

Санкт-Петербургский государственный университет

*Посысоев Максим Владимирович*  
Выпускная квалификационная работа

*Молекулярно-динамическое моделирование распределения  
агрегатов по размерам для неионного ПАВ в неполярном  
растворителе*

Уровень образования: магистратура  
Направление 03.04.02 "Физика"  
Образовательная программа ВМ.5511.2020 "Физика"

Научный руководитель:  
к.ф.-м.н., доцент кафедры  
статистической физики СПбГУ,  
Волков Николай Александрович

Рецензент:  
Федеральное государственное бюджетное  
учреждение науки  
Объединенный институт высоких  
температур РАН (ОИВТ РАН),  
старший научный сотрудник,  
Кондратюк Николай Дмитриевич

# Оглавление

<b>Введение</b> . . . . .	<b>3</b>
<b>1. Структурные свойства обратных мицелл</b> . . . . .	<b>6</b>
1.1. Обратная мицелла C12E4 . . . . .	6
1.2. Радиальные функции распределения . . . . .	6
<b>2. Распределение мицелл по числам агрегации</b> . . . . .	<b>10</b>
2.1. Работа образования мицеллы . . . . .	10
2.2. Процедура вычисления распределений по числам агрегации . . . . .	10
2.3. Результаты получения работы агрегации и распределений по числам агрегации . . . . .	11
<b>3. Электростатический вклад в работу образования</b> . . . . .	<b>21</b>
3.1. Энергетическая выгода от мицеллообразования . . . . .	21
3.2. Результаты вычисления энергетической выгоды . . . . .	23
<b>4. Коэффициенты присоединения и отсоединения</b> . . . . .	<b>32</b>
4.1. Диффузионное приближение . . . . .	32
4.2. Процедура вычисления коэффициентов присоединения и отсоединения . . . . .	33
4.3. Предварительные результаты . . . . .	33
4.4. Проблемы алгоритма кластеризации . . . . .	39
4.5. Результаты полученные вручную . . . . .	40
4.6. Алгоритм очистки временных рядов, основанный на вычислении скользящей энтропии . . . . .	42
4.7. Результаты полученные с помощью алгоритма очистки . . . . .	44
<b>Заключение</b> . . . . .	<b>47</b>
<b>Приложение А. Определение радиуса обрезания при кластеризации</b> . . . . .	<b>48</b>
<b>Список литературы</b> . . . . .	<b>52</b>

## Введение

Данная работа посвящена исследованию полярных и неполярных растворов неионных поверхностно-активных веществ (ПАВ) [1] с помощью метода молекулярной динамики [2] [3] с использованием полноатомных моделей веществ. В качестве неионного ПАВ было выбрано вещество C12E4. Особенностью поверхностно-активных веществ является их способность образовывать мицеллы находясь в растворителе при достижении определённой концентрации, которая называется критическая концентрация мицеллообразования (ККМ). Мицеллы – это агрегаты поверхностно-активного вещества, которые находятся в растворителе, но не на поверхности. Это происходит из-за гидрофобного эффекта [4].

Существует два разных вида мицелл: прямые и обратные. Разница между ними в конфигурации. Чтобы её понять, для начала взглянем на одну молекулу C12E4. Как видно из рисунка 1, молекула может быть условно разделена на две части. Первая часть (слева), содержащая атомы кислорода, является гидрофильной. Её обычно называют головой. Вторая часть молекулы (справа) представляет собой углеводородный хвост и является гидрофобной. Её обычно называют хвостом. Схематично прямые и обратные мицеллы выглядят так (см Рисунок 2):

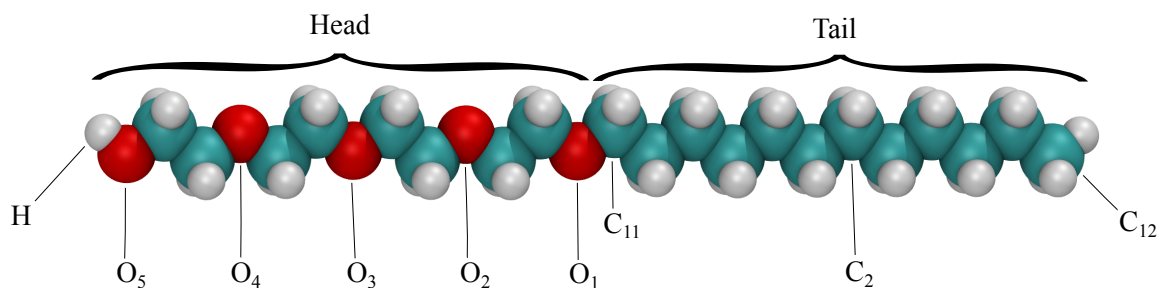


Рис. 1 Молекула вещества C12E4.

Отметим что механизм образования прямых и обратных мицелл различен. Об этом пойдет речь в Разделе 3.

Значительная часть ВКР посвящена вопросу нахождения работы образования мицелл и различным вкладам в эту работу. Также была предпринята попытка вычислить коэффициенты присоединения и отсоединения для обратных мицелл.

### Актуальность темы

Поверхностно-активные вещества встречаются повсеместно. Мы ежедневно ими пользуемся. Например, поверхностно-активные вещества входят в состав мыла, зубной пасты, стиральных порошков. Важной областью применения ПАВ является нефтедобыча. Поэтому изучение систем содержащих ПАВ является актуальной темой.

### Степень разработанности темы исследования

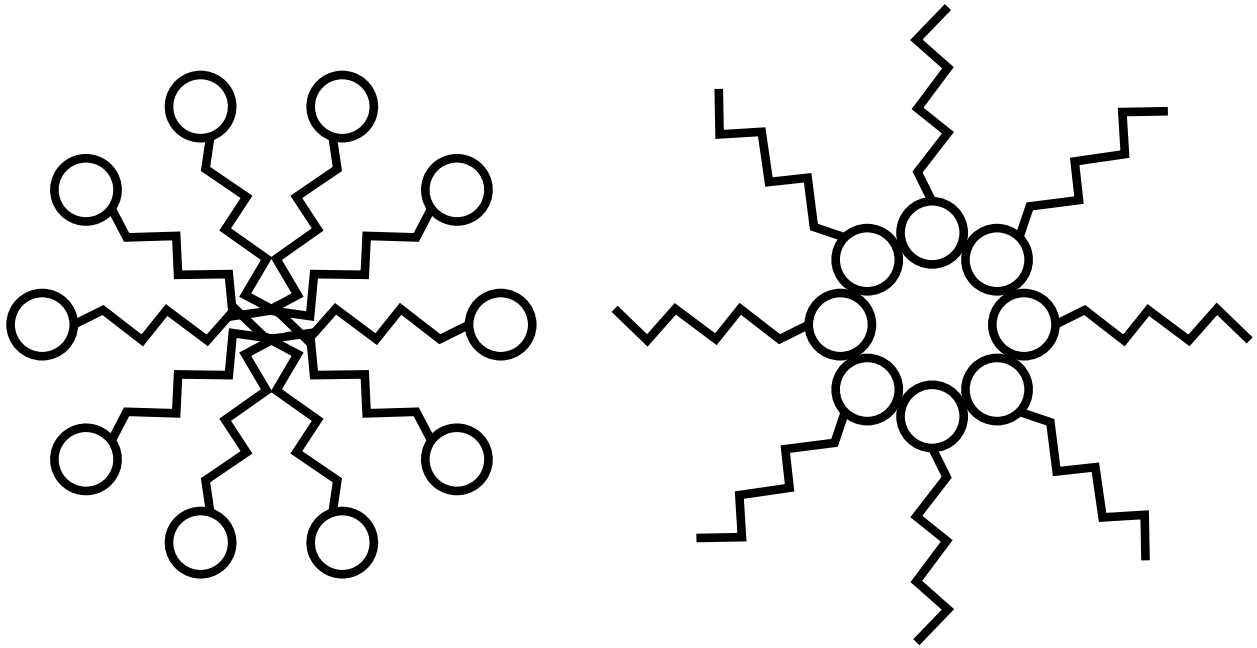


Рис. 2 Прямая мицелла в полярном растворителе (слева, в  $H_2O$ ) и обратная мицелла в неполярном растворителе (справа, в  $C_7H_{16}$ ) мицеллы.

Изучением ПАВ занимаются достаточно давно [5]. Системы для прямых мицелл достаточно хорошо изучены. В то же время число работ по изучению обратных мицелл методами молекулярного моделирования гораздо меньше, чем для прямых мицелл.

**Научная новизна.** Данная работа в основном посвящена изучению обратных мицелл. В Разделе 3 затрагиваются и прямые мицеллы, но лишь в контексте сравнения с результатами полученными для обратных мицелл.

**Основной целью** данной работы является получение работы образования  $W(n)$ , как функции числа агрегации  $n$  (количество мономеров агрегате), а также вычисление электростатического вклада в работу образования. Поставленная цель требует решения следующих задач:

- I. Провести кластерный анализ молекулярно-динамических траектории и получить распределения мицелл по числам агрегации
- II. Написать программы для расчета электростатического вклада в работу образования

Так же в работе исследовались кинетические свойства мицеллярного раствора. В частности вычислялись коэффициенты присоединения и отсоединения мономеров к мицелле.

**Теоретическая и практическая значимость.** Результаты компьютерного моделирования могут быть полезны для уточнения аналитических теорий. Методы молекулярного моделирования позволяют на уровне отдельных атомов лучше понять процессы в рассматриваемых системах.

**Методология и методы исследования.** Все молекулярно-динамические траектории были получены с помощью программы GROMACS [6], с использованием силового

поля CHARMM36 [7] и CGenFF v. 4.4. Программы для анализа траекторий были написаны на языках программирования Python и Rust.

**Достоверность полученных результатов.** Распределения полученные в данной работе очень похожи на распределения для других систем. [8]. Вид работы образования похож на теоретические модели [9]. К сожалению, сравниться с экспериментальными данными на данном этапе развития эксперимента невозможно.

**Апробация работы.** Результаты и положения работы докладывались и обсуждались на следующих научных конференциях:

1. Международная научная конференция "Современная химическая физика на стыке физики, химии и биологии сборник тезисов, стр 418, 2021 год;

Также планируется представить результаты работы на XXIII Международной конференции по химической термодинамике в России (RCST-2022, Казанский (Приволжский) федеральный университет, 22-27 августа 2022).

Работа выполненная в рамках данной ВКР была поддержана грантом РФФИ "Развитие термодинамики и кинетики мицеллообразования в неполярных средах".

**Объем и структура работы.** Работа состоит из введения, четырёх разделов и заключения. Полный объём работы составляет 53 страницы с 33 рисунками и 5 таблицами. Список литературы содержит 20 наименований.

- Во **Введении** дано краткое обзор определение исследуемой системы, обоснована актуальность проделанной работы, сформулирована цель и аргументирована научная новизна исследований, описаны методология, методы и степень разработанности темы исследования, а также показана практическая значимость полученных результатов. Представлены выносимые на защиту научные положения.
- В **Разделе 1** были вычислены радиальные функции распределения плотности относительно центра масс
- В **Разделе 2** были вычислены распределения по числам агрегации
- В **Разделе 3** был вычислен электростатический вклад в работу образования для прямых и обратных мицелл
- В **Разделе 4** было вычислено количество присоединений мономеров к мицелле в единицу времени
- В **Заключении** представлены основные результаты и выводы

# 1. Структурные свойства обратных мицелл

В этом разделе речь пойдет о системе, о её структурных свойствах и процедуре вычисления радиальных функций распределения.

## 1.1. Обратная мицелла C12E4

Системы, которые были выбраны для данной работы, делятся на два типа, в зависимости от того, какой растворитель использовался, а само поверхностно-активное вещество (далее ПАВ) оставалось неизменным. Оно представляет собой вытянутую молекулу  $C_{20}H_{42}O_5$  (рисунок 1), у этой формулы есть сокращенная запись  $C_{12}E_4$ , где  $E$  – это сокращение одного из звеньев цепочки. Далее в тексте будет использоваться сокращенная форма записи химической формулы.

Поведение этого вещества оказывается различным, в зависимости от того в каком растворителе оно находится. Так, например, в воде ( $H_2O$ ) система собирается в мицеллы, которые называются прямыми. Происходит это вследствие гидрофобного эффекта, который вносит основной вклад в мицеллообразование. В неполярном растворителе, например в гептане ( $C_7H_{16}$ ), вещество  $C_{12}E_4$  собирается в мицеллы, которые называются обратными 2.

## 1.2. Радиальные функции распределения

### Что это такое?

Радиальные функции распределения (англ. Radial Distribution Function – RDF) являются важным инструментом для изучения структурных и термодинамических свойств молекулярных систем (широко распространено обозначение  $g(r)$ ). По виду RDF можно понять, в каком агрегатном состоянии находится система. В данном разделе мы занимались вычислением радиальной функции распределения разных атомов мицеллы относительно её центра масс. Этот объект позволяет понять структуру мицеллы. Формально  $g(r)$  определяется следующим образом [10]:

$$\frac{N}{V} = \rho, \quad (1)$$

$$\rho^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n) = \rho^n g^{(n)}(\mathbf{r}_1, \dots, \mathbf{r}_n), \quad (2)$$

где  $\rho^{(n)}$  – это  $n$ -частичная функция распределения. В случае нас интересует парная корреляционная функция  $g^{(2)}(r_1, r_2)$ .

### Как это вычислить?

На первом шаг необходимо вычислить центр масс мицеллы для каждого момента времени по следующей формуле:

$$\vec{r}_c = \frac{\sum m_i \vec{r}_i}{\sum m_i}, \quad (3)$$

где  $m_i$  – масса  $i$  атома,  $r_i$  – радиус-вектор  $i$  атома. Вычисление центра масс сопряжено с трудностями. Периодические граничные условия (ПГУ) могут частично пересекать мицеллу, перенося какие-либо части в другой край ячейки моделирования. Это делает некорректным прямое применение формулы (3). Поэтому перед тем, как вычислить положение центра масс, необходимо сдвинуть мицеллу так, чтобы её не пересекали ПГУ, а после вычисления центра масс учать сделанный сдвиг.

На следующем шаге необходимо вычислить расстояния между центром масс и атомами, принадлежащими мицелле. Результаты сохраняются в массив.

На третьем шаге вычисляется гистограмму ( $n(r)$ ) для массива с предыдущего шага для количества частиц в дифференциальном слое  $dr$  в зависимости от расстояния, а так же учесть нормировки.

$$g(r) = \frac{V}{4\pi r^2 dr N} n(r), \quad (4)$$

где  $4\pi r^2 dr$  – это дифференциальный объем слоя, на расстоянии  $r$ ,  $N$  – количество частиц.

### Что получилось?

Система, которая была проанализирована в данном разделе, была смоделирована при температуре  $T = 223$  К в растворе из 1000 молекул гептана. Количество молекул ПАВ мы варьировали: 16, 25, 28, 32.

Были посчитаны RDF: для всех атомов мицеллы; только для кислорода, входящего в состав; только для углерода, входящего в состав хвоста каждой из молекул ПАВ. И, как можно видеть, из полученных графиков поведение у RDF сильно меняется в зависимости от размера мицеллы. Так, например, при числах агрегации 16 и 25 графики имеют ожидаемый нами вид. Большая часть атомов кислорода сконцентрирована в центре мицеллы, атомы углерода с хвостов мономеров находятся за пределами этого ядра. Однако при числах агрегации 28 и 32 ядро перестает быть однородным, внутрь ядра, судя по всему, начинают попадать мономеры целиком, в том числе с углеводородным хвостом. Об этом, например, свидетельствует пик желтого цвета на рисунке 5, который находится левее пика концентрации атомов кислорода.

Данные расчеты могут быть полезны для построения теоретических моделей. Например, данные системы можно с какой-то степенью точности считать двумя концентрическими сферами. Меньшая из них является ядром. Так поступают в теории диффузионного приближения, в ней мицеллы аппроксимируют сферами, речь о которой пойдет в другом разделе.

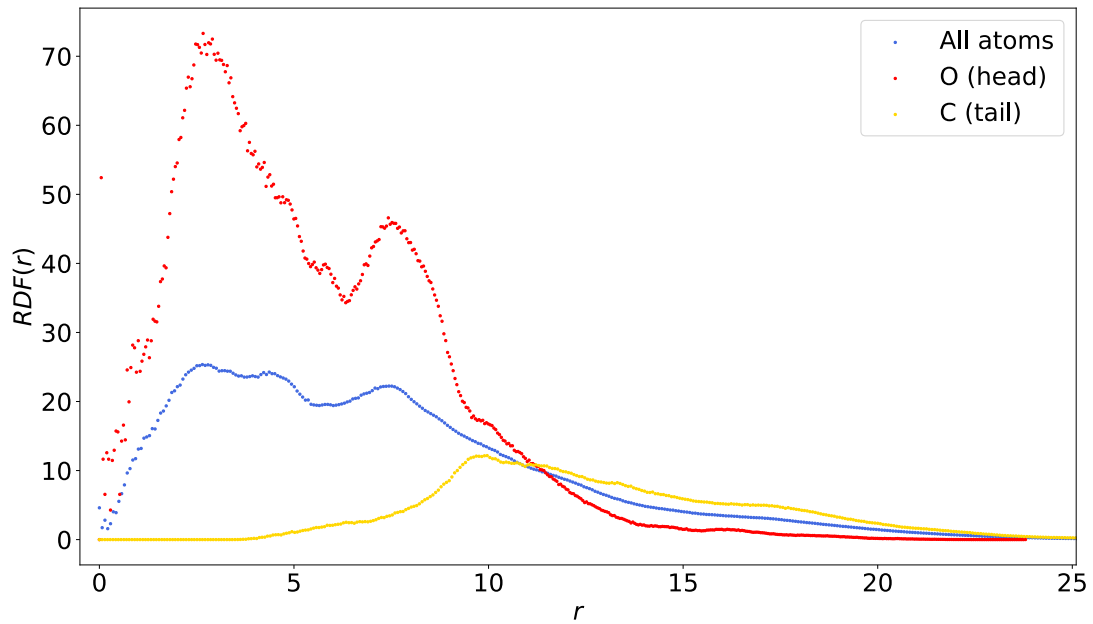


Рис. 3 Функция радиального распределения атомов относительно центра масс мицеллы с числом агрегации 16.

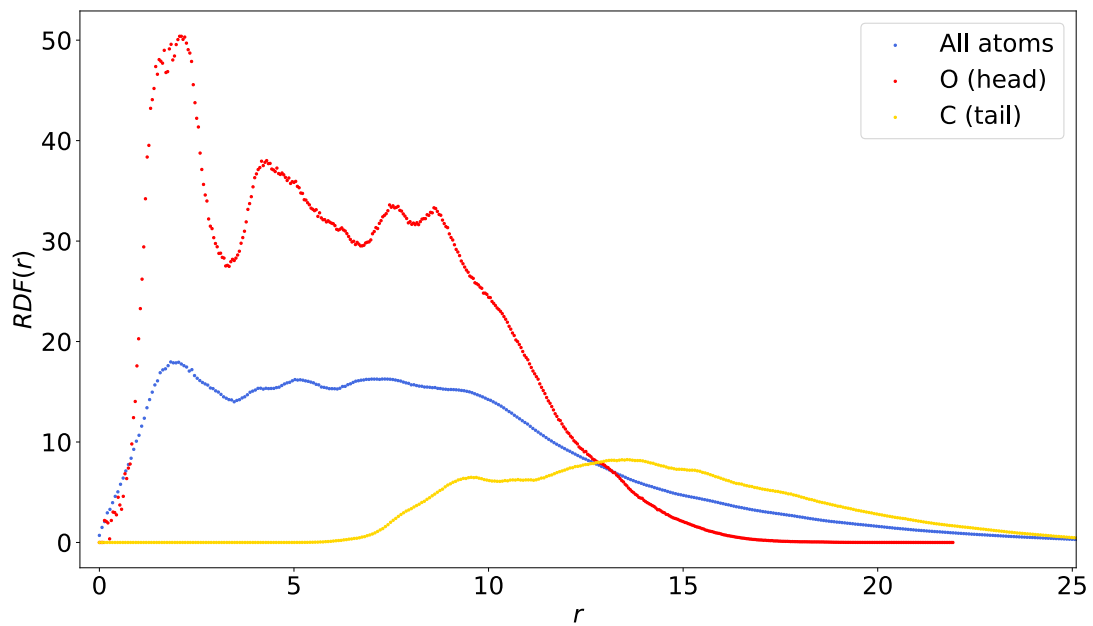


Рис. 4 Функция радиального распределения атомов относительно центра масс мицеллы с числом агрегации 25.



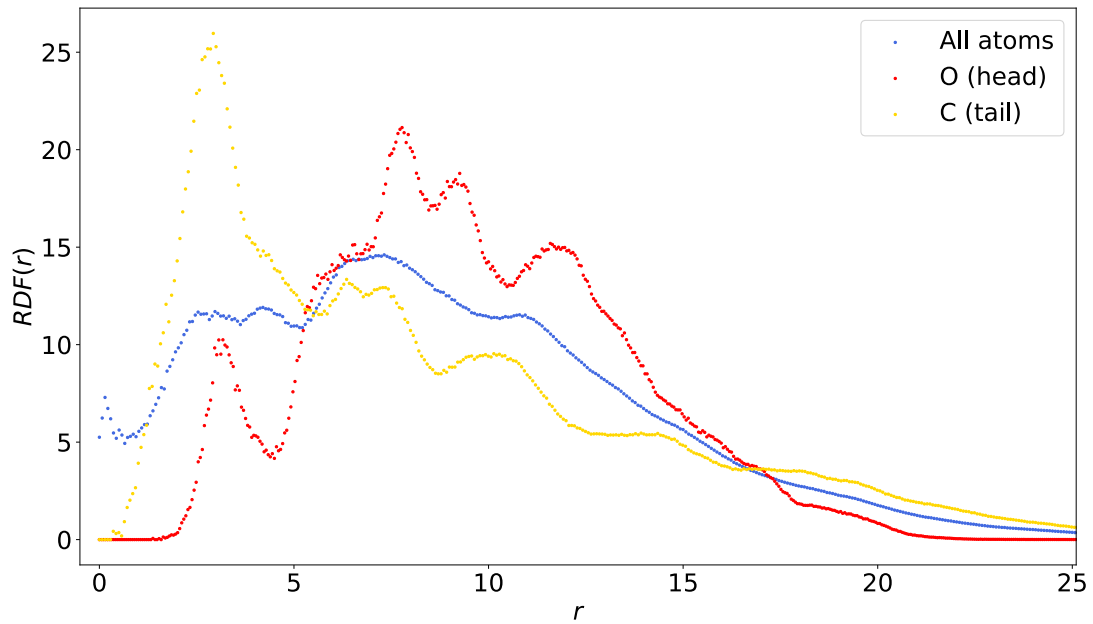


Рис. 5 Функция радиального распределения атомов относительно центра масс мицеллы с числом агрегации 28.

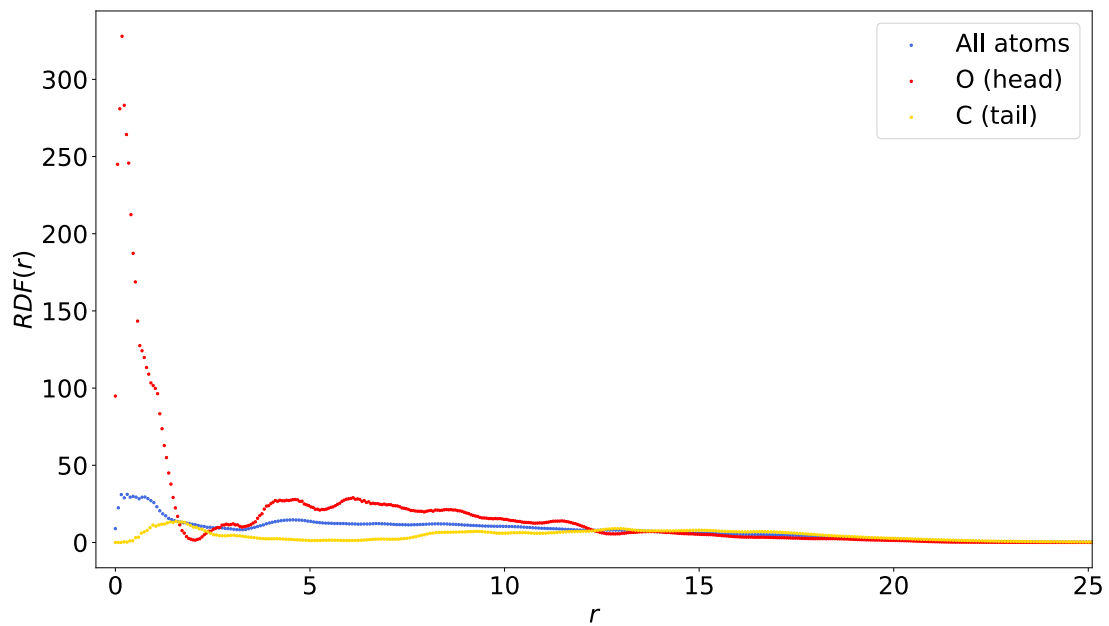


Рис. 6 Функция радиального распределения атомов относительно центра масс мицеллы с числом агрегации 32.

## 2. Распределение мицелл по числам агрегации

В этом разделе пойдет речь про распределение мицелл по числам агрегации и взаимосвязь между этим распределением и работой образования мицелл.

### 2.1. Работа образования мицеллы

Основной целью данной работы является получение работы образования обратных мицелл. Эта величина является необходимой для построения аналитической теории мицеллообразования. С помощью молекулярного моделирования можно оценить величину и узнать характер поведения этой функции. Существует формула, которая связывает распределение по числам агрегации с работой образования, поэтому для получения этой оценки, необходимо получить это распределение [8].

Формула (5) отражает эту связь, где  $c_n$  – это концентрация агрегатов в растворе размера  $n$ ,  $c_1$  – концентрация мономеров, а  $W(n)$  – искомая работа образования.

$$c_n = c_1 e^{-W(n)} \quad (5)$$

Путём несложных преобразований можно получить следующее выражение:

$$W(n) \sim -\ln(c_n) \quad (6)$$

Знак пропорциональности написан не случайно. Он говорит о том, что в данной работе мы интересуемся только характером поведения функции.

### 2.2. Процедура вычисления распределений по числам агрегации

Для получения распределения по числам агрегации для обратных мицелл была выбрана система состоящая из 25 мономеров  $C_{12}E_4$  в растворе гептана  $C_7H_{16}$ . Была проведена симуляция при достаточно низкой температуре  $T = 223$  К. При такой температуре мицелла является стабильной, от неё не отделяются момеры. А значит, это позволит определить радиус обрезания  $r_{cutoff}$  для алгоритма single linkage из встроенной в GROMACS [6] утилиты `gmx clustsize`.  $r_{cutoff}$  – это максимальное расстояние, при котором две частицы ещё считаются одним кластером. Если расстояние больше, то это разные кластеры. Алгоритм работает по цепочке. Лучше всего он иллюстрируется Рисунком 7.

Красной окружностью обозначается расстояние  $r_{cutoff}$  считая от центра атома. Условные частицы, обозначенные одним цветом, находятся в одном кластере.

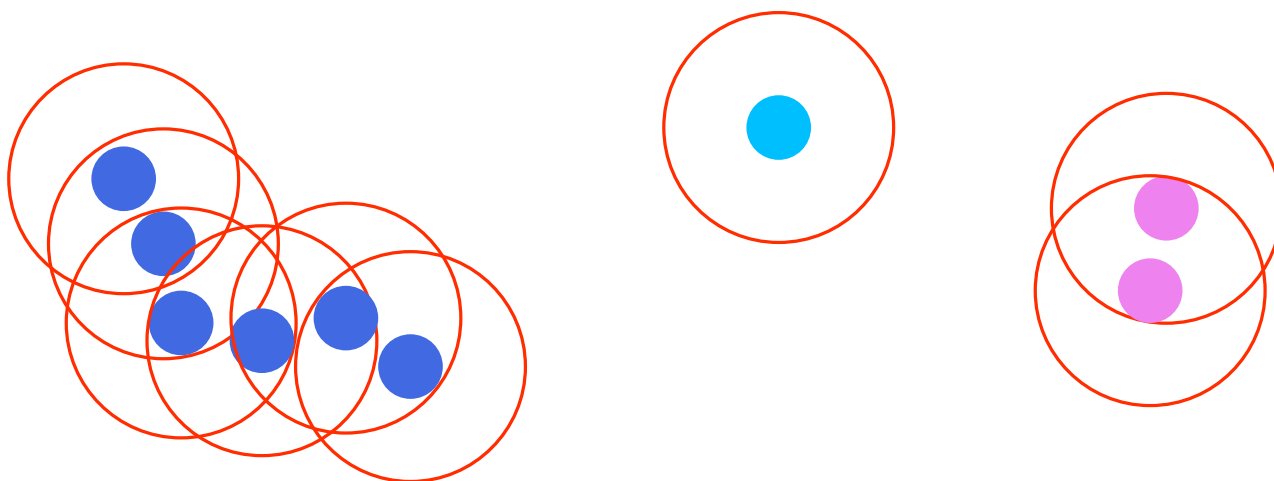


Рис. 7 Процедура кластеризации.

### 2.3. Результаты получения работы агрегации и распределений по числам агрегации

Итак, изначально мы пытались добиться распределения вида как показано на рисунке 8 для системы, состоящей из 25 мономеров при температуре  $T = 223$  К, при которой все 25 мономеров собираются в один агрегат. Путем нескольких итераций было получено значение  $r_{cutoff} = 4.6 \text{ \AA}$ , при котором все 25 мономеров образуются в один агрегат размером 25 в 100% моментов времени всей молекулярно-динамической траектории. На самом деле, это значение не является идеально точным и подходящим для любой ситуации, а является оценкой снизу. Меньше этого значения брать не стоит, поскольку в стабильной "замороженной" мицелле, от которой, точно известно, не отделяются мономеры, при радиусах меньше алгоритм находит отделившиеся атомы, что неверно. При повышении температуры подвижность мономеров, входящих в состав мицеллы растет. Из-за движений хвостами иногда алгоритм считает некоторые особо подвижные мономеры отделившимися, что тоже неверно. Очень трудно дать строгий и однозначный критерий определения радиуса обрезания, поэтому разумно эмпирически выбирать значение чуть больше минимальной границы. Слишком большие значения – это тоже плохо. Поэтому стоит действовать, исходя из древнекитайской концепции середины пути. (Чуть позже в этой работе пойдет речь ещё раз об алгоритме кластеризации, о проблеме определения радиуса обрезания. Будет дана гипотеза о том, как стоит поступать, чуть более конкретно). [см. Приложение А]

#### Распределение по числам агрегации для системы из 25 мономеров при разных температурах

После этого была произведена серия симуляций при различных температурах, и для полученных данных были снова получены распределения по числам агрегации (см. Рисунок 9). На этом графике 9 легко заметить ожидаемую закономерность: чем больше

температура системы, тем выше пик в области мономеров ( $n = 1$ ) и меньше пик в области мицеллы находящейся в системе ( $n \approx 20 - 25$ ).

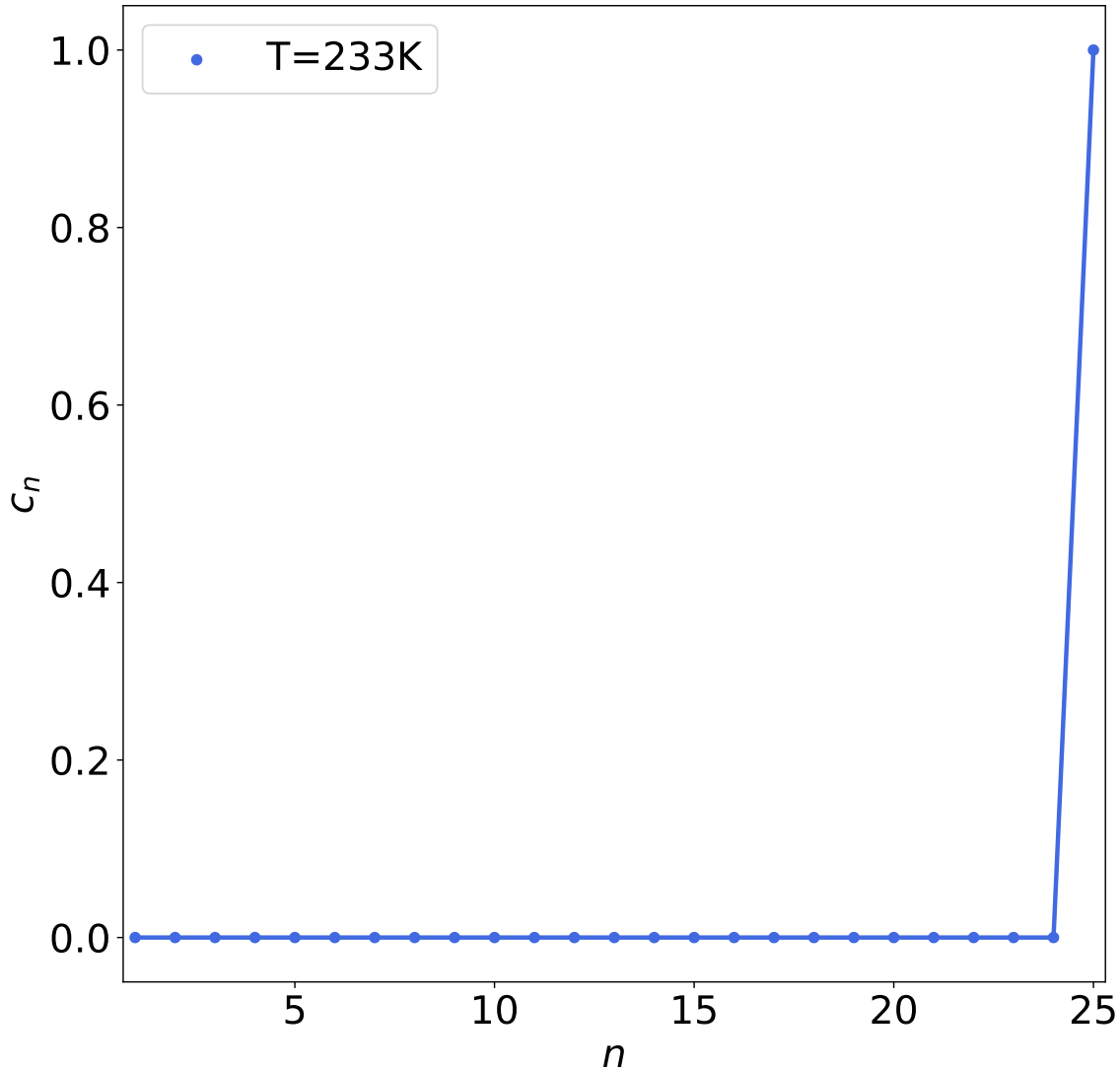


Рис. 8 Распределение по числам агрегации для "замороженной" мицеллы с числом агрегации 25 при температуре  $T = 223$  К.

$$\sum_{n=1}^N c_n \cdot n = N \quad (7)$$

Примечание: У каждого из этих графиков существует нормировка вида (7), поэтому можно всегда проверить, что получается распределение с тем же самым смыслом.

**Распределение по числам агрегации для систем состоящих из 100 и 200 мономеров.**

У представленных в предыдущем разделе графиков есть недостатки. Может вызывать вопрос факт, что пик в области мицеллы стремится к правому краю графика и упи-

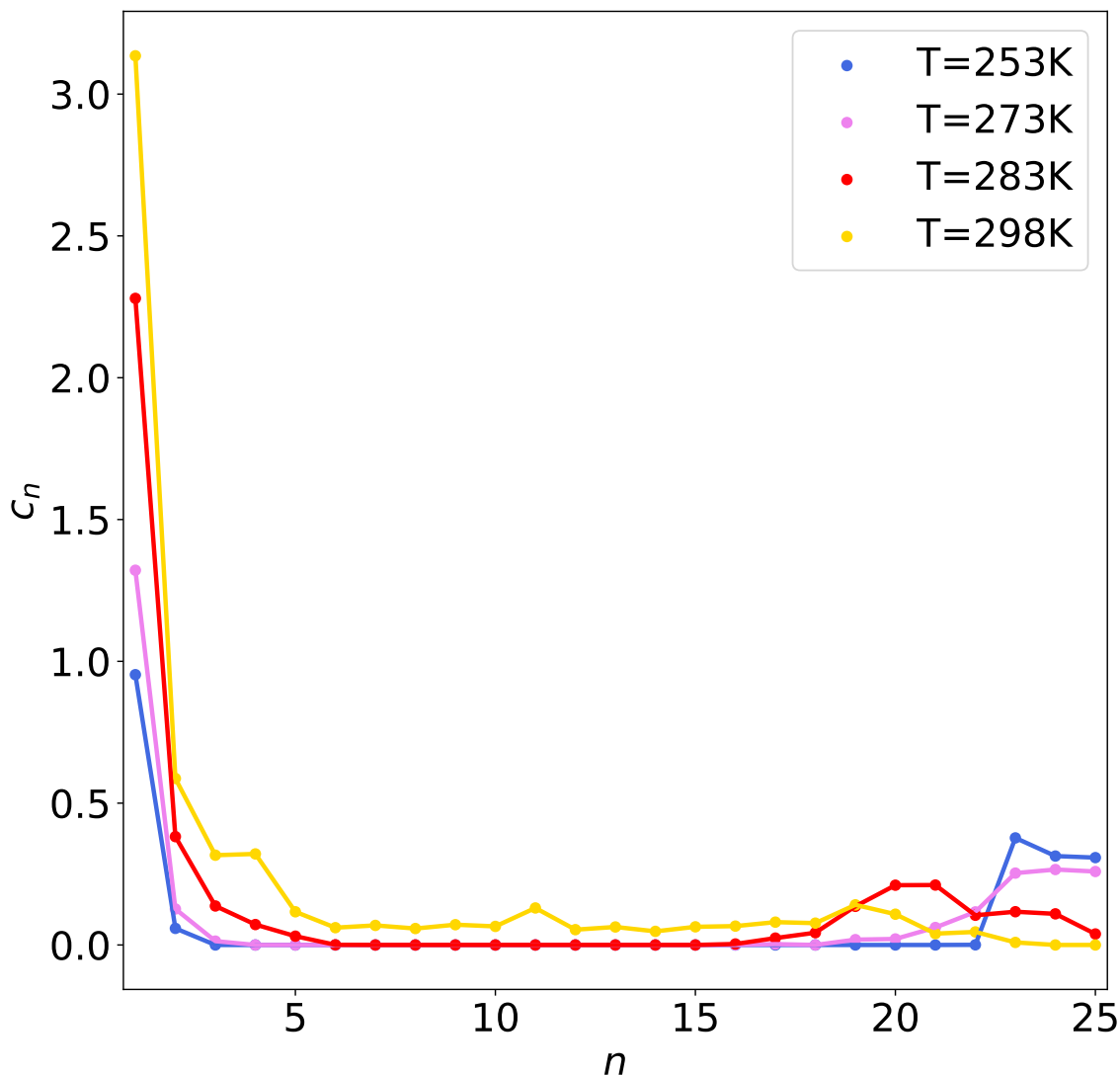


Рис. 9 Распределение по числам агрегации для системы 25 мономеров  $C_{12}E_4$  при различных температурах.

рается в него. Для того, чтобы избавиться от обусловленности размерными эффектами, была выбрана температура, которая даёт наиболее плавное распределение в области мицелл  $T = 283$  К. Для выбранной температуры было проведено моделирование для системы с размером 100 молекул  $C_{12}E_4$ . Полученная траектория была длиной в 1 микросекунду. Мы ожидали, что этого будет достаточно, для того, чтобы получить достаточно гладкое распределение.

Получив траекторию, мы отбросили первые 200 нс, поскольку эти времена соответствуют неравновесному состоянию системы. Критерием отброса именно этого времени послужил график 10. Траектория была разделена на отрезки по 50 нс, поэтому мы построили график распределения для каждого куска по отдельности. На этом графике 10 синим цветом отмечены 4 линии, соответствующие первым 4 файлам (т.е. первым 200 нс).

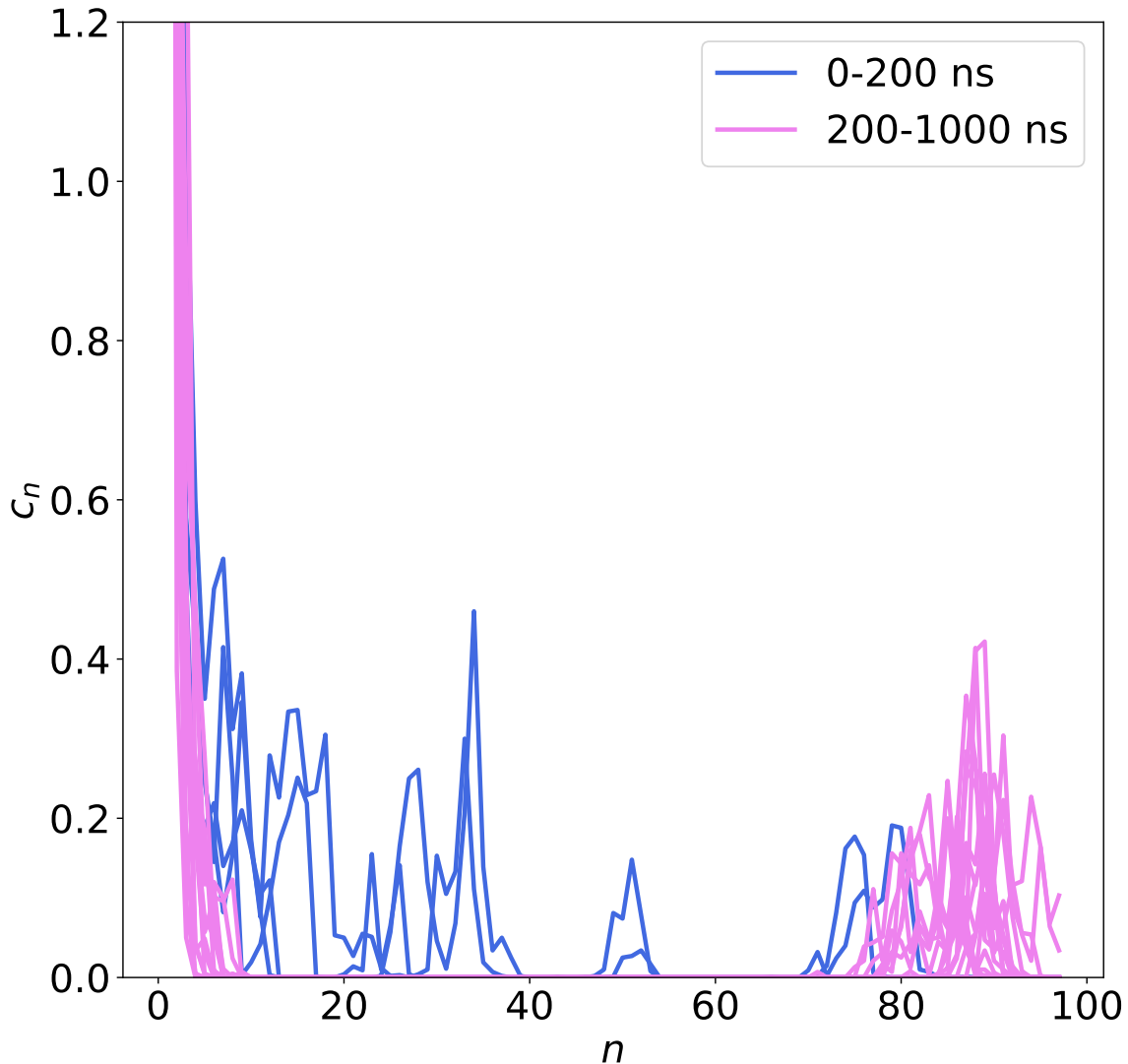


Рис. 10 Распределение по числам агрегации для системы 100 мономеров  $C_{12}E_4$  при температуре  $T = 283$  К, кусочки траектории по отдельности по 50 нс.

Как можно видеть, их поведение соответствует стадии формирования мицеллы. Фиолетовым цветом отмечены линии равновесного состояния системы. Поэтому разумно делать усреднение только по этим файлам, которые не имеют пиков в области  $n \approx 15-70$ .

На графике 11 представлено распределение по числам агрегации для системы, состоящей из 100 мономеров. На графике 12 представлено то же самое распределение, однако масштаб увеличен.

После получения распределения по числам агрегации можно вычислить  $-\ln(x)$  от функции распределения и получить вид работы образования (6). Результаты логарифмирования представлены на рисунке 13.

Аналогичная процедура была проведена для системы, состоящей из 200 мономеров ПАВ и 8000 молекул гептана. Была получена траектория длиной 900 нс, для уравни-

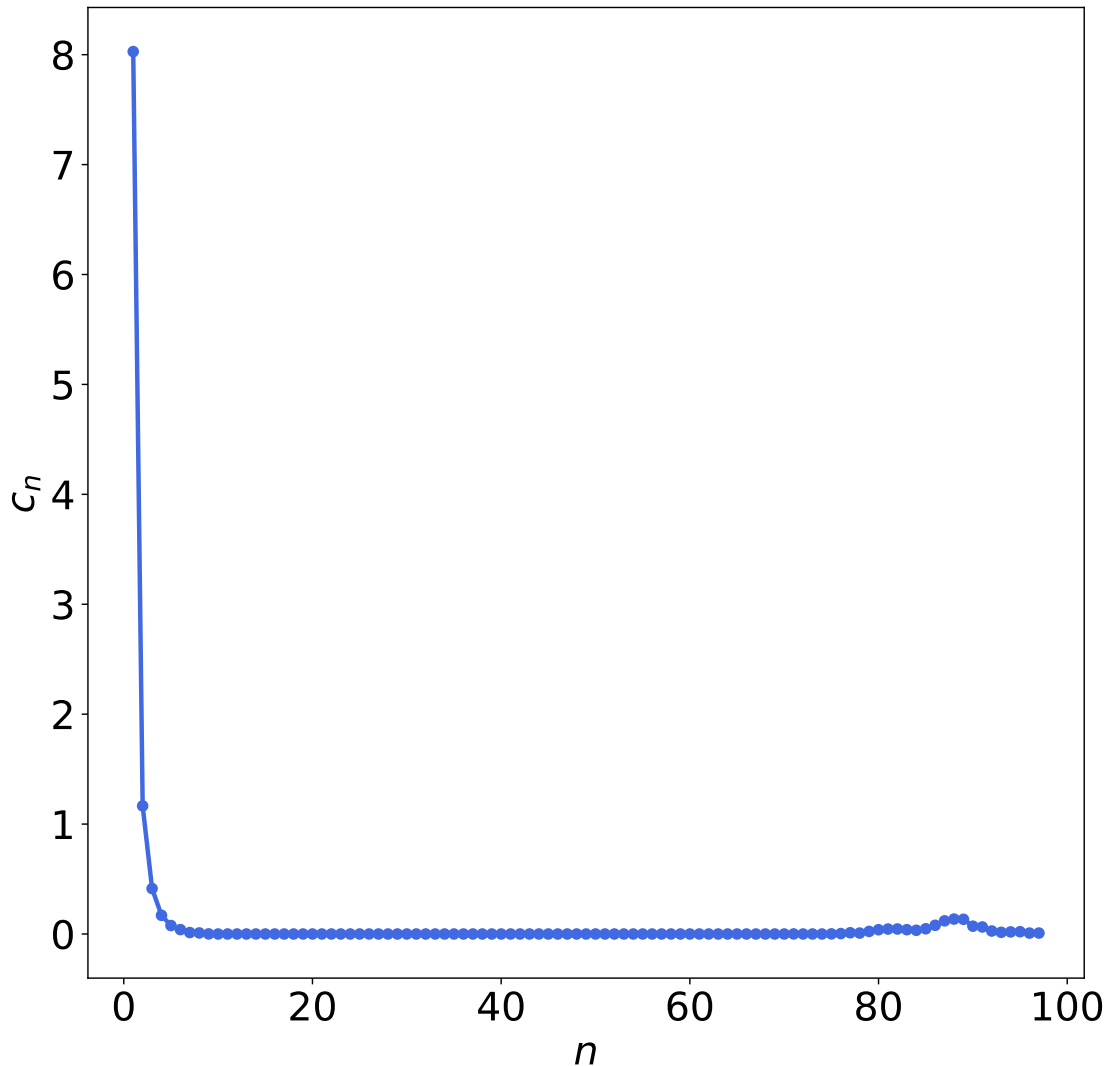


Рис. 11 Распределение по числам агрегации для системы состоящей из 100 мономеров  $C_{12}E_4$  при температуре  $T = 283$  К в растворе 4000 молекул гептана.

вания были отброшены 100 нс. Построено усредненное распределение по числам агрегации (Рисунки 14 и 15). После этого вычислена работа образования (Рисунок 16).

Небольшой комментарий. На графике для 200 мономеров наблюдаются два пика. Причем, если посмотреть на каждый участок по 50 нс по отдельности, видно, что такая картина наблюдается почти на каждом отрезке. Это может говорить о том, что длина молекулярно-динамической траектории недостаточно большая. Однако размер мицеллы крайне редко оказывается больше 100. Это, по всей вероятности, связано с тем, что существует некий потенциальный барьер; или просто работа образования становится большой и из-за этого энергетически менее выгодно формировать конфигурации с большим размером. Поэтому вычисления для систем размером в 100 частиц можно считать достоверными

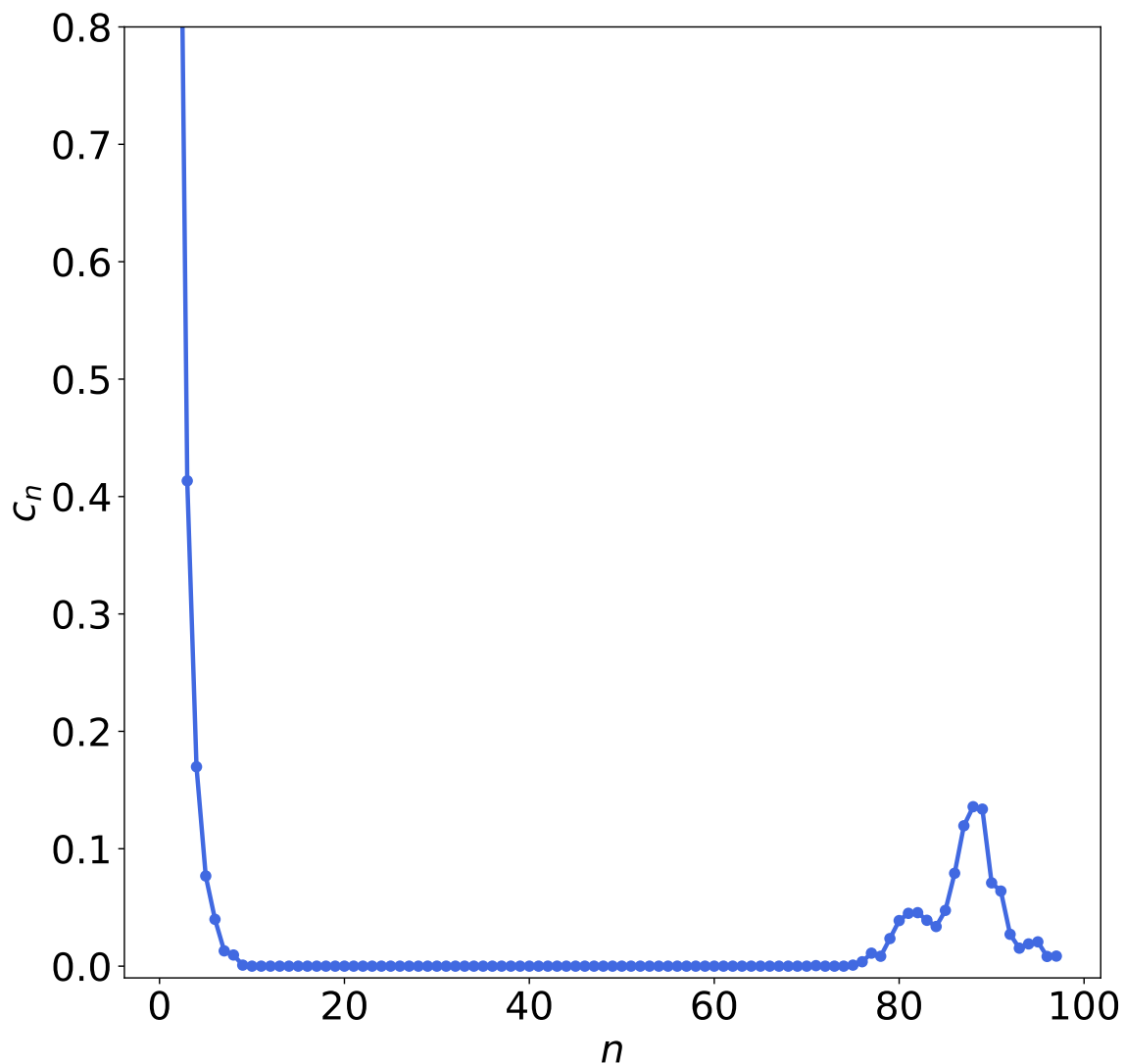


Рис. 12 Распределение по числам агрегации для системы состоящей из 100 мономеров  $C_{12}E_4$  при температуре  $T = 283$  К в растворе 4000 молекул гептана.

(в плане размера системы и длительности траектории). В целом результаты этого раздела согласуются с результатами других работ [11; 12].



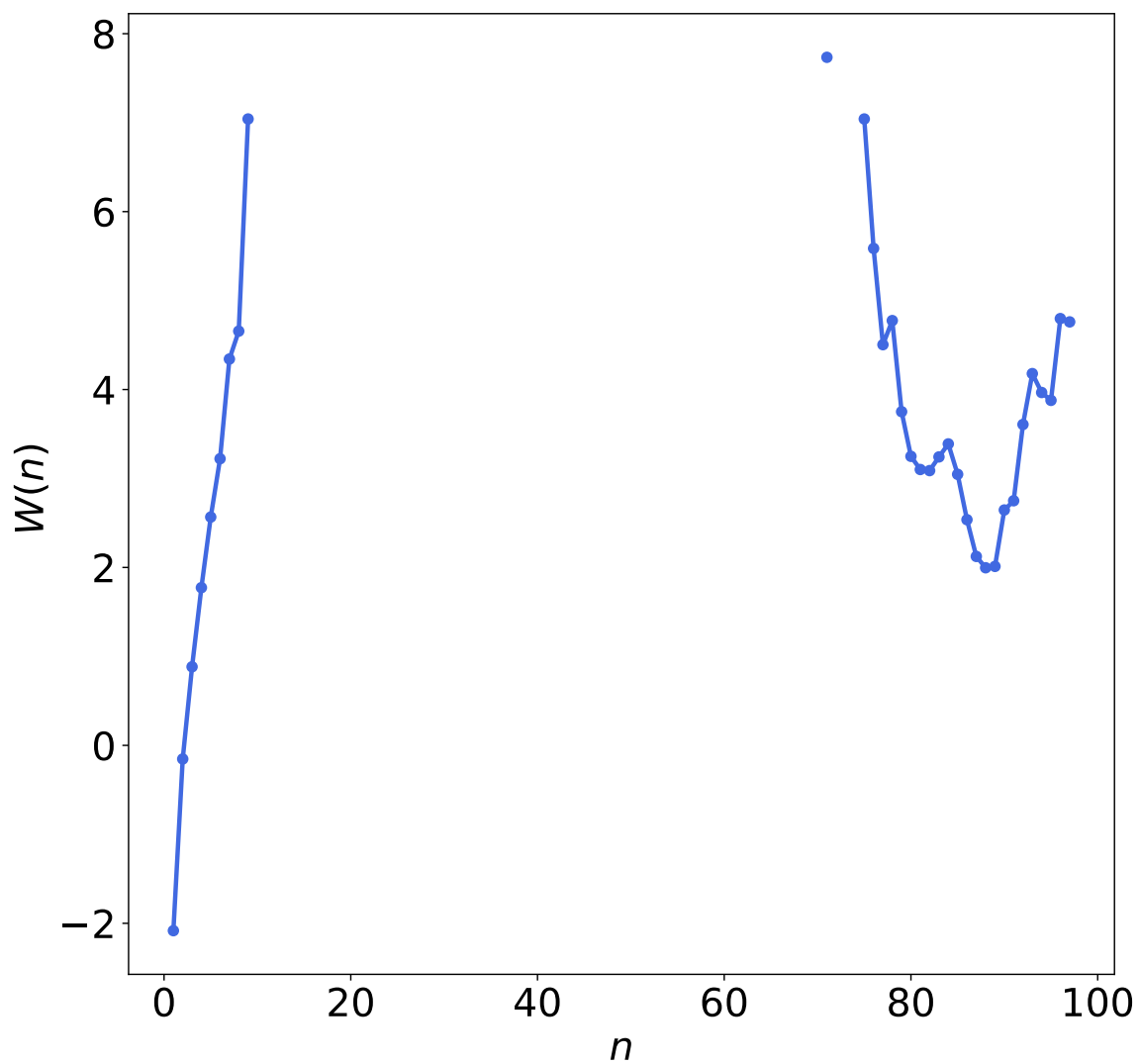


Рис. 13 Работа образования обратной мицеллы в зависимости от размера агрегата для системы 100 мономеров  $C_{12}E_4$  при температуре  $T = 283$  К.

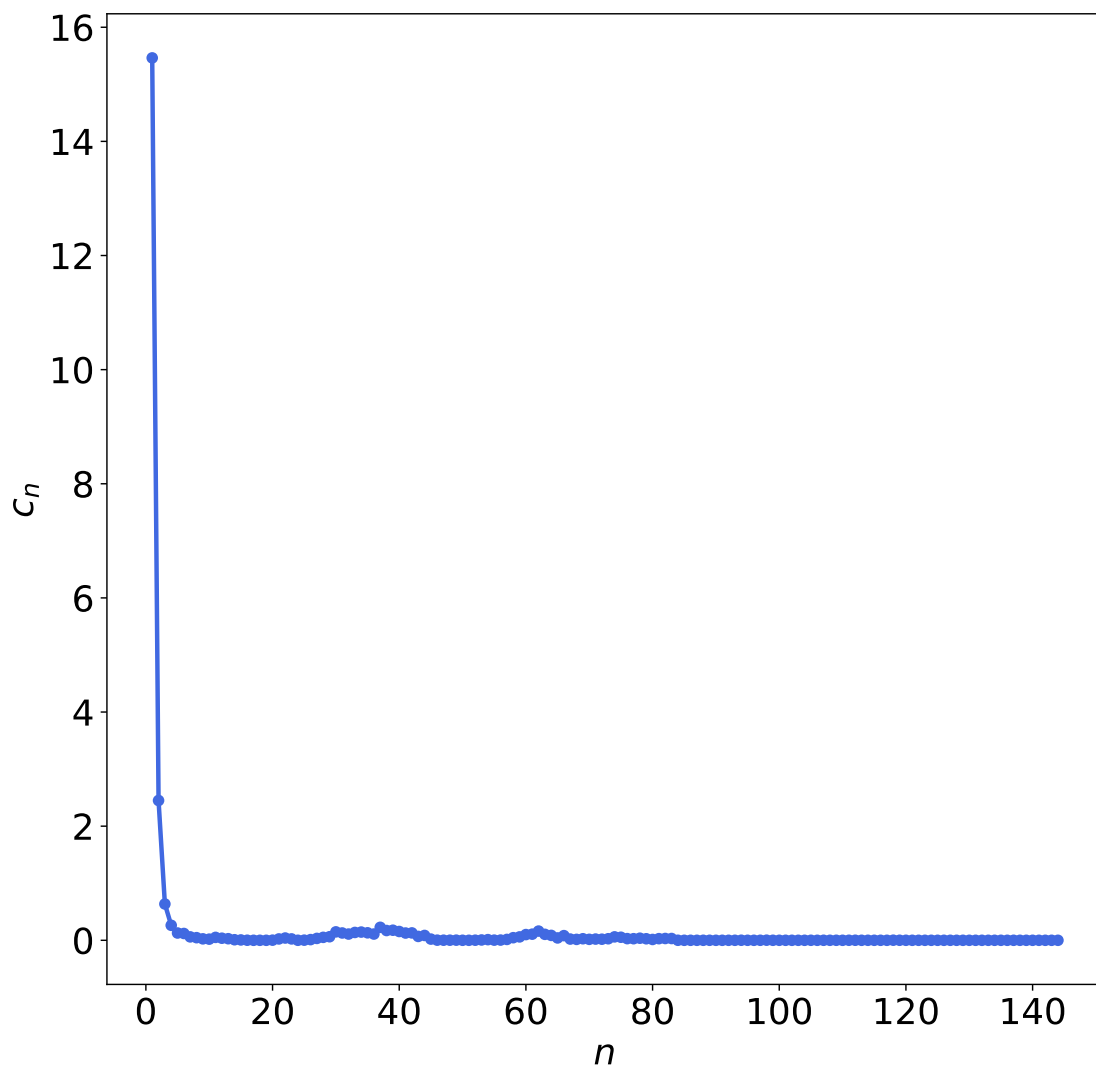


Рис. 14 Распределение по числам агрегации для системы 200 мономеров  $C_{12}E_4$  при температуре  $T = 283$  К.

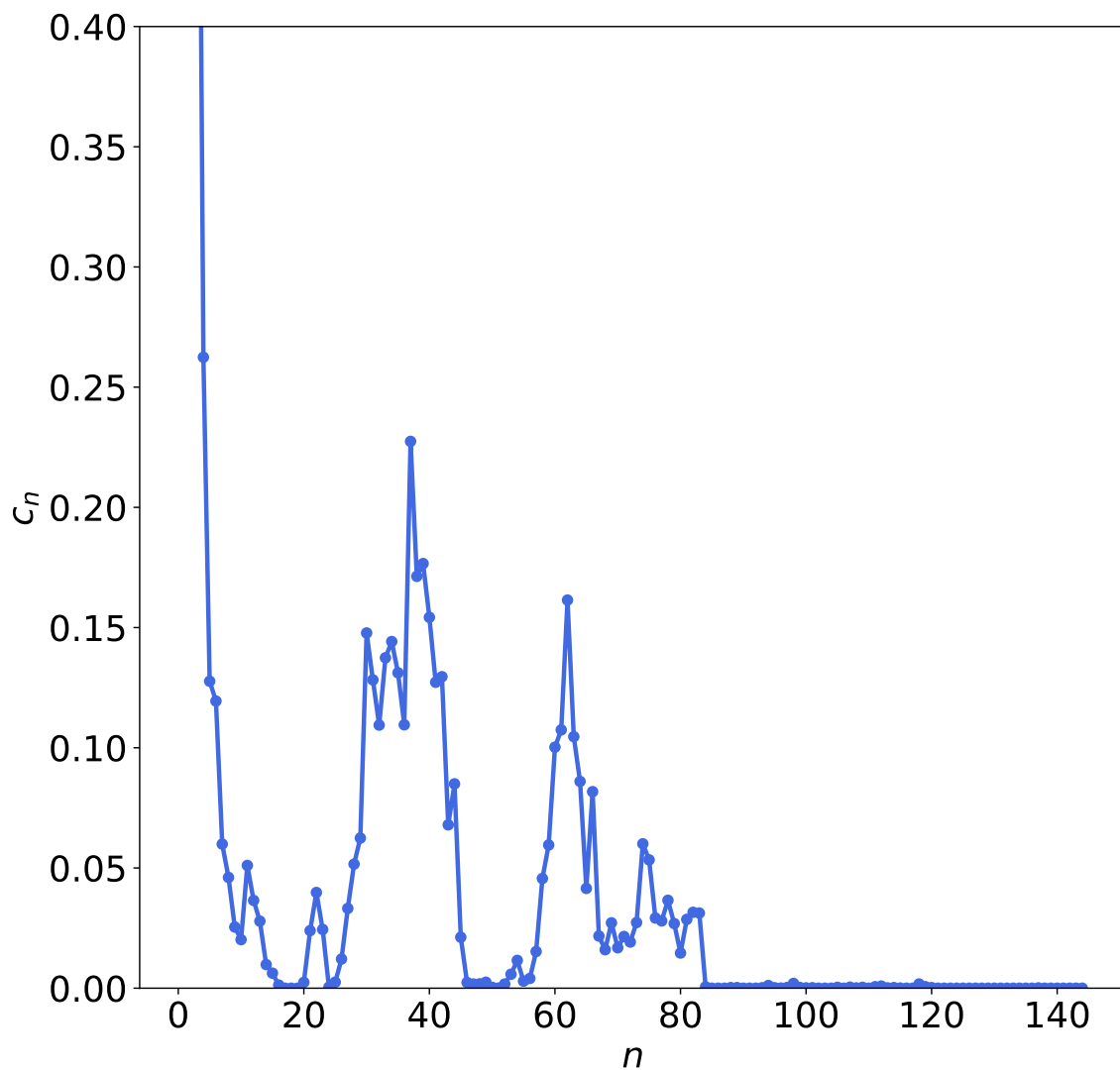


Рис. 15 Распределение по числам агрегации для системы 200 мономеров  $C_{12}E_4$  при температуре  $T = 283$  К.

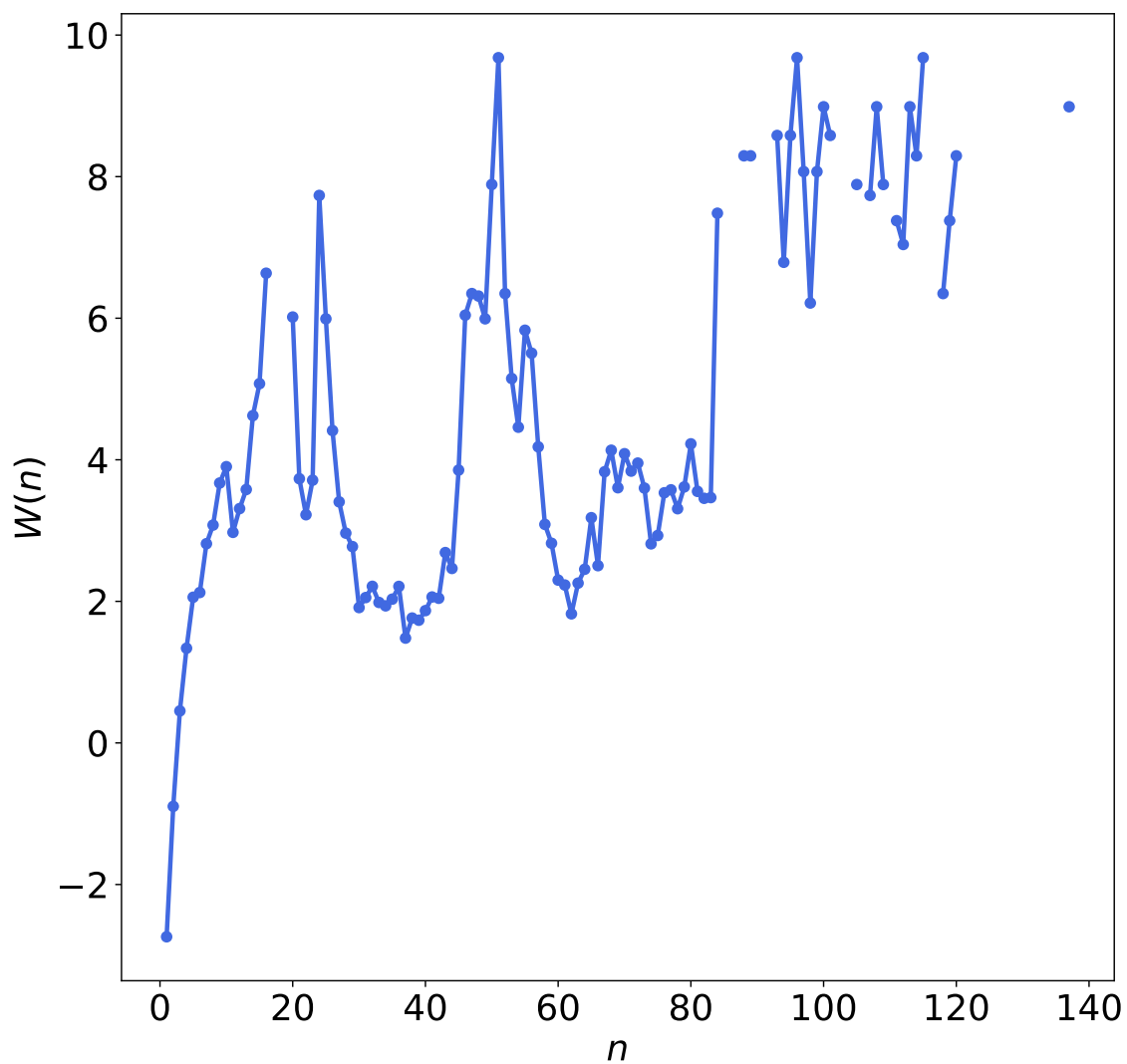


Рис. 16 Работа образования обратной мицеллы в зависимости от размера агрегата для системы 200 мономеров  $C_{12}E_4$  при температуре  $T = 283$  К.

### 3. Электростатический вклад в работу образования

В этом разделе речь пойдет про электростатический вклад в работу образования.

#### 3.1. Энергетическая выгода от мицеллообразования

##### Связь с работой образования

Обычно аналитические теории мицеллообразования выражают работу образования в виде формулы:  $W = A + B + C \dots$ , где  $A$ ,  $B$ ,  $C$  и т.д. – это различные вклады от гидрофобного эффекта, от электростатического взаимодействия и вклад от поверхностного натяжения агрегата. [9] В этом разделе мы сконцентрируемся на вычислении электростатического вклада.

##### Процедура вычисления

В данном случае рассматривалась отдельно плавающая мицелла в бесконечном пространстве в рамках полноатомной модели. Диэлектрическая проницаемость ядра и короны принималась равной 1. Вычисление электростатического вклада происходило по следующей формуле:

$$E = \frac{1}{4\pi\epsilon_0\epsilon} \frac{1}{2} \sum_{i,j} \frac{q_i q_j}{r_{ij}}, \quad (8)$$

где  $\epsilon$  – диэлектрическая постоянная среды, которую мы принимали равной единице,  $\epsilon_0$  – универсальная диэлектрической постоянной,  $r_{ij}$  – расстояние между частицами с номерами  $i$  и  $j$ ,  $q_i$  – заряды локализованные в центре частицы, взятые из модельного потенциала CGenFF v 4.4. из семейства силовых полей CHARMM36 [7].  $1/2$  в формуле возникает из-за повторений в суммировании.

Для того, чтобы вычислить электростатический вклад в работу образования, для систем с прямыми и обратными мицеллами, была реализована следующая процедура:

1. Вычислить полную электростатическую энергию
  - это означает, что нужно вычислить сумму всех попарных взаимодействий между всеми атомами входящими в систему
2. Вычислить сумму электростатических энергий попарных взаимодействий ТОЛЬКО между атомами входящими в состав отдельного мономера
  - для того, чтобы понять, что это означает обратимся к рисунку 17.
  - на ней схематично изображен агрегат, состоящий из двух молекул разного цвета, каждая из которых состоит из двух атомов.
  - вычислить полную энергию означает посчитать ВСЕ попарные взаимодействия. То есть: (1,2), (1,3), (1,4), (2,3), (2,4), (3,4).

- вычислить энергию мономеров означает посчитать: (1,2) и (3,4).
- 3. Далее нужно вычислить разницу между полной энергией и энергией мономеров
  - это означает, что останется энергия, которая появляется только из-за взаимодействия между разными мономерами в составе мицеллы.
  - То есть остаются только: (1,3), (1,4), (2,3), (2,4).

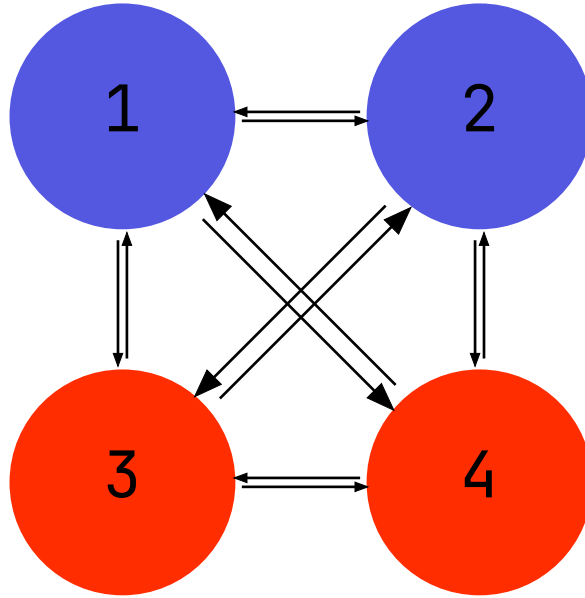


Рис. 17 Иллюстрация взаимодействий между частицами. Разными цветами показаны разные мономеры.

Формулой эту мысль можно выразить так:

$$E_b = E_{full} - E_{mono}, \quad (9)$$

где  $E_{full}$  – это полная энергия, а  $E_{mono}$  – энергия всех мономеров.

### Какой это имеет физический смысл?

Строго говоря, электростатический вклад в работу образования нужно вычислять так: 1) создать ячейку моделирования, в которой все мономеры плавают отдельно, 2) проводить моделирование до тех пор, пока все мономеры не соберутся в одну мицеллу, 3) вычислять  $\Delta E$  разницу между энергиями двух состояний, когда все мономеры собрались в мицеллу, и когда они плавают отдельно (поэтому электростатический вклад можно определить, как некую энергетическую выгоду от слияния в единый агрегат.  $E_b$  – от слова benefit). Однако мы вычисляем разницу между энергиями для уже сформированной мицеллы, что эквивалентно ситуации, в которой мономеры плавали бы отдельно, но были изогнуты точно в такую же форму, как если бы они находились в мицелле. Это обстоятельство нужно учитывать, поскольку мы искусственно вносим погрешность в наши расчеты, чтобы получать результаты быстрее. Иначе это требует больших вычислительных мощностей. Спасает это положение тот факт, что энергия свободно плавающего мономера в растворе и энергия мономера уже в составе мицеллы почти не отличаются (см. Таблицу

1). Даже несмотря на то, что мономер меняет свою форму (обычно мономеры в составе мицеллы сильно изгибаются).

### 3.2. Результаты вычисления энергетической выгоды

В Таблице 1 представлены результаты вычислений для системы состоящей из  $n$  молекул  $C_{12}E_4$  и 4000 молекул гептана  $C_7H_{16}$ .

ГЕПТАН   $n$	1	2	4	16
Полная энергия	$-551 \pm 5$	$-1111 \pm 7$	$-2240 \pm 10$	$-9000 \pm 17$
Энергия мономера		$-547 \pm 5$	$-546 \pm 6$	$-542 \pm 6$
Энергетическая выгода		$-17 \pm 6$	$-56 \pm 12$	$-321 \pm 22$
Выгода на один мономер		$-8.5 \pm 3$	$-14 \pm 3$	$-20 \pm 1.4$
$n$ (размер)		25	28	32
Полная энергия		$-14092 \pm 22$	$-15739 \pm 23$	$-18010 \pm 27$
Энергия мономера		$-541 \pm 6.58$	$-543 \pm 6$	$-544 \pm 7$
Энергетическая выгода		$-563 \pm 24$	$-516 \pm 23$	$-599 \pm 25$
Выгода на один мономер		$-22.5 \pm 0.96$	$-18.4 \pm 0.8$	$-18.7 \pm 0.8$

Таблица 1: Таблица значений электростатической энергии для  $C_{12}E_4$  в растворе  $C_7H_{16}$  (гептан). Все значения представлены в единицах  $kT$ .

Аналогичная процедура была проделана для прямых мицелл, состоящих из того же вещества ( $C_{12}E_4$ ), находящихся в воде (12500 молекул). Здесь была выбрана более высокая температура  $T = 298$  К, потому что в полярном растворителе мицелла оказывалась намного более стабильной. Результаты представлены в Таблице 2.

ВОДА   $n$	1	2	4	25
Полная энергия	$-395 \pm 5$	$-792 \pm 8$	$-1589 \pm 12$	$-9992 \pm 30$
Энергия мономера		$-395 \pm 6$	$-396 \pm 6$	$-399 \pm 6$
Энергетическая выгода		$-0.5 \pm 1.4$	$-2.0 \pm 2.8$	$-46.4 \pm 14.2$
Выгода на один мономер		$-0.25 \pm 0.7$	$-0.5 \pm 0.4$	$-1.8 \pm 0.5$

Таблица 2: Таблица значений электростатической энергии для  $C_{12}E_4$  в полярном растворителе  $H_2O$  (в воде). Все значения представлены в единицах  $kT$ .

#### Что ожидалось получить?

В таблицах 1, 2 следует обратить внимание стоит на: 1) если посмотреть на строчку "полная энергия" в первом столбце (для мономера) и на строчку средняя "энергия мономера" в составе агрегата, то можно увидеть, что отдельно плавающий мономер не сильно отличается от мономера в составе агрегата, хотя, конечно, некоторые различия присутствуют. Это касается обеих таблиц 1, 2. 2) Важно посмотреть на строчки "Выгода на один мономер" главное в них – это знак. В аналитических теориях до этого считалось, что электростатический вклад является небольшим и положительным [13]. Оказалось, что

он большой и отрицательный в случае обратных сухих мицелл (*сухие означает в отсутствие воды*). И небольшой, но тоже отрицательный в случае прямых мицелл. Основной причиной мицеллообразования в воде является гидрофобный эффект. А вот с обратными ситуация не была такой очевидной. Судя по источникам [14], единственным отрицательным вкладом (из известных) в образование обратных сухих мицелл является электростатический вклад. Понимание того, что основной причиной образования сухих обратных мицелл является электростатика, является основным результатом этого раздела.

Долгое время оставался открытым вопрос о возможности существования сухих обратных мицелл. Результаты этой ВКР подтверждают возможность их существования.

На графиках 18, 19, 20, 21, 22, 23, 24, 25 представлены результаты из таблиц 1, 2.

Отметим, что линейность в функции энергетической выгоды может нарушаться.

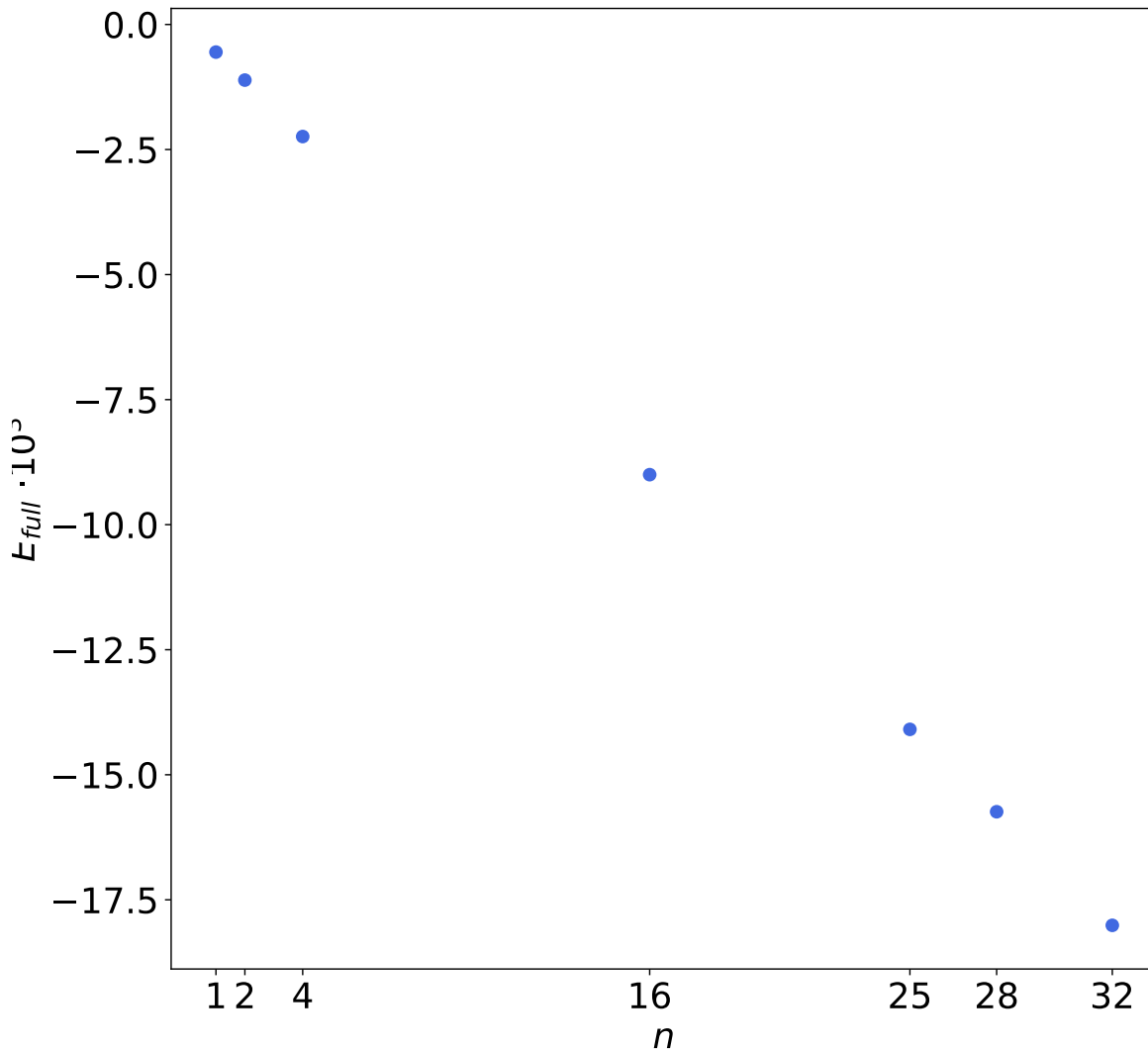


Рис. 18 Полная электростатическая энергия для агрегатов в растворе гептана.



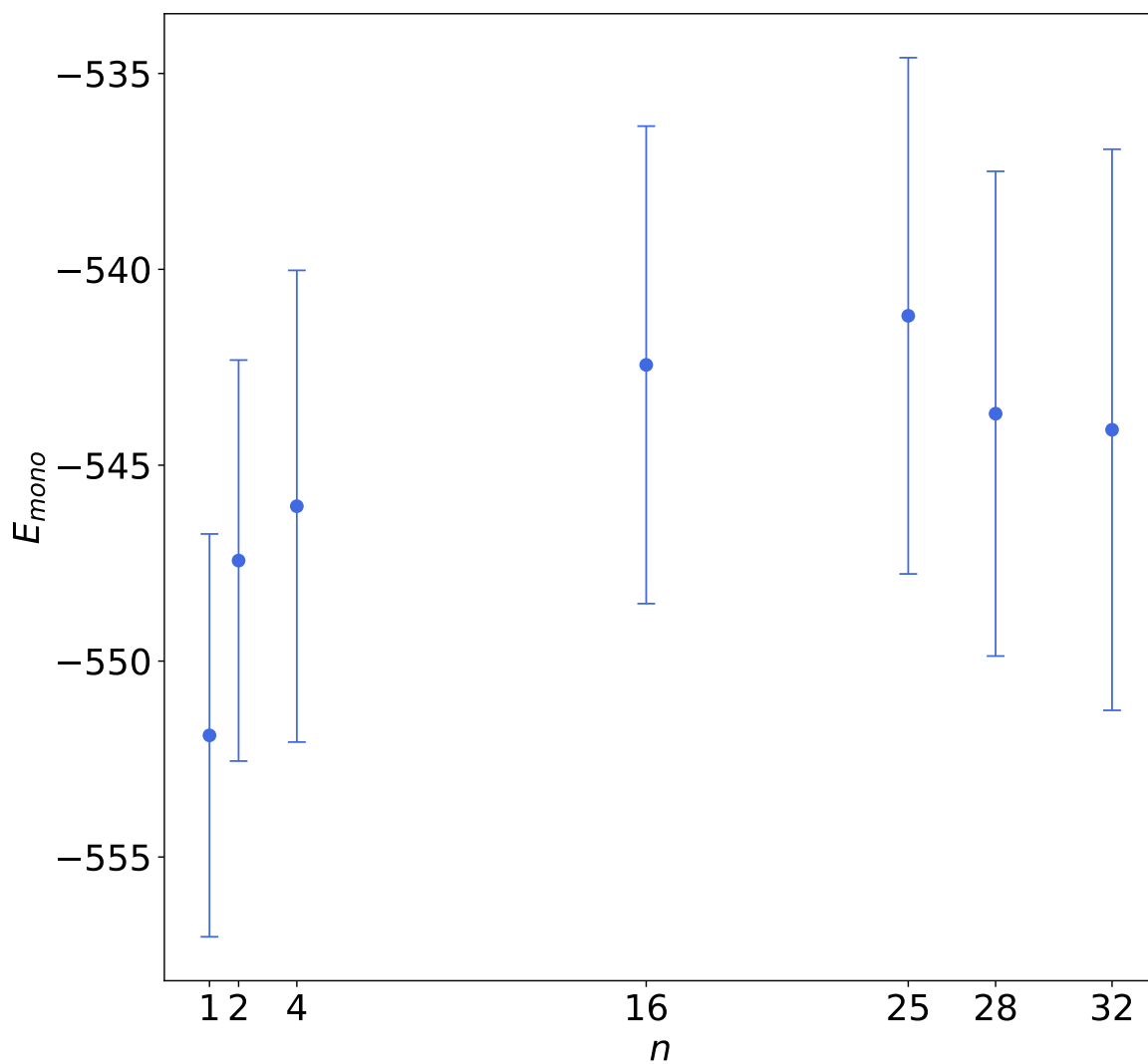


Рис. 19 Средняя электростатическая энергия для мономеров в составе агрегатов в растворе гептана.

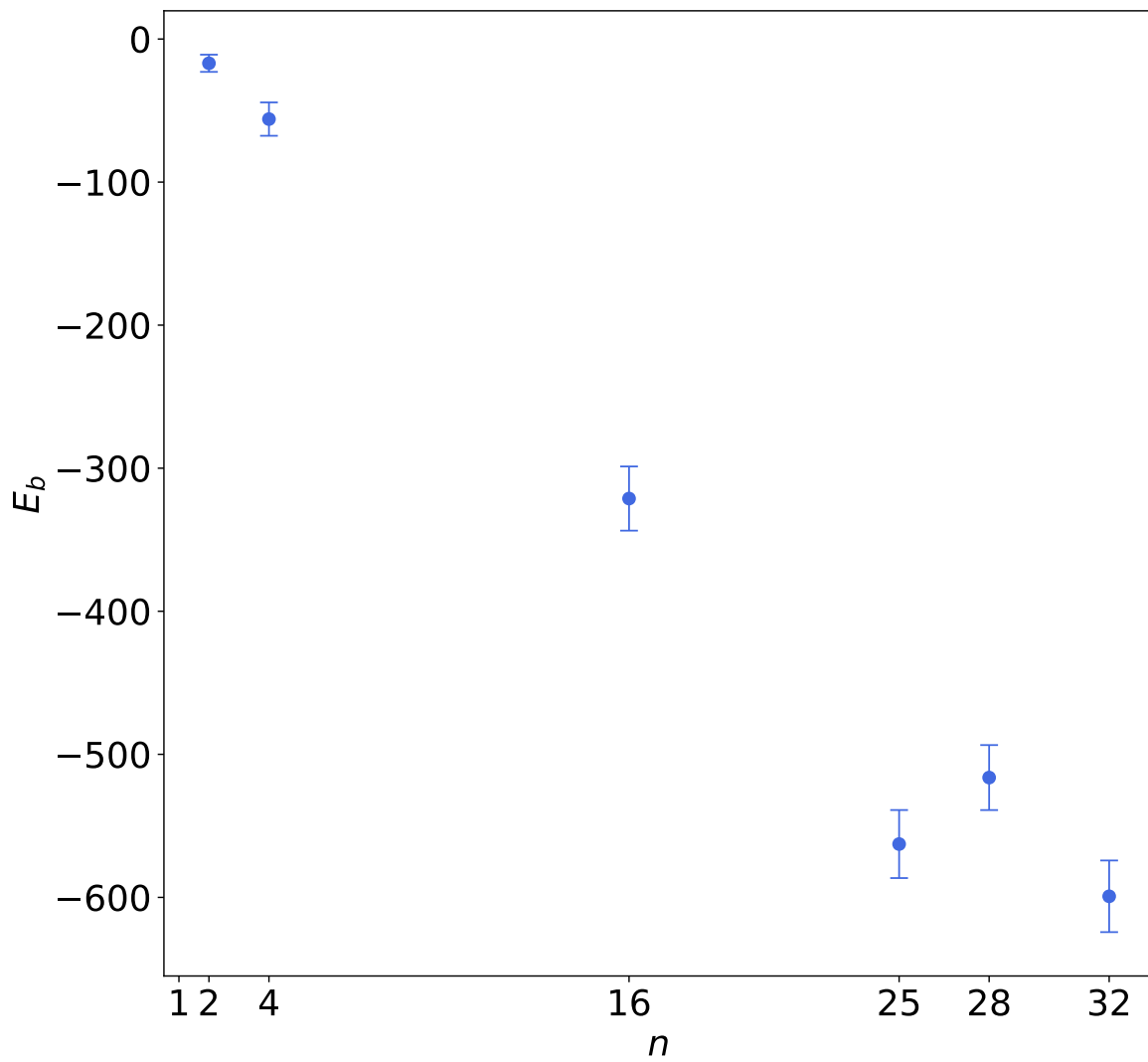


Рис. 20 Энергетическая выгода от образования агрегатов в растворе гептана.

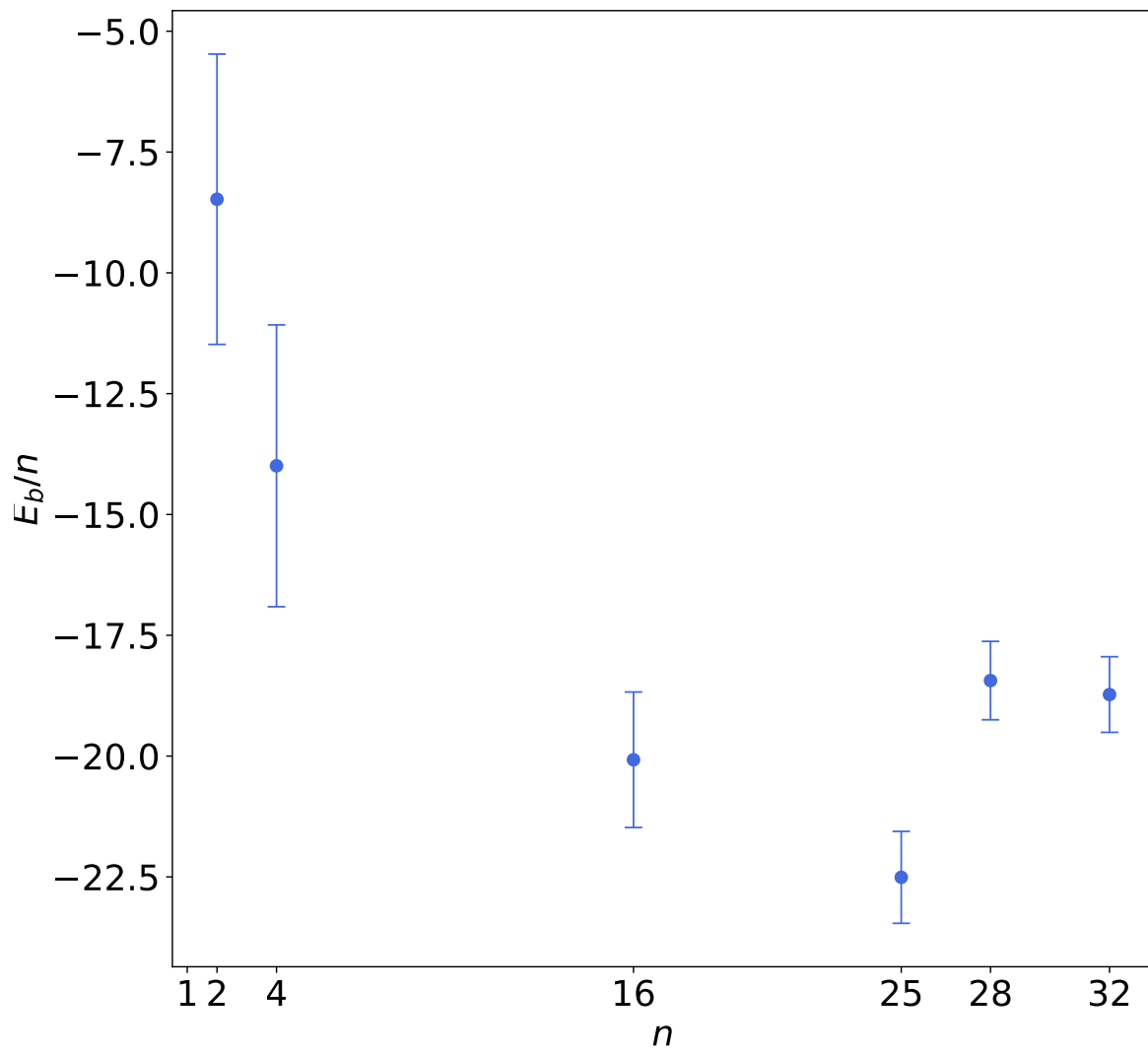


Рис. 21 Энергетическая выгода от образования агрегатов на один мономер в растворе гептана.

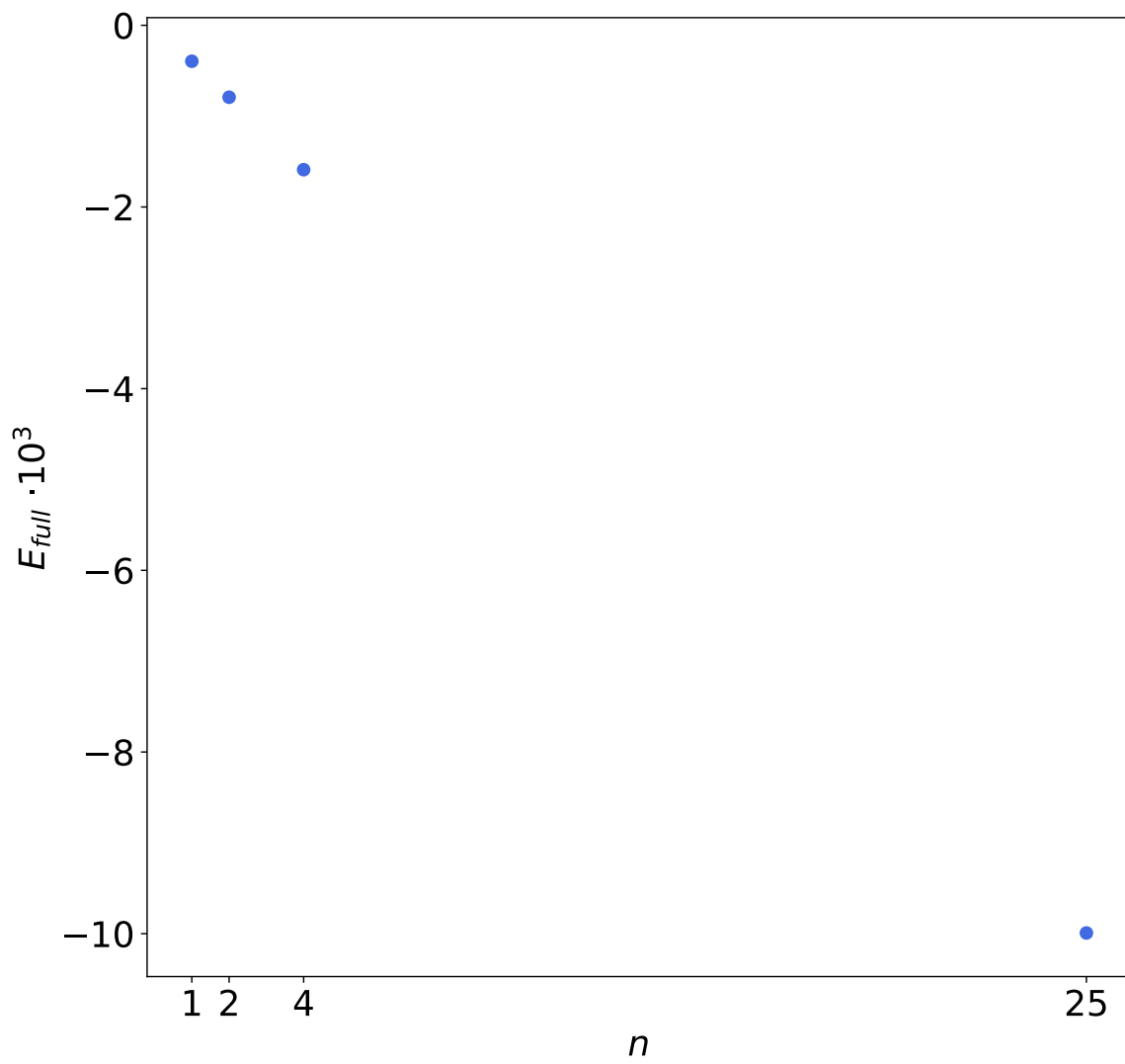


Рис. 22 Полная электростатическая энергия для агрегатов в воде.

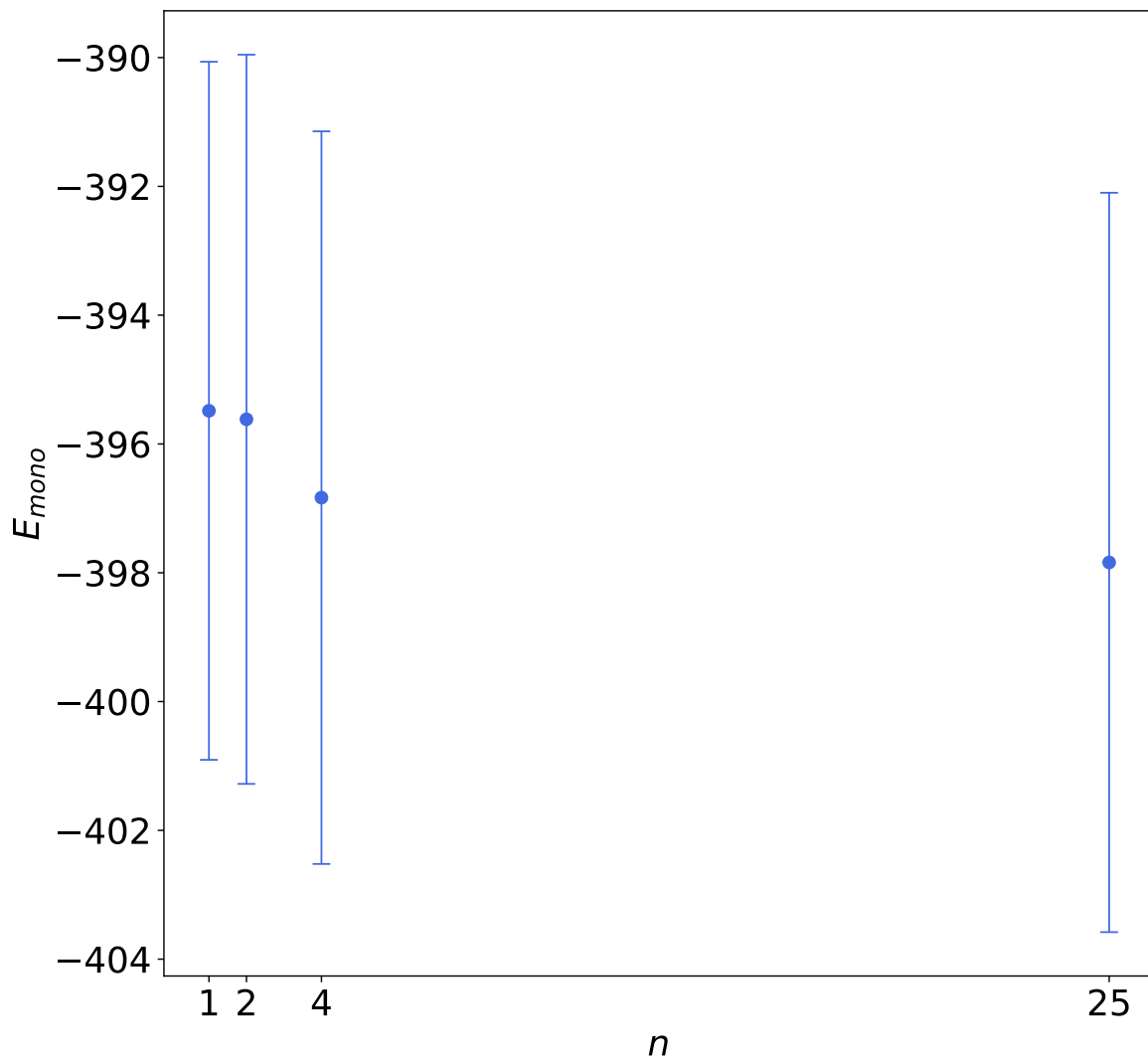


Рис. 23 Средняя электростатическая энергия для мономеров в составе агрегатов в воде.

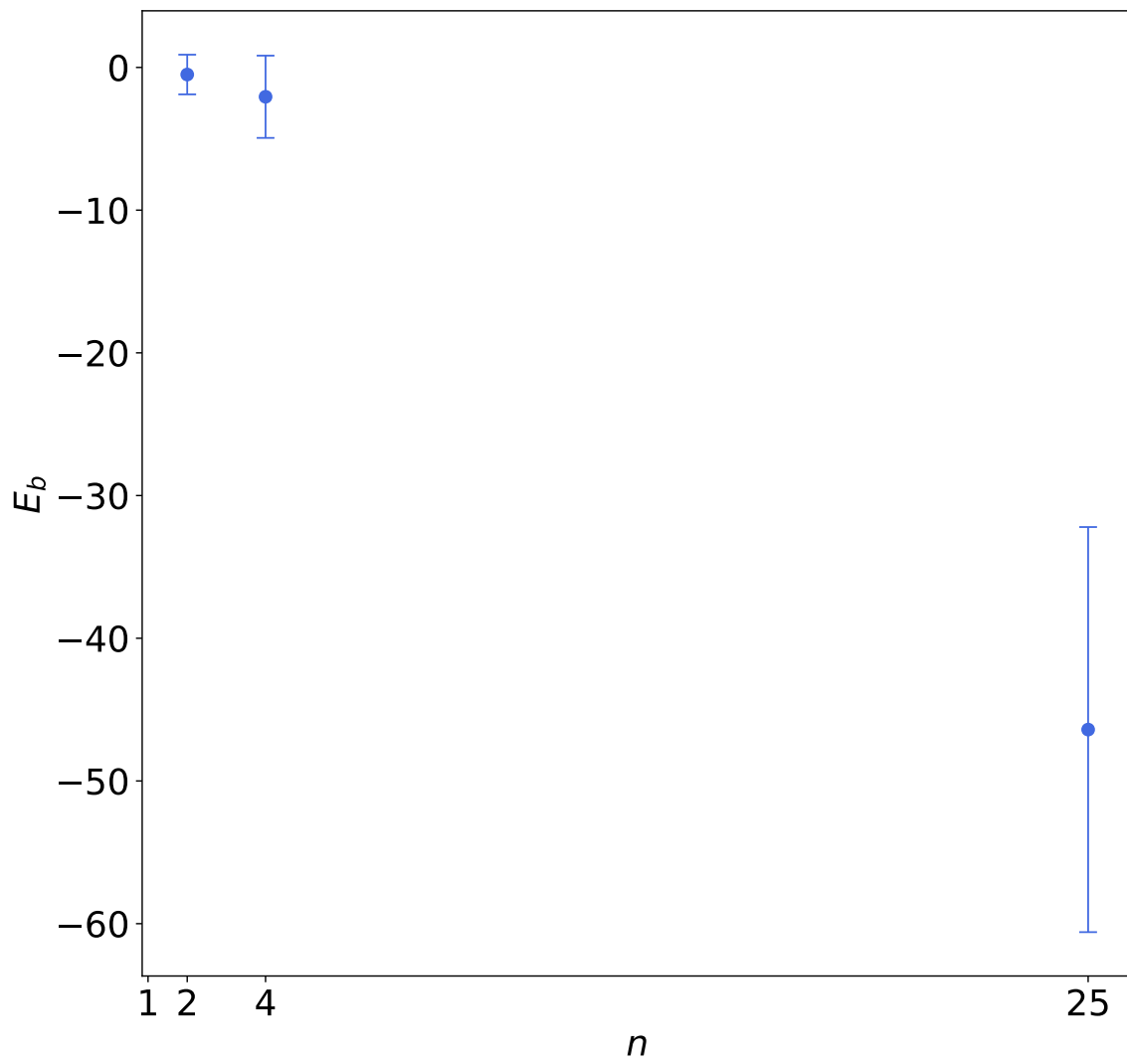


Рис. 24 Энергетическая выгода от образования агрегатов в воде.

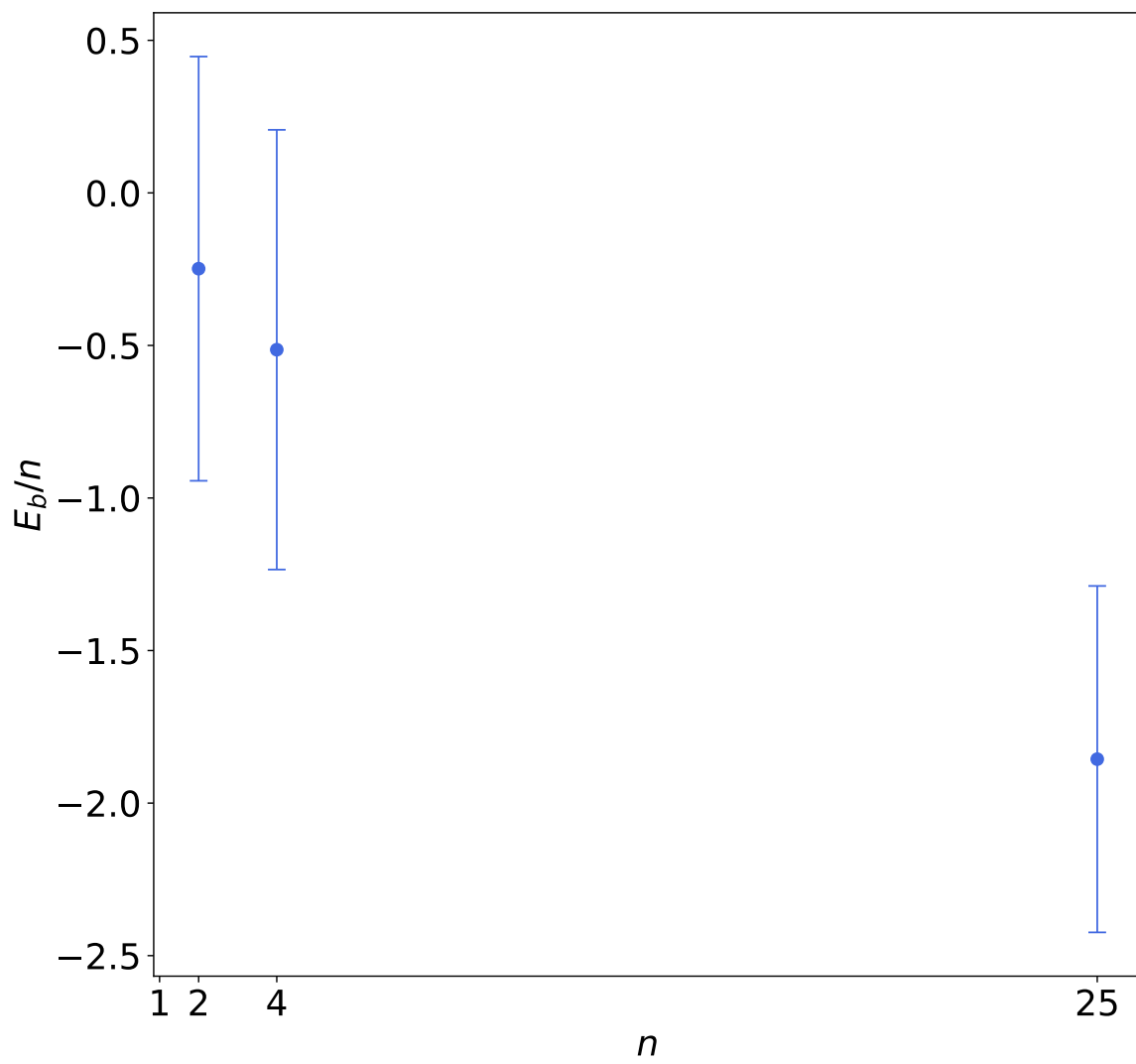


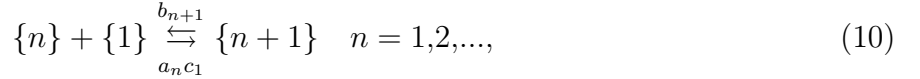
Рис. 25 Энергетическая выгода от образования агрегатов на один мономер в воде.

## 4. Коэффициенты присоединения и отсоединения

### 4.1. Диффузионное приближение

#### Уравнение Беккера-Дёринга

В аналитических теориях кинетики мицеллообразования возникают уравнения, которые моделируют динамику коагуляции и фрагментации кластеров идентичных частиц. Например, уравнения Беккера-Дёринга (Becker-Döring Equations) [15] [8] описывают следующий механизм присоединения и отсоединения мономеров:



где величина  $a_n c_1$  равна количеству мономеров поглощаемых агрегатом с размером  $n$  в единицу времени. Величина  $b_{n+1}$  соответствует числу мономеров испускаемых агрегатом с размером  $n+1$  в раствор в единицу времени.

Уравнения Беккера-Дёринга имеют следующий вид [9]:

$$\frac{\partial c_n}{\partial t} = -(J_n - J_{n-1}) \quad n = 2, 3, \dots, \quad (11)$$

$$\frac{\partial c_1}{\partial t} = -J_1 - \sum_{k=1}^{\infty} J_k, \quad (12)$$

где величины  $J_n$  – это потоки, определяются они так:

$$J_n = a_n c_1 c_n - b_{n+1} c_{n+1}. \quad (13)$$

Уравнения Беккера-Дёринга работают в условиях так называемого диффузионного приближения, когда считаются возможными присоединения и отсоединения только с мономерами. Это означает, что мицелла размером допустим 80 может соединиться с мономером и стать агрегатом с размером 81, но не может соединиться с димером. Такие переходы описываются через последовательное присоединение двух мономеров (т.е.  $80 + 2 = 80 + 1 + 1$ ). Ещё это приближение подразумевает, что любой агрегат имеет сферическую форму.

**Соглашение.** В тексте много раз встретится словосочетание "присоединения и отсоединения", поэтому далее мы будем использовать аббревиатуру "ПНО".

Давайте взглянем на уравнение (11) поподробней и мысленно подставим туда уравнение (13), получается уравнение, которое описывает взаимосвязь между скоростями ПНО  $a_n$  и  $b_n$ , и концентрациями агрегатов  $c_n$ , которые по сути своей являются распределениями из раздела (2). Для полной картины не хватает  $a_n$  и  $b_n$ .

Задача с точки зрения алгоритмов непростая. Если мы хотим получить все коэффициенты, тогда нам необходимо знать полную картину, то есть знать в каждый момент



времени размеры всех находящихся в системе кластеров. Также у каждого кластера в системе должен быть свой уникальный `id`, чтобы мы могли отследить изменения в системе: какой именно агрегат и какого размера к какому конкретно агрегату присоединился. В работе [16] проводились расчеты коэффициентов ПИО для более простой модели. В этой работе мы вычисляем часть коэффициентов, которые относятся к области с номерами  $n \approx 80-90$ , для системы состоящей из 100 мономеров C12E4 и 4000 молекул гептана при температуре  $T = 283$  К.

## 4.2. Процедура вычисления коэффициентов присоединения и отсоединения

Стандартная утилита, встроенная в GROMACS с названием `gmx clustsize`, позволяет получать информацию о размере самого большого агрегата в системе с течением времени. Назовём этот временной ряд функцией  $\text{size}(t)$ . Значит, вычисляя разность между значениями этой функции в соседние моменты времени ( $\text{size}(t) - \text{size}(t-1)$ ), можно вычислить, когда и на какую величину случались изменения размера мицеллы. Следовательно, можно посчитать количество ПИО, которые случались с мицеллой.

Сравнить эти численные результаты можно с формулой, которая выводится в рамках диффузионного приближения [17] [18]:

$$A_n = \frac{2}{3} \frac{kT}{\eta} \frac{(1 + n^{1/3})^2}{n^{1/3}} c_1, \quad (14)$$

где  $A_n$  – это количество мономеров, которые присоединяются к мицелле с размером  $n$  в единицу времени (так как система находится в равновесии, количество отсоединений считается таким же),  $c_1$  – это концентрация свободно плавающих мономеров,  $\eta$  – вязкость раствора.

Можно сказать, что  $A_n$  из (14) при некотором  $n$  равно  $a_n$  из (13), однако нужно быть осторожными. Для того, чтобы набрать достаточно хорошую статистику, нам приходится усреднять по отрезкам траектории. Поэтому мы вычисляем на самом деле величину  $a_{\langle n \rangle}$  и сравнивать с результатом формулы (14) от среднего размера  $A_{\langle n \rangle}$ .

## 4.3. Предварительные результаты

Для этих вычислений была взята самая длинная молекулярно-динамическая траектория из имеющихся у нас. 1 микросекунда системы состоящей из 100 мономеров C12E4 (это соответствует 20 файлам по 50 нс, 5 первых файлов из которых мы отбросили, так как система тогда находилась не в равновесном состоянии). С помощью встроенной утилиты `gmx clustsize` были получены файлы с названиями `maxclust`. Это текстовые файлы, в которых в первой колонке указан момент времени, во второй колонке размер агрегата. Если

вычислить разницу между соседними значениями ( $\text{size}(t) - \text{size}(t - 1)$ ), получится массив с числами примерно от  $-7$  до  $+7$ . После этого необходимо посчитать, сколько раз встречается каждое уникальное значение в этом массиве. Нули будут соответствовать ситуациям, когда размер мицеллы не менялся.

Проделаем эту процедуру. В таблице 3 представлены результаты вычислений для одного из кусков траектории. (Результаты представленные в этой таблице представлены ДО очистки от шума, поэтому результаты имеют большую погрешность).

Изменение $n$	Количество (за 50 нс)	Количество в ед. времени
-6	0	0
-5	27	$54.0 \cdot 10^7$
-4	64	$128.0 \cdot 10^7$
-3	3	$6.0 \cdot 10^7$
-2	86	$172.0 \cdot 10^7$
-1	1294	$2588.0 \cdot 10^7$
0	97043	$194086.0 \cdot 10^7$
1	1316	$2632.0 \cdot 10^7$
2	73	$146.0 \cdot 10^7$
3	3	$6.0 \cdot 10^7$
4	64	$128.0 \cdot 10^7$
5	26	$52.0 \cdot 10^7$
6	1	$2.0 \cdot 10^7$

Таблица 3: Количество изменений размера мицеллы на участке траектории длиной 50 нс. 0 – соответствует моментам времени, когда размер не изменялся, 2 – соответствует присоединению агрегата размером 2 (димера),  $-4$  – отсоединение тетрамера и т.д.

Сравним результаты, представленные в таблице 3, с результатами, которые дает формула выведенная в рамках диффузионного приближения (14). Средний размер мицеллы на протяжении участка траектории равнялся  $\langle n \rangle = 88.663813$ , концентрацию свободно плавающих мономеров  $c_1$  можно оценить как  $(100 - \langle n \rangle)/V = 1.1336187 \cdot 10^{25}$ , вязкость  $\eta = 0.466 \cdot 10^{-3}$  мПа·с при температуре  $T = 283$  К [19],  $V = (100 \cdot 10^{-10})^3$  м<sup>3</sup>. Подставим эти значения в формулу (14) и получим:  $A_{\langle n \rangle} \approx 42.35 \cdot 10^7$ . Теперь это нужно сравнить с результатами моделирования. Существует два варианта: сравнить с количеством присоединенных мономеров  $2632.0 \cdot 10^7$ , тогда получается **расхождение в 62 раза** (результат моделирования больше), либо учесть тот факт, что в диффузионном приближении присоединение димера считается за последовательное присоединение двух мономеров. Значит, надо сравнивать с суммой всех присоединений, умноженных на величину изменения  $3726 \cdot 10^7$ , в этом случае **расхождение в 88 раз**. (На других кусочках траектории наблюдаются расхождения от 58 до 178 раз).

Чем вызвано такое сильное расхождение? Давайте взглянем на графики временных рядов в разных масштабах. Ниже представлены 4 графика одного и того же временного ряда. Первый график в полном размере, а каждый следующий это увеличенный предыдущий.

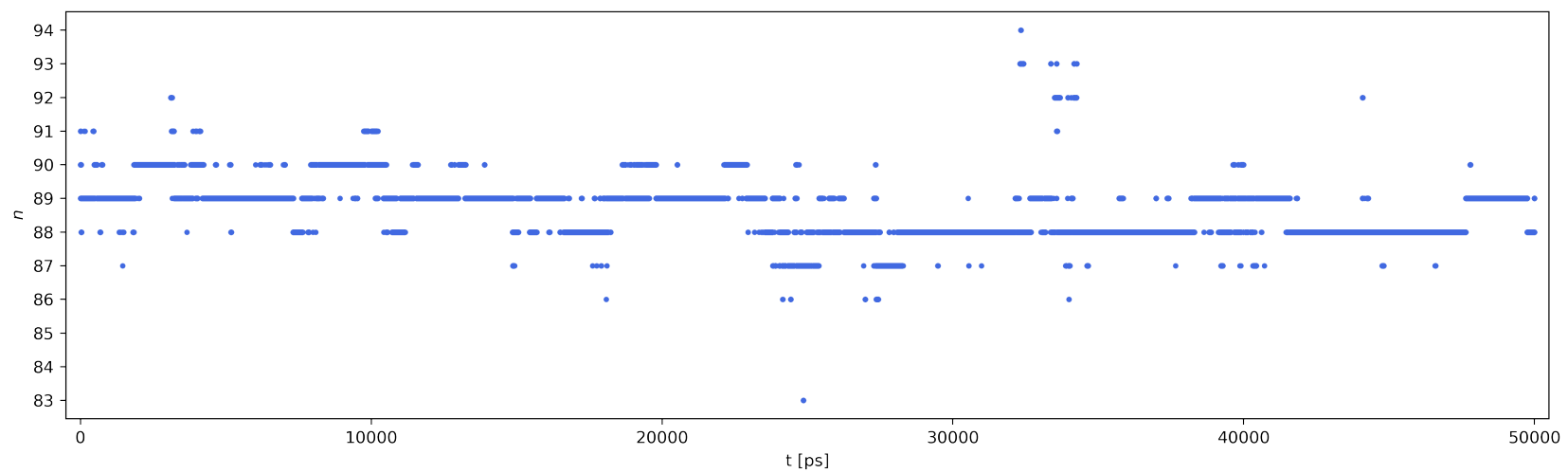


Рис. 26 Временной ряд для размера мицеллы в системе состоящей из 100 мономеров.

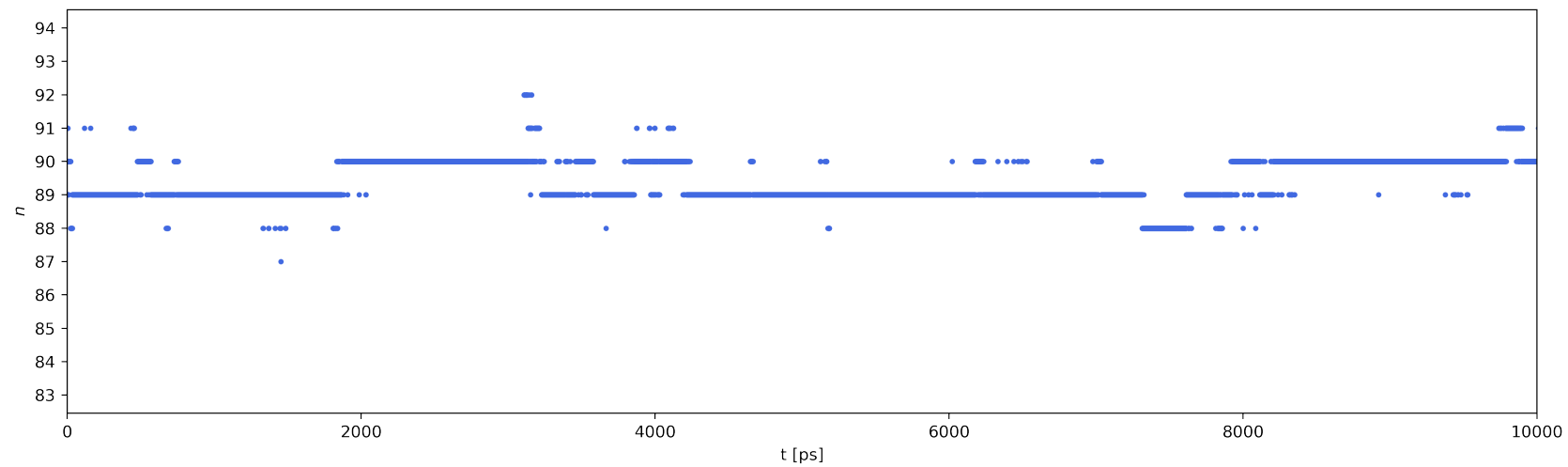


Рис. 27 Временной ряд для размера мицеллы в системе состоящей из 100 мономеров.

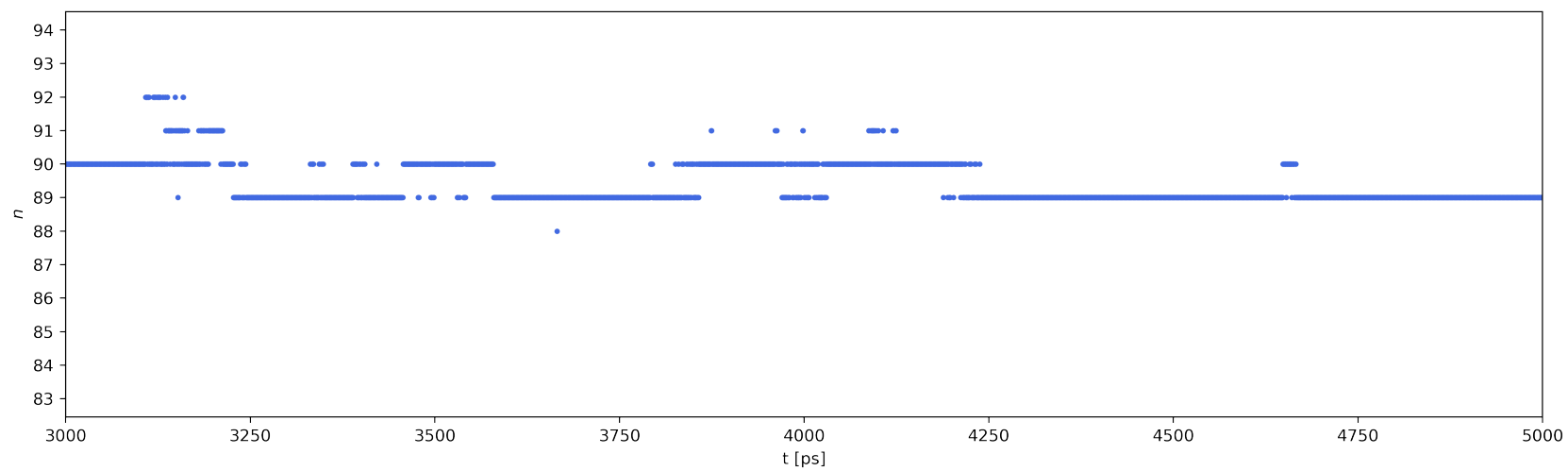


Рис. 28 Временной ряд для размера мицеллы в системе состоящей из 100 мономеров.

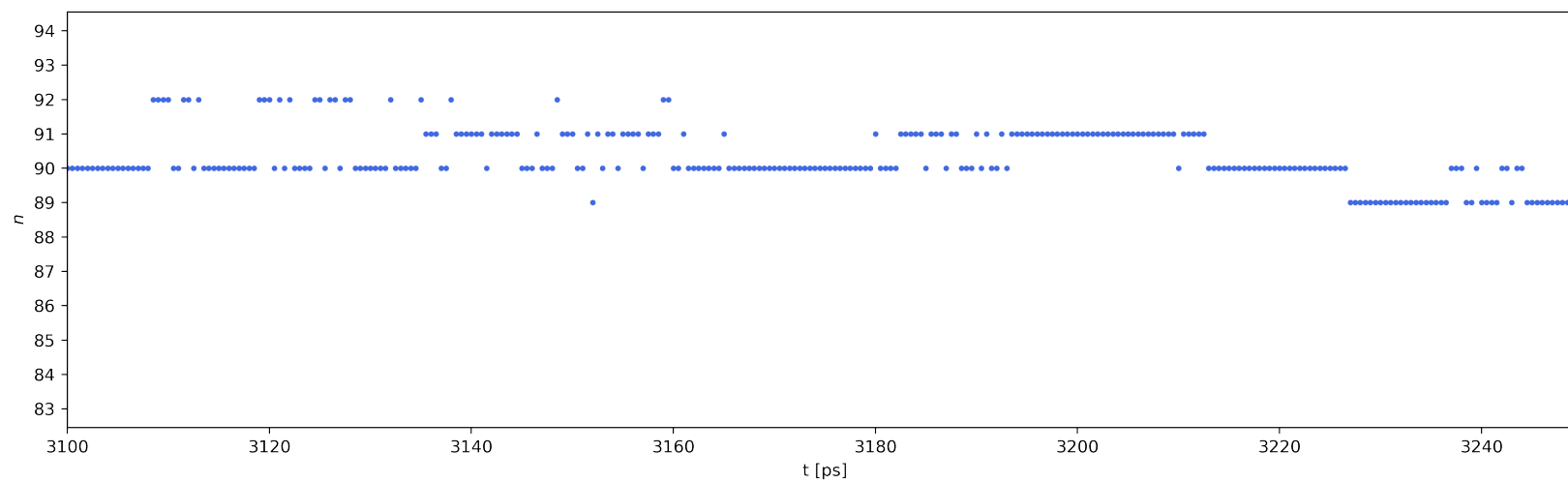


Рис. 29 Временной ряд для размера мицеллы в системе состоящей из 100 мономеров.

На рисунке 26 у линий, который образуются из-за последовательного прибывания в одном состоянии (с одинаковым размером  $n$ ), наблюдаются пересечения. Чего быть не должно. Из-за чего возникают эти "пересечения"? Если взглянуть на график 29, то можно увидеть, что никаких пересечений не бывает, зато явно наблюдаются частые переходы из одного состояния в другое. Насколько эти частые переходы реалистичны? Утверждается, что такое поведение временных рядов совершенно не физично! Давайте разбираться.

#### 4.4. Проблемы алгоритма кластеризации

Кто виноват?

---

А. И. Герцен

Основное утверждение заключается в том, что каждое состояние должно быть хоть сколько-нибудь стабильным (как минимум 100 пикосекунд). Слишком частые ПИО невозможны, но на этих графиках они возникают как артефакт работы алгоритма кластеризации.

Дело в том, что встроенный в стандартную утилиту GROMACS `gmx clustsize` алгоритм `linkage` пользуется единственным критерием включения в кластер. Это расстояние  $r_{cut}$ . А теперь представим такую ситуацию: к поверхности мицеллы приблизился мономер, но не вошел в структуру мицеллы. Алгоритм, вероятно, распознает его как часть кластера с мицеллой. После этого, в следующий момент времени мономер отплыл, алгоритм уже не включает его в состав мицеллы. После нескольких повторений мономер всё-таки встраивается в структуру мицеллы, и теперь уже всё в порядке. Такая ситуация не столько портит сами конфигурации (поэтому Раздел 2 можно считать корректным), сколько портит количество переходов между ними.

Можно представить себе ещё множество различных ситуаций, когда какие-то изменения, подплывания, отплывания могли вносить помехи во временные ряды, создавая нефизичные переходы, завышая их количество. При этом сам алгоритм кластеризации, основанный на радиусе обрезания, выглядит как самое очевидное решение этой задачи. Однако он недостаточно хорошо работает с вытянутыми молекулами, которые могут лишь касаться одним из своих атомов мицеллы, но не встраиваться в её структуру. Наверное, этот алгоритм не так уж и плох для упрощенных моделей, которые состоят из сферических частиц, но, увы, для длинных молекул не годится.

Автор хочет построить гипотезу о том, как должен в идеале выглядеть алгоритм, который выполняет кластеризацию правильно. Вместо того, чтобы вокруг каждого атома каждой молекулы "рисовать" виртуальную сферу, а после искать пересечения этих сфер, и принимать решение только из-за наличия этого пересечения, необходимо вокруг каждой молекулы описывать объем, который равен объединению сфер описанных вокруг центров каждого из атомов входящих в её состав. После чего искать пересечения между такими объемами и вычислять их величину, а после принимать решение о включении в один агрегат

только если, эти объемы пересеклись на 50-70%. Предполагаю, что такой алгоритм вносил бы гораздо меньше погрешностей во временные ряды.

#### 4.5. Результаты полученные вручную

Что делать?

---

Н. Г. Чернышевский

Наиболее точным средством вычисления количества переходов являются визуальный анализ. Для этого вся траектория была разбита на кусочки по 2000 пикосекунд для удобства восприятия, и каждый график обрабатывался вручную. Мысленно убирались всплески, кривая спрямлялась участками длиной, как минимум, 100 пикосекунд. Сравнение результатов этой процедуры со значением полученным по формуле (14) от среднего  $\langle n \rangle$  представлено в таблице 4.

Главными колонками в таблице 4 являются "Соотношение<sup>1</sup>" и "Соотношение<sup>2</sup>". Они показывают, во сколько раз результат, полученные путем численного моделирования, расходится с результатом формулы (14). Количество присоединений и отсоединений в единицу времени в диффузионном приближении считается одинаковым, поскольку система находится в равновесии. По результатам численного моделирования это не совсем так, поэтому существуют две разные колонки для присоединений и отсоединений отдельно. Количество присоединений и отсоединений считалось как  $\sum_i i \cdot \Delta n_i$ , иными словами, присоединение димера учитывалось как последовательных присоединения мономера. Разные участки оказываются неоднородными, результаты разнятся, но в среднем они таковы: **присоединений в  $(1.88 \pm 0.56)$  больше, отсоединений в  $(1.82 \pm 0.64)$  больше, чем в диффузионном приближении.** Согласие результатов улучшилось приблизительно в 50 раз.



$t$ [ns]	$\langle n \rangle$	$A_{\langle n \rangle}$	Присоединения	Соотношение <sup>1</sup>	Отсоединения	Соотношение <sup>2</sup>
200 – 250	86.675473	495408506	1020000000	2.059	700000000	1.413
250 – 300	90.164718	368737293	740000000	2.007	640000000	1.736
300 – 350	88.663813	423501960	340000000	0.803	360000000	0.850
350 – 400	85.263617	546013566	660000000	1.209	720000000	1.319
400 – 450	82.077399	658797968	900000000	1.366	840000000	1.275
450 – 500	81.835752	667269950	820000000	1.229	880000000	1.319
500 – 550	82.865871	631073804	1100000000	1.743	1080000000	1.711
550 – 600	88.021790	446800978	1240000000	2.775	1140000000	2.551
600 – 650	88.323917	435846340	900000000	2.065	980000000	2.248
650 – 700	87.424306	468414613	1020000000	2.178	940000000	2.007
700 – 750	89.368816	397829516	880000000	2.212	920000000	2.313
750 – 800	87.804992	454651231	580000000	1.276	580000000	1.276
800 – 850	93.413176	248811560	500000000	2.010	380000000	1.527
850 – 900	91.911651	304479235	780000000	2.562	900000000	2.956
900 – 950	83.808862	597755068	1680000000	2.811	1920000000	3.212
950 – 1000	83.132449	621672564	1100000000	1.769	900000000	1.448

Таблица 4:  $t$  – кусок траектории считая с начала в наносекундах;  $\langle n \rangle$  – средний размер агрегата на участке траектории;  $A_{\langle n \rangle}$  – результат полученный из теории диффузионного приближения; Присоединения и Отсоединения – это количество вычисленное вручную, путем анализа временных рядов; Соотношение<sup>1</sup> – это отношение колонок Присоединения/ $A_{\langle n \rangle}$  (т.е. во сколько раз результат моделирования больше, чем аналитическая теория); Соотношение<sup>2</sup> – отношение колонок Отсоединения/ $A_{\langle n \rangle}$ .

#### 4.6. Алгоритм очистки временных рядов, основанный на вычислении скользящей энтропии

Кому на Руси жить хорошо?

---

Н. А. Некрасов

Самое время задать закономерный вопрос: неужели нельзя было применить какой-либо из известных методов фильтрации данных, попытаться убрать всплески и уже программными методами вычислить количество ПИО? И да, и нет.

Для начала давайте попробуем оценить насколько плоха исходная ситуация. Чтобы это сделать, посчитаем количество изменений временного ряда, не интересуясь тем, в сторону увеличения или уменьшения происходили переходы и какой они были величины. Для сырых, никак не обработанных данных, это число равно 81423. Количество переходов вычисленных вручную, равно 991  $\Rightarrow 991/81423 \cdot 100\% \approx 1.21\%$ . Это означает, что среди всех переходов почти 99% являются неправильными. Это обстоятельство само по себе является большой проблемой для фильтров.

Основная идея устройства большинства фильтров данных построена на скользящем среднем и среднеквадратичном отклонении от этих средних значений. Но в нашей ситуации они не подходят ещё вот по какой причине. Давайте представим, что из одного стабильного состояния произошёл переход +2 в другое стабильное состояние, это ситуация является для нас идеальная. Но скользящее среднее сгладит этот переход до двух последовательных +1 переходов. Это, конечно, может быть удобно для сравнения с диффузионным приближением, но тогда мы теряем важную информацию о количестве переходов +2. Вторая проблема заключается в следующем. Если взять слишком большой размер окна, то физически значимые переходы могут просто исчезнуть. Третья проблема: представим, что в стабильном состоянии внезапно случился ошибочный переход большой величины (встречались +9), продержался некоторое время, и график вернулся в исходное стабильное состояние. Физически это отвечает тому, что к мицелле приблизился достаточно большой агрегат, но не слился с ней. Скользящее среднее может резко отреагировать на такие скачки, сильно завышая среднее. И вместо одного ложного перехода, (допустим +9), среднее сделает несколько последовательных переходов на +1, которых вообще быть не должно.

Основой для постройки фильтра подходящего под эту ситуацию является структура самих рядов. Выделим два пункта:

1. Количество состояний конечно и их немного.
2. Переходы сопровождаются чередованием.

Если взглянуть на временной ряд в достаточно крупном масштабе (Рисунок 29), то можно увидеть, что временной ряд состоит из стабильных участков, где чаще всего нет никаких всплесков, и участков, которые отвечают переходам. Из-за нестабильности алгоритма кластеризации при переходах случаются чередования.

Иными словами в этих участках последовательностей наблюдается **повышение хаотичности**.

В связи с этим очень разумно попробовать вычислить для временных рядов **скользящую энтропию**. Энтропия является прекрасным индикатором повышения хаотичности в последовательностях. Ведь, допустим у нас есть последовательность длиной 3, каждый элемент этой последовательности может принимать одно из трёх состояний: 1, 2, 3. Тогда энтропия будет вести себя так:

– всё одинаковые

$$\mathbb{E}([1,1,1]) = \mathbb{E}([2,2,2]) = \mathbb{E}([3,3,3]) = 0$$

– два одинаковых, одно отличается

$$\mathbb{E}([1,2,2]) = \mathbb{E}([1,3,3]) = \dots \approx 0.6360$$

– всё три разные

$$\mathbb{E}([1,2,3]) = \mathbb{E}([1,3,2]) = \dots \approx 1.098$$

Для вычисления энтропии использовалась формула Шеннона [20]:

$$\mathbb{E}(X) = - \sum p_i \log_2(p_i), \quad (15)$$

где  $p_i$  – это вероятность встретить состояние  $i$ .

Теперь давайте применим следующий алгоритм к нашим временным рядам:

#### **Алгоритм**

1. Загрузить данные
2. Найти минимум массива
3. Вычесть минимальное значение из каждого элемента массива
4. Найти максимальное значение. Это число будет равняться возможному количеству состояний
5. Вычислить массив со значениями скользящей энтропии (размер окна 100 пикосекунд)
6. Создать массив-маску, который равен нулю, там где скользящая энтропия равна нулю, и равен единице, там где энтропия отлична от нуля.
7. Умножить массив с исходными данными на массив-маску (останутся только участки с повышенной энтропией)
8. Вычесть из исходных данных массив результат предыдущего шага (останутся только стабильные участки) – назовём этот массив *data*
9. Заполнить все нули в получившемся массиве по такому правилу:  
Если  $data(t) = 0$ , тогда  $data(t) = data(t - 1)$

#### 4.7. Результаты полученные с помощью алгоритма очистки

Жить стало лучше, жить  
стало веселей

---

И. В. Сталин

Давайте посмотрим, что получается в результате работы этого алгоритма (см. Рисунок 30).

Как видно в "спокойных" участках? скользящая энтропия равна нулю, а там где происходят изменения она повышается. После обработки алгоритмом график ведёт себя так, как мы ожидали.

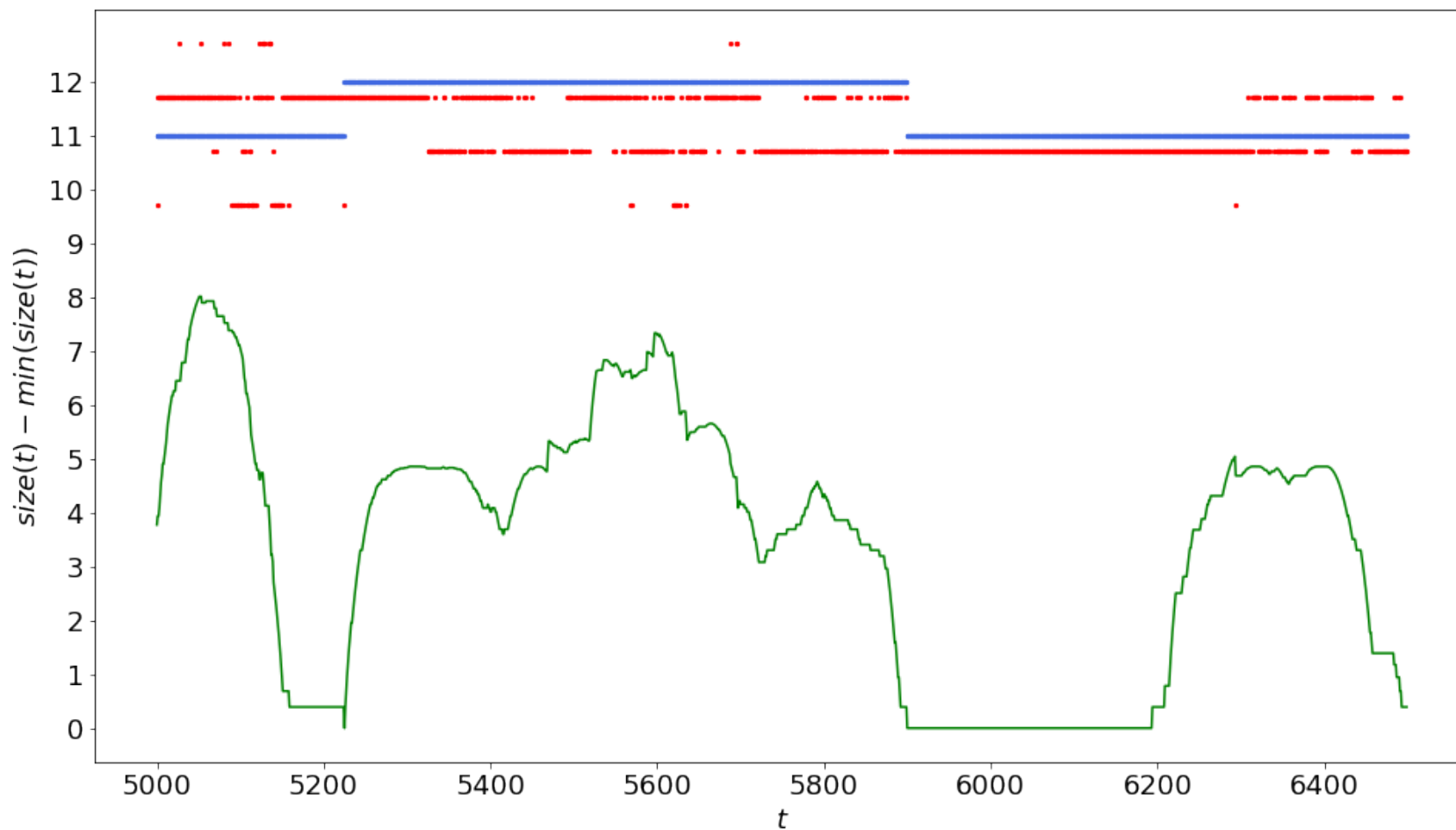


Рис. 30 Временной ряд для размера мицеллы в системе, состоящей из 100 мономеров, с вычтенным минимальным значением. **Красными** точками изображен исходный ряд (точки слегка опущены, для наглядности), **Синими** точками изображен ряд после обработки алгоритмом, **Зеленым** цветом изображено значение скользящей энтропии.

Переход	Через алгоритм	Вручную	Соотношение <sup>1</sup>
-9	2500000	1250000	2
-8	0	1250000	–
-7	1250000	3750000	0.33
-6	1250000	2500000	0.5
-5	2500000	3750000	0.66
-4	6250000	3750000	1.66
-3	22500000	28750000	0.78
-2	96250000	112500000	0.85
-1	480000000	460000000	1.04
1	471250000	458750000	1.02
2	92500000	111250000	0.83
3	26250000	30000000	0.875
4	8750000	8750000	1
5	1250000	5000000	0.25
6	2500000	1250000	2
7	1250000	0	–
8	1250000	3750000	0.33
9	2500000	2500000	1

Таблица 5: Количество переходов в единицу времени, подсчитанных через алгоритм, основанный на скользящей энтропии, и подсчитанных вручную. Соотношение<sup>1</sup> – это отношение столбцов "алгоритм"/"вручную".

Значения, получаемые с помощью алгоритма, сильно зависят от заданного размера окна скользящего. Результаты таблицы 5 были получены при значении 100 пикосекунд. Возможно, это число слегка завышено, поскольку энтропия может стать равна нулю, только когда 100 точек подряд были одинаковыми. При этом значении хорошее согласование результатов для ПИО мономеров, а для больших размеров расхождение больше.

Давайте вспомним наши оценки количества переходов. В необработанных данных оно равно 81423, при подсчете вручную – 991, после применения алгоритма остаётся 976, что даже меньше, чем необходимо.

## Заключение

ВКР посвящена изучению работы неионной мицеллы в неполярном растворителе при помощи молекулярного моделирования. Основные результаты работы можно сформулировать в виде следующих пунктов:

1. Были получены функции радиального распределения плотности (RDF) для обратных мицелл при различных числах агрегации. Эти результаты могут быть полезны для построения моделей для аналитических теорий.
2. Были получены распределения по числам агрегации: а) при различных температурах; б) при различных размерах системы. На основе полученных распределений были сделаны оценки работы образования, построены графики функции  $W(n)$ .
3. Были произведены оценки электростатического вклада в работу образования для прямых и обратных мицелл при различных размерах агрегатов. Получены зависимости энергетической выгоды от размера мицеллы.
4. Проведено вычисление коэффициентов присоединений и отсоединений для обратных мицелл. Полученные результаты продемонстрировали хорошее согласие с аналитической теорией.
5. Предложен алгоритм устранения шума при переходах мицеллы из одного числа агрегации в другое. Применение алгоритма продемонстрировало хорошее согласие с результатами ручной обработки.

## Приложение А

### Определение радиуса обрезания при кластеризации

В этом приложении обсуждается проблема определения радиуса обрезания при кластеризации  $r_{cutoff}$ . Сама проблема заключается в том, что иногда результаты оказываются чувствительными к выбранному радиусу. Также радиус обрезания может зависеть от температуры, потому что при повышении температуры, повышается подвижность молекул входящих в состав кластера.

Индикатором того, что радиус выбран неправильно можно служить количество изменений числа агрегации с течением времени. Если построить график зависимости числа изменений в зависимости от разного радиуса обрезания, то получится результат представленный на рисунке 31.

Наша гипотеза заключается в том, что этот график состоит из двух конкурирующих вкладов. Первый из них, обозначенный на графике 31 красным цветом, лучше всего описывается словами: при увеличении радиуса, алгоритм кластеризации начинает захватывать лишнее. Второй вклад, обозначенный желтым цветом, описывается словами: при увеличении радиуса, алгоритм перестает реагировать на колебание молекул в мицелле. Поэтому, для того, чтобы минимизировать ошибку, необходимо выбирать радиус таким, чтобы количество изменений было минимальным.

На рисунках 32, 33 необработанные очисткой от шума временные ряды для различных радиусов. Можно заметить, что при радиусе 60 нм, линии более разреженные.



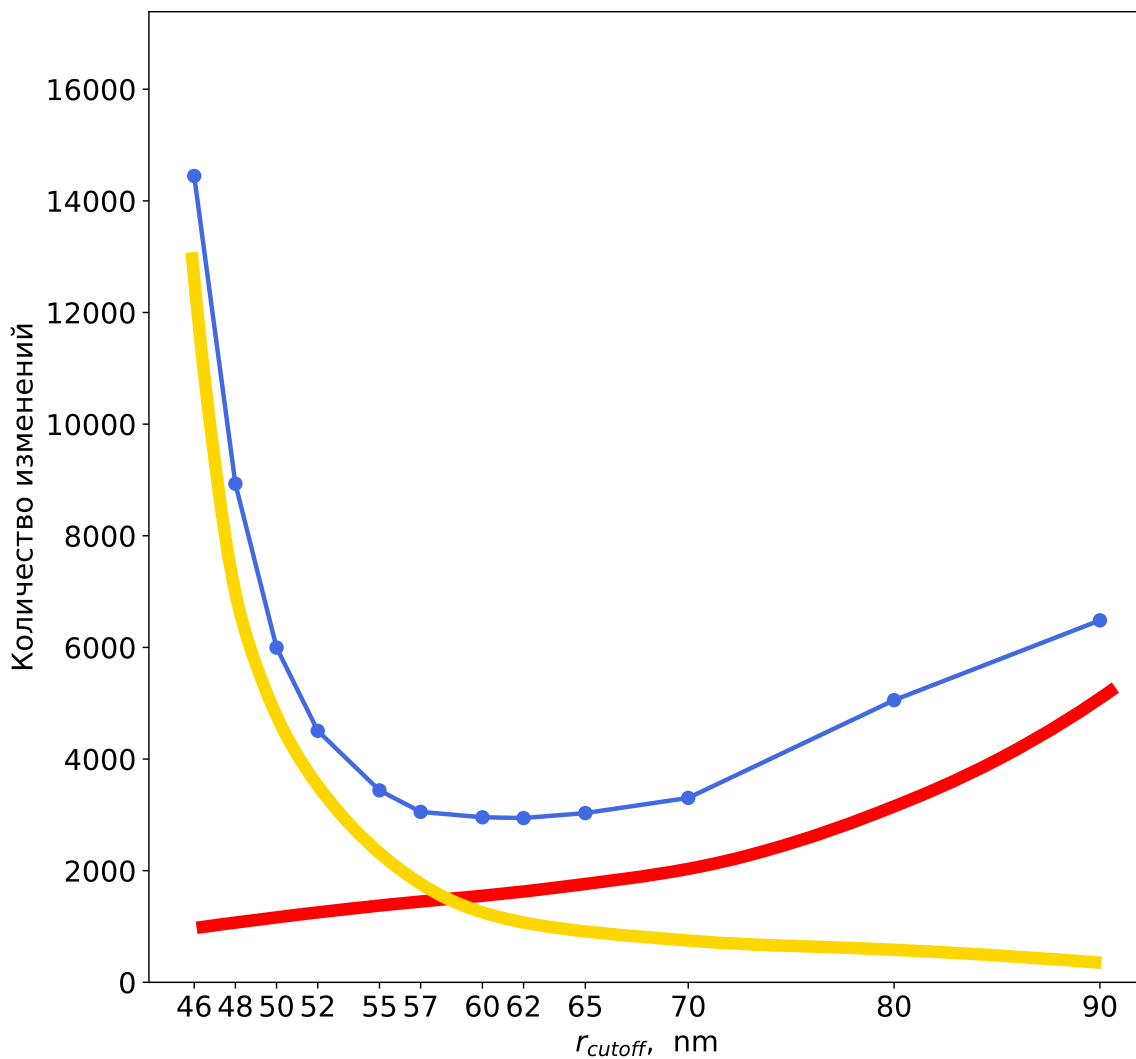


Рис. 31 **Синяя линия** – Зависимость количества изменения числа агрегации от радиуса обрезания  $r_{cutoff}$ . **Красная** и **Жёлтая** линии – два конкурирующих вклада (схематично).

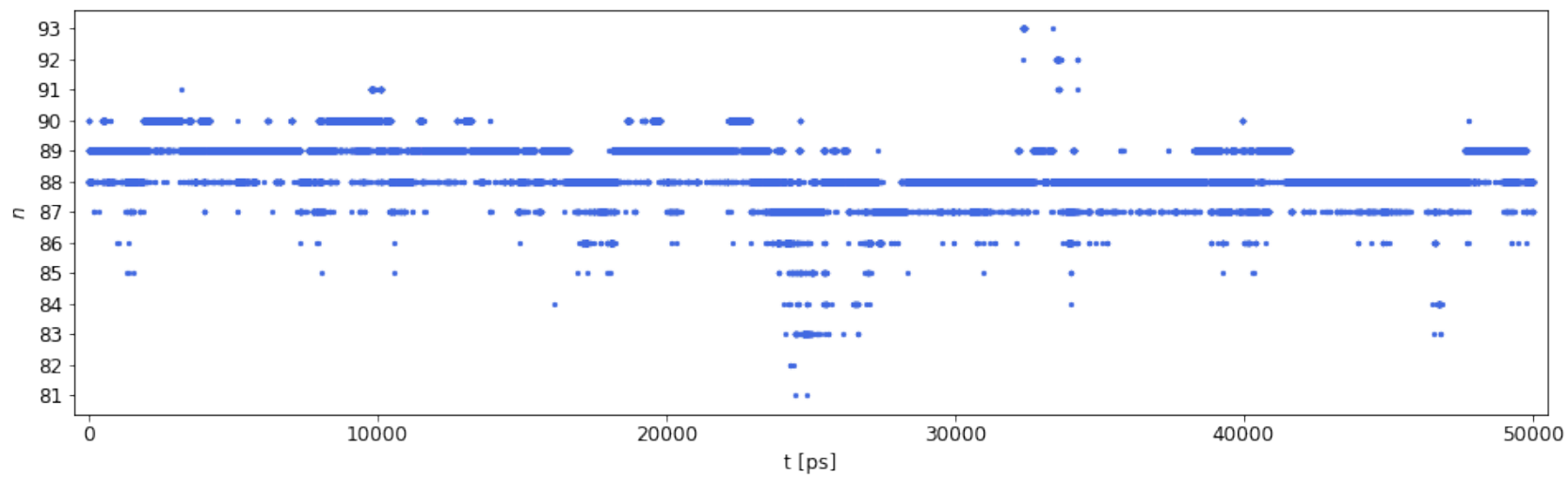


Рис. 32 Зависимость числа агрегации мицеллы в системе состоящей из 100 мономеров от времени при радиусе обрезания  $r_{cutoff} = 46$  нм.

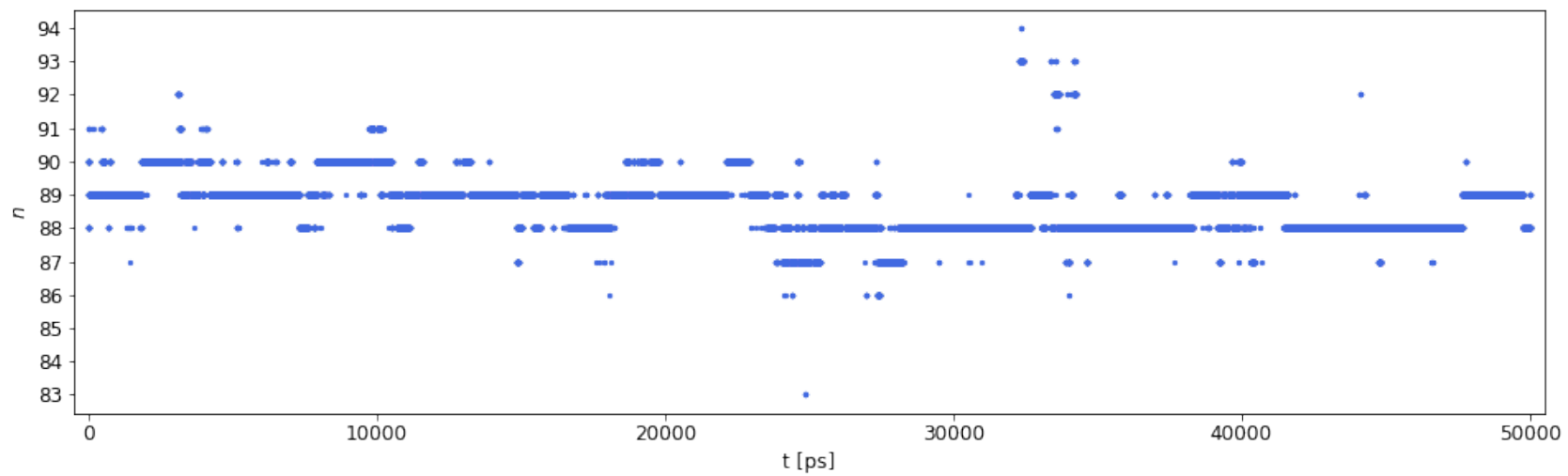


Рис. 33 Зависимость числа агрегации мицеллы в системе состоящей из 100 мономеров от времени при радиусе обрезания  $r_{cutoff} = 60$  нм.

## Список литературы

- [1] Volkov N. A., Posysoev M. V., Shchekin A. K. The Effect of Simulation Cell Size on the Diffusion Coefficient of an Ionic Surfactant Aggregate // *Colloid Journal*. — 2018. — Vol. 80, no. 3. — Pp. 248–254.
- [2] Frenkel Daan, Smit Berend. Understanding Molecular Simulation. — 2nd edition. — Orlando, FL, USA: Academic Press, Inc., 2001.
- [3] Allen M. P., Tildesley D. J. Computer Simulation of Liquids. — USA: Clarendon Press, 1989.
- [4] Walker C.H. The Hydrophobic Effect: Formation of Micelles and Biological Membranes // *FEBS Letters*. — 1981. — Vol. 124, no. 1. — Pp. 127–127. <https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/0014-579328812980071-3>.
- [5] Brodskaya E. N. Computer simulations of micellar systems // *Colloid Journal*. — 2012. — Vol. 74, no. 2. — Pp. 154–171. [www.scopus.com](http://www.scopus.com).
- [6] GROMACS - A PARALLEL COMPUTER FOR MOLECULAR-DYNAMICS SIMULATIONS / H BEKKER, HJC BERENDSEN, EJ DIJKSTRA at al. // PHYSICS COMPUTING '92 / edited by RA DeGroot, J Nadrchal. — World Scientific Publishing, 1993. — Pp. 252–256. — 4th International Conference on Computational Physics (PC 92) ; Conference date: 24-08-1992 Through 28-08-1992.
- [7] CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields / K. Vanommeslaeghe, E. Hatcher, C. Acharya et al. // *Journal of Computational Chemistry*. — 2010. — Vol. 31, no. 4. — Pp. 671–690. <http://dx.doi.org/10.1002/jcc.21367>.
- [8] Adzhemyan L. T., Shchekin A. K., Babintsev I. A. The “fine structure” of the slow micellar relaxation mode and the aggregation rates in the range between a potential hump and well in the work of aggregation // *Colloid Journal*. — 2017. — Vol. 79, no. 3. — Pp. 295–302. [www.scopus.com](http://www.scopus.com).
- [9] Бабинцев И. А. Исследование кинетики мицеллообразования и релаксации сферических и цилиндрических мицелл на основе уравнения Беккера-Дёринга // *Кандидатская диссертация, СПбГУ, 2014 год*.
- [10] Hill Terrell L. An introduction to Statistical Thermodynamics. — Addison-Wesley publishing company, inc., 1960.
- [11] Burov S. V., Shchekin A. K. Aggregation work at polydisperse micellization: Ideal solution and dressed micelle models comparing to molecular dynamics simulations // *Journal of Chemical Physics*. — 2010. — Vol. 133, no. 24. [www.scopus.com](http://www.scopus.com).

- [12] *Kshevetskiy M. S., Shchekin A. K.* The aggregation work and shape of molecular aggregates upon the transition from spherical to globular and cylindrical micelles // *Colloid Journal*. — 2005. — Vol. 67, no. 3. — Pp. 324–336. [www.scopus.com](http://www.scopus.com).
- [13] *A. K. Shchekin F. M. Kuni A. P. Grinin, Rusanov A. I.* Nucleation in micellization processes // *In Journ W.P. Schmelzer, editor, Nucleation theory and application, chapter 9, pages 312–374. Wiley, New York, 2005.*
- [14] Thermodynamic Characteristics of Micellization in the Droplet Model of Surfactant Spherical Molecular Aggregate / A. I. Rusanov, F. M. Kuni, A. P. Grinin, A. K. Shchekin // *Colloid Journal*. — 2002. — Sep. — Vol. 64, no. 5. — Pp. 605–615. <https://doi.org/10.1023/A:1020670228275>.
- [15] *Becker R., Döring W.* Kinetische Behandlung der Keimbildung in übersättigten Dämpfen // *Ann. der Physik, V. F.* — 1935. — Vol. 24. — Pp. 719–752.
- [16] *Burov S V., Vanin A. A., Brodskaya E. N.* Principal Role of the Stepwise Aggregation Mechanism in Ionic Surfactant Solutions Near the Critical Micelle Concentration. Molecular Dynamics Study // *The Journal of Physical Chemistry B*. — 2009. — Vol. 113, no. 31. — Pp. 10715–10720. — PMID: 19591445. <https://doi.org/10.1021/jp9025305>.
- [17] On the predictions and limitations of the Becker-Doring model for reaction kinetics in micellar surfactant solutions / I M Griffiths, C D Bain, C J W Breward et al. // *J Colloid Interface Sci.* — 2011. — . — Vol. 360, no. 2. — Pp. 662–671.
- [18] Theory of the kinetics of micellar equilibria and quantitative interpretation of chemical relaxation studies of micellar solutions of ionic surfactants / E. A. G. Aniansson, S. N. Wall, M. Almgren et al. // *The Journal of Physical Chemistry*. — 1976. — Apr. — Vol. 80, no. 9. — Pp. 905–922. <https://doi.org/10.1021/j100550a001>.
- [19] *Бобылёв В. Н.* "Физические свойства наиболее известных химических веществ" РХТУ им. Менделеева, 2004г.
- [20] *Shannon C.E. Weaver W.* The Mathematical Theory of Communication, Univ. of Illinois Press. ISBN 0-252-72548-4. — (1949),.