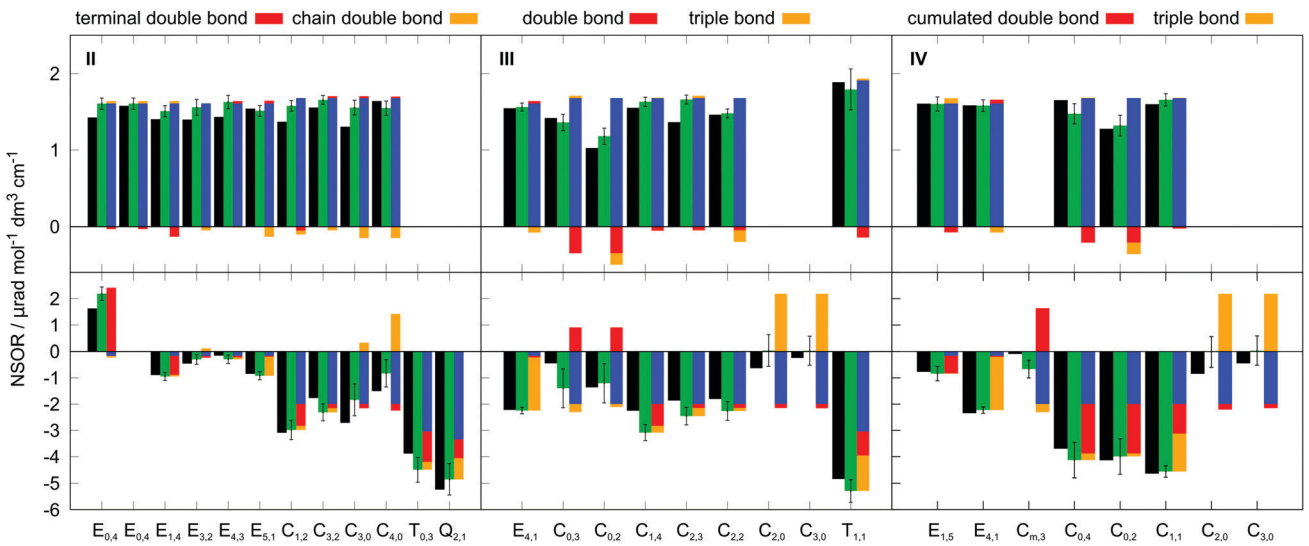

Fig. 8 Results of calculations of ^{1}H (top) and ^{13}C (bottom) NSOR for the test molecules **II** (left), **III** (middle) and **IV** (right). The black bars show NSOR calculated directly for the whole molecule, and the green bars correspond to the estimates from the additive model, together with the standard deviation. The stacked bars show the contributions from which the green bar was constructed: the blue bars represent the baseline values of alkanes, and the red and orange are contributions from the chromophores labeled above each panel. The subscripts in nucleus labels indicate the distance from the chromophores in order or red contribution, orange contribution.

Overall a good agreement is found for all three molecules. Similarly as in all other cases, the ^{13}C NSOR covers a wider range of values compared to ^{1}H, which is dominated by the alkane contribution and is only slightly changed by the influence of the chromophores. On the other hand, ^{13}C nuclei close to the unsaturated systems see significant shifts from the alkane baseline. In general, the corrections shift the NSOR towards the values obtained from the full calculations for majority of cases. The most notable exceptions are the $\mathbf{C}_{3,0}$ and $\mathbf{C}_{4,0}$ nuclei in molecule II. This disagreement can be attributed to the discussed wide spread of NSOR values of $\mathbf{C}_0$ in alkenes, which can shift towards both positive and negative with respect to alkanes. In this particular case the influence of the double bond is in the opposite direction than the approximate average contribution for $\mathbf{C}_0$ used in the model. This is also reflected in the large error bar for these nuclei. Nevertheless, the model provides a reasonable agreement with the full quantum-chemical calculation and can be used for pre-screening estimation of NSOR in molecules of potential experimental interest before expensive response function-based computations.

4 Conclusion

Nuclear spin-induced optical rotation (NSOR) of ^{1}H and ^{13}C nuclei was investigated by density functional theory computations on a set of hydrocarbon molecules with various structural motifs. The results were discussed with respect to the local chemical environment of the nuclei. It is shown that NSOR is sensitive in a systematic manner to the position of the nucleus in the carbon backbone, isomerism, presence of a particular chromophore in the molecule and the distance of the nucleus from it. Different combinations of these features give rise to a wide range of NSOR responses and characteristic spectroscopic features for molecular structure elucidation. Further patterns might be revealed by extending the study beyond hydrocarbons to molecules containing functional groups with heteroatoms.

In addition, it has been demonstrated that NSOR can be to a good approximation modelled as an additive property, where the signal of each nucleus can be constructed by adding contributions from the neighbouring chemical groups. This allows one to quickly estimate NSOR with a reasonable accuracy without a need for extensive quantum-chemical calculations.

The results provide a proof of concept that the NSOR effect correlates with the underlying molecular structure and therefore could be experimentally exploited as a basis for a viable spectroscopic method.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The author is grateful for the financial support from Academy of Finland (Grant 316180). The support from the Kvantum institute (University of Oulu) and grants of computer capacity from the Finnish Grid and Cloud Infrastructure (persistent identifier urn:nbn:fi:research-infras-2016072533) are also acknowledged. The author thanks Sarah Mailhiot for reviewing the manuscript for clarity.

Notes and references

1 J. Keeler, *Understanding NMR Spectroscopy*, John Wiley & Sons, 2010.
2 P. T. Callaghan, *Translational Dynamics and Magnetic Resonance*, Oxford University Press, 2011.
3 E. Haake, R. Brown, M. Thompson and R. Venkatesan, *Magnetic Resonance Imaging. Physical Principles and Sequence Design*, John Wiley & Sons, Inc., 1999.
4 I. M. Savukov, S. Lee and M. V. Romalis, *Nature*, 2006, **442**, 1021–1024.
5 D. Pagliero, W. Dong, D. Sakellariou and C. A. Meriles, *J. Chem. Phys.*, 2010, **133**, 154505.
6 D. Pagliero and C. A. Meriles, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 19510–19515.
7 I. M. Savukov, H. Chen, T. Karaulanov and C. Hilty, *J. Magn. Reson.*, 2013, **232**, 31–38.
8 Y. Zhu, Y. Gao, S. Rodocker, I. Savukov and C. Hilty, *J. Phys. Chem. Lett.*, 2018, **9**, 3323–3327.
9 I. Savukov, D. Filin, Y. Zhu, R. Castro and C. Hilty, *Eur. Phys. J. D*, 2019, **73**, 156.
10 J. Shi, S. Ikäläinen, J. Vaara and M. V. Romalis, *J. Phys. Chem. Lett.*, 2013, **4**, 437–441.
11 P. Štěpánek and A. M. Kantola, *J. Phys. Chem. Lett.*, 2019, **10**, 5458–5462.
12 S. Ikäläinen, P. Lantto and J. Vaara, *J. Chem. Theory Comput.*, 2012, **8**, 91–98.
13 T. S. Pennanen, S. Ikäläinen, P. Lantto and J. Vaara, *J. Chem. Phys.*, 2012, **136**, 184502.
14 S. Ikäläinen, M. V. Romalis, P. Lantto and J. Vaara, *Phys. Rev. Lett.*, 2010, **105**, 153001.
15 G. Yao, M. He, D. Chen, T. He and F. Liu, *Chem. Phys.*, 2011, **387**, 39–47.
16 J. Vähäkangas, P. Lantto and J. Vaara, *J. Phys. Chem. C*, 2014, **118**, 23996–24005.
17 F. Chen, G. Yao, T. He, D. Chen and F. Liu, *Chem. Phys.*, 2014, **453**, 57–61.
18 Y. Zhu, C. Hilty and I. Savukov, *eMagRes*, 2019, **8**, 205–214.
19 J. Vaara, A. Rizzo, J. Kauczor, P. Norman and S. Coriani, *J. Chem. Phys.*, 2014, **140**, 134103.
20 M. Straka, P. Štěpánek, S. Coriani and J. Vaara, *Chem. Commun.*, 2014, **50**, 15228–15231.
21 F. Chen, G. Yao, Z. Zhang, F. Liu and D. Chen, *ChemPhysChem*, 2015, **16**, 1954–1959.
22 P. Štěpánek, S. Coriani, D. Sundholm, V. A. Ovchinnikov and J. Vaara, *Sci. Rep.*, 2017, **7**, 46617.
23 P. Štěpánek and S. Coriani, *Phys. Chem. Chem. Phys.*, 2019, **21**, 18082–18091.
24 T. Lu, M. He, D. Chen, T. He and F. Liu, *Chem. Phys. Lett.*, 2009, **479**, 14–19.

25 L. Fu and J. Vaara, *J. Chem. Phys.*, 2013, **138**, 204110.
26 L. Fu and J. Vaara, *ChemPhysChem*, 2014, **15**, 2337–2350.
27 G. Yao, M. He, D. Chen, T. He and F. Liu, *ChemPhysChem*, 2012, **13**, 1325–1331.
28 L. Fu and J. Vaara, *J. Chem. Phys.*, 2014, **140**, 024103.
29 L. Fu, A. Rizzo and J. Vaara, *J. Chem. Phys.*, 2013, **139**, 181102.
30 R. Ahlrichs, M. Bär, M. Häser, H. Horn and C. Kölmel, *Chem. Phys. Lett.*, 1989, **162**, 165–169.
31 TURBOMOLE V7.4 2019, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
32 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
33 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
34 A. Schäfer, C. Huber and R. Ahlrichs, *J. Phys. Chem.*, 1994, **100**, 5829.
35 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
36 DALTON, a molecular electronic structure program, Release DALTON2017 (2017), see http://daltonprogram.org.
37 K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekstrom, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernandez, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hattig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. HjertenÃ¦s, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. KjÃ¦rgaard, W. Klopper, S. Knecht, R. K. H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. LutnÃ¦s, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawlowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Sałek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski and H. Ågren, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 269–284.
38 P. Sałek, O. Vahtras, T. Helgaker and H. Ågren, *J. Chem. Phys.*, 2002, **117**, 9630.
39 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.
40 P. Manninen and J. Vaara, *J. Comput. Chem.*, 2006, **27**, 434–445.
41 J. Vähäkangas, S. Ikäläinen, P. Lantto and J. Vaara, *Phys. Chem. Chem. Phys.*, 2013, **15**, 4634–4641.
42 L. Frediani, H. Ågren, L. Ferrighi and K. Ruud, *J. Chem. Phys.*, 2005, **123**, 144117.
43 P. Štěpánek, Dataset: Nuclear spin-induced optical rotation of functional groups in hydrocarbons, http://urn.fi/urn:nbn:fi:att:ff3a6b18-e02d-4df1-b7a3-2b4295df4055, 2020.
44 E. F. Pettersen, T. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and R. E. Ferrin, *J. Comput. Chem.*, 2004, **13**, 1605–1612.