PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Phys. Chem. Chem. Phys.*, 2022, **24**, 20164

Prediction of anisotropic NMR data without knowledge of alignment medium structure by surface decomposition†

Yizhou Liu, *[a] Ikenna E. Ndukwe, ‡[b] Mikhail Reibarkh,[b] Gary E. Martin §[b] and R. Thomas Williamson ¶[b]

Received 9th June 2022,
Accepted 11th August 2022

DOI: 10.1039/d2cp02621j

rsc.li/pccp

Prediction of anisotropic NMR data directly from solute-medium interaction is of significant theoretical and practical interest, particularly for structure elucidation, configurational analysis and conformational studies of complex organic molecules and natural products. Current prediction methods require an explicit structural model of the alignment medium: a requirement either impossible or impractical on a scale necessary for small organic molecules. Here we formulate a comprehensive mathematical framework for a parametrization protocol that deconvolutes an arbitrary surface of the medium into several simple local landscapes that are distributed over the medium's surface by specific orientational order parameters. The shapes and order parameters of these local landscapes are determined *via* fitting that maximizes the congruence between experimentally determined anisotropic NMR measurables and their predicted counterparts, thus avoiding the need for an *a priori* knowledge of the global medium morphology. This method achieves substantial improvements in the accuracy of predicted anisotropic NMR values compared to current methods, as demonstrated herein with sixteen natural products. Furthermore, because this formalism extracts structural commonalities of the medium by combining anisotropic NMR data from different compounds, its robustness and accuracy are expected to improve as more experimental data become available for further re-optimization of fitting parameters.

1. Introduction

Anisotropic NMR data from solutes dissolved in nematic liquid crystalline (LC) solvents or dilute lyotropic LC (LLC) solutions have long been utilized for a variety of scientific purposes, including structural geometry determination of small solute molecules,[1] physical chemical investigations on intermolecular interactions and LC phase transitions,[2] conformational determination of biomolecules,[3] and stereochemical and constitutional studies of complex organic molecules and natural products.[4–6] Theoretical prediction of anisotropic NMR parameters based on solute–mesogen interactions is also of long-standing interest to the scientific community, playing an important role in understanding intermolecular interactions in nematic phases[7–9] and studying molecular geometries when experimental determination of solute order parameters, in particular those based on the singular value decomposition (SVD) method,[10] is ineffective. Examples of the latter include, characterizing a homo-oligomer,[11,12] analysing conformational dynamics of flexible molecules,[13–18] and structure determination using extremely sparse experimental data.[19] Solvated polymeric gels are also widely utilized as alignment media for solute geometry determination,[20,21] but theoretical investigation into solute–gel interactions involved in molecular alignment is rarely explored.[22]

Orientational order of a solute can arise from a combination of steric, electrostatic, and dispersive interactions with the alignment medium. Although highly desirable from both a theoretical and practical standpoint, deconvoluting the relative importance of different interactions in a real alignment system is extremely challenging. In particular, the partition function of a many-body interacting system cannot be factored into individual partition functions associated with specific energy types.

[a] *Analytical Research and Development, Pfizer Worldwide Research and Development, 445 Eastern Point Road, Groton, CT, 06340, USA. E-mail: Yizhou.Liu@pfizer.com*

[b] *Analytical Research and Development, Merck & Co. Inc., 126 E. Lincoln Ave., Rahway, NJ, 07065, USA*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2cp02621j

‡ Current address: Pivotal Attribute Sciences, Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA 91320, USA.

§ Current address: Seton Hall University, Department of Chemistry and Biochemistry, 400 South Orange Ave., South Orange, NJ 07079, USA.

¶ Current address: University of North Carolina Wilmington, Department of Chemistry and Biochemistry, 5600 Marvin K. Moss Lane, Wilmington, NC 28409, USA.

Consequently, energy components of different interactions have intertwined effects on order parameters. Over the past three decades, various experimental investigations[23–26] and computational simulations[27–31] have primarily concluded that steric interactions tend to represent a dominant mechanism of solute ordering in apolar nematic solvents while electrostatic interactions are less impactful (note: apolarity by the LC nomenclature refers to the head-tail symmetry of a nematic phase instead of no electric dipole moment in the nematogen). In a somewhat parallel line of research, anisotropic NMR data prediction of macromolecules ordered in dilute aqueous LC systems have also attracted substantial interest and found useful applications in the past two decades. The prevalent prediction method called PALES originally considered only the steric obstruction mechanism but nevertheless yielded remarkably accurate predictions even for charged proteins dissolved in neutral phospholipid bicelle solutions.[11] A later expanded version of PALES incorporated both steric and electrostatic interactions to predict protein/DNA alignment in dilute charged LC media such as the mesophase of the filamentous phage Pf1, in which electrostatic interaction plays a more significant role than in charge-neutral media.[32]

In light of these earlier studies, our goal in this article was to formulate a method to overcome the current challenges in modelling steric interaction. Accurate account of steric contribution to solute ordering is not only important for its own sake, but also provides a solid basis for further evaluation of electrostatic and dispersive components. Although a mean-field anisotropic potential resulting from decoupling of the relative position and orientation of two interacting bodies is often used to approximate long-range interactions, applying this approach to short-range steric interaction can be problematic due to the strong interdependence of position and orientation.[9] Methods that have proven effective so far resort to numeric calculation of solute order parameters by integrating all solute positions and orientations near a medium structural model, which avoids decoupling the positional and orientational variables or using any mean-field potentials.[11,19,27] A major limitation of this approach, however, is that an appropriate geometric model for the medium must be available. For large biomolecules, a highly simplified medium model often suffices, such as an infinite plane for the phospholipid bicelle medium or a cylinder for a rod-shaped medium like bacteriophage as used in PALES. Small organic compounds, however, can sample much finer structural details on the medium surface, thus demanding more comprehensive medium models. Clearly, MD simulations or force-field calculations are promising approaches as atomistic models can be employed for both medium and solute. Unfortunately, an atomistic model may not be feasible for some alignment media. For example, many types of polymeric gels have stereocenters with undefined chirality along the chain. If each stereocenter has an equal probability for *R* and *S* as in an atactic polymer, even a short six-unit fragment can have forty-eight stereo-configurations (considering head-tail symmetry of polymers having opposite end-chain chiralities), each of which can interact differently with a solute. Obviously, modelling all possible stereoisomers at the atomistic level will be computationally impractical. The situation is exacerbated by conformational heterogeneity due to lack of regular secondary structure in the gel polymer. With these variabilities, a microscopic model is out of the question even for a fully relaxed gel, not to mention a gel deformed by mechanical strain. Furthermore, unlike in the LC mesophase, the anisotropic phenomenon of a strained gel lacks quantitative descriptors, obfuscating the physical picture of how anisotropy can be further transferred to a solute. It is tempting to depict the anisotropy of a strained gel by analogy to that of the LC, but fundamental differences between them forbid direct theoretical translation. For example, the LC is a spatially homogeneous solution of repeating structural units (mesogens), whereas the solvated gel is a heterogeneous environment containing an immobile solid phase with no apparent repeating structural units. Consequently, certain key elements in the LC theory, such as mean-field approximation and description of system anisotropy by mesogen order parameters, are not applicable to the gel.

Herein, we develop a decomposition method that approximates the gel surface manifold as a positional and orientational arrangement of several repetitive local landscapes, with each having a different shape. This decomposition allows introduction of order parameters, defined for each local landscape based on its orientational distribution on the entire gel surface, into the description of the gel anisotropy, thus leading to an orientational order transfer formalism similar to that of the LC system. As far as anisotropic NMR prediction is concerned, only the five rank two order parameters of each landscape are relevant. In the special case of a rod-shaped mesogen or a disc-shaped mesogen aggregate (*e.g.*, a lipid bicelle), only one rank two order parameter that describes the alignment amplitude is sufficient thanks to their axial symmetry. Here, however, we do not impose cylindrical symmetry on local landscapes, thus allowing more diverse landscape types to be generated. With proper parametrization of landscape shape, two or three rank two order parameters and two or three shape parameters are required for each landscape, depending on whether medium chirality is considered. Thus, the surface decomposition method avoids not only the inadequacy of an oversimplified model that has only one landscape of axial symmetry such as a rod or a plane, but also the unmanageable complexity of an atomistic medium model that would require an astronomical number of parameters to be considered if completely unknown. The total number of local landscapes required for an alignment medium and their shapes and order parameters can be determined and optimized by maximizing the agreement between predicted and measured anisotropic NMR data from experimental databases and performing statistical analysis on the outcome. The modeled alignment medium can be used to predict the expected anisotropic NMR data of other compounds of interest and then compared with their experimental values acquired in the same type of alignment medium to confirm or refute proposed compound structures. The surface decomposition concept was examined using

experimental data from sixteen complex natural products, with results displaying increasingly improved agreement as higher levels of parametrization, *i.e.* larger numbers of local landscapes, are used. The statistical significance of employing high levels of parametrization is analysed by *F*-tests. The robustness of this method was further cross-validated by leaving out data points of two challenging molecules during the parameter optimization stage. The anisotropic NMR data for these two molecules were predicted with the optimized parameters using the reduced training dataset and compared to experimental data. The results from this cross-validation confirms good prediction accuracy and tolerance for over-parametrization of the surface decomposition method proposed. Direct order parameter prediction with the overall degree of accuracy observed in this work promises to overcome a known limitation in SVD based analyses when applied to some aforementioned scenarios.

2. Theory

In Section 2.1, we first develop a formalism for parametrization of sterically-induced alignment in a dilute lyotropic LC (LLC) medium. Our derivation is based on classical molecular statistical theory of LCs and aims to establish a relationship between the order parameters of the solute and the mesogen, which we refer to as an order transfer equation (OTE). In Section 2.2, we develop the theoretical counterpart for sterically-induced alignment in a polymeric gel. Proof for all equations in the theory section is given in the ESI.†

2.1 Liquid crystal with a mesogen model

We consider a simplified LC alignment system containing N mesogens and only one solute, ignoring solute–solute interaction. We also assume that mesogen ordering is not perturbed by the solute at a low concentration. Deduction based on the molecular statistical theory under the usual mean-field approximation leads to the following equation for the orientational distribution function (ODF) of the solute:

$$f_s(\boldsymbol{a}) \approx \frac{[V - \langle V_{ex}(\boldsymbol{a})\rangle]^N}{\int d\boldsymbol{a}[V - \langle V_{ex}(\boldsymbol{a})\rangle]^N} \quad (1)$$

The ODF $f_s(\boldsymbol{a})$ describes the probability density of finding the solute at an orientation $\boldsymbol{a}$ with respect to (wrt) the laboratory frame. V is the total sample volume. $\langle V_{ex}(\boldsymbol{a})\rangle$ is the average excluded volume between one solute and one mesogen, which can be considered the average volume excluded from a mesogen with an ODF of $f_m(\boldsymbol{a}_m)$ due to the presence of a solute at orientation $\boldsymbol{a}$, or equivalently the average volume excluded from a solute at orientation $\boldsymbol{a}$ due to the presence of a mesogen with an ODF of $f_m(\boldsymbol{a}_m)$:

$$\langle V_{ex}(\boldsymbol{a})\rangle = \int d\boldsymbol{a}_m f_m(\boldsymbol{a}_m)\int d\boldsymbol{r}\{1 - \exp[-\beta U^{HB}(\boldsymbol{a}, \boldsymbol{a}_m, \boldsymbol{r})]\} \quad (2)$$

Here, $\boldsymbol{a}_m$ is the mesogen orientation wrt the laboratory frame. $f_m(\boldsymbol{a}_m)$ is the mesogen's ODF. $\boldsymbol{r}$ is the solute–mesogen relative position vector. β is the inverse product of Boltzmann constant and the temperature, $(k_BT)^{-1}$. $U^{HB}(\boldsymbol{a}, \boldsymbol{a}_m, \boldsymbol{r})$ is the hard-body potential, which is infinite if the solute and mesogen overlap, or zero otherwise, thus leading to a "0 or 1" binary outcome for the Boltzmann factor $\exp[-\beta U^{HB}(\boldsymbol{a}, \boldsymbol{a}_m, \boldsymbol{r})]$. Since the denominator in eqn (1) is a normalization factor independent of $\boldsymbol{a}$, eqn (1) shows that the solute prefers an orientation that minimizes $\langle V_{ex}(\boldsymbol{a})\rangle$. Note that $\langle V_{ex}(\boldsymbol{a})\rangle$ is on the order of single molecular volume, which is much smaller than V. For a dilute LLC solution, if we further assume $N\langle V_{ex}(\boldsymbol{a})\rangle \ll V$ such that $\left[1 - \frac{\langle V_{ex}(\boldsymbol{a})\rangle}{V}\right]^N \approx 1 - \frac{N\langle V_{ex}(\boldsymbol{a})\rangle}{V}$ by truncation after the 1st order term, we obtain:

$$\left.\begin{aligned} f_s(\boldsymbol{a}) &\approx \frac{V - N\langle V_{ex}(\boldsymbol{a})\rangle}{8\pi^2(V - NV_{ex}^r)} \\ V_{ex}^r &= \frac{1}{8\pi^2}\int d\boldsymbol{a}\langle V_{ex}(\boldsymbol{a})\rangle \end{aligned}\right\} \quad (3)$$

V_{ex}^r is $\langle V_{ex}(\boldsymbol{a})\rangle$ averaged over all solute orientations, which is equivalent to the excluded volume calculated as if the mesogen or the solute are randomly rotating around each other (see the ESI†). Calculating V_{ex}^r between two arbitrary anisotropic bodies is a well-known challenge.[33] Here, we can ignore the effect of V_{ex}^r since the dilute LLC approximation satisfying $NV_{ex}^r \ll V$ leads to:

$$f_s(\boldsymbol{a}) \approx \frac{1}{8\pi^2} - \frac{N}{V}\frac{\langle V_{ex}(\boldsymbol{a})\rangle}{8\pi^2} \quad (4)$$

Not surprisingly, $\langle V_{ex}(\boldsymbol{a})\rangle$ remains the most critical part. Noting that the steric potential only depends on relative orientation and position, we can replace the integration over the mesogen orientation wrt the laboratory-frame ($\boldsymbol{a}_m$ in eqn (2)) with that over its orientation wrt the solute frame, thus obtaining an alternative form for eqn (2):

$$\langle V_{ex}(\boldsymbol{a})\rangle = \int d\boldsymbol{\Omega} f_m(\hat{R}\boldsymbol{a})\int d\boldsymbol{r}\{1 - \exp[-\beta U^{HB}(\boldsymbol{\Omega}, \boldsymbol{r})]\} \quad (5)$$

The orientational relations between the solute, medium and the laboratory-frame are depicted in Fig. 1. Here we focus on the uniaxial order scenario, because the biaxial mesophase is rarely used for NMR. The mesogen orientation can then be defined by the position of the uniaxial director (red vector $\boldsymbol{a}_m$) in the mesogen frame (blue frame) using polar and azimuthal angles θ_M and φ_M (Fig. 1a). Likewise, the solute orientation can be defined by the director position in the solute frame (green frame $\boldsymbol{a}$) using θ_S and φ_S (Fig. 1b). The uniaxial director in most experiments coincides with the laboratory-frame Z axis, except for a variable-angle NMR experiment with a spinning LC sample[34] or a constrained gel.[35] The relative orientation is depicted by Euler angles α, β, and γ that rotate the mesogen frame onto the solute frame by the intrinsic Z–Y'–Z'' convention (Fig. 1c). Accordingly, the integral $\int d\boldsymbol{\Omega}$ is given by $\int_0^{2\pi} d\alpha \int_0^{\pi} d\beta \sin\beta \int_0^{2\pi} d\gamma$. The director position in mesogen and solute frames is related by $\boldsymbol{a}_m = \hat{R}\boldsymbol{a}$, where the rotation operator $\hat{R}$ corresponds to the solute–mesogen frame transformation. Substituting eqn (5) into eqn (4), performing multipole

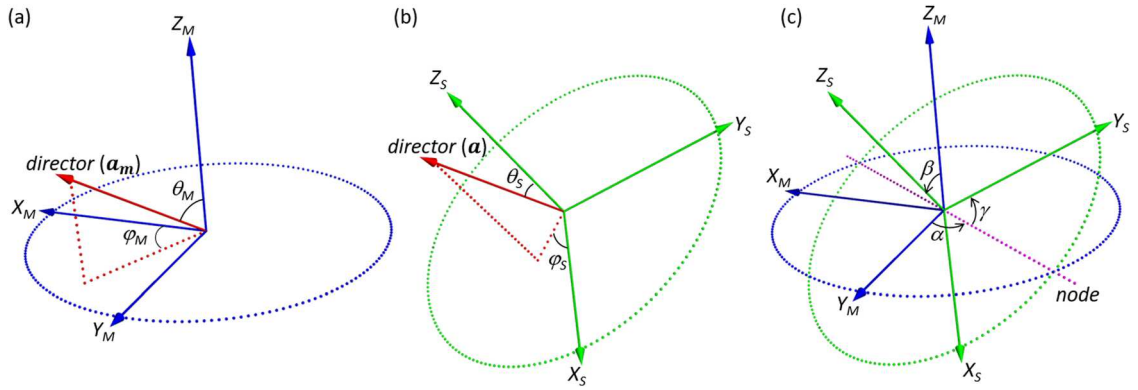

Fig. 1 Orientational relationships of the mesogen and the solute in the laboratory frame. (a) orientation of the mesogen in the laboratory frame, described by the director vector $\boldsymbol{a}_m$; (b) orientation of the solute in the laboratory frame, described by the director vector $\boldsymbol{a}$; (c) relative orientation $\boldsymbol{\Omega}$ between the mesogen and the solute, described by Euler angles α, β, and γ.

expansion for both $f_s(\boldsymbol{a})$ and $f_m(\hat{R}\boldsymbol{a})$, and equating the coefficients of the 2nd order spherical harmonics, yields the OTE below:

$$\sqrt{\frac{4\pi}{5}}\langle Y_2^m\rangle_S \approx \sum_{m'=-2}^{2}\sqrt{\frac{4\pi}{5}}\left\langle Y_2^{m'}\right\rangle_M\left\{\frac{N}{8\pi^2 V}\int d\boldsymbol{\Omega} D^2_{m'm}(\hat{R}) \times \int d\boldsymbol{r}\exp\left[-\beta U^{HB}(\boldsymbol{\Omega},\boldsymbol{r})\right]\right\} \quad (6)$$

The coefficients $\sqrt{\frac{4\pi}{5}}\langle Y_2^m\rangle_S$ and $\sqrt{\frac{4\pi}{5}}\langle Y_2^{m'}\rangle_M$ are the rank two order parameters of the solute and the mesogen, respectively, where the angle bracket denotes orientational averaging. The connection between solute and mesogen order parameters given in eqn (6) is useful because it shows how orientational order is transferred from the mesogen to the solute through their anisotropic interaction potential. We should point out that the relatively simple relationship shown in eqn (6) is made possible by the dilute LLC approximation; otherwise, their connection is rather complicated, involving combinatorial terms of order parameters of different ranks (see the ESI†). Because anisotropic NMR only reports on the 2nd moment of the ODF, we focus on rank two (l = 2) order parameters here, although eqn (6) is also valid for higher ranks. As shown in the ESI,† the order parameter here is in fact the complex conjugate of the conventional definition, chosen as such to simplify the notation. Rod or disc like mesogens have axisymmetric order, for which only the $\sqrt{\frac{4\pi}{5}}\langle Y_2^0\rangle_M$ term, *i.e.*, $\frac{1}{2}(3\cos^2\theta_M - 1)$, is relevant leading to further simplification, but here we retain the general asymmetric version because of its connection to the surface decomposition method for polymeric gels to be introduced later. The Wigner D matrix $D^2_{m'm}(\hat{R})$ is in the α, β, and γ order (Fig. 1). Note that the double integral in the curly bracket is a volumetric quantity after normalization by $8\pi^2$, and thereby the entire item inside the curly bracket is a pure number. The double integral over $\boldsymbol{\Omega}$ and $\boldsymbol{r}$ has some remarkable characteristics. In the isotropic sample space where all relative orientations are allowed, the Boltzmann factor is 1 and the integral is zero due to the symmetry of $D^2_{m'm}(\hat{R})$. This integral is also zero at positions where overlapping occurs at all orientations because the Boltzmann factor is zero. This integral builds up only over regions with partially allowed orientations. These characters simply reflect that the solute acquires ordering only within a shell immediately outside the mesogen's van der Waals (vdw) surface, as expected. The spherical harmonic order parameters are related to the Saupe ordering matrix, which is more frequently used in NMR studies, by eqn (7).

$$\left.\begin{aligned}
S_{xx} &= \sqrt{\frac{3\pi}{10}}\left(\langle Y_2^2\rangle + \langle Y_2^{-2}\rangle\right) - \frac{1}{2}\langle Y_2^0\rangle\\
S_{yy} &= -\sqrt{\frac{3\pi}{10}}\left(\langle Y_2^2\rangle + \langle Y_2^{-2}\rangle\right) - \frac{1}{2}\langle Y_2^0\rangle\\
S_{zz} &= \langle Y_2^0\rangle\\
S_{xy} &= S_{yx} = -i\sqrt{\frac{3\pi}{10}}\left(\langle Y_2^2\rangle - \langle Y_2^{-2}\rangle\right)\\
S_{xz} &= S_{zx} = -\sqrt{\frac{3\pi}{10}}\left(\langle Y_2^1\rangle - \langle Y_2^{-1}\rangle\right)\\
S_{yz} &= S_{zy} = i\sqrt{\frac{3\pi}{10}}\left(\langle Y_2^1\rangle + \langle Y_2^{-1}\rangle\right)
\end{aligned}\right\} \quad (7)$$

All anisotropic NMR parameters can be readily calculated from the Saupe ordering matrix.

Clearly, evaluating the double integral in eqn (6) is a crucial step in predicting solute alignment. Although implemented somewhat differently in the original reports,[11,19] the numerical simulation strategy as used in PALES, P3D, and other related methods can be understood based on the following procedure. First, we identify the minimal enclosing ball (MEB) of the solute, which is the smallest sphere that encloses the solute's vdw volume. Then we consider the process of rolling the MEB over the medium's vdw surface (vdwS, shown as the innermost purple surface in Fig. 2) and mapping out the closed surface traced by the MEB center. We refer to this surface as the MEB-accessible surface (MEB-AS, shown as the outermost gray mesh in Fig. 2). Another surface resulting from this process is the

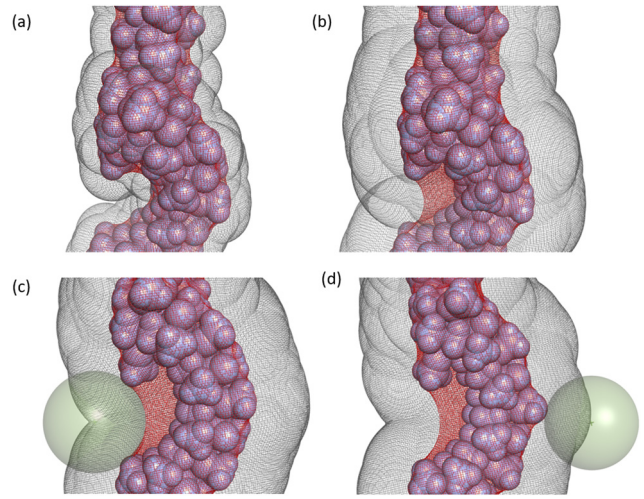

Fig. 2 Relationship between three surfaces: MEB-AS (gray), MEB-ES (red), and vdwS (purple). (a) and (b): comparison of MEB-AS and MEB-ES generated with r_{MEB} of 2 and 5 Å, respectively; (c): a pocket of high concave curvature on the vdwS is smoothed to a concave surface on the MEB-ES with a curvature of $-r_{MEB}^{-1}$; (d): a convex surface on the vdwS has the same curvature with the corresponding region on the MEB-ES.

MEB-excluded surface (MEB-ES, shown as a red mesh in Fig. 2), representing the mesogen's surface boundary against the MEB. Obviously, the MEB-AS and MEB-ES are identical to the well-known solvent-accessible surface (SAS) and solvent-excluded surface (SES), respectively, except that here the probe radius is the MEB radius (r_{MEB}) of the solute *in lieu* of the solvent. The MEB-AS and *vdw*S divide the space surrounding the mesogen into three zones. The space outside the MEB-AS is an isotropic zone where the solute freely rotates. As previously discussed, this zone is irrelevant for steric alignment calculation although it may be of interest for long-range electrostatic interactions. The space enclosed by the vdwS is a forbidden zone, which also makes no contribution to solute ordering. However, the space between the MEB-AS and vdwS constitutes the interaction zone, which encompasses the positional grids that a numeric simulation algorithm should sample. In the general case of an asymmetric mesogen, the double integral in eqn (6) is evaluated by a six-dimensional integration that fixes the orientation and position of the mesogen, moves the MEB center of the solute on the 3D grid points inside the interaction zone (three-dimensional Cartesian integration over $\boldsymbol{r}$) while uniformly sampling all solute orientations at each grid point (three-dimensional spherical integration over $\boldsymbol{\Omega}$), and sums up $D_{m'm}^{2}(\hat{R})\exp\left[-\beta U^{HB}(\boldsymbol{\Omega},\boldsymbol{r})\right]$ (only $D_{0m}^{2}(\hat{R})\exp\left[-\beta U^{HB}(\boldsymbol{\Omega},\boldsymbol{r})\right]$ in case of axisymmetric order) at each orientation–position combination. Integration over reduced positional dimensions is possible by taking advantage of mesogen structural symmetry, as in the cylindrical and planar models of PALES.

Note that the MEB-ES is not involved in these methods where explicit medium models are used, but it plays an essential role in our surface decomposition method as shown later. A serious limitation with these methods is that an appropriate mesogen model is needed, *i.e.*, the purple vdwS in Fig. 2 must be known during numeric integration. Unfortunately, even a qualitatively correct model can be unavailable for a polymeric gel due to its chemical complexity as previously mentioned. To address this issue, we will develop a parametrization-based formalism for alignment prediction in polymeric gels or mesogens of unknown structures.

2.2 Polymeric Gel

As the simplest approximation, we treat the cross-linked gel as a solid matrix of rigid structure. The effect of internal motions will be qualitatively discussed at the end of this section.

By analogy with eqn (1)–(3), we can obtain the following ODF for a solute in a constrained gel:

$$\left.\begin{aligned} f_s(\boldsymbol{a}) &= \frac{V - V_{ex}(\boldsymbol{a})}{8\pi^2\left(V - V_{ex}^{r}\right)} \\ V_{ex}(\boldsymbol{a}) &= \int d\boldsymbol{r}\left\{1 - \exp\left[-\beta U^{HB}(\boldsymbol{a},\boldsymbol{r})\right]\right\} \\ V_{ex}^{r} &= \frac{1}{8\pi^2}\int d\boldsymbol{a} V_{ex}(\boldsymbol{a}) \end{aligned}\right\} \quad (8)$$

There are some noteworthy differences between LLC's and gels that are manifested in these equations. First, the entire cross-linked gel is treated as one giant molecule that remains stationary in the laboratory frame, thus $N = 1$ and the mean-field approximation used in the LC system is no longer needed. Consequently, the equation for $f_s(\boldsymbol{a})$ in eqn (8) is exact. Second, unlike a LC mesogen whose position and orientation are in a dynamic equilibrium, the solid gel matrix is considered immobile, so the volume excluded from a solute of orientation $\boldsymbol{a}$ is not averaged over the medium's orientation, differently from eqn (5). Third, eqn (8) does not explicitly need the dilute medium approximation, *i.e.*, neither $V_{ex}(\boldsymbol{a}) \ll V$ nor $V_{ex}^{r} \ll V$ is required.

The primary focus of this work is to generate a method for calculating $V_{ex}(\boldsymbol{a})$ in the absence of a global medium model. The immediate challenge is that the gel surface, described by the vdwS, is unknown and potentially highly complex with various convex and concave features (see purple surface in Fig. 2). At least in principle, with an extensive experimental database, it should be possible to extract key structural parameters of the gel, which can in turn be used to predict the alignment of other compounds of interest. From a parametrization standpoint, it is highly desirable to decouple $V_{ex}(\boldsymbol{a})$ using parameters that depend only on the solute or the gel. Unfortunately, such decoupling is extremely challenging, in no small part due to the presence of concave features on the gel surface, and any method of practical utility likely entails drastic simplification or oversimplification. In the surface decomposition method, we approximate the unknown vdwS of the gel with a MEB-ES and further decompose the MEB-ES as an orientational-positional ensemble of a few representative local landscapes (LS_{rep}). Each LS_{rep} has a unique shape idealized as a paraboloid of two principal curvatures or a twisted paraboloid with an additional twist curvature (described in Section 3). Each LS_{rep} follows a certain orientational and positional distribution

on the gel surface. This treatment leads to tremendous variable deduction, because: (1) the positional distribution of a LS_{rep} does not affect solute orientational order; (2) only the 2nd moment of a LS_{rep}'s orientational distribution, involving at most five, and with appropriate landscape representation, only two or three order parameters, contributes to the observation of anisotropic NMR data.

MEB-ES is essentially a vdwS smoothed by a rolling sphere that has some remarkable properties associated with the pros and cons of this approach.

Property 1: The MEB-ES superimposes well with the vdwS in most cases, except at cavities of comparable size to the solute. For example, in Fig. 2a and b the MEB-ES (red) is generated over the same vdwS (purple) with a r_{MEB} of 2 and 5 Å, respectively. The MEB-ES and vdwS superimpose well in Fig. 2a, where the vdwS is either convex, concave with small cavities, or concave with a large cavity capable of fully accommodating the 2 Å MEB (see lower-left side). Note that the quality of superposition is relative to the MEB size, *i.e.*, deviation much smaller than r_{MEB} is not expected to cause significant error in $V_{ex}(\boldsymbol{a})$ calculation. In Fig. 2b, however, a larger MEB of 5 Å that is comparable in size to the lower-left cavity causes pronounced deviation between the MEB-ES and vdwS. Such deviation overestimates $V_{ex}(\boldsymbol{a})$ if the solute has high shape anisotropy that allows it to enter a cavity with certain but not all orientations, which is a shortcoming of our method. For example, a $2 \times 2 \times 5$ Å ellipsoid (r_{MEB} = 5 Å) can conceivably enter the lower-left cavity along its long axis, but such entry is denied by the corresponding MEB-ES (Fig. 2b). Despite this shortcoming, the MEB-ES is clearly a much more faithful representation of the vdwS than other far oversimplified models such as a cylinder.

Property 2: The MEB-ES is dependent on r_{MEB} but not the shape of the solute. This property is obvious because the MEB-ES is generated with a rolling sphere. As solute shape is decoupled from the MEB-ES, the mere dependence on r_{MEB} is relatively easy to cope with. For example, an average MEB-ES can be parametrized with a database containing solutes of similar sizes such that variations due to r_{MEB} are small. In this work, our database contains 16 natural products with r_{MEB} varying from 5.5 to 7.8 Å, covering the typical size range of small molecules. It is also worth noting that the dependence on r_{MEB} is small if the medium surface mostly has either very small or very large cavities, according to Property 1.

Property 3: The MEB-ES cannot have a concave curvature higher than ${r_{MEB}}^{-1}$. This is because any greater concave curvature on the vdwS is smoothed to ${r_{MEB}}^{-1}$ by the rolling MEB; see for example Fig. 2c. There is no such restriction for a convex curvature because rolling the MEB does not change convex curvatures; see for example Fig. 2d. Property 3 is a specific example of Property 2 that requires consideration in constructing a consistent model for surface decomposition.

Property 4: Point-to-point (p2p) mapping exists between MEB-AS and MEB-ES, established through the normal vector to the MEB-ES. This property provides the basis for using the MEB-ES for surface decomposition. The p2p mapping is easily seen for a convex MEB-ES region. As shown in Fig. 3a, a MEB placed at a point P_A on the MEB-AS must make one and only one contact P_E with a local convex surface of the MEB-ES. The vector $\overrightarrow{P_E P_A}$ is the normal vector of MEB-ES at P_E. This p2p correspondence is also obvious in a concave region with a curvature lower than ${r_{MEB}}^{-1}$. However, mapping is not p2p for a concave curvature equal to ${r_{MEB}}^{-1}$. As discussed in Property 3, any concave curvature higher than ${r_{MEB}}^{-1}$ is smoothed to ${r_{MEB}}^{-1}$, so this particular curvature can occur frequently on the MEB-ES. For example, Fig. 3b shows a MEB of r_{MEB} placed at P_A contacts a patch of surface on the MEB-ES of curvature ${r_{MEB}}^{-1}$ instead of one point as in the convex case. The 1-to-∞ correspondence between MEB-AS and MEB-ES causes a "division by zero" problem for the numeric method to be described later. However, this problem is avoidable by slightly modifying the surface generation process: we first create a somewhat smoother exclusion surface MEB-ES' using a ball of radius r'_{MEB} that is slightly larger than r_{MEB}, and then generate the associated accessible surface MEB-AS' by rolling a ball of radius r_{MEB} over MEB-ES'. According to Property 3, any concave curvature on the MEB-ES' must be lower than ${r_{MEB}}^{-1}$ because $r'_{MEB} > r_{MEB}$. Consequently, a p2p mapping between the MEB-AS' and MEB-ES' exists for a ball of r_{MEB}. The outcome of these operations is illustrated in Fig. 3c, in which MEB-AS' and MEB-ES' are shown as gray and red meshes, respectively. Differently from Fig. 3b, now the same MEB only contacts a single point on the MEB-ES', establishing the p2p mapping. To facilitate visualization, Fig. 3c uses a r'_{MEB} substantially larger than r_{MEB} (7 *vs.* 5 Å), but any arbitrarily small increment (set to 0.1 Å in this work) suffices to avoid the 1-to-∞ problem. Therefore, the MEB-ES' and MEB-AS' are virtually identical to those generated with the exact r_{MEB} for any meaningful precision, and we do not make explicit distinction in what follows.

With the gel surface now approximated by the MEB-ES, we calculate $V_{ex}(\boldsymbol{a})$ by moving and rotating the solute between the MEB-AS and MEB-ES. The p2p mapping suggests a polar sampling method. We can visualize the MEB-AS as a polygonal mesh, such as the gray mesh displayed in Fig. 2. If the MEB-AS is approximated with N surface polygons, there will also be N corresponding local landscapes on the MEB-ES. According to Property 4, a MEB placed at P_A within each polygon unit must be tangential to a corresponding local landscape at P_E (see Fig. 3a and c). A polar positional sampling scheme can then be constructed to calculate $V_{ex}(\boldsymbol{a})$, by moving the MEB center of a solute of fixed orientation $\boldsymbol{a}$ from P_A towards P_E along the normal vector and summing up all overlap-causing positions. For overlap detection, the vdw boundary of the solute is generated with the atomic vdw radius obtained by OPLS-AA.[36] If the N local landscapes are further clustered into n LS_{rep}'s based on shape similarity, $V_{ex}(\boldsymbol{a})$ is approximated by:

$$V_{ex}(\boldsymbol{a}) \approx \sum_{i=1}^{n} A_i \sum_{j=1}^{N_i} \int_0^{r_{MEB}} \mathrm{d}r\, w_i(r)\left\{1 - \exp\left[-\beta U_i^{HB}\left(\boldsymbol{a}, \boldsymbol{a}_{ij}, r\right)\right]\right\} + V_{MEB\_ES} \quad (9)$$

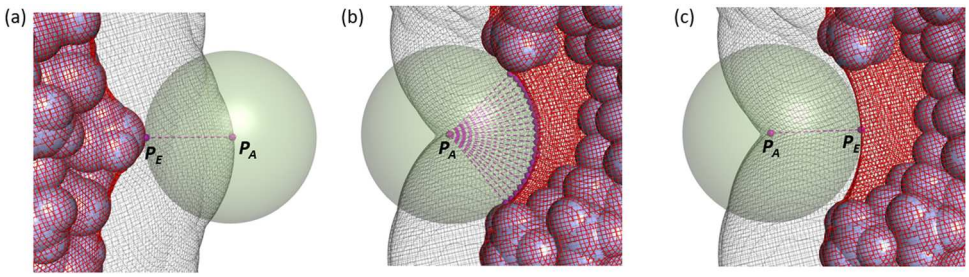

Fig. 3 Point-to-point mapping between MEB-AS (gray) and MEB-ES (red) (a): a p2p correspondence exists at a convex region; (b): smoothing a high concave curvature leads to a 1-to-∞ correspondence; (c): smoothing the vdwS in (b) using a slightly larger MEB ensures a p2p correspondence.

The i-th $\mathrm{LS_{rep}}$ occurs N_i times on the MEB-ES (obviously $\sum_{i=1}^{n} N_i = N$). Its corresponding MEB-AS polygon has an area of A_i. One can consider dividing the MEB-AS into unit polygons of equal area, but this is unnecessary, as it will become evident that only the percentage area of the i-th $\mathrm{LS_{rep}}$ over the total MEB-AS area is relevant. The integration boundary for the one-dimensional variable r is between P_A and P_E along the normal vector, with r being the distance of the MEB center to P_E. $w_i(r)$ is a pure number that dictates positional sampling density along the normal vector. $w_i(r)$ is needed for the polar sampling method because for a convex surface at P_E, going from P_E to P_A is accompanied by spatial expansion, whereas for a concave surface it is accompanied by spatial contraction. Accordingly, as r increases, $w_i(r)$ should increase in the former case but decrease in the latter case to ensure uniform spatial sampling. For example, if the MEB-ES is a cylinder of radius R, $w_i(r)$ should scale by $\frac{R+r}{R+r_\mathrm{MEB}}$; not employing $w_i(r)$ would over-sample positions closer to the cylinder. The exact form of $w_i(r)$ depends on the shape of the $\mathrm{LS_{rep}}$ and will be given in the "Landscape Parametrization" section. The term $\boldsymbol{a}_{ij}$ is the laboratory-frame orientation of the j-th occurrence of the i-th $\mathrm{LS_{rep}}$. The steric potential U_i^HB depends on the three-dimensional structure of the solute, the shape of the i-th landscape, and their relative orientation and position. $V_\mathrm{MEB\_ES}$ is the volume enclosed by the MEB-ES (red mesh in Fig. 2). Clearly, eqn (9) approximates the total excluded volume as the sum of the excluded volume between the MEB-AS and MEB-ES (evaluated by the sum of integrals) and the volume enclosed by MEB-ES ($V_\mathrm{MEB\_ES}$), which is considered completely inaccessible to the solute.

In the next step, we consider that the N_i occurrences of the i-th $\mathrm{LS_{rep}}$ on the MEB-ES follow a laboratory-frame orientational distribution described by an ODF f_i. Of course, such an ODF can always be defined based on an orientational histogram, regardless of whether the nature of the distribution is dynamic as for a mesogen in the LC or static as for the $\mathrm{LS_{rep}}$ for a gel. Then eqn (9) can be equivalently written as:

$$V_\mathrm{ex}(\boldsymbol{a}) \approx \sum_{i=1}^{n} A_i N_i \int \mathrm{d}\boldsymbol{a}_\mathrm{m}\, f_i(\boldsymbol{a}_\mathrm{m}) \int_0^{r_\mathrm{MEB}} \mathrm{d}r\; w_i(r) \times \left\{1 - \exp\left[-\beta U_i^\mathrm{HB}(\boldsymbol{a}, \boldsymbol{a}_\mathrm{m}, r)\right]\right\} + V_\mathrm{MEB\_ES} \tag{10}$$

To highlight its similarity to eqn (2), we reuse the label $\boldsymbol{a}_\mathrm{m}$ here to represent the laboratory-frame orientation of a $\mathrm{LS_{rep}}$, which was earlier used for mesogen orientation in eqn (2) (here we denote the subscript m as "medium"). The transition from eqn (9) and (10) is easy to see by realizing that $N_i f_i(\boldsymbol{a}_\mathrm{m})\mathrm{d}\boldsymbol{a}_\mathrm{m}$ is the number of times that the i-th $\mathrm{LS_{rep}}$ occurs at orientation $\boldsymbol{a}_\mathrm{m}$. Like the LC case, we can recast eqn (10) into an alternative form based on the relative orientation $\boldsymbol{\Omega}$ between the solute and $\mathrm{LS_{rep}}$:

$$V_\mathrm{ex}(\boldsymbol{a}) \approx \sum_{i=1}^{n} A_i N_i \int \mathrm{d}\boldsymbol{\Omega} f_i(\hat{R}\boldsymbol{a}) \int_0^{r_\mathrm{MEB}} \mathrm{d}r\; w_i(r) \times \left\{1 - \exp\left[-\beta U_i^\mathrm{HB}(\boldsymbol{\Omega}, r)\right]\right\} + V_\mathrm{MEB\_ES} \tag{11}$$

The orientation relation is depicted in Fig. 1 like the LC case, except that here the blue frame is associated with a $\mathrm{LS_{rep}}$ in *lieu* of a mesogen. Note that in the LC equations, $\boldsymbol{r}$ is a three-dimensional position vector associated with Cartesian positional sampling, whereas in eqn (9)–(11) $\boldsymbol{r}$ is a one-dimensional length associated with polar sampling. Because $A_i N_i$ is the subsurface area on the MEB-AS due to the i-th $\mathrm{LS_{rep}}$ and the double integral over $\boldsymbol{\Omega}$ and $\boldsymbol{r}$ is the average exclusion length between MEB-AS and MEB-ES, their product corresponds to the excluded volume associated with the i-th $\mathrm{LS_{rep}}$. Substituting eqn (11) into eqn (8), performing multipole expansion for $f_\mathrm{s}(\boldsymbol{a})$ and $f_i(\hat{R}\boldsymbol{a})$ on both sides of the equation, and equating the coefficients of the 2nd order spherical harmonics leads to the following OTE for the gel, analogous to eqn (6) for the LC:

$$\sqrt{\frac{4\pi}{5}}\left\langle Y_2^m \right\rangle_S \approx \sum_{m'=-2}^{2} \sum_{i=1}^{n} \sqrt{\frac{4\pi}{5}}\left\langle Y_2^{m'} \right\rangle_i \left\{ \frac{SP_i}{8\pi^2 (V - V_\mathrm{ex}^r)} \int \mathrm{d}\boldsymbol{\Omega} {D_{m'm}}^2(\hat{R}) \times \int_0^{r_\mathrm{MEB}} \mathrm{d}r\; w_i(r) \exp\left[-\beta U_i^\mathrm{HB}(\boldsymbol{\Omega}, r)\right] \right\} \tag{12}$$

Here $\sqrt{\frac{4\pi}{5}}\left\langle Y_2^{m'} \right\rangle_i$ are the order parameters of the i-th $\mathrm{LS_{rep}}$. S is the total surface area of the MEB-AS $\left(S = \sum_{i=1}^{n} A_i N_i\right)$, and P_i is the percentage of surface area due to the i-th landscape ($SP_i = A_i N_i$). Note that the term inside the curly bracket is a pure

number that connects the solute order parameters to those of LS_{rep}'s.

We should note here that the order parameters of a LS_{rep} are fundamentally different from those of a LC. First, the LC is in a dynamic equilibrium with the sample in heat exchange with the environment and its order parameters are averaged over all system microstates over time. In contrast, the gel matrix is treated as a static structure, somewhat resembling a particular LC microstate with uniaxial mesogen distribution that is frozen in time. Second, the LC order parameters are naturally defined because the system contains repeating units, *i.e.*, the mesogens, whereas the gel landscape order parameters can be defined only because we artificially create repeating units by representing certain local surfaces with the same LS_{rep}. A simple example is shown in Fig. 4, which shows part of the cross-section of an elliptic cylindrical MEB-ES (red dashed line), the corresponding MEB-AS cross-section (gray line), and the MEB (green circle). Three P_A–P_E pairs are selectively displayed, highlighting three different landscape curvatures and orientations (blue frames with *Z*-axis pointing towards the eyes) at P_E. If we represent all local landscapes on the dashed ellipse with the same LS_{rep} (cross-section shown in red solid line at the three P_E's), we can then define the order parameters of this LS_{rep} based on its orientational distribution along the elliptic circumference. Third, a LC is often attributed with axisymmetric order because its repeating unit (mesogen) has approximate cylindrical symmetry. In contrast, a LS_{rep} is a local surface which cannot be assumed to have cylindrical symmetry and consequently axisymmetric order cannot be assumed, although a LS_{rep} can have lower symmetry that enables reduced order parameters as described later. Note that the lack of axial symmetry does not contradict uniaxial order, which is observed by a constrained gel with no orientational preference within the laboratory-frame *XY* plane.

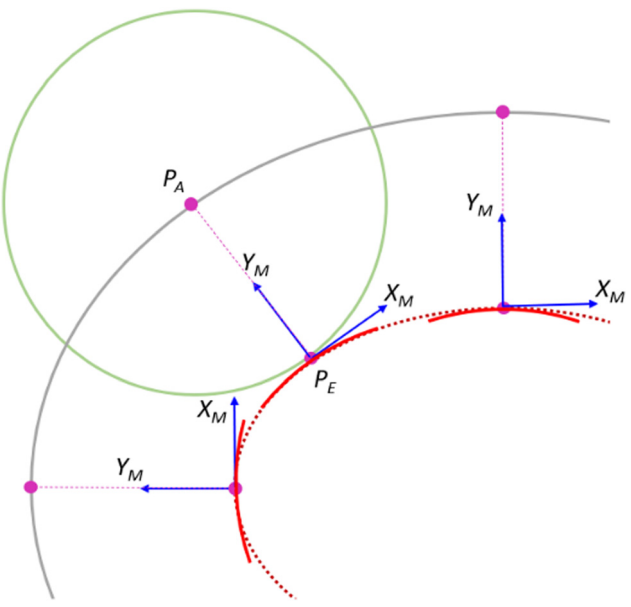

Fig. 4 Approximation of an elliptic cylinder with a single LS_{rep}.

As previously mentioned in Property 2, in this work we focus on gel surface parameterization for solutes of similar sizes, due to the dependence of the MEB-ES on r_{MEB}. Accordingly, all r_{MEB} – dependent parameters, namely $\sqrt{\frac{4\pi}{5}}\langle Y_2^{m'}\rangle_i$, SP_i, and V_{ex}^r, can be considered identical for all solutes during database fitting, allowing further simplifications. First, P_i can be absorbed into $\sqrt{\frac{4\pi}{5}}\langle Y_2^{m'}\rangle_i$ because only their product is important. As a result, the modified order parameters, still labeled as $\sqrt{\frac{4\pi}{5}}\langle Y_2^{m'}\rangle_i$ and simply referred to as order parameters in what follows, reflect both orientational order and relative importance of each LS_{rep} (due to P_i). Second, $\frac{S}{8\pi^2(V-V_{ex}^r)}$ in eqn (12) can be replaced by the inverse of a normalization factor *Z*. Hence a simplified equation is obtained as below:

$$\sqrt{\frac{4\pi}{5}}\langle Y_2^m\rangle_S \approx \frac{1}{Z}\sum_{m'=-2}^{2}\sum_{i=1}^{n}\sqrt{\frac{4\pi}{5}}\langle Y_2^{m'}\rangle_i\left\{\int d\boldsymbol{\Omega} D_{m'm}{}^2(\hat{R}) \times \int_0^{r_{MEB}} dr\, w_i(r)\exp\left[-\beta U_i^{HB}(\boldsymbol{\Omega},r)\right]\right. \quad (13)$$

The double integral in the curly bracket is numerically calculated based on landscape curvatures that are determined together with landscape order parameters during database fitting. The LS_{rep} curvatures and order parameters can then be used to predict the order parameters of other solutes of similar sizes to database compounds. If all data are collected under identical alignment conditions, the same normalization factor, *Z*, applies to all solutes and thereby can be determined from database fitting. In this work, however, we combine data collected with both compressed and stretched gels polymerized at different concentrations and cross-linking ratios to maximize available data for parametrization. Hence, we are not concerned with predicting the alignment amplitude but only alignment asymmetry and orientation, which are minimally affected by polymerization conditions and straining methods. Accordingly, the normalization factor *Z* is ignored (by setting *Z* to 1 Å) during fitting, thus the quality of prediction is evaluated by the scale-invariant coefficient of determination, r^2, instead of the *Q*-factor.

An argument could be made that a parametrization with a single average MEB-ES may become problematic if a wide range of solute sizes is incorporated in the database because of high variability in individual MEB-ES'es for compounds of different sizes. But to what extent this poses an issue also depends on the medium surface structure. As mentioned, in regard to Property 1, the MEB-ES resembles the actual vdwS regardless of solute size if the *vdw*S is predominantly convex or concave with either very small or very large cavities. Only when the *vdw*S contains many medium-sized cavities that are accessible to some but not other solutes in the database, the MEB-ES becomes critically dependent on solute size and the validity of applying eqn (13) to database fitting degrades. Clearly, the susceptibility of the

method to solute size distribution is closely related to the shortcoming in approximating the *vdw*S with the MEB-ES for a solute of high shape anisotropy as discussed in Property 1.

Finally, the order transfer relationship (eqn (12) and (13)) is derived based on a static gel, but it is also applicable if the polymer undergoes slow mainchain motion or fast sidechain motion, the rate being compared to the rotational and translational diffusion of the solute. In case of slow mainchain motion, one can imagine constructing an ensemble of MEB-ES'es associated with different dynamic states of the gel, which after surface decomposition still reduce to several LS_{rep}'s and associated order parameters just like a single MEB-ES (population amplitudes can be absorbed into LS_{rep} order parameters), although a larger number of LS_{rep}'s may be necessary for parametrizing multiple dynamic states. In case of fast sidechain motion, one can imagine employing a single MEB-ES to represent the average exclusion surface, to which the surface decomposition method can be applied without change. However, our method does not account for intermediate timescale motions because they can interfere with solute diffusion in the interaction zone.

2.3 Liquid crystal without a mesogen model

Surface decomposition can also be applied to a dilute LLC system in which a suitable mesogen model is unavailable for direct excluded volume calculation. In this case, the *vdw*S of a single mesogen is approximated by a MEB-ES and decomposed into an orientational and positional arrangement of several LS_{rep}'s. eqn (11) can then be used to evaluate the excluded volume due to one orientationally fixed mesogen. Next, $\langle V_{ex}(\boldsymbol{a})\rangle$ in eqn (5) can be calculated by averaging eqn (11) over all mesogen orientations in all microstates, resulting in an equation formally identical to eqn (11) except that the landscape ODF f_i is accounted for over all microstates. All subsequent steps are the same as those previously described in Section 2.2, eventually leading to the same OTE in eqn (12) and its simplified version in eqn (13) under the uniform solute size approximation. Consequently, and usefully, the surface decomposition method provides a unified formalism to parametrize solute alignment in both LCs and gels. It is interesting to note that two different ODFs can be defined for a LS_{rep} on the mesogen surface: besides the ODF f_i defined in the laboratory frame and averaged over all microstates, a second ODF f_i' can be defined in the mesogen frame and averaged over only one mesogen surface. Not surprisingly, the two ODFs derived for the LS_{rep} can be correlated by the ODF of the mesogen f_m, and consequently the corresponding order parameters can be related as well. Additional details can be found in the ESI.†

3. Landscape parametrization

Steric interaction between a solute and a LS_{rep} is calculated by numeric integration over $\boldsymbol{r}$ and $\boldsymbol{\Omega}$ using the polar sampling strategy. This simulation must proceed on the basis of a proposed LS_{rep} shape. In this section, we focus on methods to parametrize three-dimensional surfaces using a minimal set of shape parameters. These shape parameters all have the dimension of a curvature, *i.e.*, length^{-1}. Like the order parameters, these shape parameters will also be determined by fitting to an experimental database.

3.1 Paraboloid landscape

The simple paraboloid as described by eqn (14) can generate many qualitatively different landscape types (Fig. 5).

$$y = -\frac{1}{2}\left(k_x x^2 + k_z z^2\right) \tag{14}$$

Here, k_x and k_z are the principal curvatures at the origin, whose principal directions are along the X_M and Z_M axes, respectively. The origin coincides with P_E on the MEB-ES. The Y_M axis is the normal vector to the MEB-ES at P_E. By varying the principal curvatures, various landscapes can be generated, including a plane ($k_x = k_z = 0$), a ridge ($k_x > 0$, $k_z = 0$), a valley ($k_x < 0$, $k_z = 0$), a hump ($k_x > 0$, $k_z > 0$), a basin ($k_x < 0$, $k_z < 0$), or a saddle ($k_x > 0$, $k_z < 0$). Note that a positive or negative $k_{x/z}$ is associated with a convex or concave curvature, respectively. The ridge and the valley are the convex and concave faces of a parabolic cylinder, respectively; the hump and the basin are the convex and concave faces of an elliptic paraboloid, respectively; the saddle has both convex and concave features and is a face of a hyperbolic paraboloid—note that both faces of a hyperbolic paraboloid are saddle surfaces. It should be further noted that x^2 and z^2 are symmetric in eqn (14), so exchanging the values of x and z rotates the landscape by 90° but does not change its shape. Different orientations of a landscape are accounted for through landscape order parameters as previously mentioned.

The principal curvature frame (PCF) X_M–Y_M–Z_M in Fig. 5 corresponds to the blue frame with the same axis labeling in Fig. 1. The normal vector Y_M is the polar sampling axis. The location of the solute during simulation is indicated in Fig. 5. To calculate the double-integral in eqn (13), the solute's MEB center is moved between 0 and r_{MEB} on the Y_M axis with a 0.2 Å step-size while a total of 52 488 solute orientations are sampled at each positional step by uniformly rotating the solute around its MEB center. Orientational uniformity is confirmed by solute order parameters being effectively zero when the steric interaction potential is omitted in eqn (12). The scaling factor for positional sampling density, *i.e.*, $w(r)$ in eqn (10)–(13) (the subscript i is omitted here without causing ambiguity), is given by:

$$w(r) = \frac{(1 + k_x r)(1 + k_z r)}{(1 + k_x r_{MEB})(1 + k_z r_{MEB})} \tag{15}$$

The term r is the distance of the MEB center to the origin. According to Property 3, any concave curvature must be lower than r_{MEB}^{-1}, *i.e.*, the inequality must be satisfied: $1 + k_{x/z} r_{MEB} > 0$. The "division by zero" problem with $k_{x/z} = -r_{MEB}^{-1}$ is avoided by generating the MEB-ES using a slightly enlarged MEB as described in Property 4. This operation is implemented by

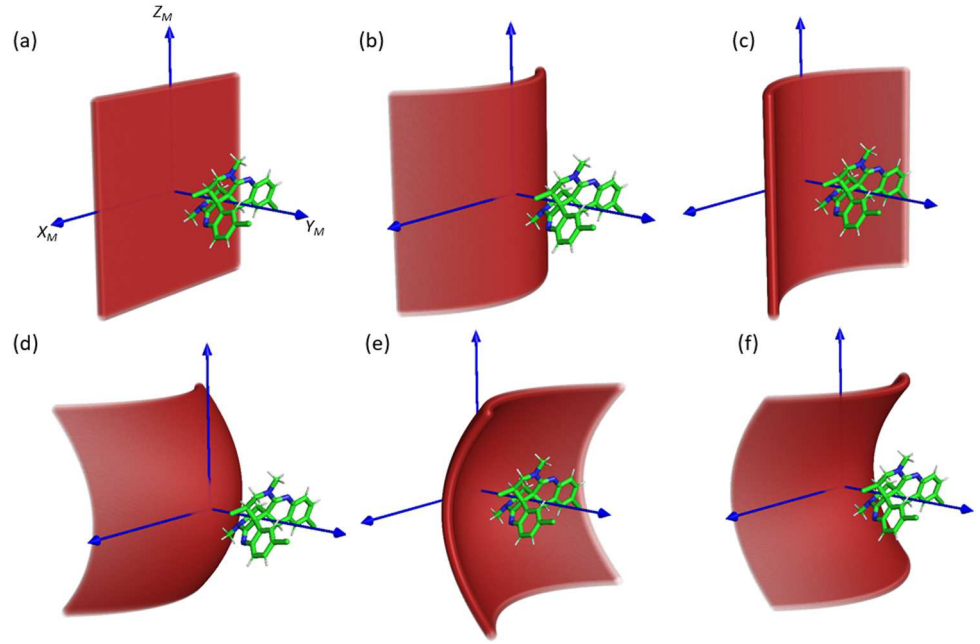

Fig. 5 Paraboloid landscapes. (a) A plane with $k_x = k_z = 0$ Å^{-1} (b): a ridge with $k_x = 0.1$ Å^{-1} and $k_z = 0$ Å^{-1} (c): a valley with $k_x = -0.1$ Å^{-1} and $k_z = 0$ Å^{-1} (d): a hump with $k_x = 0.1$ Å^{-1} and $k_z = 0.1$ Å^{-1} (e): a basin with $k_x = -0.1$ Å^{-1} and $k_z = -0.1$ Å^{-1} (f): a saddle with $k_x = 0.1$ Å^{-1} and $k_z = -0.1$ Å^{-1}. A molecule of caulamidine A is displayed to show the solute location during the polar sampling simulation.

constraining $1 + k_{x/z}r_{\mathrm{MEB}} > 0$ during database fitting. For example, the minimization process automatically flattens a high curvature valley of $k_x = -0.150$ Å^{-1} to -0.099 Å^{-1} for a solute of $r_{\mathrm{MEB}} = 10$ Å (which is enlarged by 0.1 Å to a r_{MEB} of 10.1 Å thus leading to the highest possible concave curvature of -0.099 Å^{-1}, as described in Property 4), but no flattening is performed for a solute of $r_{\mathrm{MEB}} = 5$ Å. By doing so, different $\mathrm{LS}_{\mathrm{rep}}$'s can be generated in accordance with solute size when needed. Although there is no direct restriction on a positive (convex) curvature (Property 3), the highest realistic curvature is one generated by a hydrogen atom of the medium. Therefore, besides $1 + k_{x/z}r_{\mathrm{MEB}} > 0$, we also require $k_{x/z} \leq 0.83$ Å^{-1} corresponding to a hydrogen *vdw* radius of 1.2 Å.

Because a paraboloid has no cylindrical symmetry unless $k_x = k_z$, the orientational ordering is generally not axisymmetric. However, a paraboloid has at least two reflection planes of symmetry, namely the X_{M}–Y_{M} and Z_{M}–Y_{M} plane, which also allows order parameter reduction. For a paraboloid residing on an achiral or atactic medium, the orientational principal axis frame (PAF) must coincide with its PCF (the orthogonal frame formed by two principal directions of curvature and the normal vector). This is because a paraboloid with an orientation defined by $(\theta_{\mathrm{M}},\varphi_{\mathrm{M}})$ (Fig. 1) has an equal probability of adopting three other orientations, namely $(\pi - \theta_{\mathrm{M}},\varphi_{\mathrm{M}})$, $(\theta_{\mathrm{M}}, \pi - \varphi_{\mathrm{M}})$, and $(\pi - \theta_{\mathrm{M}}, \pi - \varphi_{\mathrm{M}})$, associated with simultaneous reflection of the entire medium and the director about the X_{M}–Y_{M} and/or Z_{M}–Y_{M} planes. Straightforward calculation based on eqn (7) shows that this 4-fold equivalency nullifies the off-diagonal Saupe elements: S_{xy}, S_{xz}, and S_{yz}, *i.e.*, the Saupe ordering matrix is diagonalized in the PCF. As a result, only two diagonal elements are needed to depict the ordering of a paraboloid: S_{yy} and S_{zz} (as the Saupe ordering matrix is traceless, S_{xx} is a dependent parameter).

In terms of parameter fitting, solute ordering by a paraboloid landscape can be parametrized with four variables, including two curvatures: k_x and k_z, and two landscape ordering matrix parameters: S^L_{yy} and S^L_{zz} (the superscript L indicates the parameters are related to a $\mathrm{LS}_{\mathrm{rep}}$ but not a solute). Hence n L_{rep}'s requires $4n$ variables. But since we are not concerned with absolute alignment amplitude, we can fix the S^L_{zz} of the first landscape to 1 and optimize its S^L_{yy} and all order parameters in a relative sense, leading to a total of $4n - 1$ variables for n L_{rep}'s.

3.2 Twisted paraboloid landscape

Due to the reflection symmetry of a paraboloid, surface decomposition solely based on paraboloid landscapes is not expected to fully describe solute alignment in a chiral medium. In fact, even an achiral molecule with conformational flexibility generally lacks internal reflection symmetry. Note that being achiral only means that the molecule has two equally probable conformations related by mirror reflection, but neither conformation itself necessarily has internal reflection symmetry.[37] For example, if the exchange between the two conformers is hindered, the two conformers become atropisomers. Considering the restricted mobility in a cross-linked gel matrix, even an achiral gel can therefore have local structural asymmetry despite global reflection symmetry. Of course, such asymmetry is even more expected for an atactic medium which has local chemical chirality. To take local structural asymmetry into account in a most simple fashion, we consider a twisted paraboloid by adding a twist curvature to eqn (14), as described

by the following parametric equation:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos(k_t z') & -\sin(k_t z') & 0 \\ \sin(k_t z') & \cos(k_t z') & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ y' + k_x{}^{-1} + \frac{1}{2}k_z z'^2 \\ z' \end{pmatrix} - \begin{pmatrix} 0 \\ k_x{}^{-1} + \frac{1}{2}k_z z'^2 \\ 0 \end{pmatrix} \quad (16)$$

Here, (x', y', z') is a point on a paraboloid as described by eqn (14). Twisting is generated by rotating the parabolic cross-sections at different layers along the Z_M axis by a phase of $k_t z'$ around a vertical axis located at $x' = 0$ and $y' = -k_x{}^{-1} - \frac{1}{2}k_z z'^2$. The twist curvature, k_t, describes how rapidly a cross-section twists along the Z_M axis, with $k_t = 0$ indicating no twisting, in which case the landscape reduces to a regular paraboloid as in Fig. 5. At different Z layers, both the amount of rotation and the rotation axis are different, because they are dependent on z' (note that $z' = z$ in eqn (16)). Some representative landscapes generated by eqn (16) are displayed in Fig. 6. In each landscape, seven parabolic cross-sections along Z_M are marked in dark red, and the parabolic vertex at each layer is traced by a cyan curve to facilitate visualization. Fig. 6a–e were generated by twisting the ridge, valley, hump, basin, and saddle shown in Fig. 5b–f, respectively, by a twist curvature k_t of 0.05 Å^{-1}. Note that twisting along Z_M destroys the equivalence between X_M and Z_M axes, *i.e.*, exchanging the x and z values in eqn (16) leads to a different shape, unlike the paraboloid case. For example, twisting a saddle with $k_x < 0$, $k_z > 0$ or a saddle with $k_x > 0$, $k_z < 0$ leads to qualitatively different landscapes, as shown in Fig. 6e and f, respectively (corresponding comparisons for other landscape types are not shown).

An important consideration in parametrizing an asymmetric landscape is that Property 3 must be satisfied to ensure the generated landscape is a valid local surface on the MEB-ES. In the ESI,† we show that a landscape created with eqn (16) has principal curvatures of k_x and k_z at the origin, regardless of k_t. Furthermore, like the regular paraboloid, the normal vector is the Y_M axis and the principal directions are the X_M and Z_M axes at the origin. Therefore, the same curvature constraint as in the regular paraboloid case, *i.e.*, $1 + k_{x/z} r_{MEB} > 0$, ensures the resulting twisted paraboloid is a valid local surface on the MEB-ES.

Twisting "handedness" is dictated by the sign of k_t. All landscapes in Fig. 6 were created with $k_t = 0.05$ Å^{-1}. Their enantiomers can be created with $k_t = -0.05$ Å^{-1} (figures not shown). For a chiral medium, twisted paraboloids generated by eqn (16) with the correct sign of k_t can reasonably be employed as LS_{rep}'s. For an achiral or atactic medium, the left- and right-handed landscapes are both present and must occur with equal probability, in which case eqn (13) should be expanded according to landscape handedness:

$$\sqrt{\frac{4\pi}{5}}\langle Y_2^m\rangle_S \approx \frac{1}{Z}\sum_{m'=-2}^{2}\sum_{i=1}^{n}\sum_{e=R,S}\sqrt{\frac{4\pi}{5}}\left\langle Y_2^{m'}\right\rangle_{i,e}\left\{\int d\boldsymbol{\Omega} D^2_{m'm}(\hat{R}) \times \int_0^{r_{MEB}} dr\, w_i(r)\exp\left[-\beta U^{HB}_{i,e}(\boldsymbol{\Omega}, r)\right]\right. \quad (17)$$

Each LS_{rep} can be either left- or right-handed, indicated by the subscript e. Opposite enantiomers interact differently with a

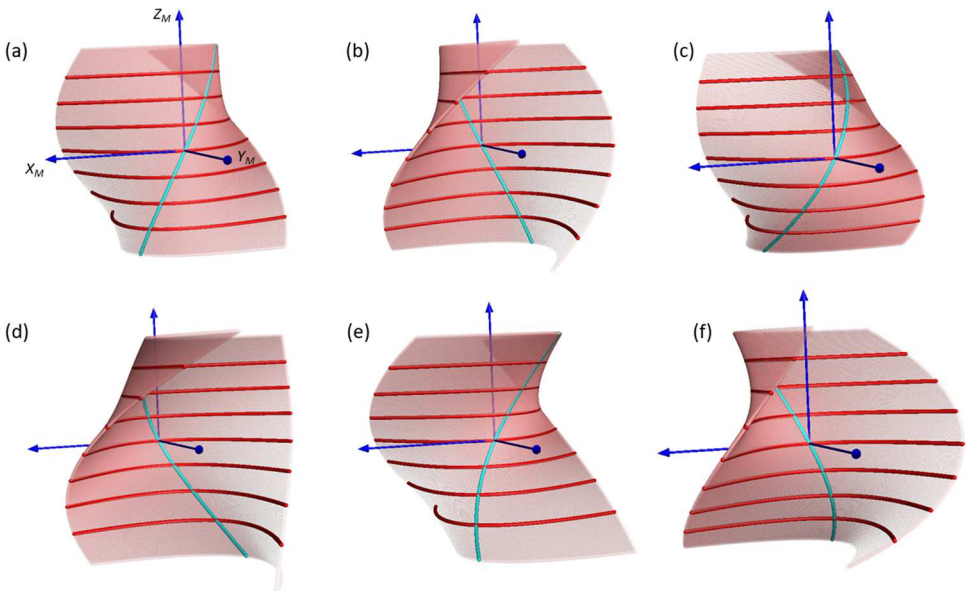

Fig. 6 Twisted paraboloid landscapes. (a): a twisted ridge with $k_x = 0.1$ Å^{-1}, $k_z = 0$ Å^{-1}, and $k_t = 0.05$ Å^{-1} (b): a twisted valley with $k_x = -0.1$ Å^{-1}, $k_z = 0$ Å^{-1} and $k_t = 0.05$ Å^{-1} (c): a twisted hump with $k_x = 0.1$ Å^{-1}, $k_z = 0.1$ Å^{-1}, and $k_t = 0.05$ Å^{-1} (d): a twisted basin with $k_x = -0.1$ Å^{-1}, $k_z = -0.1$ Å^{-1}, and $k_t = 0.05$ Å^{-1} (e): a twisted saddle with $k_x = 0.1$ Å^{-1}, $k_z = -0.1$ Å^{-1}, and $k_t = 0.05$ Å^{-1} (f): a twisted saddle with $k_x = -0.1$ Å^{-1}, $k_z = 0.1$ Å^{-1}, and $k_t = 0.05$ Å^{-1}.

chiral solute, described by steric potential U_{ie}^{HB} (e = R or S). Previously we have shown that for a regular paraboloid, the ordering PAF coincides with its PCF. A twisted paraboloid is somewhat different, namely, the normal vector Y_M is still an ordering principal axis but the principal curvature directions X_M and Z_M are not necessarily so. Like the paraboloid, four orientations of equal probability exist for any frame residing on an achiral or atactic medium surface, namely (θ_M,φ_M),$(\pi - \theta_M,\varphi_M)$, $(\theta_M,\pi - \varphi_M)$, and $(\pi - \theta_M,\pi - \varphi_M)$. Of these, (θ_M,φ_M) and $(\pi - \theta_M,\pi - \varphi_M)$ are associated with one landscape, while $(\pi - \theta_M,\varphi_M)$ and $(\theta_M,\pi - \varphi_M)$ are associated with its enantiomer. Straightforward calculation shows that the order parameters $\langle {Y_2}^{\pm 1,2}\rangle_{i,R/S}$ have zero imaginary part, which leads to $S_{xy}^{R/S} = S_{yz}^{R/S} = 0$, *i.e.*, the normal vector Y_M is an ordering principal axis for both left- and right-handed landscapes. In contrast, $S_{xz}^{R/S}$ may not be zero, suggesting X_M and Z_M are not necessarily ordering principal axes, although the sum of S_{xz}^{R} and S_{xz}^{S} is zero. Calculation also shows that the Saupe ordering matrices of left- and right-handed landscapes have identical diagonal elements: $S_{xx}^{R} = S_{xx}^{S}$, $S_{yy}^{R} = S_{yy}^{S}$, and $S_{zz}^{R} = S_{zz}^{S}$. Based on these relations, the Saupe ordering matrix for a twisted paraboloid can be parametrized by:

$$\hat{S}^{R/S} = \widehat{R_y}(+/-\beta_S)\begin{pmatrix} -S_{yy}^{L} - S_{zz}^{L} & 0 & 0 \\ 0 & S_{yy}^{L} & 0 \\ 0 & 0 & S_{zz}^{L} \end{pmatrix} \quad (18)$$

where $\widehat{R_y}(+/-\beta_S)$ is a rotation operator about the Y-axis by an angle of $+\beta_S$ and $-\beta_S$ for left- and right-handed landscapes, respectively (β_S is between 0 and π). The Saupe ordering matrix can be easily converted to spherical harmonic order parameters using the reverse relationships in eqn (7).

In terms of parameter fitting, three curvatures, namely k_x, k_z, and k, and three parameters related to the Saupe ordering matrix, namely S_{zz}^{L}, S_{yy}^{L}, and β_S, need to be determined for each LS_{rep} of the twisted paraboloid type. For an achiral or atactic medium, contribution from opposite handedness is calculated by simultaneously inverting the signs of k and β_S and combined using eqn (17). For a chiral alignment medium (not studied here), only the correct handedness should be chosen for calculation by eqn (13). For n different LS_{rep}'s, $6n - 1$ parameters need determination after fixing the S_{zz} of the first LS_{rep} (or the first enantiomeric pair of LS_{rep}'s in case of an achiral or atactic medium) to 1.

4. Parameter optimization by database fitting

Since the solute order parameters to be predicted are parametrized as functions of shape and order parameters of a series of representative landscapes, the prediction accuracy critically depends on obtaining reliable landscape parameters. In this section we describe the computational workflow that yields the optimal curvatures and order parameters of representative landscapes using an experimental database.

Landscape curvatures and order parameters are optimized by the Nelder–Meade (N–M) simplex method[38] that minimizes the differences between predicted NMR data and those in an experimental database. Only rigid molecules are included in the database; including flexible molecules would require optimizing the geometries of major conformational states and calculating their Boltzmann populations, which would add unnecessary uncertainty to parameter optimization. Of course, parameters optimized from rigid molecules can later be used to predict alignment of flexible molecules. We will follow up on this topic in a separate communication. Importantly, all experimental data should be measured with the same type of alignment medium. For example, RDCs collected in poly-methyl methacrylate (PMMA) gels and poly-hydroxyethylmethacrylate (poly-HEMA) gels should not be mixed in the same database, because they correspond to different landscape parameters. In this study we only utilize data collected in PMMA gels. We include data collected with different PMMA concentrations and cross-linking ratios because these factors only influence alignment amplitude but not alignment asymmetry and orientation.[21] We also include data from both stretched and compressed gels because different straining methods yield data that are simply correlated through a scaling factor.[21,39] Data from compressed gels are sign inverted for compatibility with data from stretched gels. We included RDC and/or RCSA data based on availability.

Fig. 7 summarizes the key steps in the optimization of LS_{rep} shape and order parameters. First, the number of LS_{rep}'s to use for surface decomposition must be specified. Obviously, using too few LS_{rep}'s may not adequately depict complex surface features exhibited by the medium, while using too many LS_{rep}'s not only consumes excessive computational resources but also increases the risk of over-parametrization. Therefore, we start with only one LS_{rep}, whose Saupe order matrix and curvatures are randomly initiated, except that its S_{zz}^{L} is set to 1. Inside the N–M simplex block (dotted green box in Fig. 7), a series of automated steps are undertaken to maximize the agreement of predicted data with experimental data by varying landscape curvatures and order parameters. When the minimization is completed, a set of optimized order parameters and curvatures are obtained for the 1st LS_{rep}. To sufficiently sample the minimization surface, the simplex block is independently executed 400 times, each time with a random set of initial values for order parameters and curvatures. Results from the best run, *i.e.*, the run that yields the highest agreement with experimental data, are accepted for the final optimal parameters. Next, we repeat the same process but with two LS_{rep}'s. All parameters are optimized by 400 independent executions of the simplex block, each time initializing the parameters for the 1st LS_{rep} using values optimized in the first round but randomly initializing parameters for the 2nd LS_{rep}. Results from the best of the 400 runs are accepted. If including the 2nd LS_{rep} significantly improves agreement with experimental values, a new round with three LS_{rep}'s is performed. This process continues iteratively, each time conducting 400 all-parameter minimizations after initializing existing LS_{rep}'s with optimized

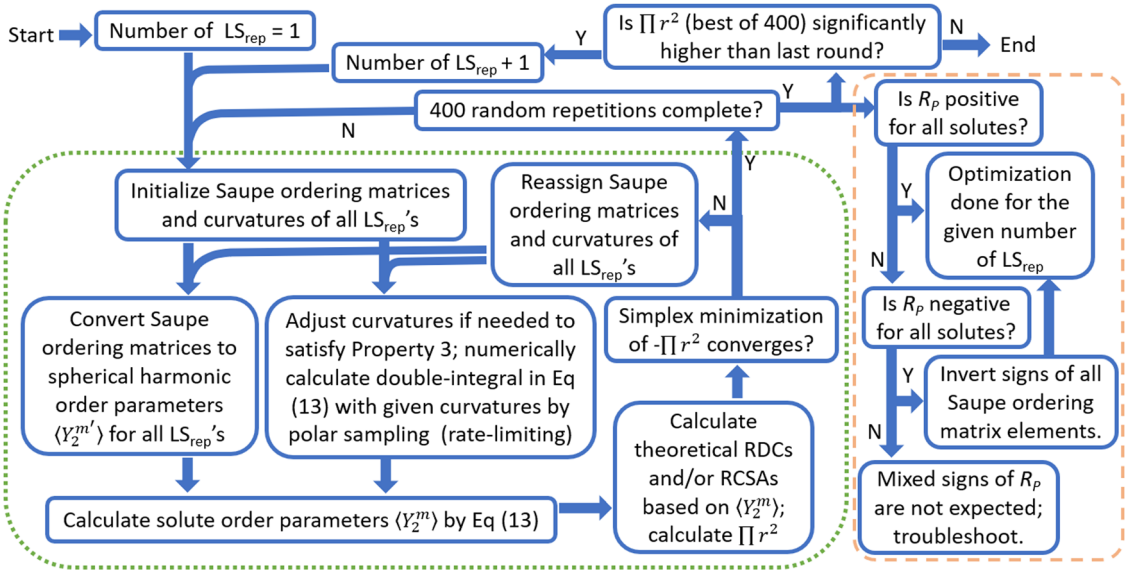

Fig. 7 Flow chart to determine landscape parameters by simplex minimization.

values from the previous round and randomly initializing the newly added LS$_{rep}$, until the improvement in agreement from adding a new LS$_{rep}$ becomes insignificant. To evaluate the likelihood of over-fitting as the number of LS$_{rep}$ increases, the entire process described above is cross validated by using a subset of data for parameter optimization and the remaining data for testing.

Steps inside the dotted simplex block are mostly based on descriptions in the theory section. First, initial values of landscape curvatures (k_x and k_z for a paraboloid, or k_x, k_z, and k_t for a twisted paraboloid) and Saupe ordering matrix parameters (S_{yy}^L and S_{zz}^L for a paraboloid, or S_{yy}^L, S_{zz}^L, and β_S for a twisted paraboloid) are assigned for each L_{rep} as previously described. Next, the Saupe ordering matrix is converted to spherical harmonic order parameters using the reverse relationships in eqn (7). In an independent process, the double-integral in eqn (13) is numerically calculated with the polar sampling method, based on landscapes generated with the initially assigned curvatures. Note that prior to this calculation, concave curvatures may need to be adjusted to satisfy Property 3. Calculating the double-integral is the rate-limiting step in the workflow, which involves a four-dimensional integration (three orientational dimensions and one positional dimension). Next, solute order parameters are calculated using eqn (13) (the normalization factor Z is set to 1 Å as previously described) if paraboloid-type landscapes are used, or eqn (17) if enantiomeric pairs of twisted paraboloid landscapes are used. After this step, RDCs and RCSAs can be predicted either directly from the spherical harmonic order parameters, or from Saupe ordering matrices after conversion by eqn (7). Agreement between predictions and experimental values is evaluated by the coefficient of determination r^2 based on a single regressor $y = \beta x$, in which x and y are the predicted and experimental values, respectively. No y-intercept is allowed because the regression line between predicted and measured data is expected to pass through the origin. Coefficients of determination from all solutes in the database are multiplied together and sign inverted, labeled as $-\prod r^2$, which is used as the minimization target to achieve best overall prediction. Landscape order parameters and curvatures are reassigned by the N–W simplex method after each iteration to minimize $-\prod r^2$. The updated parameters then enter the next iteration until a convergence criterion is met that indicates $-\prod r^2$ has reached a minimum.

Finally, we note that arbitrarily setting S_{zz}^L of the first LS$_{rep}$ to 1 seemingly causes an issue when a negative S_{zz}^L is called for. However, because r^2 does not differentiate positively *vs.* negatively correlated agreement owing to the sign flexibility of the slope (β) in the single regressor, setting the wrong sign for the first S_{zz} will simply strengthen a negative correlation after simplex minimization. Hence, a few extra steps after each round of 400 runs are undertaken (dashed orange block in Fig. 7), which employ the Pearson's R (R_p) to determine the sign of correlation. If a positive R_p is seen for all solutes, the optimized Saupe ordering matrices are kept as they are as the final parameters regarding the given number of LS$_{rep}$'s. However, if a negative R_p is seen for all solutes, the signs of all Saupe ordering matrix elements should be inverted for the final record. Note that R_p should be either all positive or all negative among different compounds. Presence of both positive and negative signs would indicate either method failure or inconsistency in the database, *e.g.*, combing data from compressed gels without sign inversion with data from stretched gels can cause mixed signs of R_p.

5. Results and discussion

The surface decomposition method was tested on an experimental database including RDC and/or RCSA data of 16 natural products of various structural complexity, including strychnine

(**1**),[39] estrone (**2**), retrorsine (**3**), aquatolide (**4**), caulamidine A (**5**), 10-*epi*-8-deoxycumambrin B (**6**),[40] mefloquine (**7**),[39] menthol (**8**),[39] ludartin (**9**),[41] parthenolide (**10**), yohimbine (**11**),[42] santonin (**12**),[43] sesquiterpenoid-13 (**13**),[44] artemisinin (**14**),[45] 19-OH-eburnamonine (**15**),[46] and eburnamonine (**16**).[42] Most of these compounds are highly rigid with a single predominant conformation. Retrorsine has three low energy rotamers regarding the hydroxymethyl group, but because the major rotamer constitutes over 90% of the population in chloroform and all three rotamers have very similar global geometries, only the major rotamer was considered during data fitting. All data were measured in PMMA gels cross-linked with ethylene glycol dimethacrylate (EGDMA). RDC and/or RCSA data for most compounds are taken from existing publications, and data for **2**, **3**, **4**, **5**, and **10** were collected during this study and are listed in the ESI.†

Prediction methods based on surface decomposition are referred to as either P*i* or TP*i*, where P or TP indicates a paraboloid landscape or a twisted paraboloid landscape, respectively, and the number i denotes the number of LS_{rep}'s employed. For example, P2 stands for a method using two paraboloid landscapes. Because the PMMA gel is atactic, the TP model engages pairs of landscape enantiomers. For example, TP3 represents a method using three enantiomeric pairs of twisted paraboloids. For comparison, we also examined predictions based on the simple cylindrical medium model, included in Table 1 as "cylinder". The cylinder radius is the only adjustable parameter in this model. The plane model was not explicitly considered because it represents a special case of the cylinder model, *i.e.*, a cylinder with an infinite radius, and therefore is accessible by the cylinder model through simplex minimization. Landscape curvatures and order parameters in P and TP models, or the cylinder radius in the cylinder model, are optimized as described in Section 4. The final prediction results and optimized landscape parameters are summarized in Table 1.

Although a crude approximation, the cylinder model coupled with axisymmetric order predicts reasonably well—many compounds display r^2 above 0.8 and all compounds have positive and significant R_p correlations (Table 1). Simplex minimization converges to a high cylindrical curvature of 0.827 $Å^{-1}$ (corresponding to a radius of 1.21 Å), thus favoring a thin rod model over a plane model, as expected for a PMMA polymer. However, as shown in Fig. 8a (blue curve), r^2 is not very consistent among different compounds, displaying some outliers significantly below average, such as **2** (r^2 = 0.64) and **5** (r^2 = 0.43).

Next, we evaluated different P models, starting with P1 and increasing the LS_{rep} number by one at a time up to P6. With only one LS_{rep}, P1 already achieves significantly better agreement with experimental data than the cylinder model, improving $\prod r^2$ from 0.06 to 0.23. As shown in Fig. 8a (orange *vs.* blue), r^2 is greatly improved for compounds that are poorly predicted by the cylinder model (see **1**, **2**, **3**, and **5** for example), although degraded r^2 is seen for **12**. The agreement consistency is also greatly improved, with r^2 falling in a narrower range of 0.8–1.0 for all sixteen compounds (orange curve). The improvement in both quality and consistency of agreement is owing to two additional fitting parameters, one associated with local surface curvature on an orthogonal direction, namely, k_z, and the other associated with order asymmetry of the medium, namely, S^L_{yy} (note that the cylinder model assumes an axisymmetric order). Not surprisingly, fitting with increased numbers of LS_{rep}'s from P1 to P6, which adds two curvature parameters and two order parameters with each LS_{rep} increment, boosts overall agreement progressively. As clearly shown in Fig. 8b (green *vs.* orange), P6 displays substantially better agreement over P1 in terms of both quality and consistency. Correlation plots of P6-predicted *vs.* experimental RDCs (red) and/or RCSAs (blue) for all sixteen compounds are shown in Fig. 9.

We further performed *F*-tests to evaluate whether the improved fitting by employing more LS_{rep}'s is statistically significant. The degree of freedom (DF) is calculated as 80 (5×16) minus the number of parameters in each model, giving 79, 77, 73, 69, 65, 61, and 57 for the cylinder and P1–P6 models, respectively. Note that DF calculation is not based on the total number of datapoints, which is 283 for all 16 compounds, but $5m$, with m being the number of compounds present (m = 16). This is because data in the same compound are not independent variables but interconnected through the 5 solute order parameters, *i.e.*, if the 5 order parameters are known, all data can be determined based on the fixed 3D geometry such that each compound can have at most 5 degrees of freedom. Based on this reasoning, the sum of squared errors (SSE) is calculated as the sum of average squared errors of each compound in order to give equal weightings to all compounds, which yields 673.1, 592.1, 481.0, 389.0, 359.0, 324.2, and 296.9 Hz^2 for the cylinder and P1–P6 models, respectively. Based on this statistical setup, the resulting *P*-value between the simpler model and the next model is 0.007 for cylinder *vs.* P1, 0.004 for P1 *vs.* P2, 0.005 for P2 *vs.* P3, 0.258 for P3 *vs.* P4, 0.178 for P4 *vs.* P5, and 0.277 for P5 *vs.* P6. The very small *P*-values from the cylinder model up to P3 clearly justifies the usage of more complex models. Starting from P4, however, the statistical significance of adding additional fitting parameters drops, indicating potential over-parametrization.

To further assess the likelihood of over-parametrization, we next cross-validated the prediction accuracy by omitting the experimental data of **2** and **5** from the landscape parameter training set, reserving these data for testing the trained parameters. Compounds **2** and **5** are chosen as challenging cases since they are poorly predicted by the cylinder model. Fig. 10 shows the r_{free}^2 of **2** (blue) and **5** (orange) and the geometric mean r^2 of the remaining fourteen training compounds ($\langle r^2 \rangle$, gray). The cylinder and P1–P6 models are evaluated for cross validation purposes. Prediction accuracy as reflected by r_{free}^2 clearly improves from the cylinder model to P3, although P2 yielded lower accuracy than P1 possibly due to instabilities from insufficient fitting parameters. From P4 to P6, prediction accuracy gradually degrades presumably because of over-parametrization, which is consistent with the *F*-test results, but the degradation is modest. The $\langle r^2 \rangle$ for the training set increases monotonically with more sophisticated models as

Table 1 Prediction results and optimized landscape parameters by different methods

Compound		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
r_{MEB} (Å)		6.9	7.2	6.5	7.0	6.5	5.8	7.1	5.9	5.5	6.0	6.0	7.8	5.9	5.9	6.9	7.2
Cylinder	R_p	0.86	0.80	0.87	0.96	0.66	0.94	0.98	0.97	0.98	0.92	0.98	0.97	0.99	0.99	0.92	0.94
	r^2	0.75	0.64	0.75	0.92	0.43	0.88	0.96	0.94	0.95	0.85	0.96	0.93	0.99	0.98	0.85	0.88
	$\prod r^2 = 0.06$																
	$k_x = 0.827$ Å^{-1}, $S^L_{zz} = 1.000$																
P1	R_p	0.96	0.90	0.92	0.95	0.94	0.93	0.99	0.97	0.97	0.94	0.97	0.91	0.99	0.99	0.97	1.00
	r^2	0.92	0.80	0.84	0.90	0.88	0.87	0.97	0.94	0.94	0.89	0.94	0.83	0.98	0.98	0.94	0.99
	$\prod r^2 = 0.23$																
	LS_{rep} #1: $k_x = 0.347$ Å^{-1}, $k_z = 0.580$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = -0.043$																
P2	R_p	0.97	0.94	0.97	0.96	0.93	0.94	0.99	0.97	0.98	0.96	0.96	0.92	0.99	0.99	0.97	0.98
	r^2	0.93	0.88	0.93	0.91	0.86	0.89	0.98	0.94	0.95	0.92	0.93	0.84	0.99	0.98	0.93	0.97
	$\prod r^2 = 0.30$																
	LS_{rep} #1: $k_x = 0.265$ Å^{-1}, $k_z = 0.697$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = 0.185$																
	LS_{rep} #2: $k_x = 0.827$ Å^{-1}, $k_z = 0.821$ Å^{-1}, $S^L_{zz} = -0.119$, $S_{yy} = -0.287$																
P3	R_p	0.98	0.97	0.96	0.96	0.94	0.95	0.99	0.98	0.98	0.98	0.98	0.94	0.99	0.98	0.97	0.99
	r^2	0.96	0.93	0.92	0.93	0.89	0.91	0.99	0.96	0.96	0.96	0.96	0.88	0.99	0.96	0.95	0.98
	$\prod r^2 = 0.40$																
	LS_{rep} #1: $k_x = 0.155$ Å^{-1}, $k_z = 0.679$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = 0.230$																
	LS_{rep} #2: $k_x = 0.827$ Å^{-1}, $k_z = 0.821$ Å^{-1}, $S^L_{zz} = -0.127$, $S^L_{yy} = -0.198$																
	LS_{rep} #3: $k_x = 0.830$ Å^{-1}, $k_z = 0.039$ Å^{-1}, $S^L_{zz} = -0.246$, $S^L_{yy} = -0.090$																
P4	R_p	0.98	0.97	0.95	0.96	0.95	0.97	1.00	0.98	0.98	0.98	0.98	0.95	0.99	0.97	0.98	0.99
	r^2	0.97	0.94	0.90	0.92	0.90	0.94	0.99	0.97	0.96	0.96	0.95	0.91	0.98	0.95	0.95	0.97
	$\prod r^2 = 0.43$																
	LS_{rep} #1: $k_x = 0.154$ Å^{-1}, $k_z = 0.682$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = 0.123$																
	LS_{rep} #2: $k_x = 0.827$ Å^{-1}, $k_z = 0.825$ Å^{-1}, $S^L_{zz} = -0.130$, $S^L_{yy} = -0.189$																
	LS_{rep} #3: $k_x = 0.829$ Å^{-1}, $k_z = 0.040$ Å^{-1}, $S^L_{zz} = -0.246$, $S^L_{yy} = -0.090$																
	LS_{rep} #4: $k_x = 0.648$ Å^{-1}, $k_z = 0.290$ Å^{-1}, $S^L_{zz} = -0.258$, $S^L_{yy} = 0.112$																
P5	R_p	0.98	0.97	0.97	0.97	0.96	0.97	1.00	0.98	0.98	0.99	0.98	0.95	0.99	0.98	0.98	0.99
	r^2	0.96	0.95	0.95	0.93	0.91	0.94	0.99	0.97	0.96	0.97	0.96	0.91	0.99	0.95	0.96	0.97
	$\prod r^2 = 0.48$																
	LS_{rep} #1: $k_x = 0.155$ Å^{-1}, $k_z = 0.682$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = 0.123$																
	LS_{rep} #2: $k_x = 0.775$ Å^{-1}, $k_z = 0.823$ Å^{-1}, $S^L_{zz} = -0.034$, $S^L_{yy} = -0.190$																
	LS_{rep} #3: $k_x = 0.823$ Å^{-1}, $k_z = 0.040$ Å^{-1}, $S^L_{zz} = -0.246$, $S^L_{yy} = -0.090$																
	LS_{rep} #4: $k_x = 0.652$ Å^{-1}, $k_z = 0.260$ Å^{-1}, $S^L_{zz} = -0.262$, $S^L_{yy} = 0.112$																
	LS_{rep} #5: $k_x = 0.710$ Å^{-1}, $k_z = 0.827$ Å^{-1}, $S^L_{zz} = 0.182$, $S^L_{yy} = 0.004$																
P6	R_p	0.97	0.98	0.98	0.97	0.96	0.98	1.00	0.99	0.98	0.99	0.98	0.96	0.99	0.98	0.98	0.98
	r^2	0.95	0.96	0.96	0.94	0.92	0.96	0.99	0.97	0.96	0.97	0.96	0.92	0.98	0.95	0.96	0.96
	$\prod r^2 = 0.50$																
	LS_{rep} #1: $k_x = 0.154$ Å^{-1}, $k_z = 0.682$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = 0.126$																
	LS_{rep} #2: $k_x = 0.780$ Å^{-1}, $k_z = 0.825$ Å^{-1}, $S^L_{zz} = -0.039$, $S^L_{yy} = -0.186$																
	LS_{rep} #3: $k_x = 0.821$ Å^{-1}, $k_z = 0.042$ Å^{-1}, $S^L_{zz} = -0.249$, $S^L_{yy} = -0.083$																
	LS_{rep} #4: $k_x = 0.652$ Å^{-1}, $k_z = 0.260$ Å^{-1}, $S^L_{zz} = -0.307$, $S^L_{yy} = 0.120$																
	LS_{rep} #5: $k_x = 0.717$ Å^{-1}, $k_z = 0.827$ Å^{-1}, $S^L_{zz} = 0.169$, $S^L_{yy} = 0.003$																
	LS_{rep} #6: $k_x = 0.382$ Å^{-1}, $k_z = 0.400$ Å^{-1}, $S^L_{zz} = 0.137$, $S^L_{yy} = -0.019$																
TP1	R_p	0.95	0.89	0.94	0.97	0.96	0.94	0.98	0.98	0.96	0.96	0.96	0.94	0.99	0.99	0.98	0.99
	r^2	0.91	0.79	0.89	0.94	0.92	0.88	0.97	0.95	0.93	0.92	0.93	0.88	0.97	0.97	0.96	0.98
	$\prod r^2 = 0.27$																
	LS_{rep} enantiomer pair #1: $k_x = 0.113$ Å^{-1}, $k_z = 0.712$ Å^{-1}, $k_t = \pm 0.188$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = -0.022$, $\beta_S = \pm 0.727$																
TP2	R_p	0.97	0.95	0.97	0.97	0.92	0.93	0.99	0.97	0.98	0.97	0.97	0.94	1.00	0.99	0.97	0.98
	r^2	0.94	0.91	0.93	0.93	0.85	0.87	0.98	0.94	0.95	0.93	0.94	0.88	0.99	0.97	0.94	0.97
	$\prod r^2 = 0.32$																
	LS_{rep} enantiomer pair #1: $k_x = 0.269$ Å^{-1}, $k_z = 0.542$ Å^{-1}, $k_t = \pm 0.210$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = 0.185$, $\beta_S = \pm 0.270$																
	LS_{rep} enantiomer pair #2: $k_x = 0.775$ Å^{-1}, $k_z = 0.826$ Å^{-1}, $k_t = \mp 0.013$ Å^{-1}, $S^L_{zz} = -0.422$, $S^L_{yy} = -0.285$, $\beta_S = \pm 0.206$																
TP3	R_p	0.98	0.97	0.96	0.96	0.95	0.95	0.99	0.98	0.98	0.98	0.98	0.94	0.99	0.98	0.98	0.99
	r^2	0.96	0.94	0.91	0.93	0.90	0.91	0.99	0.96	0.96	0.97	0.96	0.88	0.99	0.96	0.95	0.98
	$\prod r^2 = 0.41$																
	LS_{rep} enantiomer pair #1: $k_x = 0.155$ Å^{-1}, $k_z = 0.682$ Å^{-1}, $k_t = \pm 0.000$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = 0.231$, $\beta_S = \pm 3.126$																
	LS_{rep} enantiomer pair #2: $k_x = 0.822$ Å^{-1}, $k_z = 0.824$ Å^{-1}, $k_t = \pm 0.002$ Å^{-1}, $S^L_{zz} = -0.126$, $S^L_{yy} = -0.197$, $\beta_S = \pm 0.173$																
	LS_{rep} enantiomer pair #3: $k_x = 0.830$ Å^{-1}, $k_z = 0.039$ Å^{-1}, $k_t = \pm 0.000$ Å^{-1}, $S^L_{zz} = -0.248$, $S^L_{yy} = -0.091$, $\beta_S = \pm 0.004$																
TP4	R_p	0.97	0.97	1.00	0.97	0.95	0.97	1.00	0.97	0.98	0.98	0.99	0.94	0.99	0.98	0.97	0.99
	r^2	0.94	0.94	0.99	0.93	0.90	0.95	0.99	0.94	0.96	0.97	0.97	0.88	0.98	0.95	0.94	0.97
	$\prod r^2 = 0.44$																
	LS_{rep} enantiomer pair #1: $k_x = 0.153$ Å^{-1}, $k_z = 0.639$ Å^{-1}, $k_t = \pm 0.070$ Å^{-1}, $S^L_{zz} = -1.000$, $S^L_{yy} = 0.128$, $\beta_S = \pm 0.052$																
	LS_{rep} enantiomer pair #2: $k_x = 0.826$ Å^{-1}, $k_z = 0.824$ Å^{-1}, $k_t = \pm 0.014$ Å^{-1}, $S^L_{zz} = -0.132$, $S^L_{yy} = -0.190$, $\beta_S = \pm 1.559$																
	LS_{rep} enantiomer pair #3: $k_x = 0.827$ Å^{-1}, $k_z = 0.037$ Å^{-1}, $k_t = \pm 0.040$ Å^{-1}, $S^L_{zz} = -0.430$, $S^L_{yy} = -0.090$, $\beta_S = \pm 0.431$																
	LS_{rep} enantiomer pair #4: $k_x = 0.650$ Å^{-1}, $k_z = 0.239$ Å^{-1}, $k_t = \pm 0.070$ Å^{-1}, $S^L_{zz} = -0.255$, $S^L_{yy} = 0.112$, $\beta_S = \pm 0.004$																

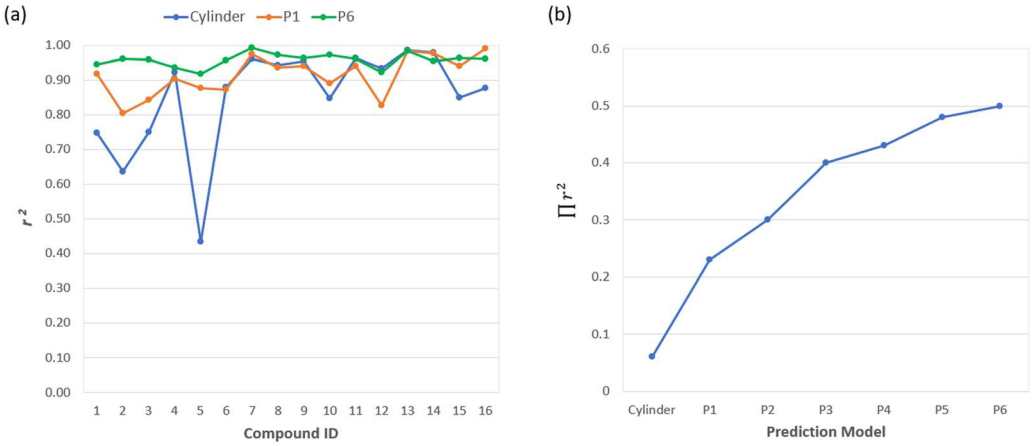

Fig. 8 Agreement between prediction and measurement. (a): Congruence between predicted and experimental data as reflected by r^2 for sixteen compounds based on the cylinder, P1, and P6 models; (b): Overall congruence for all sixteen compounds as reflected by $\prod r^2$ from using prediction models of increasing levels of parametrization.

expected and appears to reach a plateau after P4. These results suggest that P3 is the most suitable model for the fourteen training compounds because ${r_{\text{free}}}^2$ and $\langle r^2 \rangle$ are both high and similar.

To take local chirality of PMMA into account, we also evaluated four TP-based models: TP1-4. TP*i* is expected to yield higher agreement with experimental data than the corresponding P*i*, because it contains two additional fitting parameters for each LS_{rep} employed, namely a twist curvature k_t (eqn (16)) and an *Euler* angle β_S (eqn (18)). The improvement is observed as expected but tends to be moderate, especially when more than two LS_{rep}'s are employed, *e.g.*, $\prod r^2$ increases from 0.23 in P1 to 0.27 in TP1, from 0.30 in P2 to 0.32 in TP2, from 0.40 in P3 to 0.41 in TP3, and from 0.43 in P4 to 0.44 in TP4 (Table 1). It is also clear from Table 1 that as larger numbers of LS_{rep}'s are used, the twist curvature optimizes to very small numbers for all LS_{rep}'s, which suggests that several regular paraboloid-type landscapes are adequate surface descriptors for the PMMA gel. The lack of benefit from introducing surface chirality is clearly shown by *F*-tests as well, with *P*-values of 0.057 for TP1 *vs.* P2, 0.510 for TP2 *vs.* P2, 0.985 for TP3 *vs.* P3, and 0.568 for TP4 *vs.* P4. This result is presumably because the stereocenters on an atactic PMMA polymer have random chirality and therefore cannot combine constructively to form a local surface of significant asymmetry.

It is worth noting that for all P*i* and TP*i* models, fixing the S^L_{zz} of the first LS_{rep} to 1 during simplex optimization yielded negative R_p's for all compounds, which indicates that a negative sign should be assigned to this S^L_{zz}. As previously mentioned, assigning a wrong sign to the first S^L_{zz} does not cause any issue when $\prod r^2$ is used as the target of minimization. However, all S^L_{zz} and S^L_{yy} values should be sign flipped after simplex optimization to effect a positive correlation between predicted and measured data. For this reason, Table 1 only reports the sign inverted Saupe ordering parameters that yield positive R_p's for all compounds.

Landscape curvatures determined from simplex minimization can provide insight into gel surface structure. Notably, only positive curvatures are obtained after minimization (Table 1), suggesting that the MEB-ES associated with the PMMA gel surface is dominated by convex landscapes. Based on curvatures from different P*i* models, the MEB-ES appears to contain two general types of landscapes. One type is a nearly symmetrical hump (Fig. 5d) with high curvatures on both *X* and *Z* directions, such as LS_{rep} #2 in P2-P6 (see Table 1). The other type is a highly asymmetric hump with a high curvature on one direction and a low curvature on the other direction, visually resembling a ridge (Fig. 5b), such as LS_{rep} #1 and #3 in P3–P6. P1 contains only one LS_{rep} with intermediate curvatures, likely due to the need to satisfy both landscape types with limited parameters, but as more LS_{rep}'s are employed for surface decomposition, the segregation into hump-like and ridge-like landscapes becomes clearer. In the context of PMMA polymer structure, it is tempting to speculate that the hump-like landscape is related to a bulging side group such as the methyl or methyl ester, whereas the ridge-like landscape is related to the main polymer chain. The lack of concave representatives in the optimized LS_{rep}'s is likely because the solute molecules in the database are large in comparison to the gel surface cavities such that concave surfaces are poorly accessible (compare Fig. 2a and b). No correlation is observed between r_{MEB} and r^2 (Table 1), *i.e.*, the agreement between prediction and measurement is independent of solute size, which is also consistent with a predominantly convex MEB-ES (Property 1).

Finally, we should emphasize that the surface decomposition method uses medium structural descriptors as fitting variables to reproduce solute anisotropic NMR data based on a steric interaction model. It is not expected to achieve the same level of agreement with experimental data as other fitting methods that directly parametrize solute order parameters, in particular, the SVD method. However, SVD cannot make use of commonalities in a compound database but only ensures the best agreement for each individual compound, whereas the surface decomposition method simultaneously achieves the best overall agreement for all compounds in the database using

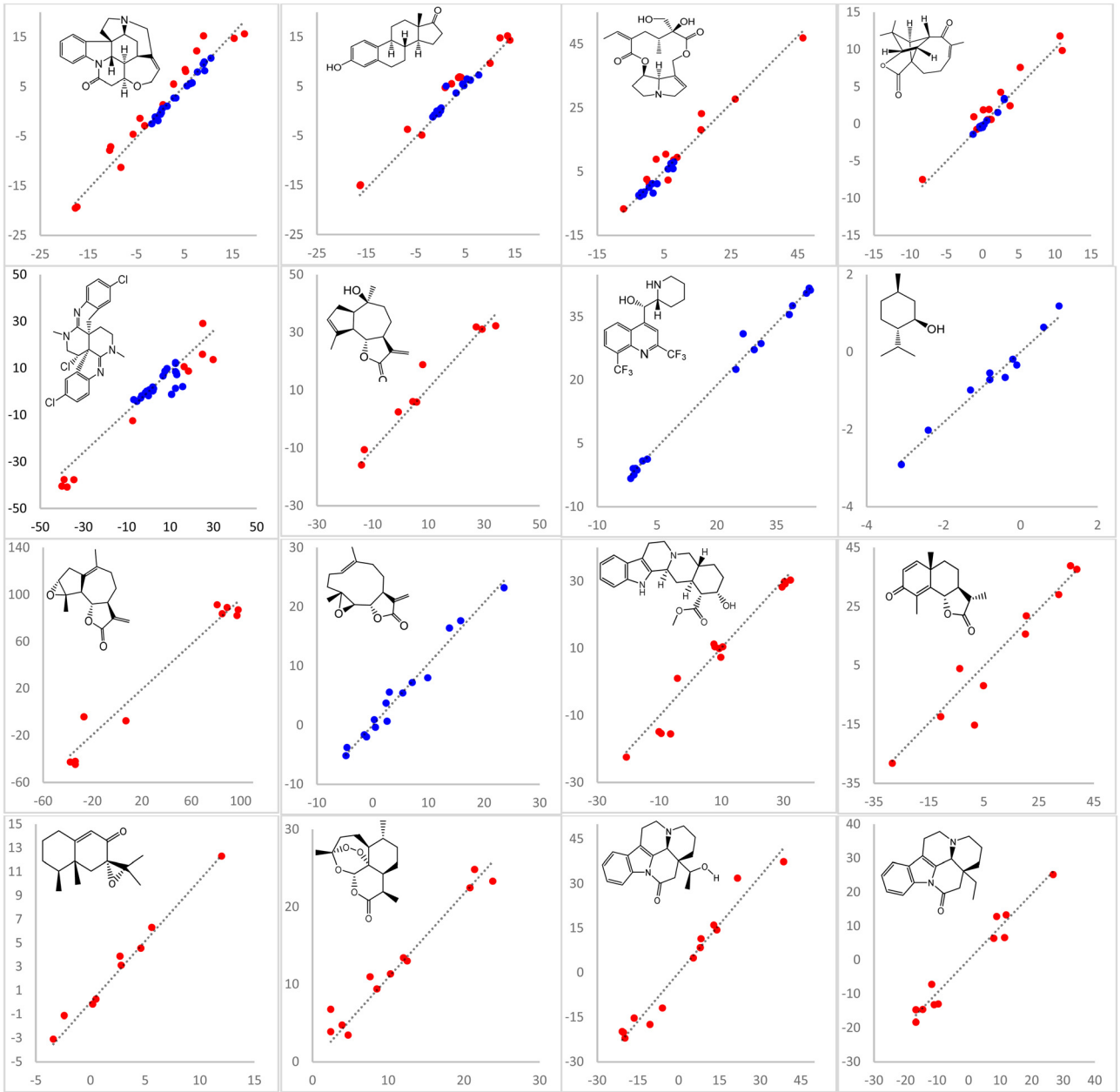

Fig. 9 Correlation between P6-predicted (Y) and measured (X) RDCs (red) and RCSAs (blue). The predicted data are rescaled into approximately the same range as the experimental data by a factor of $(exp_{max} - exp_{min})/(pred_{max} - pred_{min})$ to facilitate inspection.

medium parameters as common fitting variables. From a data fitting standpoint, the surface decomposition method utilizes fitting variables more efficiently than SVD or other methods that directly parametrize solute order parameters. For example, parametrizing solute order parameters directly requires Mm fitting variables for m molecule, where M is 1, 3, or 5 depending on the molecular symmetry.[47–49] In the general case of an anisotropic molecular structure, $5m$ fitting variables are needed, or if the alignment amplitude is not considered as in this work, $4m$ fitting variables depicting alignment asymmetry and orientation are needed. In contrast, if the paraboloid medium model of the surface decomposition method is adopted, $4n - 1$ fitting variables are required (*vide supra*), where n is the number of LS_{rep}'s. When a database containing a significant number of compounds is available, $4n$ can be much smaller than $4m$ (in this work $n \leq 6$ while $m = 16$) such that surface decomposition allows dramatic variable reduction. Perhaps the most interesting application of surface decomposition is to differentiate structural candidates of flexible compounds. The challenge with analysing a flexible molecule of m conformations is somewhat analogous to simultaneously fitting m compounds in a database, but even greater than fitting a database because only one set of NMR data representing the average of m conformations are experimentally

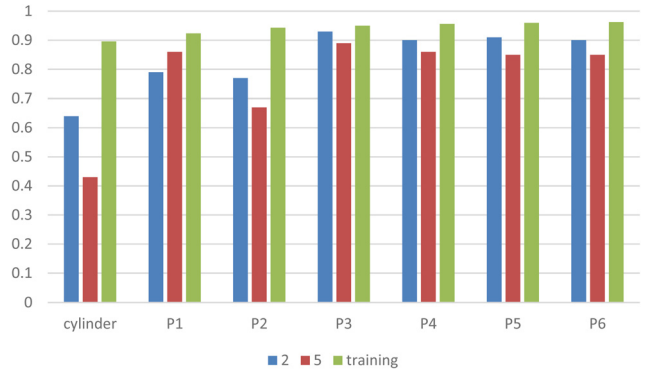

Fig. 10 Cross validation of prediction accuracy by the cylinder and P1–P6 models. The *Y* axis indicates r_{free}^2 for **2** and **5** or $\langle r^2 \rangle$ for the other fourteen compounds used for training.

available. With medium parameters determined from a sufficient database of rigid compounds, the surface decomposition method can potentially allow flexible structural differentiation without employing any fitting variables, which can either complement SVD-based methods[50–52] as an independent means of verification or provide analyses when SVD-based methods are inapplicable (*vide supra*).

6. Conclusion

In this study we proposed an alignment prediction method for sterically ordered compounds. This method does not require a global structural model for the alignment medium. It works by decomposing the alignment medium surface into simple elemental landscapes, whose shapes and orientations are parametrized by surface curvatures and order parameters, respectively. An order transfer equation was developed to connect these landscape parameters to the order parameters of the solute, which therefore allowed prediction of anisotropic NMR data. The landscape parameters, *i.e.*, curvatures and order parameters, determined by fitting predictions to an experimental database can be used to predict the alignment of an arbitrary compound of interest. The surface decomposition method represents a general approach to predict steric alignment because it can handle a wide variety of surface landscapes and arbitrary medium symmetry. The idealized medium models, such as the cylinder and disc models of axisymmetric order as commonly adopted for dilute LLC media, are special cases, which the surface decomposition method can reduce to under simplified conditions. Not surprisingly, for the structurally complex PMMA gel, the surface decomposition method achieves greatly improved agreement between predictions and experimental measurements. Importantly, this method is not prone to over-parametrization and predicts alignment significantly more accurately than the cylinder model according to cross validation tests. The mathematical framework developed here is general and could likely be applied to other popular alignment media when steric interaction is the dominant alignment mechanism. The ability to accurately predict the alignment allows a researcher to minimize the number of anisotropic NMR measurements in order to determine the structure and the stereochemical configuration of organic small molecules and natural products. Even more importantly, this approach potentially expands the scope of anisotropic NMR applications to highly flexible molecules that have traditionally been considered intractable for anisotropic NMR.

Finally, the limitations of this method should also be mentioned. First, the prediction considers steric interaction between the alignment medium and a monomeric solute structure but does not account for solute–solute interactions that can result in solute oligomers or aggregates that can interact with the medium differently than a monomeric solute. We previously noted that the experimentally obtained alignment orientation of estrone in a PMMA gel was unexpected based on a qualitative assessment of its molecular shape likely due to the low solubility of estrone in chloroform.[39] In this work, a mixed solvent system with 95% $CDCl_3$ and 5% DMSO-d_6 (v/v) was used for estrone and retrorsine (see SI), which improved solubility and yielded anisotropic NMR data that are consistent with predictions based on a monomeric solute structure (Table 1). Second, ignoring electrostatic interactions may be a source of error for some medium-solvent-solute systems. While nonspecific charge-charge interactions are expected to be weak for neutral media and solutes, the formation of hydrogen-bond can cause more significant orientational preference that is not predicted by a steric model. In this work, we have considered optimizing for an energy factor to facilitate a potential hydrogen-bond donor on the solute to face towards the landscape surface, assuming that the donor may form a hydrogen bond with the PMMA acrylate group, but we did not observe improvement over a pure steric interaction model, suggesting at least in the PMMA gel, hydrogen bonding with the solute is unlikely a significant mechanism of alignment. From a practical standpoint, stable hydrogen bonding of the solute with the medium contributes to resonance line broadening and represents an undesirable interaction. Whether a pure steric interaction model is also applicable to other organic alignment systems requires further investigation, but nevertheless, adequately accounting for the dominant steric interactions using the surface decomposition method allows more meaningful evaluation on the relevance of other intermolecular interactions.

Conflicts of interest

There are no conflicts to declare.

References

1 J. W. Emsley and J. C. Lindon, *NMR spectroscopy using liquid crystal solvents*, Pergamon Press, Oxford, New York, 1st edn, 1975.
2 E. E. Burnell and C. A. De Lange, *NMR of ordered liquids*, Kluwer Academic Publishers, Dordrecht, Boston, 2003.

3 N. Tjandra and A. Bax, *Science*, 1997, **278**, 1697.
4 C. Aroulanda, V. Boucard, F. Guibe, J. Courtieu and D. Merlet, *Chem. – Eur. J.*, 2003, **9**, 4536–4539.
5 G. Kummerlowe and B. Luy, *Ann. Rep. NMR Spectrosc.*, 2009, **68**(68), 193–232.
6 R. R. Gil, C. Griesinger, A. Navarro-Vázquez and H. Sun, *Structure Elucidation in Organic Chemistry The Search for the Right Tools Preface*, Wiley-VCH, 2015.
7 E. E. Burnell and C. A. de Lange, *Chem. Rev.*, 1998, **98**, 2359–2387.
8 A. F. Terzis and D. J. Photinos, *Mol. Phys.*, 1994, **83**, 847–865.
9 J. W. Emsley, W. E. Palke and G. N. Shilstone, *Liq. Cryst.*, 1991, **9**, 643–648.
10 J. A. Losonczi, M. Andrec, M. W. F. Fischer and J. H. Prestegard, *J. Magn. Reson.*, 1999, **138**, 334–342.
11 M. Zweckstetter and A. Bax, *J. Am. Chem. Soc.*, 2000, **122**, 3791–3792.
12 C. A. Bewley and G. M. Clore, *J. Am. Chem. Soc.*, 2000, **122**, 6009–6016.
13 H. F. Azurmendi and C. A. Bush, *Carbohydr. Res.*, 2002, **337**, 905–915.
14 J. A. Marsh, J. M. Baker, M. Tollinger and J. D. Forman-Kay, *J. Am. Chem. Soc.*, 2008, **130**, 7804–7805.
15 G. Nodet, L. Salmon, V. Ozenne, S. Meier, M. R. Jensen and M. Blackledge, *J. Am. Chem. Soc.*, 2009, **131**, 17908–17918.
16 J. R. Huang, V. Ozenne, M. R. Jensen and M. Blackledge, *Angew. Chem., Int. Ed.*, 2013, **52**, 687–690.
17 J. C. Xia and D. A. Case, *Biopolymers*, 2012, **97**, 276–288.
18 M. Martin-Pastor, A. Canales, F. Corzana, J. L. Asensio and J. Jimenez-Barbero, *J. Am. Chem. Soc.*, 2005, **127**, 3589–3595.
19 A. Ibanez de Opakua, F. Klama, I. E. Ndukwe, G. E. Martin, R. T. Williamson and M. Zweckstetter, *Angew. Chem., Int. Ed.*, 2020, **59**, 6172–6176.
20 Y. Liu, J. Sauri, E. Mevers, M. W. Peczuh, H. Hiemstra, J. Clardy, G. E. Martin and R. T. Williamson, *Science*, 2017, **356**, 43.
21 Y. Z. Liu, A. Navarro-Vazquez, R. R. Gil, C. Griesinger, G. E. Martin and R. T. Williamson, *Nat. Protoc.*, 2019, **14**, 217.
22 A. O. Frank, J. C. Freudenberger, A. K. Shaytan, H. Kessler and B. Luy, *Magn. Reson. Chem.*, 2015, **53**, 213–217.
23 J. W. Emsley, S. K. Heeks, T. J. Horne, M. H. Howells, A. Moon, W. E. Palke, S. U. Patel, G. N. Shilstone and A. Smith, *Liq. Cryst.*, 1991, **9**, 649–660.
24 R. T. Syvitski and E. E. Burnell, *Chem. Phys. Lett.*, 1997, **281**, 199–206.
25 R. T. Syvitski and E. E. Burnell, *J. Chem. Phys.*, 2000, **113**, 3452–3465.
26 A. F. Terzis, C. D. Poon, E. T. Samulski, Z. Luz, R. Poupko, H. Zimmermann, K. Muller, H. Toriumi and D. J. Photinos, *J. Am. Chem. Soc.*, 1996, **118**, 2226–2234.
27 T. Dingemans, D. J. Photinos, E. T. Samulski, A. F. Terzis and C. Wutz, *J. Chem. Phys.*, 2003, **118**, 7046–7061.
28 A. Pizzirusso, M. B. Di Cicco, G. Tiberio, L. Muccioli, R. Berardi and C. Zannoni, *J. Phys. Chem. B*, 2012, **116**, 3760–3771.
29 A. C. J. Weber, A. Pizzirusso, L. Muccioli, C. Zannoni, W. L. Meerts, C. A. de Lange and E. E. Burnell, *J. Chem. Phys.*, 2012, **136**, 174506.
30 J. M. Polson and E. E. Burnell, *Mol. Phys.*, 1996, **88**, 767–782.
31 J. M. Polson and E. E. Burnell, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1997, **55**, 4321–4337.
32 M. Zweckstetter, G. Hummer and A. Bax, *Biophys. J.*, 2004, **86**, 3444–3460.
33 J. M. Rallison and S. E. Harding, *J. Colloid Interface Sci.*, 1985, **103**, 284–289.
34 J. Courtieu, J. P. Bayle and B. M. Fung, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1994, **26**, 141–169.
35 G. Kummerlowe, S. L. Grage, C. M. Thiele, I. Kuprov, A. S. Ulrich and B. Luy, *J. Magn. Reson.*, 2011, **209**, 19–30.
36 W. L. Jorgensen, D. S. Maxwell and J. TiradoRives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
37 I. E. Ndukwe, A. Brunskill, D. R. Gauthier, Y. L. Zhong, G. E. Martin, R. T. Williamson, M. Reibarkh and Y. Z. Liu, *Org. Lett.*, 2019, **21**, 4072–4076.
38 J. A. Nelder and R. Mead, *Comput. J.*, 1965, **7**, 308–313.
39 N. Nath, M. Schmidt, R. R. Gil, R. T. Williamson, G. E. Martin, A. Navarro-Vazquez, C. Griesinger and Y. Z. Liu, *J. Am. Chem. Soc.*, 2016, **138**, 9548–9556.
40 C. Gayathri, N. V. Tsarevsky and R. R. Gil, *Chem. – Eur. J.*, 2010, **16**, 3622–3626.
41 R. R. Gil, C. Gayathri, N. V. Tsarevsky and K. Matyjaszewski, *J. Org. Chem.*, 2008, **73**, 840–848.
42 E. Troche-Pesqueira, C. Anklin, R. R. Gil and A. Navarro-Vazquez, *Angew. Chem., Int. Ed.*, 2017, **56**, 3660–3664.
43 F. Hallwass, R. R. Teles, E. Hellemann, C. Griesinger, R. R. Gil and A. Navarro-Vazquez, *Magn. Reson. Chem.*, 2018, **56**, 321–328.
44 S. J. Castro, M. E. Garcia, J. M. Padron, A. Navarro-Vazquez, R. R. Gil and V. E. Nicotra, *J. Nat. Prod.*, 2018, **81**, 2329–2337.
45 A. Navarro-Vazquez, R. R. Gil and K. Blinov, *J. Nat. Prod.*, 2018, **81**, 203–210.
46 P. Trigo-Mourino, R. Sifuentes, A. Navarro-Vazquez, C. Gayathri, H. Maruenda and R. R. Gil, *Nat. Prod. Commun.*, 2012, **7**, 735–738.
47 I. E. Ndukwe, Y. H. Lam, S. K. Pandey, B. E. Haug, A. Bayer, E. C. Sherer, K. A. Blinov, R. T. Williamson, J. Isaksson, M. Reibarkh, Y. Z. Liu and G. E. Martin, *Chem. Sci.*, 2020, **11**, 12081–12088.
48 A. Saupe, *Angew. Chem., Int. Ed. Engl.*, 1968, **7**, 97–112.
49 H. M. Al-Hashimi, P. J. Bolon and J. H. Prestegard, *J. Magn. Reson.*, 2000, **142**, 153–158.
50 C. M. Thiele, V. Schmidts, B. Bottcher, I. Louzao, R. Berger, A. Maliniak and B. Stevensson, *Angew. Chem., Int. Ed.*, 2009, **48**, 6708–6712.
51 V. M. Sanchez-Pedregal, R. Santamaria-Fernandez and A. Navarro-Vazquez, *Org. Lett.*, 2009, **11**, 1471–1474.
52 H. Sun, U. M. Reinscheid, E. L. Whitson, E. J. d'Auvergne, C. M. Ireland, A. Navarro-Vazquez and C. Griesinger, *J. Am. Chem. Soc.*, 2011, **133**, 14629–14636.

