Liquid Water and Interfacial, Cubic, and Hexagonal Ice Classification through Eclipsed and Staggered Conformation Template Matching

Golnaz Roudsari,*,† Farshad G. Veshki,‡ Bernhard Reischl,† and Olli H. Pakarinen†

†*Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, P.O. Box 64, FI-00014, Finland*

‡*Department of Signal Processing and Acoustics, School of Electrical Engineering, Aalto University, P.O. Box 11000, FI-00076, Finland*

E-mail: golnaz.roudsari@helsinki.fi

Abstract

We propose a novel method based on template-matching for the recognition of liquid water, cubic ice (ice I_c), hexagonal ice (ice I_h), clathrate hydrates, as well as different interfacial structures in atomistic and coarse-grained simulations of water and ice. The two template matrices represent the staggered and eclipsed conformations which are the building blocks of hexagonal and cubic ice , as well as clathrate crystals. The algorithm is rotationally invariant and highly robust against imperfections in the ice structure, and its sensitivity for recognizing ice-like structures can be tuned for different applications. Unlike most other algorithms, it can discriminate between cubic-, hexagonal-, clathrate-, mixed-, and other interfacial ice types, and is therefore well-suited to study complex systems and heterogeneous ice nucleation.

1 Introduction

Freezing of water is a ubiquitous phenomenon which plays an important role in nature and in technological applications. As pure water only freezes homogeneously when it is cooled more than ~ 38 K below the melting point[1,2], ice mostly forms through heterogeneous nucleation, where the formation of a critical ice cluster is aided by the presence of a surface. The atomistic details of nucleation mechanisms and the atomic-level structure of a critical ice cluster are difficult to study experimentally, due to the limits of a spatial and temporal resolution.[3,4] Computer simulations of homogeneous and heterogeneous ice nucleation using atomistic or coarse-grained models of water can provide valuable insight. However, the simulation-based studies require accurate and efficient algorithms for distinguishing the liquid and crystalline phases, and identifying different structures of ice crystals.[5–9] Depending on the temperature and pressure, more than 18 distinct ice polymorphs can exist.[10] However, under atmospheric conditions, only cubic ice (ice I_c) and hexagonal ice (ice I_h) are viable.[11–13]. Therefore, most available ice recognition algorithms are developed to enable the identification of these crystal polymorphs.[14–16]

A majority of commonly used ice structure recognition methods are based on measuring a bond order parameter using spherical harmonics analysis.[14,15,17–21] Specifically, in these methods, a local orientational parameter vector is calculated for each atom, by averaging over the spherical harmonics coefficients corresponding to the bond vectors between that atom and its four nearest neighbours. The inner product of the local orientational parameter vectors of pairs of neighbouring atoms is used as an alignment measure for analysis of the local structure of the ice crystal.[14,15,21,22] Such an alignment measure is used for the identification of staggered and eclipsed conformations in the ice structure, which can be effectively used for the recognition of cubic and hexagonal ice.[14,15] In addition, the average Euclidean norm of the local orientational parameter vectors of all of the atoms is used as a global orientational parameter for approximation of the ratio of crystal and liquid phases in the whole system.[17]

Other approaches for ice structure recognition include methods based on bond angle anal-

ysis. For example, Brukhno *et al.* suggest a method for recognition of cubic and hexagonal ice with known and fixed orientations, based on an average maximum correlation between the bonds in a tetrahedral ice crystal and fourteen predetermined director vectors.[23] However, this method is not rotationally invariant. Thus, it cannot be used when the ice crystal can grow in arbitrary orientations or structures.[14]

In another work, Geiger *et al.*[24] employed a set of symmetry functions for extracting features from the local ice crystal structures. An artificial neural network is then trained over the extracted features and used for classification of different ice structures. However, the choice of suitable symmetry functions is not straight-forward in this method. Specifically, selection of the symmetry functions is computationally expensive and should be performed partly manually. An ice recognition method based on deep learning is also proposed by Fulford *et al.*[25] where concatenations of several representations of the local ice structures (including Cartesian coordinates, spherical coordinates, Fourier transform of histogram of Cartesian coordinates, and spherical harmonics of degrees 2, 3 and 4) are used as inputs to multiple parallel deep neural networks. The outputs of the parallel networks are then concatenated and used as inputs for another fully connected deep neural network trained for recognition of ice from liquid water (the ice structures are not identified in this method).

Maras *et al.*[26] extended the common neighbour analysis (CNA) method[27,28] for identification of cubic and hexagonal diamond structures (17-atom structures). Larsen *et al.*[29] proposed a template matching approach for detection of different lattice structures in the crystalline solids, which can be also employed for identification of hexagonal and cubic diamond structures in ice. In a different approach, ring analysis based on topological features of cubic and hexagonal ice has been used by Haji-Akbari *et al.*[16] for recognition of ice structures.

In this paper, we propose a novel conformation template matching approach for identification of cubic ice and hexagonal ice structures, different interfacial structures, as well as clathrate hydrates ("LICH-TEST"). The existence of different ice polymorphs is investigated based on the conformations of a given water oxygen atom and its four nearest oxygen neigh-

bours. Specifically, the proposed method identifies staggered and eclipsed ice conformations by determining the similarity between the local structure and two templates representing the two aforementioned conformations.

The conformation templates used in the algorithm contain rotationally invariant information about local structures, involving at most eight water molecules; moreover, the template matching stage does not require four neighbours to be present. As a result, the proposed method is very versatile and well-suited to classify interfacial and defected crystal structures. In addition, since our method is based on straightforward and computationally inexpensive operations in the Cartesian coordinate system, it is analytically interpretable and computationally efficient. The selectivity of the template matching can be easily tuned by a single parameter.

The remainder of this paper is organised as follows: In section 2, we describe the eclipsed and staggered geometries and in section 3 we provide the details of the template matching algorithm. In section 4, we present applications of the LICH-TEST algorithm to different examples: heterogeneous ice nucleation at a mineral surface, ice recognition of an antifreeze protein, and clathrate hydrates. We discuss the sensitivity of the algorithm and benchmark against the CHILL+ algorithm.[15] Section 5 summarises and concludes this work.

2 Eclipsed and staggered conformation templates

In perfect cubic and hexagonal ice crystals and clathrate hydrates, each water molecule's oxygen atom is at the center of a regular tetrahedron formed by its four nearest oxygen atom neighbors; moreover, each pair of tetrahedra which have neighbouring oxygen atoms have either symmetric or anti-symmetric arrangements with respect to the O-O bond between the two neighbouring (or "central") atoms. These symmetric and anti-symmetric arrangements define the eclipsed and staggered conformations,[23] respectively, and are illustrated in Fig. 1. In cubic ice, all four bonds between an oxygen atom and its four nearest neighbours have

staggered conformations. In hexagonal ice, only three of these bonds have staggered conformations and one bond has an eclipsed conformation. In clathrate hydrates, all four bonds have eclipsed conformations.[15]

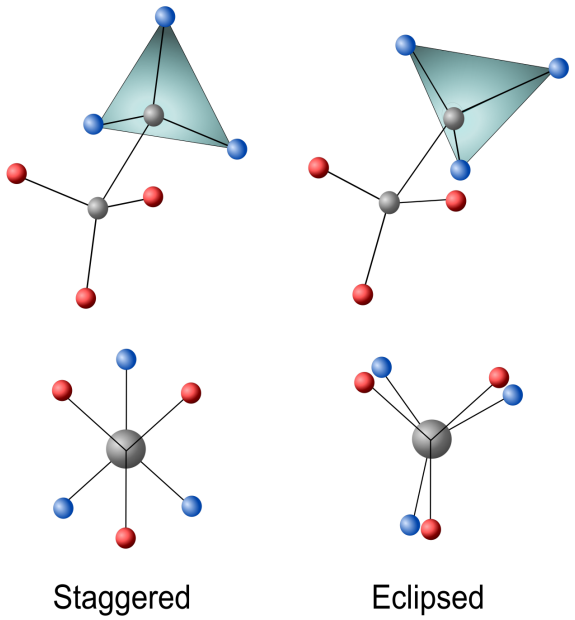

Figure 1: Sawhorse (top) and Newman (bottom) projections of staggered (left) and eclipsed (right) conformations.

In Fig. 2, the construction of an ice crystal from the connection of two tetrahedral oxygen structures is presented, where the unit vectors $\mathbf{u}_i$ and $\mathbf{v}_i$, $i = 1, \ldots, 4$, denote the bond directions originating at the two neighbouring central oxygen atoms. Such an ice structure can be captured using a rotationally invariant representation matrix $\mathbf{T}$ whose elements are the inner products of the bond directions $\mathbf{u}_i$ and $\mathbf{v}_i$. Since $\mathbf{u}_i$ and $\mathbf{v}_i$ are unit vectors, the inner product gives the cosine of the angles between the bond directions.

We form the structure representation matrix $\mathbf{T}$ as follows:

$$\mathbf{T} = \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{v}_1 \rangle & \langle \mathbf{u}_1, \mathbf{v}_2 \rangle & \langle \mathbf{u}_1, \mathbf{v}_3 \rangle & \langle \mathbf{u}_1, \mathbf{v}_4 \rangle \\ \langle \mathbf{u}_2, \mathbf{v}_1 \rangle & \langle \mathbf{u}_2, \mathbf{v}_2 \rangle & \langle \mathbf{u}_2, \mathbf{v}_3 \rangle & \langle \mathbf{u}_2, \mathbf{v}_4 \rangle \\ \langle \mathbf{u}_3, \mathbf{v}_1 \rangle & \langle \mathbf{u}_3, \mathbf{v}_2 \rangle & \langle \mathbf{u}_3, \mathbf{v}_3 \rangle & \langle \mathbf{u}_3, \mathbf{v}_4 \rangle \\ \langle \mathbf{u}_4, \mathbf{v}_1 \rangle & \langle \mathbf{u}_4, \mathbf{v}_2 \rangle & \langle \mathbf{u}_4, \mathbf{v}_3 \rangle & \langle \mathbf{u}_4, \mathbf{v}_4 \rangle \end{bmatrix}, \tag{1}$$

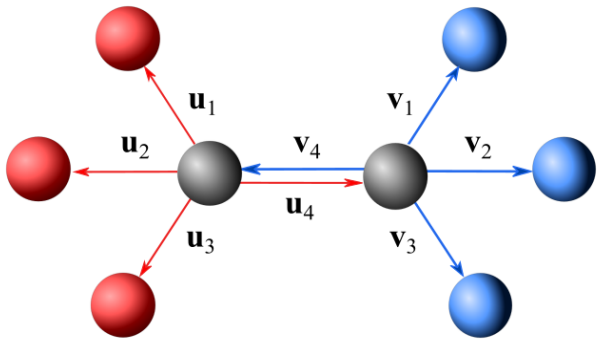

Figure 2: Construction of ice crystal from connection of two tetrahedral structures.

where $\langle \mathbf{u}_i, \mathbf{v}_j \rangle$ denotes the inner product between the vectors $\mathbf{u}_i$ and $\mathbf{v}_j$. We refer to the representation matrices $\mathbf{T}$ corresponding to the staggered and eclipsed conformations as templates denoted by $\mathbf{T}_\mathrm{s}$ and $\mathbf{T}_\mathrm{e}$, respectively:

$$\mathbf{T}_\mathrm{s} = \begin{bmatrix} -1 & 0.5 & 0.5 & 0.5 \\ 0.5 & -1 & 0.5 & 0.5 \\ 0.5 & 0.5 & -1 & 0.5 \\ 0.5 & 0.5 & 0.5 & -1 \end{bmatrix}, \quad \text{and} \quad \mathbf{T}_\mathrm{e} = \begin{bmatrix} 0.5 & -0.5 & -0.5 & 0.5 \\ -0.5 & 0.5 & -0.5 & 0.5 \\ -0.5 & -0.5 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0.5 & -1 \end{bmatrix}. \tag{2}$$

We aim to identify the staggered and eclipsed conformations in the ice crystals by searching for the structures that match the templates in eq. 2. Since the bond direction vectors are ordered arbitrarily, *i.e.*, the numbers $1, \ldots, 4$ are assigned randomly, the templates $\mathbf{T}_\mathrm{s}$ and $\mathbf{T}_\mathrm{e}$ are not unique and any row- or column-wise permutations of these templates represent the same conformations. However, considering the symmetric structure of the templates, it is sufficient to search only among the column or only the row permutations of the templates. This is because for a symmetric matrix, the set of all column-wise permutations and the set of all row-wise permutations are equal. The common rows and columns in both templates are highlighted in gray color.

3 Algorithm

3.1 Structure representation matrix

In order to form the structure representation matrix $\mathbf{T}$ for a pair of neighboring oxygen atoms with an unknown conformation, it is enough to find their corresponding bond directions. The bond directions are defined between an oxygen atom and a maximum of four nearest neighbouring oxygen atoms within a cutoff distance of 0.35 nm, corresponding to the minimum between the nearest and second-nearest neighbour peaks in the radial distribution function of ice, and including the first solvation shell in liquid water. The matrix $\mathbf{T}$ can be efficiently calculated as follows:

$$\mathbf{T} = \mathbf{U}^T\mathbf{V} \tag{3}$$

where $(\cdot)^T$ denotes the matrix transpose operation, and the matrices $\mathbf{U}$ and $\mathbf{V}$ are defined as:

$$\mathbf{U} = [\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3\ \mathbf{u}_4] \quad \text{and} \quad \mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ \mathbf{v}_3\ \mathbf{v}_4]. \tag{4}$$

Here, note that $\mathbf{u}_{1-4}$ and $\mathbf{v}_{1-4}$ are column vectors and have unit norms. When the number of neighbouring atoms is less than four, the bond directions corresponding to the missing neighbours are considered as zero vectors and stored in the last column(s). The number of nonzero columns in $\mathbf{U}$ and $\mathbf{V}$, *i.e.*, the number of neighbours for the corresponding oxygen atoms, are denoted by n_{U} and n_{V}, respectively.

3.2 Template matching

In order to make the templates more distinct, we first remove their common components, that is the set of entries in $\mathbf{T}_{\mathrm{s}}$ and $\mathbf{T}_{\mathrm{e}}$ associated with the bond directions between the two central atoms. In Fig. 1, notice that the angles between the central O-O bond direction originating at a first atom ($\mathbf{u}_4$ in Fig. 2, for example) and the bond directions corresponding to the second atom ($\mathbf{v}_{1-4}$ in Fig. 2) are identical in the staggered and eclipsed conformations.

The common component of $\mathbf{T}_\mathrm{s}$ and $\mathbf{T}_\mathrm{e}$, which is highlighted with grey color in eq. 2, consists of a row and a column intersecting at an entry equal -1 which indicates $\mathbf{u}_4$ and $\mathbf{v}_4$ are the antiparallel vectors associated with the same bond (see Fig. 2). The modified templates are then written as

$$\mathbf{T}'_\mathrm{s} = \begin{bmatrix} -1 & 0.5 & 0.5 \\ 0.5 & -1 & 0.5 \\ 0.5 & 0.5 & -1 \end{bmatrix}, \quad \text{and} \quad \mathbf{T}'_\mathrm{e} = \begin{bmatrix} 0.5 & -0.5 & -0.5 \\ -0.5 & 0.5 & -0.5 \\ -0.5 & -0.5 & 0.5 \end{bmatrix}. \tag{5}$$

The row and column which intersect at the entry equal to -1 should be also eliminated from the measured representation matrix $\mathbf{T}$, to form the reduced version $\mathbf{T}'$. If there is no element equal to -1 in $\mathbf{T}$, it means one of the two central oxygen atoms does not consider the other as one of its four nearest neighbours. It is obvious that such a structure cannot be staggered or eclipsed. When the matrix $\mathbf{T}'$ is formed, the squares of its Euclidean distances from $\mathbf{T}'_\mathrm{s}$ and $\mathbf{T}'_\mathrm{e}$, denoted by d^2_s and d^2_e, respectively, can be calculated by solving

$$\begin{aligned} d^2_\mathrm{s} &= \min_t \|\mathbf{T}'(1:n'_\mathrm{U}, 1:n'_\mathrm{V}) - \mathbf{T}'_\mathrm{s}(1:n'_\mathrm{U}, \cdot)\mathbf{P}_t(\cdot, 1:n'_\mathrm{V})\|^2_F, \quad t = 1, \ldots, 6 \\ \text{and} \quad d^2_\mathrm{e} &= \min_t \|\mathbf{T}'(1:n'_\mathrm{U}, 1:n'_\mathrm{V}) - \mathbf{T}'_\mathrm{e}(1:n'_\mathrm{U}, \cdot)\mathbf{P}_t(\cdot, 1:n'_\mathrm{V})\|^2_F, \quad t = 1, \ldots, 6 \end{aligned} \tag{6}$$

where $\mathbf{P}_t$ are the six possible 3×3 permutation matrices, and $n'_\mathrm{U} = n_\mathrm{U} - 1$ and $n'_\mathrm{V} = n_\mathrm{V} - 1$ are the number of nonempty rows and the number of nonempty columns in $\mathbf{T}'$, respectively. The Frobenius norm of a matrix $\mathbf{A}$ is calculated as $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} \mathbf{A}(i,j)^2}$. Note that the distances can be measured using eq. 6 only if $n'_\mathrm{U} > 0$ and $n'_\mathrm{V} > 0$ (this is also essential for considering any structure). Equation 6 can be solved by searching for a permutation matrix that minimizes the distance. Such a search can be efficiently done by forming the squared distance matrices $\mathbf{D}^2_\mathrm{s}$ and $\mathbf{D}^2_\mathrm{e}$ so that

$$\begin{aligned} \mathbf{D}^2_\mathrm{s}(i,j) &= \|\mathbf{T}'(1:n'_\mathrm{U}, i) - \mathbf{T}'_\mathrm{s}(1:n'_\mathrm{U}, j)\|^2_2, \quad i = 1, \ldots, n'_\mathrm{V} \text{ and } j = 1, 2, 3 \\ \text{and} \quad \mathbf{D}^2_\mathrm{e}(i,j) &= \|\mathbf{T}'(1:n'_\mathrm{U}, i) - \mathbf{T}'_\mathrm{e}(1:n'_\mathrm{U}, j)\|^2_2, \quad i = 1, \ldots, n'_\mathrm{V} \text{ and } j = 1, 2, 3. \end{aligned} \tag{7}$$

Then d_{s}^2 and d_{e}^2 can be calculated as

$$\begin{aligned} d_{\mathrm{s}}^2 &= \min_t \sum_{\substack{i=1,\ldots,n'_{\mathrm{V}} \\ j=1,2,3}} \mathbf{D}_{\mathrm{s}}^2(i,j)\mathbf{P}_t(i,j), \quad t=1,\ldots,6 \\ \text{and} \quad d_{\mathrm{e}}^2 &= \min_t \sum_{\substack{i=1,\ldots,n'_{\mathrm{V}} \\ j=1,2,3}} \mathbf{D}_{\mathrm{e}}^2(i,j)\mathbf{P}_t(i,j), \quad t=1,\ldots,6. \end{aligned} \tag{8}$$

Since each of the permutation matrices has only 3 nonzero elements, computation of the squared distances using eq. 8 is computationally efficient. When d_{s}^2 and d_{e}^2 are calculated, the similarity scores corresponding to the staggered and eclipsed conformations, denoted by S_{s} and S_{e}, respectively, are assigned as follows:

$$S_{\mathrm{s}} = \exp\left(-\frac{d_{\mathrm{s}}^2}{\lambda n'_{\mathrm{U}} n'_{\mathrm{V}}}\right) \quad \text{and} \quad S_{\mathrm{e}} = \exp\left(-\frac{d_{\mathrm{e}}^2}{\lambda n'_{\mathrm{U}} n'_{\mathrm{V}}}\right). \tag{9}$$

The similarity scores in eq. 9 are close to 1 if the distance is small and become closer to 0 as distance increases. The parameter λ is employed to tune the distribution of the scores between 0 and 1. In addition, the squared distances are averaged over the number of nonzero elements in $\mathbf{T}'$ which is equal to $n'_{\mathrm{U}} n'_{\mathrm{V}}$, and thus possible vacancies in the ice structure do not have any effect on the calculated similarity scores. Based on the measured scores S_{s} and S_{e}, the conformation represented by $\mathbf{T}$ is labeled as follows:

$$\begin{cases} \text{Staggered}, & \text{if } S_{\mathrm{s}} > \max(S_{\mathrm{e}}, S_{\min}) \\ \text{Eclipsed}, & \text{if } S_{\mathrm{e}} > \max(S_{\mathrm{s}}, S_{\min}) \\ \text{Eclipsed/Staggered}, & \text{if } S_{\mathrm{s}} = S_{\mathrm{e}} > S_{\min}. \end{cases} \tag{10}$$

As can be seen in eq. 10, only scores larger than the predefined minimum value $S_{\min}$ are taken into account. That means in order to assign any conformation label to a structure represented by $\mathbf{T}$, one of the measured squared distances, d_{s}^2 and d_{e}^2, should be smaller than $-\lambda n'_{\mathrm{U}} n'_{\mathrm{V}} \ln(S_{\min})$. Parameter $S_{\min}$ determines the minimum acceptable similarity between

the given structure and the templates, thus it can be used as a means for tuning the selectivity of the algorithm. In addition, with a sufficiently large choice of $S_{\min}$, the third possibility in eq. 10 can only occur when $\mathbf{T}'$ has a single nonzero element (*i.e.*, $n'_{\mathrm{V}} = n'_{\mathrm{U}} = 1$) and the value of that element is close to 0.5 (the common element in $\mathbf{T}'_{\mathrm{s}}$ and $\mathbf{T}'_{\mathrm{e}}$).

3.3 Parameter selection

Using the distance measure introduced in eq. 6, the squared Euclidean distance between $\mathbf{T}'_{\mathrm{s}}$ and $\mathbf{T}'_{\mathrm{e}}$ is 3.75. In order to have more distinct distributions for the similarity scores, we set the parameter $\lambda = 0.15$, which gives $S_{\mathrm{s}} = S_{\mathrm{e}} = 0.5$ when $d_{\mathrm{s}}^2 = d_{\mathrm{e}}^2 = 3.75/2$, *i.e.*, when $\mathbf{T}'$ represents a structure exactly between $\mathbf{T}'_{\mathrm{s}}$ and $\mathbf{T}'_{\mathrm{e}}$. This choice of parameter guarantees that S_{s} and S_{e} cannot be larger than 0.5 simultaneously.

Figure 3 shows the probability density functions (PDF) and the cumulative distribution functions (CDF) of the similarity scores S_{s} and S_{e} for hexagonal and cubic ice, clathrate hydrates and liquid water systems. Cubic and hexagonal ice simulations were carried out at temperature $T = 243$ K, containing over 3500 water molecules. The liquid water system was equilibrated at $T = 298$ K and contained around 5000 water molecules. Simulation of the cubic sI clathrate hydrate system was performed at $T = 263$ K and contained around 1000 water molecules. All simulations were carried out at constant volume and the TIP4P/Ice model[30] for water was used.

We observed that in the cubic ice system all the measured S_{s} values are significantly larger than S_{e} values, which agrees with the fact that in cubic ice all the conformations are staggered. It can be also seen that in the hexagonal ice system, 75 % of the S_{s} values and 25 % of the S_{e} values have considerably large values. Moreover, the CDF curve of clathrate hydrates shows that all the S_{e} values are larger than 0.6 while the S_{s} values are always smaller than 0.2. These observations are also consistent with the assumptions on the hexagonal ice and clathrate hydrates structures, explained earlier in section 2. The considerable differences between the peaks in PDF curves of S_{s} and S_{e} for different ice structures, also small similarity

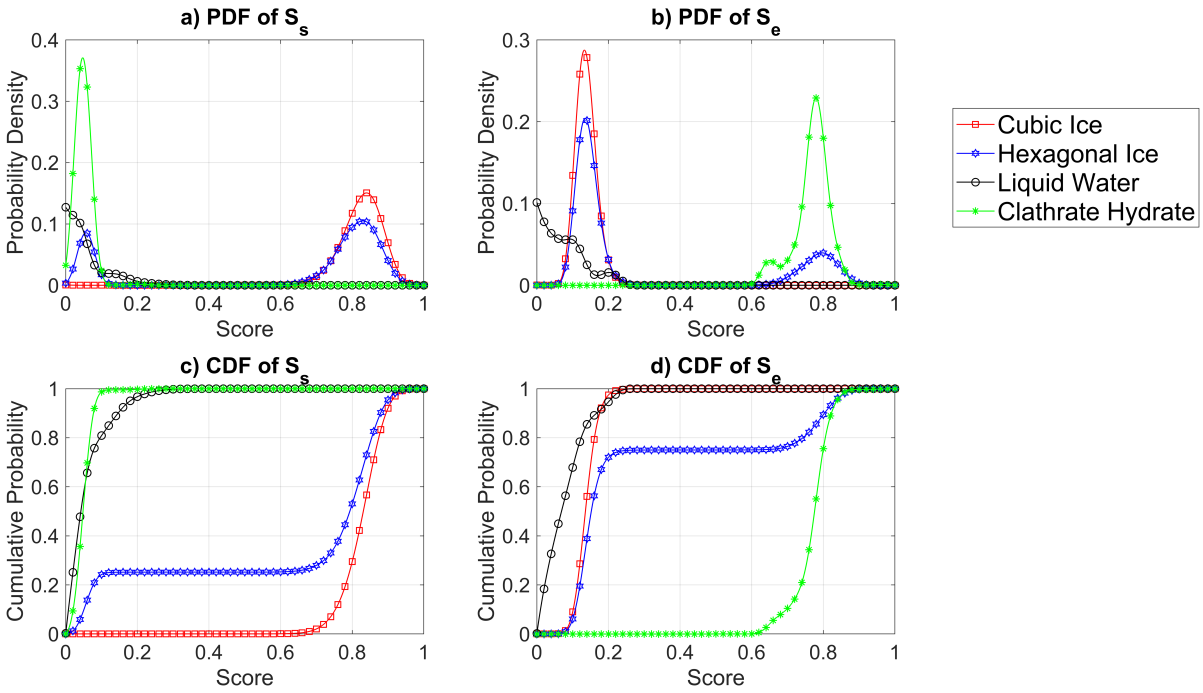

Figure 3: (a, b) Probability density function (PDF) and (c, d) cumulative distribution function (CDF) of S_s and S_e values in cubic ice, hexagonal ice, clathrate hydrates and liquid water structures.

scores measured for liquid water guarantee a very low probability of mislabeling or false detection of ice by the proposed method.

3.4 Structure classification rules

The template-matching algorithm presented here allows the classification of liquid water, cubic or hexagonal ice, all possible interfacial structures, *i.e.*, cubic-interfacial, hexagonal-interfacial, mixed-interfacial, and other interfacial structures, as well as clathrate hydrates, using the following rules for water oxygen atoms:

- Atoms with four staggered conformations are cubic ice,
- Atoms with three staggered conformations and one eclipsed conformation are hexagonal ice,

- Atoms with four eclipsed conformations are clathrate hydrates.

By construction, water molecules at interfaces are not classified as "bulk" ice, even if they exhibit ice like structure. Using the above classifications, different interfacial ice types can now be distinguished based on the identity of their neighbours:

- Atoms which are not cubic ice, hexagonal ice, or clathrate hydrates, but have at least two neighbours with ice structures of different types (cubic, hexagonal and clathrate hydrate) are considered mixed-interfacial ice,

- Atoms which are not cubic ice, hexagonal ice, or clathrate hydrates, but have at least one cubic ice neighbour and no hexagonal ice or clathrate hydrate neighbour are considered cubic-interfacial ice,

- Atoms which are not cubic ice, hexagonal ice, or clathrate hydrates, but have at least one hexagonal ice neighbour and no cubic ice or clathrate hydrate neighbour are considered hexagonal-interfacial ice,

- Atoms which are not cubic ice, hexagonal ice, or clathrate hydrates, but have at least one clathrate hydrate neighbour and no cubic or hexagonal ice neighbour are considered clathrate-interfacial ice,

- Atoms which are not from any of the aforementioned groups of ice, but have at least one staggered or eclipsed conformation with a mixed-interfacial, cubic-interfacial, hexagonal-interfacial, or clathrate-interfacial, are considered other interfacial ice.

Finally, atoms that are not part of any of the above-mentioned categories are considered liquid. The classification rules are illustrated in Fig. 4.

3.5 Code availability

Implementations of the LICH-TEST algorithm in MATLAB and Python are openly available in a GitHub repository.[31]

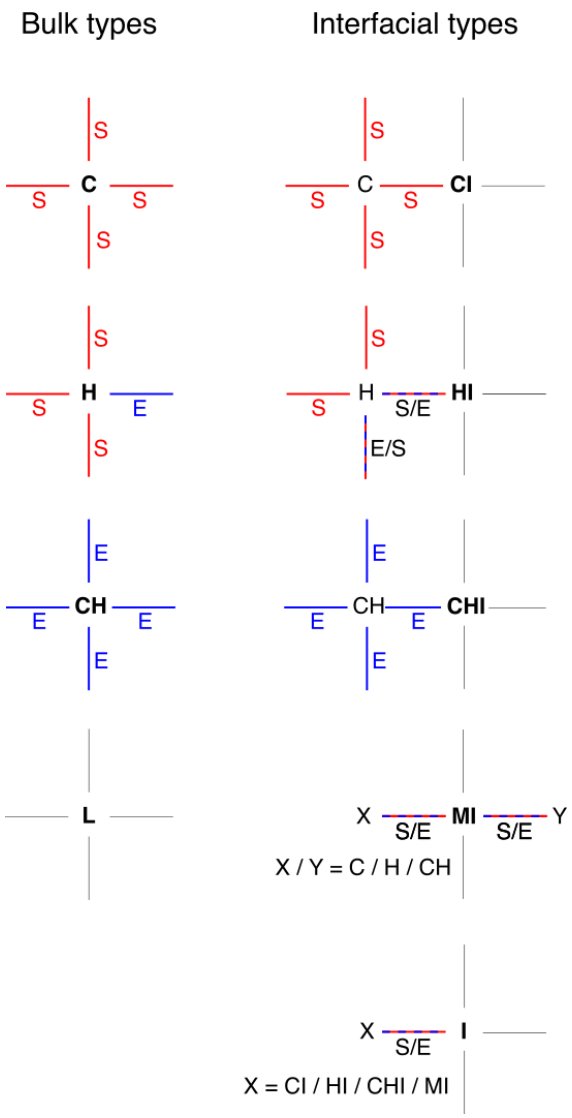

Figure 4: LICH-TEST classification rules for bulk and interfacial types. The classification rule applies for the highlighted water oxygen atom (C: cubic ice, H: hexagonal ice, CH: clathrate hydrate, L: liquid, CI: cubic-interfacial, HI: hexagonal-interfacial, CHI: clathrate-interfacial, MI: mixed-interfacial, I: other interfacial ice). S and E denote staggered and eclipsed conformations, respectively. In interfacial ice structure (I), the S/E bonds include staggered/eclipsed conformation defined in Eq. 10.

4 Results and discussion

4.1 Case study 1: heterogeneous ice nucleation and growth on a flat mineral surface

We applied the LICH-TEST algorithm to identify ice structures in an atomistic molecular dynamics simulation of nucleation and growth of ice on the Ag-terminated AgI (0001) surface.

Specifically, the system contained 20438 TIP4P/Ice water molecules at $T = 263$ K on an AgI slab measuring 10.076×10.313 nm^2. Further simulation details can be found in Ref.[9] We investigate the performance of the proposed method using different values of S_{min}. In addition, we compare our method to a commonly used ice structure recognition algorithm based on spherical harmonics analysis, the CHILL+ algorithm,[15] in terms of ice structure recognition accuracy and computational efficiency. Both algorithms were implemented in MATLAB 2019b on a PC equipped with Intel(R) Core(TM) i5-8265U CPU (1.6 GHz) and common elements such as file I/O or neighbour lists were identical.

The labeling approach for different ice structures in the CHILL+ and LICH-TEST algorithms are not the same. In CHILL+, there is only one type of interfacial ice, whereas in LICH-TEST, such molecules are further sub-categorised as cubic-interfacial (CI), hexagonal-interfacial (HI), clathrate-interfacial (CHI), mixed-interfacial (MI), or other interfacial ices (I).

4.1.1 Selectivity tuning

In Fig. 5, the ice structure recognition results obtained using the LICH-TEST algorithm with five different similarity cut-offs ($S_{\min} = 0.50$, 0.55, 0.60, 0.65, and 0.70) are compared to those obtained using the CHILL+ algorithm. In addition, the numbers of different ice structures obtained in the simulations are summarized in Table 1. Using larger values for $S_{\min}$ imposes harder requirements for the recognition of staggered or eclipsed structures and thus, as can be seen in Table 1, decreases the total number of detected ice structures. Conversely, smaller $S_{\min}$ values can be chosen on purpose to help identify molecules in the liquid, which exhibit ice-like structure. For example, in the context of ice nucleation studies, simulations are usually hindered by the long time scale, which requires the use of enhanced sampling techniques.[8,16,32] Here, smaller $S_{\min}$ values could be used to detect pre-critical fluctuations in the undercooled liquid, and to define a collective variable along which the system could be biased in order to initiate nucleation events. However, based on the simulation results,

including the hexagonal and cubic lattice structures in Sec. 3.3, we observed that a good choice for the minimum considerable similarity score can be $0.5 \leq S_{\text{min}} \leq 0.6$.

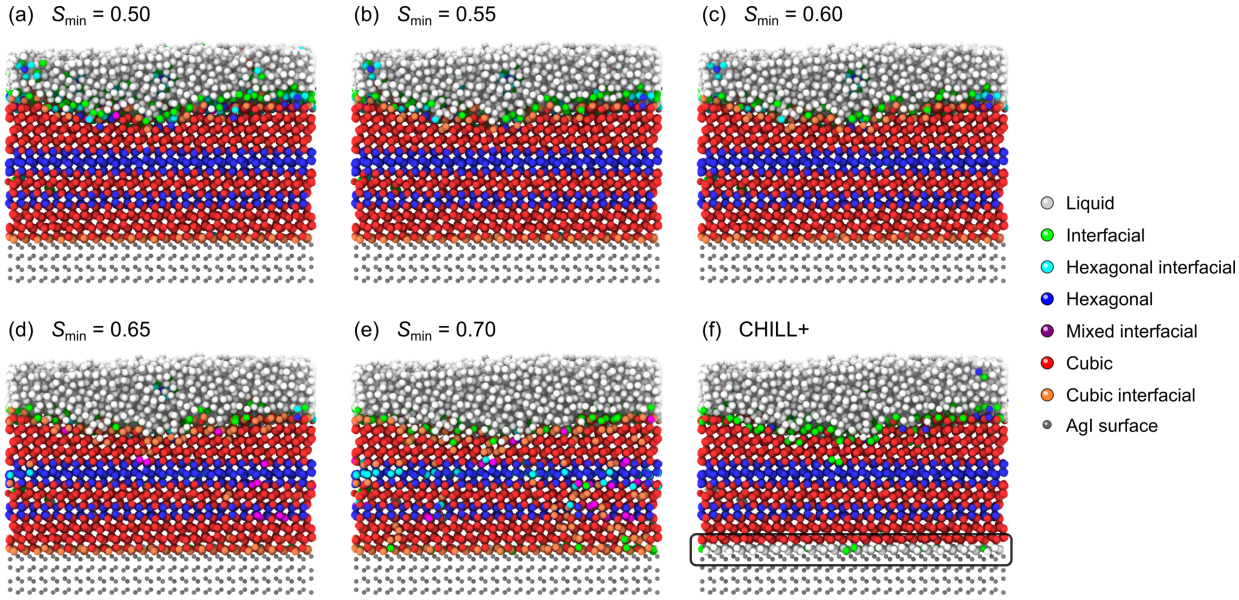

Figure 5: Ice structure recognition at the AgI (0001) – H_2O interface using LICH-TEST with similarity score cut-offs of (a) $S_{\text{min}} = 0.50$, (b) $S_{\text{min}} = 0.55$, (c) $S_{\text{min}} = 0.60$, (d) $S_{\text{min}} = 0.65$, (e) $S_{\text{min}} = 0.70$, and (f) using the CHILL+ algorithm. Water molecules are color-coded according to their structure and surface atoms are shown as small grey spheres.

4.1.2 Performance and sensitivity

A significant advantage of LICH-TEST over the CHILL+ algorithm is the detection of defected ice structures. In Fig. 5f, CHILL+ classifies the majority of atoms in the hydration layer on top of the AgI (0001) surface as liquid water and a small number as interfacial ice, as highlighted by the black frame in Fig. 5f. Since the CHILL+ method requires the existence of four neighbors inside the first coordination shell for calculation of local bond order parameters, it cannot identify ice structures with broken symmetry, for example at interfaces.[6] The proposed method measures the similarities with respect to the existing bonds, thus it successfully identifies the defected ice structures. For values of $S_{\text{min}} < 0.70$, LICH-TEST correctly labels the interfacial ice at the mineral surface, as well as the cubic and hexagonal ice molecules in stacking disordered ice I, which is the preferred structure of

Table 1: Numbers of cubic (C), cubic-interfacial (CI), hexagonal (H), hexagonal-interfacial (HI), mixed interfacial (MI), or interfacial (I) molecules in the systems shown in Fig. 5 using the LICH-TEST algorithm with different S_{min} values and the CHILL+ algorithm. Note the different definitions of 'interfacial' molecules in the two algorithms.

	Cubic		**Hexagonal**		**Interfacial**		**Total**
	C	CI	H	HI	MI	I	
$S_{\mathrm{min}} = 0.50$	9570	1092	3766	535	128	757	15848
$S_{\mathrm{min}} = 0.55$	9432	1209	3622	309	97	209	15178
$S_{\mathrm{min}} = 0.60$	9230	1372	3507	208	101	373	14791
$S_{\mathrm{min}} = 0.65$	8732	1736	3273	161	201	318	14421
$S_{\mathrm{min}} = 0.70$	7456	2649	2638	436	436	460	14075
CHILL+	8804		3575		965		13344

ice freshly grown from supercooled water.[11,33,34] Correct identification of ice-like hydration layer structures can be important in detecting the formation of a critical ice nucleus in heterogeneous nucleation. We note that the additional details revealed by LICH-TEST do not come at the price of higher computational cost; in fact, in our implementation of both codes, LICH-TEST required only ~ 75 % of the wall time required by CHILL+.

It is important to point out that the similarity score in LICH-TEST, calculated from the squared Euclidean distance between a given structure and the perfect template structures, has a different sensitivity to deviations from the reference structure than the order parameter calculated from a projection onto spherical harmonics, as used in *e.g.* the CHILL+ algorithm. Moreover, since the calculation of the similarity scores includes normalization of the measured squared distances with respect to the number of atoms in the structure, the defects in the ice structure, in terms of missing atoms, do not affect the performance of the LICH-TEST.

4.2 Case study 2: ice recognition by an anti-freeze protein

In section 4.1 we have benchmarked LICH-TEST against the CHILL+ algorithm and shown how the sensitivity of our algorithm can be tuned by varying S_{min}. The planar AgI-water

interface considered there is a relevant geometry for heterogenous ice nucleation and growth on solid surfaces. However, single biomolecules or bacteria can also enhance or inhibit the nucleation or growth of ice, even though the interface is "soft" and finite on the nanoscale. As an example of such a system, we consider a spruce budworm anti-freeze protein (sbwAFP) in supercooled water, halting an advancing front of hexagonal ice, as illustrated in Fig. 6. Based on the work of Kuiper et al.,[35] the system consisted of one sbwAFP, two Cl^- ions, and 17393 water molecules, described by the CHARMM[36] and TIP4P[37] forcefields, in a simulation box measuring $5.2042 \times 12.6656 \times 16.0$ nm^3. The timestep was 1 fs, and stochastic velocity rescaling was used to keep a temperature of 225 K, 5 K below the melting point of TIP4P water. Water oxygen atoms at the bottom of cell were harmonically restrained to the ice I_h lattice positions. The a-axis of the ice seed was tilted by $\sim 4°$ with respect to the xy plane, allowing for continuous step growth across the periodic boundary conditions in x and y directions. A vacuum gap of ~ 8 nm below the ice seed ensured no interactions with the initially liquid water through periodic boundaries along z.

The analysis of the water and ice structure around the protein using LICH-TEST is illustrated in Fig. 6c and d. Once the advancing ice front has reached the protein, the algorithm directly detects the line of interfacial hexagonal molecules hydrogen bonded to the protein's threonine residues, highlighted in the inset of Fig. 6b and d, which Kuiper et al.[35] identified as the molecular ice recognition mechanism. For comparison, the CHILL+ algorithm classifies all these water molecules as liquid-like, as shown in Fig. 6e, conveying the wrong impression of a disordered interface. This again results from the fact that CHILL+ only classifies oxygen atoms with four neighbors in the first coordination shell as ice, missing the ice structures at the interface.

4.3 Case study 3: clathrate hydrate interface

Clathrate hydrates are cage-like crystalline structures that can stably enclose small molecules from liquid or gas, such as CO_2 or CH_4, by minimising the number of broken hydrogen bonds

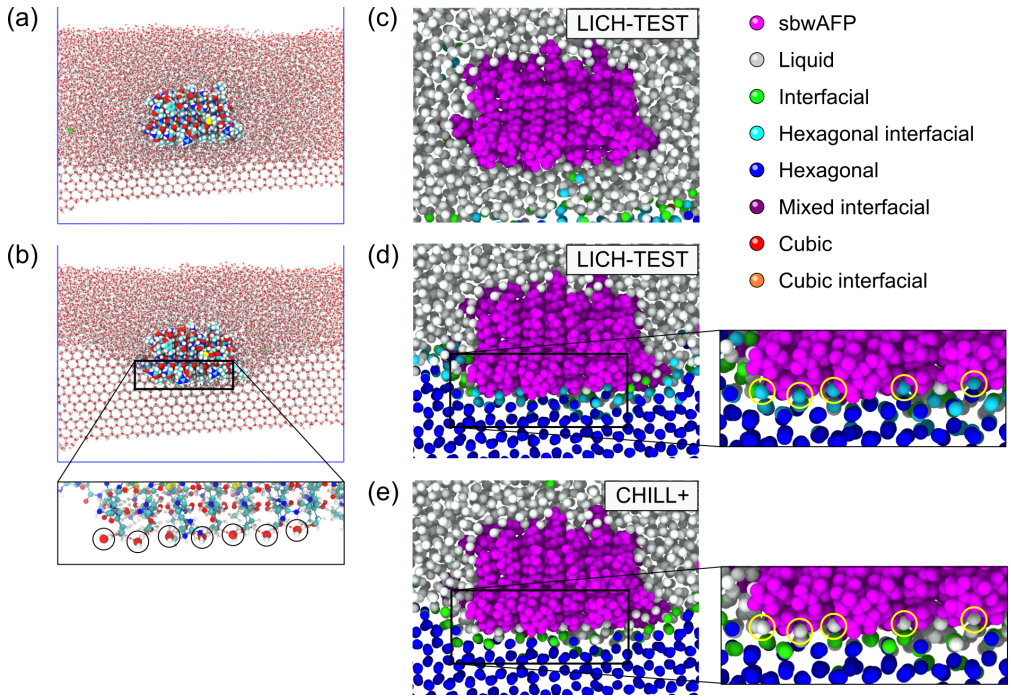

Figure 6: Ice structure recognition around a spruce budworm anti-freeze protein (sbwAFP). Simulation snapshot before (a) and after (b) the sbwAFP has bound to an advancing ice front (protein atoms are shown as large spheres, water molecules as small red and white sticks and balls). The row of water molecules hydrogen bonded to OH groups on the protein's threonine residues, which form the interface with the ice crystal, are highlighted by black circles in the inset of panel b, and hydrogen bonds are indicated by dashed red lines. Cross-section of the structure of water molecules around the protein (magenta) before (c) and after contact as determined by the LICH-TEST (d), and CHILL+ algorithms (e). Water oxygen atoms are shown as spheres color-coded according to their structure. The water molecules at the protein-ice interface, shown in panel b, are highlighted by yellow circles in the insets of panels d and e.

around the guest molecule.[38–40] The three most prevalent clathrate hydrate structures are cubic sI, cubic sII, and hexagonal sH, with space groups Pm3n, Fd3m, and P6/mmm, respectively. LICH-TEST recognises a molecule with 4 eclipsed bonds as a clathrate hydrate, and also identifies clathrate interfacial structures. We performed molecular dynamics simulations of a 3D periodic system containing a double interface of cubic sI clathrate hydrate, with argon and krypton atoms in the smaller and larger cages respectively, and an initially liquid layer of the same number of water, argon and krypton molecules at the same density. The structure of the clathrate hydrate sI structure was taken from Takeuchi *et al.*.[41] The

simulation box measured $2.4 \times 2.4 \times 4.8$ nm^3, containing 736 water molecules, 32 Ar, and 96 Kr atoms in total. A snapshot of the system is shown in Fig. 7a. Atomistic interactions were described by the TIP4P/Ice model for water[30] and Lennard-Jones potentials for argon and krypton, with Lorentz-Berthelot mixing rules ($\varepsilon_{\mathrm{Ar}}/k_{\mathrm{B}} = 116.81$, $\sigma_{\mathrm{Ar}} = 0.3401$ nm, $\varepsilon_{\mathrm{Kr}}/k_{\mathrm{B}} = 164.56$, $\sigma_{\mathrm{Kr}} = 0.3601$ nm). An NVT ensemble with a temperature of 263 K was obtained by stochastic velocity rescaling.

At this temperature, we could observe slow growth of the crystal structure, including a stacking fault. Analyses of the structure ~ 1 μs after the start of the simulation using LICH-TEST, with a minimum similarity score $S_{\mathrm{min}} = 0.5$, and CHILL+ are shown in Fig. 7b and c. The total number of clathrate and clathrate interfacial water molecules recognised by the two algorithms is shown in Tab. 2. While both algorithms identify bulk periodic clathrate crystals perfectly (not shown), there are some small differences in recognition when it comes to clathrate interfacial structure. Overall, LICH-TEST recognizes more molecules as belonging to clathrate hydrates than CHILL+, however some molecules identified as clathrates by CHILL+ are considered clathrate-interfacials by LICH-TEST. Similar small discrepancies are observed for liquid and clathrate-interfacial molecules. These small differences are in part due to the slightly different definitions of clathrate-interfacial in LICH-TEST and CHILL+, respectively, and to the tuneable sensitivity of LICH-TEST via S_{min}.

Table 2: Numbers of clathrate hydrates (CH), clathrate-interfacial (CHI) and liquid-like (L) molecules in the systems shown in Fig. 7 using the LICH-TEST algorithm with $S_{\mathrm{min}} = 0.5$ and the CHILL+ algorithm. No other types were detected by either algorithm.

	CH	**CHI**	**L**
LICH-TEST	594	101	41
CHILL+	569	103	64

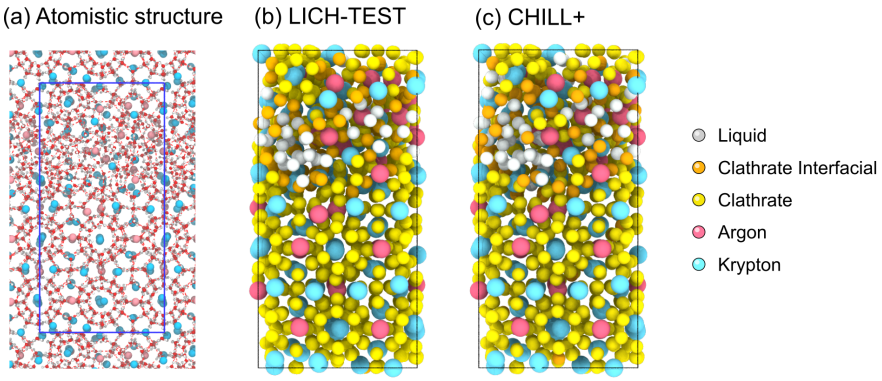

Figure 7: Interface of cubic sI clathrate hydrate. (a) Snapshot of the atomistic simulation, water molecules are shown as red and white sticks and balls, argon and krypton atoms are shown as pink and cyan spheres, respectively, and the simulation box boundaries are marked by blue lines. Structural analysis of water molecules using LICH-TEST (b) or the CHILL+ algorithm (c).

5 Summary and Conclusions

We have developed a novel algorithm, LICH-TEST, for the identification of liquid water, cubic and hexagonal ice, clathrate hydrate, as well as different interfacial ice and water structures. The method is applicable to atomistic and coarse-grained water models. The algorithm is based on template matching to staggered and eclipsed conformations, by analysing the four nearest oxygen neighbours around two neighboring oxygen atoms. The structure is classified using a similarity score based on the squared Euclidian distance between the structure and the templates. The similarity score cut-off S_{min} can be used as a parameter to tune the sensitivity of the algorithm.

We have applied LICH-TEST algorithm in three case studies: heterogeneous ice nucleation at a silver iodide surface, ice recognition by an antifreeze protein, and analysis of a clathrate hydrate interface. We have benchmarked the algorithm against the widely used CHILL+ algorithm and find overall good agreement between the two methods for $0.5 < S_{\mathrm{min}} < 0.70$. However, LICH-TEST offers more capability for the classification of interfacial ice and water molecules, at a slightly lower computational cost.

Correct identification of interfacial structures is of particular importance when studying

heterogenous interfaces of ice, for instance in ice nucleation at solid surfaces, or in the recognition of ice by biomolecules, such as anti-freeze proteins. Existing methods are often not well suited to this task.[6] LICH-TEST does not require the presence of four nearest neighbors and is robust in classifying non-ideal or defected structures. The ability to discriminate between cubic-, hexagonal-, clathrate-, mixed-, and other interfacial types can be advantageous in the analysis of complex ice structures, which would otherwise require tedious visual inspection of 3D structures.

The code has been made openly available in a GitHub repository and will be updated by the developers. We hope it will be useful to different scientific communities interested in molecular simulations involving water and ice.

Acknowledgement

This work was supported by the ERC Grant 692891-DAMOCLES, the Academy of Finland Flagship funding (grant no. 337549), the University of Helsinki, Faculty of Science ATMATH project, and the National Center of Meteorology (NCM), Abu Dhabi, UAE, under the UAE Research Program for Rain Enhancement Science. Supercomputing resources were provided by CSC–IT Center for Science, Ltd., Finland. BR is grateful to Dr. Stephen Ingram for his advice on setting up the antifreeze protein simulations. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Center of Meteorology, Abu Dhabi, UAE, funder of the research.

References

(1) Murray, B. J.; O'Sullivan, D.; Atkinson, J. D.; Webb, M. E. Ice nucleation by particles immersed in supercooled cloud droplets. *Chem. Soc. Rev.* **2012**, *41*, 65196554.

(2) Koop, T.; Luo, B.; Tsias, A.; Peter, T. Water activity as the determinant for homogeneous ice nucleation in aqueous solutions. *Nature* **2000**, *406*, 611–614.

(3) Coluzza, I.; Creamean, J.; Rossi, M.; Wex, H.; Alpert, P.; Bianco, V.; Boose, Y.; Dellago, C.; Felgitsch, L.; Fröhlich-Nowoisky, J.; et al., Perspectives on the Future of Ice Nucleation Research: Research Needs and Unanswered Questions Identified from Two International Workshops. *Atmosphere* **2017**, *8*, 1385790.

(4) Sosso, C. G.; Chen, J.; Cox, J. C.; Fitzner, M.; Pedevilla, P.; Zen, A.; Michaelides, A. Crystal Nucleation in Liquids: Open Questions and Future Challenges in Molecular Dynamics Simulations. *Chem. Rev.* **2016**, *116*, 7078–7116.

(5) Zielke, S. A.; Bertram, A. K.; Patey, G. N. A Molecular Mechanism of Ice Nucleation on Model AgI Surfaces. *J. Phys. Chem. B* **2014**, *119*, 9049–9055.

(6) Cox, S. J.; Kathmann, S. M.; Slater, B.; Michaelides, A. Molecular Simulations of Heterogeneous Ice Nucleation. I. Controlling Ice Nucleation Through Surface Hydrophilicity. *J. Chem. Phys.* **2015**, *142*, 184704.

(7) Zielke, S. A.; Bertram, A. K.; Patey, G. N. Simulations of Ice Nucleation by Kaolinite (001) with Rigid and Flexible Surfaces. *J. Phys. Chem. B* **2016**, *120*, 1726–1734.

(8) Sosso, G. C.; Li, T.; Donadio, D.; Tribello, G. A.; Michaelides, A. Microscopic Mechanism and Kinetics of Ice Formation at Complex Interfaces: Zooming in on Kaolinite. *J. Phys. Chem. Lett.* **2016**, *7*, 2350–2355.

(9) Roudsari, G.; Reischl, B.; Pakarinen, O. H.; Vehkamäki, H. Atomistic Simulation of Ice Nucleation on Silver Iodide (0001) Surfaces with Defects. *J. Phys. Chem. C* **2020**, *124*, 436–445.

(10) Tanaka, H.; Yagasaki, T.; Matsumoto, M. On the role of intermolecular vibrational mo-

tions for ice polymorphs.II. Atomic vibrational amplitudes and localization of phonons in ordered and disordered ices. *J. Chem. Phys.* **2020**, *152*, 074501.

(11) Malkin, T. L.; Murray, B. J.; Brukhno, A. V.; Anwar, J.; Salzmann, C. G. Structure of ice crystallized from supercooled water. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 1041–1045.

(12) Carignano, M. A. Formation of Stacking Faults during Ice Growth on Hexagonal and Cubic Substrates. *J. Phys. Chem. C* **2007**, *111*, 501–504.

(13) Murray, B. J.; Bertram, A. K. Formation and stability of cubic ice in water droplets. *Phys. Chem. Chem. Phys.* **2006**, *8*, 186–192.

(14) Moore, E. B.; de la Llave, E.; Welke, K.; Scherlis, D. A.; Molinero, V. Freezing, melting and structure of ice in a hydrophilic nanopore. *Phys. Chem. Chem. Phys.* **2010**, *12*, 4124–4134.

(15) Nguyen, A. H.; Molinero, V. Identification of Clathrate Hydrates, Hexagonal Ice, Cubic Ice, and Liquid Water in Simulations: the CHILL+ Algorithm. *J. Phys. Chem. B* **2015**, *119*, 9369–9376.

(16) Haji-Akbari, A.; Debenedetti, P. G. Direct Calculation of Ice Homogeneous Nucleation Rate for a Molecular Model of Water. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 10582–10588.

(17) van Duijneveldt, J. S.; Frenkel, D. Computer simulation study of free energy barriers in crystal nucleation. *J. Chem. Phys.* **1992**, *96*, 4655–4668.

(18) ten Wolde, P.-R.; Ruiz-Montero, M. J.; Frenkel, D. Simulation of homogeneous crystal nucleation close to coexistence. *Faraday Discuss.* **1996**, *104*, 93–110.

(19) Lechner, W.; Dellago, C. Accurate determination of crystal structures based on averaged local bond order parameters. *J. Chem. Phys.* **2008**, *129*, 114707.

(20) Espinosa, J. R.; Vega, C.; Valeriani, C.; Sanz, E. Seeding approach to crystal nucleation. *J. Chem. Phys.* **2016**, *144*, 034501.

(21) Reinhardt, A.; Doye, J. P. K.; Noya, E. G.; Vega, C. Local order parameters for use in driving homogeneous ice nucleation with all-atom models of water. *J. Chem. Phys.* **2012**, *137*, 194504.

(22) Stukowski, A. Structure identification methods for atomistic simulations of crystalline materials. *Model. Simul. Mater. Sci. Eng.* **2012**, *20*, 045021.

(23) Brukhno, A. V.; Anwar, J.; Davidchack, R.; Handel, R. Challenges in molecular simulation of homogeneous ice nucleation. *J. Phys.: Condens. Matter* **2008**, *20*, 494243.

(24) Geiger, P.; Dellago, C. Neural networks for local structure detection in polymorphic systems. *J. Chem. Phys.* **2012**, *139*, 164105.

(25) Fulford, M.; Salvalaglio, M.; Molteni, C. DeepIce: A Deep Neural Network Approach To Identify Ice and Water Molecules. *J. Chem. Inf. Model.* **2019**, *59*, 2141–2149.

(26) Maras, E.; Trushin, O.; Stukowski, A.; Ala-Nissila, T.; Jónsson, H. Global transition path search for dislocation formation in Ge on Si(001). *Comput. Phys. Commun.* **2016**, *205*, 13–21.

(27) Honeycutt, J. D.; Andersen, H. C. Molecular dynamics study of melting and freezing of small Lennard-Jones clusters. *J. Phys. Chem.* **1987**, *91*, 4950–4963.

(28) Faken, D.; Jónsson, H. Systematic analysis of local atomic structure combined with 3D computer graphics. *Comput. Mater. Sci.* **1994**, *2*, 279 – 286.

(29) Larsen, P. M.; Schmidt, S.; Schiøtz, J. Robust structural identification via polyhedral template matching. *Modelling and Simulation in Materials Science and Engineering* **2016**, *24*, 055007.

(30) Abascal, J. L. F.; Sanz, E.; Fernández, R. G.; Vega, C. A Potential Model for the Study of Ices and Amorphous Water: TIP4P/Ice. *J. Chem. Phys.* **2005**, *122*, 234511.

(31) Roudsari, G.; Veshki, F. G.; Reischl, B.; Pakarinen, O. H. LICH-TEST GitHub repository. https://github.com/opakarin/lich-test, Accessed: 2020-12-16.

(32) Lupi, L.; Hudait, A.; Peters, B.; Grünwald, M.; Gotchy Mullen, R.; Nguyen, A. H.; Molinero, V. Title: Role of stacking disorder in ice nucleation. *Nature* **2017**, *551*, 218.

(33) Moore, E. B.; Molinero, V. Is it cubic? Ice crystallization from deeply supercooled water. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20008–20016.

(34) Hudait, A.; Qiu, S.; Lupi, L.; Molinero, V. Free energy contributions and structural characterization of stacking disordered ices. *Phys. Chem. Chem. Phys.* **2016**, *18*, 9544.

(35) Kuiper, M. J.; Morton, C. J.; Abraham, S. E.; Gray-Weale, A. The biological function of an insect antifreeze protein simulated by molecular dynamics. *eLife* **2015**, *4*, e05142.

(36) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(38) Anderson, B. J.; Tester, J. W.; Trout, B. L. Accurate Potentials for Argon-Water and

Methane-Water Interactions via ab Initio Methods and Their Application to Clathrate Hydrates. *J. Phys. Chem. B* **2004**, *108*, 18705–18715.

(39) Jacobson, L. C.; Molinero, V. A Methane-Water Model for Coarse-Grained Simulations of Solutions and Clathrate Hydrates. *J. Phys. Chem. B* **2010**, *114*, 7302–7311.

(40) Nguyen, A. H.; Jacobson, L. C.; Molinero, V. Structure of the Clathrate/Solution Interface and Mechanism of Cross-Nucleation of Clathrate Hydrates. *J. Phys. Chem. C* **2012**, *116*, 19828–19838.

(41) Takeuchi, F.; Hiratsuka, M.; Ohmura, R.; Alavi, S.; Sum, A. K.; Yasuoka, K. Water proton configurations in structures I, II, and H clathrate hydrate unit cells. *J. Chem. Phys.* **2013**, *138*, 124504.

Graphical TOC Entry