This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
Microcanonical molecular simulations of methane hydrate nucleation and growth: evidence that direct nucleation to sI hydrate is among the multiple nucleation pathways

Zhengcai Zhang, Matthew R. Walsh, Guangjun Guo

a Key Laboratory of Earth and Planetary Physics, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, China

b Chevron Energy Technology Company, Flow Assurance, 1200 Smith St, Houston, TX 77002, USA

Abstract

The results of six high-precision constant energy molecular dynamics (MD) simulations initiated from methane-water systems equilibrated at 80 MPa and 250 K indicate that methane hydrates can nucleate via multiple pathways. Five trajectories nucleate to an amorphous solid. One trajectory nucleates to a structure-I hydrate template with long-range order which spans the simulation box across periodic boundaries despite the presence of several defects. While experimental and simulation data of hydrate nucleation with different time- and length-scales suggest there may exist multiple pathways for nucleation, including metastable intermediates and the direct formation of the globally-stable phase, this work provides the most compelling evidence that direct formation to the globally stable crystalline phase is one of the multiple pathways available to hydrate nucleation.
1. Introduction

Gas hydrates are multi-component crystalline compounds made from space-filling cages of water molecules that accommodate small guest molecules such as methane, nitrogen, carbon dioxide, and hydrogen.\(^1\) Gas hydrates have received attention because of their scientific and industrial importance – for example in energy recovery,\(^{1,2}\) climate science,\(^3\) energy storage,\(^4\) and seafloor stability.\(^5,6\) The understanding of gas hydrate formation would be beneficial for safe and efficient energy production, transportation and storage, and a description of the nucleation mechanisms of these multicomponent solids (which can be comprised of compounds of relative immiscibility) is relevant to the areas of water science (including hydrophobicity) and nucleation theory.

Understanding the formation mechanisms of crystals is an active area of research, and has been for over a century. In 1897, Ostwald proposed that via losses in free energy smaller than the overall loss in free energy from the initial to the equilibrium state, one or several metastable intermediates can form before the system reaches its most stable state.\(^7\) While not an inviolable law, this conjecture is supported for example by studies on the crystallization of proteins,\(^8\) colloids,\(^9\) inorganic compounds,\(^10\) and solids of simple molecular models,\(^11\) though studies on the nucleation of magnetite and calcite suggest there are multiple pathways to the crystalline state, that the preferred pathway depends on “many factors”, and that one such pathway is nucleation directly to the globally-stable state.\(^12\)-\(^14\)

Gas hydrate nucleation has been investigated via experiments\(^15\)-\(^20\) and molecular simulations,\(^21\)-\(^37\) and although differences in initial phases present, thermodynamic conditions, guest molecules, length-scales and times-scales probed by these experiments and simulations make direct comparisons difficult, a broad interpretation of this wealth of hydrate formation data is consistent
with previous speculations regarding hydrate nucleation (and the nucleation of other solids) that multiple pathways are available to phase change.\textsuperscript{13,30,38} That is, the initially-observed hydrate phases in these studies include amorphous packings of cages,\textsuperscript{26,27,32} partially-ordered solids containing domains of different crystal structures and long-range order together with amorphous domains,\textsuperscript{24,30,34} mixtures of different crystal structures,\textsuperscript{15-20} and the thermodynamically-preferred crystal phase.\textsuperscript{15,16} Because molecular simulations can follow detailed trajectories of a system’s molecules as phase changes occur, much of the recent work in the area of understanding hydrate nucleation mechanisms has been simulation-based. Some researchers who report on these computational investigations of hydrate nucleation have postulated a mechanism by which hydrates nucleate through an amorphous intermediate which then anneals (or grows) to the crystal phase,\textsuperscript{23,26,27,32,36} though recent computational work has shown that domains of different crystal phases (including the thermodynamically-preferred phase) can form immediately,\textsuperscript{30,34} supporting the argument that hydrates can nucleate via multiple pathways, even at the same thermodynamic conditions.\textsuperscript{30,36,38} However, computational investigations of hydrate nucleation to date have not reported conclusive evidence showing that the crystalline phase of structure-I can form so predominantly as to suppress the amorphous phase and span periodic boundaries. Here we show using multiple microcanonical MD simulations (initiated from the same thermodynamic conditions) that multiple nucleation paths are observed, including the formation of a structure-I template with a long-range order that spans the simulation box across periodic boundaries.

In this study, we performed molecular dynamics simulations in the NVE ensemble to investigate hydrate formation in the absence of a thermostat and thus without the nearly-immediate removal of the exothermic heat of formation upon nucleation and during growth. The
disadvantage of using a thermostat for hydrate nucleation studies is that the artificially fast removal of the heat of formation may skew simulation results towards certain nucleation pathways (we speculate that the artificially-fast removal of the heat of phase change will skew the system towards formation of the amorphous phase). On the contrary, the advantage of using the NVE ensemble is that the system’s temperature will rise in response to the formation exotherm (a phenomenon observed in experiment¹), and thus may allow the system to explore phase space more efficiently without becoming “locked” in local minima during the early stages of formation. It is well known that an obstacle to performing long NVE simulations is the need to keep the system energy constant; because of the imperfect computational precision of MD packages, there is often an evident energy drift during simulations. When long simulations are performed (such as those necessary for spontaneous methane hydrate formation with a clear separation of timescales between induction and formation), this energy drift can preclude a reliable definition of system conditions before phase change, a prerequisite for systematic studies. Liang and Kusalik were able to circumvent this obstacle by using a guest (H₂S) and P/T conditions for which hydrates form immediately, ³⁹ though the simulation of methane hydrate in the microcanonical ensemble had until now been an outstanding challenge. In this work, we adopted several strategies to keep the system energy constant, and performed six independent 1 μs NVE runs (labeled as A-F). Interestingly, one of the nucleation simulations formed methane hydrate with a high degree of structure-I crystallinity.

2. Methods

In this study, the GROMACS software package⁴¹ was used to perform constant-energy simulations. The systems contained 2944 TIP4P/Ice water molecules, ⁴² and 512 united atom
OPLS-UA methane molecules. Lorentz-Berthelot combining rules were used for water-methane cross interactions. The total length for each NVE simulation was 1 μs and the time step was 2 fs. In order to increase the computational precision for the constant energy, several strategies were adopted: 1) GROMACS code was compiled with double-precision; 2) a combination of Particle-Mesh Ewald and a switch function was used for calculating the long-range Columbic interactions with a switching range from 1.2 nm to 1.4 nm; 3) short-range van der Waals interactions were calculated using the shift potential with a switching range from 1.2 nm to 1.3 nm; 4) a cutoff distance for neighbor list was set as 1.65 nm. These strategies lead to a very small drift of no more than 43 kJ/mol in system energy – only about 0.03% of total energy – over every 1 μs run. Certainly, the cost of controlling system energy with high precision is lowering the computational speed (~6.5 times slower for comparable simulations for single-precision GROMACS code normally used in NVT and NPT simulations). A system with a nanobubble of methane in water was used as the initial conditions for all simulations. Similar to a previous method, we prepared a nanobubble of methane in water, that is, 64 unit cells of sI hydrate were partially melted (only the inner hydrate crystal sphere with a radius of 1.5 nm was melted) at 425 K and 100 bar for 4 ns, followed by complete melting at 305 K and 100 bar for 15 ns. After the initial configuration was prepared, we equilibrated the system in the NPT ensemble at 250 K and 80 MPa for 1 ns. Then, six independent NVE runs, labeled as A-F, were performed. The initial methane concentration in the liquid solution for the six runs is 0.0235 ± 0.0002 mole fraction, which is calculated from the system excluding the methane in bubble. The cage cluster size, the $F_{4φ}$ order parameter, the number of solid-like water molecules, the number of dominant cage types, and the degree of crystallinity were used to quantify the nucleation events. $F_{4φ}$ is calculated through
\[ F_{4\varphi} = <\cos(3\varphi)> , \]  

(1)

where \( \varphi \) is the torsion angle between two H-bonded water molecules.\textsuperscript{40} The average values of \( F_{4\varphi} \) for ice, liquid water, and hydrate are -0.4, -0.04, and 0.7, respectively. Solid-like water molecules are identified when their molecular displacements are smaller than 0.15 nm/ns. The configurations of each trajectory were analyzed with an interval of 20 ps but the results were plotted in the figures with an interval of 1 ns. For visualization, molecules were trajectory-smoothed over 1 ns. VMD was used to generate all figures.\textsuperscript{44}

3. Results and discussion

Fig. 1 shows the temperature evolution in each of the six simulated systems; the timing of the onset of the evident increase in system temperature in each case corresponds to the onset of hydrate formation, clearly showing the exothermic nature of the phase transition. The temperature increases from 250 K to 310 K in run C, ~20 K higher than other runs. Based on our interpretation of the data, this extra increase is likely due to one of two phenomena (or a combination of the two): 1) more solid formed in run C, and thus more heat was released; 2) the heat of formation for phase transitions from melt to ordered solids is generally higher than the heat of formation for transitions from melt to amorphous solids.\textsuperscript{45-47}

![Fig. 1. Evolutions of the temperature of the system for all runs.](image-url)
Fig. 2. The cage cluster size (for CCs and FSICs), the $F_{4\phi}$ order parameter, and the number of solid-like water molecules for runs A and C.

The cage cluster size, the $F_{4\phi}$ order parameter, and the number of solid-like water molecules for run A (representative of runs A, B, D, E, and F) and run C are shown in Fig. 2. Clearly, run C has a larger cluster size for both complete cages (CCs) and face-saturated incomplete cages (FSICs), $^{28}$ $F_{4\phi}$ order parameter and number of solid-like water molecules than run A, showing that more solid hydrate formed in run C. Additionally, while runs C and E show similar induction times, as indicated by their temperature response (Fig 1), cage cluster size, $F_{4\phi}$ order parameter, and the number of solid-like water molecules (Fig. S1, ESI†), a more detailed analysis of structure and cage types (cage types are shown in Fig. S2, ESI†) actually shows these two pathways to be markedly different, highlighting the importance of looking at the fine
structure of incipient hydrates when investigating nucleation pathways. A stochastic nature in induction times is expected when investigating nucleation, and thus while it is merely coincidence that runs C and E show similar induction behavior, this coincidence is helpful in pointing out the mechanistic differences between these two runs and thus highlighting one of the main messages of this work: *stochasticity is present not only in nucleation induction phenomena but also in nucleation mechanism pathways.*

![Fig. 3. Snapshots of different stages for run C. Hydrogen bonds and water molecules are shown in red. Methane molecules are represented by cyan spheres. The green and blue cages are $5^{12}6^2$ and $5^{12}$ which connect with each other with pentagon face, respectively. The dashed black square shows sI motifs at 386 ns and 392 ns.](image)
The processes of hydrate nucleation and growth in run C are shown in Fig. 3 and Movie S1 (ESI†); the structure-I ordering is visually apparent on nucleation (insofar as it is apparent when the cluster is large enough that visual identification of sI motifs is possible) and throughout growth. The first cage link between $5^{12}$ and $5^{12}6^2$ cages, being characteristic of the structure I,$^{28}$ occurs at as early as 364 ns (Fig. 3A). The dashed squares in Figs. 3B and 3C show sI motifs at 386 ns and 392 ns, respectively. Fig. 4 shows a snapshot from the [001] crystallographic face for run C at 900 ns. The sI template is clear, and the dashed lines connecting the axes of rows of face-sharing $5^{12}$ or $5^{12}6^2$ cages indicate that the long-range ordering spans periodic boundaries. Note that for perfect sI hydrate, these rows would only be $5^{12}6^2$ cages; rows of $5^{12}$ and $5^{12}6^2$ cages are indicated in Fig. S3 (ESI†). Rows of $5^{12}$ cages have previously been observed in computational investigations of hydrate growth from a crystalline sI template, with the sI template ultimately preserved upon growth beyond the axes of these rows.$^{48}$ The snapshots from other crystallographic faces are shown in Fig. S4 (ESI†), indicating a few defects in the sI template, i.e., two pieces of dislocation with ~3.6 nm in length and one spherical fluid inclusion with ~1 nm in radius.
Fig. 4. The snapshot of the [001] crystallographic face from run C at 900 ns. Hydrogen bonds and water molecules are shown in red. Methane molecules are represented by cyan spheres. The blue box represents the system box. To show the long-range ordering, the periodic images in two directions are shown, both parallel to the page. The perpendicular or parallel nature of the dashed lines guides the eye in recognizing the sI template spanning the simulation box. The right panel shows the hydrogen-bonded network of perfect sI and sII hydrates.

Fig. 5 shows the number of $5^{12}$, $5^{12}6^2$, $5^{12}6^3$, $5^{12}6^4$, $4^15^{10}6^2$, $4^15^{10}6^3$ and $4^15^{10}6^4$ cages in the simulation systems for runs A and C. These seven cage types could transform to each other\textsuperscript{30} and comprise 80% and 91% of all FS cages counted from 900 ns to 1000 ns formed in the nucleated solids for run A and run C, respectively. The $5^{12}6^2/5^{12}$ ratio in run C (~2), while lower than that of perfect sI hydrate (3), is approximately 4 times higher than that of run A, 2.5 times higher than the simulation results at 285 K from Jiménez-Ángeles and Firoozabadi,\textsuperscript{34} 2 times higher than for the most-ordered simulations of Walsh \textit{et al.},\textsuperscript{30} and 1.4 times higher than even the sI hydrate region (the region C) of the heterogeneous nucleation results of Bai \textit{et al.}\textsuperscript{50} We also calculated the $5^{12}6^2/5^{12}$ ratio for methane-filled cages in run C (Fig. 6), which is higher than the reported $5^{12}6^2/5^{12}$ ratio observed experimentally for the first 2,000 seconds after cage identification during methane hydrate formation.\textsuperscript{49} Furthermore, only three cage types – $5^{12}$, $5^{12}6^2$ and $5^{12}6^3$ –
dominate run C; it is well known that the former two are the standard sI hydrate cage types while the last one is the intermediate cage type which can facilitate contact between sI and sII structures. As shown in Fig. 5b, in the period from 600 ns to 700 ns, the number of $4^15^{10}6^2$ cages decreases accompanied by a rapid increase of the number of $5^{12}6^2$ cages. Movie S2 (running from 590-700ns, ESI†) shows some nonstandard hydrate cages melting and standard sI hydrate cages forming during this time. Less dramatic occurrences of the same phenomenon occur just before 800 and 1000 ns. Compared to run C, run A contains more $5^{12}$ cages, and is dominated with $5^{12}, 5^{12}6^2, 5^{12}6^3, 5^{12}6^4$, and $4^15^{10}6^2$ cage types. Fig. S5 (ESI†) shows the resulting solid structure after nucleation and growth for run A, which indicates an amorphous-like character.

![Graph](image)

**Fig. 5.** Evolution of seven dominant cage types in the simulation system for a) run A and b) run C.
Fig. 6. The $5^{12}6^2/5^{12}$ ratio for methane-filled cages in runs A-F. Additionally, a data-chopped version of this figure can be found in ESI† to remove the large fluctuation at shorter times for clarity (Fig. S6, ESI†).

Fig. 7. The hydrate crystallinity for runs A and C.

Fig. 7 shows the crystallinity for runs A and C. As expected, the sI crystallinity of run C is higher than that of run A. Here, the crystallinity is the ratio of the number of characteristic cage links of hydrate structures (i.e., sI, sII, or sH) to total cage links and can quantify the crystalline extent of clathrate solids. A value of 0 corresponds to a fully amorphous hydrate phase, while a value of 1 corresponds to a perfectly crystalline hydrate phase. One can see that the sI crystallinity is ~0.7 at the end of simulation C, while the sII+sH crystallinity is only ~0.03.
However, this value of crystallinity is calculated for the cage clusters composed of all FSICs, including both standard and nonstandard hydrate cages. In order to compare with the crystallinity reported by Bi and Li, we also similarly considered only four cage types – i.e., $5^{12}$, $5^{12}6^2$, $5^{12}6^3$, and $5^{12}6^4$ cages. Calculated in this manner, the sI crystallinity is as high as 0.81 ($=1090/[1090+199+56]$, from Table 1) for run C; Bi and Li reported a crystallinity value of 0.82, though it was a structure II crystal, and the sII crystalline phase annealed or grew from an amorphous phase, while the hydrate phase in this work nucleated to sI more directly. Additionally, Yang et al. reported a new hydrate structure HS-I which is composed of $5^{12}$, $5^{12}6^2$, and $5^{12}6^3$ cages. If the links related to the $5^{12}6^3$ cages are assumed to be characteristic linking structures of HS-I hydrate, and only considering the $5^{12}6^n$ cages with $n=0$, 2, 3, 4, the HS-I crystallinity is calculated as 0.15 ($=199/[1090+199+56]$, from Table 1) in run C. Furthermore, because links between $5^{12}$ cages also appear in the HS-I structure and neither the $5^{12}6^4$ cage (sII) nor the $5^{12}6^8$ and $4^56^3$ cages (sH) exist in any appreciable number in run C, it is reasonable to classify the links between $5^{12}$ cages into the HS-I structure in run C. As a result, the HS-I crystallinity would increase to 0.19. From this viewpoint, the two pieces of dislocation in Fig. S4 (ESI†) can be interpreted as the HS-I crystal, and the system in run C evolves to a hydrate crystal composed of two co-existing polymorphs (81% sI and 15-19% HS-I). As for the fluid inclusion, it contains the excess methane molecules due to the existing of empty cages (~5%) and is separated from the crystalline phase mainly by other FSICs (shown in Fig. S7, ESI†).
Table 1. The average number of different cage links after hydrate formation in run C.*

<table>
<thead>
<tr>
<th>Cage link types</th>
<th>Links for sI</th>
<th>$5^{12}6^n$ linked</th>
<th>$5^{12}-5^{12}$</th>
<th>$5^{12}-5^{12}6^4$</th>
<th>Other links</th>
</tr>
</thead>
<tbody>
<tr>
<td>For FSICs</td>
<td>1090(8)</td>
<td>239(4)</td>
<td>56(1)</td>
<td>0</td>
<td>182(4)</td>
</tr>
<tr>
<td>For $5^{12}6^n$</td>
<td>1090(8)</td>
<td>199(3)</td>
<td>56(1)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>(n = 0, 2, 3, 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Structures</td>
<td>sI</td>
<td>HS-I</td>
<td>sII, sH, or HS-I</td>
<td>sII</td>
<td></td>
</tr>
</tbody>
</table>

* The cage links are calculated from 900 to 1000 ns, and the numbers in parentheses are errors.

Finally, we compare the present results with those of Liang and Kusalik.\(^{39}\) The authors carried out simulations for H\(_2\)S hydrate formation in the NVE, NPT, and NVT ensembles, and found the hydrates which nucleated in the NVE ensemble to be more crystalline than those observed in the isothermal ensembles (NPT and NVT). Our results are consistent with theirs, insofar as the present results (obtained with NVE simulations) include what is by far the most crystalline example of methane hydrate formed to date using direct MD simulation. Interestingly, although the authors claimed their NVE simulations support the two-step nucleation mechanism – in which an amorphous structure anneals or grows into a crystalline phase\(^{26-27}\) – they also observed some large domains of regular sI, sII, and HS-I crystalline structures and motifs which formed directly in their simulations. We interpret such ordered domains (especially those such as seen in run C of the present work, in which the long-range order spans the simulation system) as evidence that the direct formation to the crystalline phase is among the multiple nucleation pathways available to hydrate formation, and that an amorphous intermediate is not always observed. These distinctions in messages may be due to different standards in assignation of the terms “amorphous” and “crystalline.” We are hesitant to call a highly-ordered solid with an
overarching sI template which spans the simulation boundaries “amorphous” simply because (as in experimentally-investigated macroscopic hydrates), it contains some imperfections in its crystal structure, though we do recognize it is not a perfect crystal. Thus, as there remain unresolved questions over definitions, the concentration of defects in the structure of imperfect (and yet non-amorphous) crystals in the early stages of experimentally-investigated nucleation events (albeit likely on millisecond or second timescales rather than on nanosecond timescales) is needed to understand the standards by which the terms “amorphous” and “crystalline” can realistically be applied to computational studies, in which every defect is clearly apparent due to the ability of tracking all molecules in the system at fine time resolutions during phase change.

4. Summary

In this work, six high-precision NVE molecular dynamics simulations were performed to investigate methane hydrate nucleation and growth in the absence of a thermostat. While five of the simulations nucleate to amorphous hydrate phase, one simulation nucleates more directly to a sI crystalline template of methane hydrate with long-range order spanning the simulation box across periodic boundaries. For the later, although the resulting solid phase is not the perfect sI structure, its sI crystallinity reaches as high as 0.7 (and as high as 0.81 if only considering 5^{12 \cdot n} cages with n=0, 2, 3, and 4). Moreover, the highly-crystalline hydrate phase nucleates directly from the melt and does not come from any dominant amorphous phase by an annealing process. By forming long-range order spanning the simulation system, this investigation reports the most compelling evidence to date that gas hydrates can nucleate not only to amorphous solids, but also to ordered solids with a high degree of crystallinity, and that the direct nucleation of the globally-stable hydrate phase is possible. This work is thus direct evidence supporting previous
speculations from computational studies that hydrates nucleate through multiple pathways.\textsuperscript{30,38} While direct comparison with experimental hydrate formation studies on the nanoscale in time and length remains an outstanding challenge, we conclude by asserting that there does exist the opportunity to explore the general concept of multiple pathways using time-resolved spectroscopy, diffraction, or microscopy for multiple experimental nucleation events.

Acknowledgements

We thank Drs. Patrick G. Lafond, David T. Wu, Carolyn A. Koh, Werner F. Kuhs, and Keith C. Hester for their valuable discussions about this work, for helping us navigate the body of kinetic and thermodynamic arguments previously forwarded to explain this intricate process, and for helpful, timely responses to our requests for clarifications and insight. We thank the Computer Simulation Lab at IGGCAS for allocation of computing time. This work was supported by National Science Foundation of China (Grant No. 41372059) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB10020301).

Notes and references

*Email: guogj@mail.igcas.ac.cn.

† Electronic Supplementary Information (ESI) available: Figs. S1-S7, and Movies S1-S2.