

Вайсман Елизавета Анатольевна

РАЗРАБОТКА МЕТОДА ИНДЕКСАЦИИ МНОЖЕСТВЕННЫХ КОНФОРМАЦИЙ ЛИГАНДОВ ДЛЯ ЗАДАЧ ВИРТУАЛЬНОГО СКРИНИНГА

Для обеспечения прогноза о наличии у лиганда конформации, пространственно комплементарной сайту связывания целевого белка, при решении задач виртуального скрининга баз химических веществ был исследован новый метод построения многоконформационных структурных 4D-QSAR дескрипторов химических соединений. В отличие от традиционных 4D-QSAR подходов, учитывающих конформационную подвижность лиганда, но зачастую страдающих от "информационного шума", возникающего при увеличении числа конформаций, данный метод позволяет бороться с информационным шумом за счет использования в своей структуре модификации фильтра Блума.

Адрес статьи: www.gramota.net/materials/1/2013/7/8.html

Статья опубликована в авторской редакции и отражает точку зрения автора(ов) по рассматриваемому вопросу.

Источник

Альманах современной науки и образования

Тамбов: Грамота, 2013. № 7 (74). С. 28-32. ISSN 1993-5552.

Адрес журнала: www.gramota.net/editions/1.html

Содержание данного номера журнала: www.gramota.net/materials/1/2013/7/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net
Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: almanac@gramota.net

Содержание учебного материала предусматривает его обновление. Это касается изучения и использования на практике новейших способов контроля измерения обрабатываемой детали [10, с. 233-245], которые не только повышают производительность труда ввиду мгновенного автоматического контроля технологического процесса при обработке детали, но и создают условия для устранения аварийной ситуации оборудования, инструмента и самой детали.

Заключение

Таким образом, обоснована и разработана педагогическая система профессиональной подготовки БКР металлообрабатывающего направления для учебно-производственного и воспитательного процесса ПТУЗа. Эта система включает учеников (БКР); педагогических работников (учителей, преподавателей, мастеров производственного обучения, воспитателей); целевой блок в зависимости от общественного заказа (цель, задание, факторы, основные подходы); процессуальный блок (организационно-педагогические условия с включением обновления содержания, методы, принципы, формы, оценивание и коррекция сформированных компонентов компетентности); блок ИРО с результативным блоком, куда входит модель формирования готовности и способностей БКР к ПД.

Список литературы

1. Беспалько В. П. Основы теории педагогических систем. Воронеж: Изд-во Воронежского ун-та, 1977. 304 с.
2. Беспалько В. П. Педагогика и прогрессивные технологии обучения // Школьные технологии. 2001. № 3.
3. Вайнтрауб М. А. Интегроване розвивальне навчання у професійній школі: монографія. К.: Т. Ключко, 2009. 179 с.
4. Вайнтрауб М. А. Критерій оцінювання якості навчання в ПТНЗ // Шлях освіти. 2010. № 3 (57). С. 27-30.
5. Вайнтрауб М. А. Метод оцінювання інтегрованого розвивального навчання в ПТНЗ // Педагогічні науки: теорія, історія, інноваційні технології. 2010. № 5 (7). С. 213-219.
6. Вайнтрауб М. А. Методичні підходи до оцінювання компетентностей сучасного майбутнього кваліфікованого робітника з металообробки у ПТНЗ // Нові технології навчання. 2011. Вип. 68. С. 69-77.
7. Вайнтрауб М. А. Методичні підходи до структурування компетентностей сучасного майбутнього кваліфікованого робітника з металообробки в ПТНЗ // Нові технології навчання. 2011. Вип. 70. С. 105-113.
8. Вайнтрауб М. А. Результативність підготовки майбутніх кваліфікованих робітників з металообробки під час експерименту у ПТНЗ // Науковий вісник Національного університету біоресурсів і природокористування України. Серія «Педагогіка. Психологія. Філософія». 2012. Вип. 175. Ч. 2. С. 188-198.
9. Вайнтрауб М. А. Створення моделі навчально-виробничого та виховного процесу при підготовці майбутнього кваліфікованого робітника з обробки металу в ПТНЗ до професійної діяльності // Наукові записки: збірн. наук. статей / Нац. пед. ун-т імені М. П. Драгоманова. 2012. Вип. С (100). С. 41-49.
10. Вайнтрауб М. А. Теорія і практика професійної підготовки майбутніх кваліфікованих робітників з обробки металу: монографія. Вид. 2-ге, доповн. К.: Т. Ключко, 2013. 328 с.
11. Викторова Л. Г. О педагогических системах. Красноярск: Изд-во КГУ, 1989. 86 с.
12. Кузьмина Н. В. Методы системного педагогического исследования. Л.: Изд-во ЛГУ, 1980. 172 с.
13. Лещенко С. Й. Педагогические условия подготовки конкурентоспособных молодых рабочих с учетом требований рынка труда (в системе начального профессионального образования): автореф. дисс. ... к. пед. н.: 13.00.01. М., 1998. 23 с.
14. Панфилов М. А. Проектирование педагогической технологии политехнической ориентации и условия ее применения в процессе трудового обучения: автореф. дисс. ... к. пед. н.: 13.00.01. Сочи, 1999. 24 с.
15. Селенков А. Н. Формирование профессиональной компетентности будущих специалистов металлообрабатывающих профессий в условиях ресурсного центра колледжа: автореф. дисс. ... к. пед. н.: 13.00.08. М., 2012. 22 с.
16. Устемиров К., Шаметов Н. Р., Васильев И. Б. Профессиональная педагогика: учебник для учащихся колледжей и студентов вузов. Алматы: ТОО «Акнур и К^о», 2005. 432 с.

УДК 004.94

Технические науки

Для обеспечения прогноза о наличии у лиганда конформации, пространственно комплементарной сайту связывания целевого белка, при решении задач виртуального скрининга баз химических веществ был исследован новый метод построения многоконформационных структурных 4D-QSAR дескрипторов химических соединений. В отличие от традиционных 4D-QSAR подходов, учитывающих конформационную подвижность лиганда, но зачастую страдающих от «информационного шума», возникающего при увеличении числа конформаций, данный метод позволяет бороться с информационным шумом за счет использования в своей структуре модификации фильтра Блума.

Ключевые слова и фразы: фильтр Блума; 4D-QSAR; множество конформаций; ансамбль конформеров; виртуальный скрининг; структурный дескриптор.

Вайсман Елизавета Анатольевна

*Новосибирский государственный университет
miina@gorodok.net*

РАЗРАБОТКА МЕТОДА ИНДЕКСАЦИИ МНОЖЕСТВЕННЫХ КОНФОРМАЦИЙ ЛИГАНДОВ ДЛЯ ЗАДАЧ ВИРТУАЛЬНОГО СКРИНИНГА[©]

Одна из проблем виртуального скрининга химических баз данных – это потенциальное несовпадение конформации лиганда, получаемой из базы данных, с конформацией, в которой он связывается с белком.

Методы, основанные на структурных 3D-QSAR дескрипторах, используют одну, часто произвольную, конформацию лиганда и поэтому могут давать ошибочно негативный прогноз относительно его связывания с мишенью. Классические же структурные 4D-QSAR методы сопряжены с большими объемами данных обо всех деталях строения и конформационного поведения исследуемых соединений, которые зачастую начинают играть роль «информационного шума», затрудняя выявление истинных взаимосвязей между их структурой и активностью. Задача данной работы состоит в разработке метода, позволяющего строить 4D-QSAR дескрипторы с высокой информационной емкостью, что позволило бы бороться с проблемой конформационного шума.

Теоретическое обоснование

Прогноз о сходстве пространственных структур тестируемого лиганда из базы данных (БД) и целевого лиганда дается на основе сравнения описывающих их QSAR дескрипторов [3] – битовых последовательностей ограниченной длины, хранящих в себе информацию о пространственном строении и конформационной изменчивости заданных химических соединений. Данный метод предполагает индексацию БД многоконформационными 4D-QSAR дескрипторами, описывающими ансамбль конформеров для каждого из тестируемых лигандов. Далее полученные дескрипторы сравниваются с 3D дескриптором целевого лиганда, содержащим только ту его конформацию, в которой он связывается с биомишенью. Успешность сравнения говорит о сходстве пространственной структуры тестируемого лиганда с целевым. Неуспешность – об обратном и, в частности (с достаточно большой долей вероятности), о том, что тестируемый лиганд не имеет конформации, в которой он пространственно комплементарен целевому белку. Ошибка в данном предположении называется ложноотрицательным срабатыванием (ЛОС) алгоритма, и небольшой процент этих ошибок свойственен всем QSAR-методам. Структурные (то есть хранящие информацию о пространственной структуре молекулы) дескрипторы, минимизирующие процент ЛОС при сравнении, должны обладать следующими свойствами:

1. Отсутствие прямой зависимости от ориентации лиганда и нумерации атомов в нем (наличие такой зависимости либо вызывало бы ЛОС при сравнении одноименных молекул с разной ориентацией или нумерацией атомов, либо непременно увеличивало бы затраты на конструирование и сравнение дескрипторов).
2. Высокие дискриминационные свойства: чем больше схожи по структуре лиганды, тем меньше отличаются их дескрипторы.
3. Возможность объединения дескрипторов различных конформаций с получением многоконформационного дескриптора без потери информации о существенных различиях конформеров.

3D-QSAR дескриптор одной конформации лиганда, обладающий первыми двумя свойствами, может быть построен на основе радиальной функции распределения межатомных расстояний в лиганде: расстояния не зависят от ориентации лиганда и порядка атомов, а совпадение наборов межатомных расстояний у двух молекул позволяет сделать вывод о потенциальном сходстве между ними. Далее посредством объединения полученных 3D-QSAR дескрипторов при помощи механизмов фильтра Блума [1] генерируется 4D-QSAR дескриптор, выполняющий свойство 3. Использование фильтра Блума также обеспечивает дальнейшую возможность сравнивать 3D-QSAR дескриптор целевого лиганда с 4D-QSAR дескриптором тестируемого лиганда. Причем, если сравнение оказалось неуспешным, то в пределах фильтра Блума гарантируется отсутствие ЛОС, а именно то, что среди множества наборов распределений межатомных расстояний для каждой взятой конформации тестируемого лиганда нет распределения межатомных расстояний такого же, как у целевого лиганда. Этот факт дает основание отсеять тестируемый лиганд из БД при первичном виртуальном скрининге, но не исключает отсутствия ЛОС метода в целом, поскольку остается вероятность того, что искомая в многоконформационном дескрипторе тестируемого лиганда конформация существует для него, но не добавлена в его фильтр Блума.

Отметим также, что, хотя фильтр Блума интенсивно используется в распределенных базах данных и сетях [2], в литературе встречается всего два упоминания о его применении в биоинформатике [4; 5].

Реализация дескриптора

Дескриптор одной конформации l -атомного лиганда состоит из $l \cdot m$ -байтовых дескрипторов каждого атома лиганда и имеет суммарную длину $l \cdot m$ байт. Дескриптор одного a_i -того атома, $i=1 \dots l$, кодирует дискретизованную радиальную функцию распределения межатомных расстояний от атома a_i до остальных атомов лиганда. Эта функция определена на интервале $[0, 20 \text{ Ангстрем}]$, то есть на характерном размере лиганда. Таким образом, дескриптор одного a_i -того атома состоит из m 8-битных ячеек и имеет вид:

00000000 00000001 00001000 00000010 ... 00000000 00000000

В k -той ячейке, $k=1 \dots m$, хранится специального вида число, равное количеству расстояний от атома a_i до других атомов лиганда a_j , удаленных от a_i не меньше, чем на $\frac{(k-1) \cdot 20}{m}$ Ангстрем, и не больше, чем на $\frac{k \cdot 20}{m}$ Ангстрем ($i, j=1 \dots l$). Вид хранения числа в ячейке обоснован необходимостью последующего применения фильтра Блума: 1 расстояние на заданном интервале – 00000001, 2 расстояния на заданном интервале – 00000010, 3 расстояния на заданном интервале – 00000100 и т.д. Максимальное количество расстояний, хранящихся в ячейке, – 8, поскольку валентность химических элементов (и, следовательно, количество соседей атома) не превышает этого числа. Заметим, что число 8 взято с запасом, так как для подавляющего большинства лигандов достаточным значением является 4 (валентность углерода).

Многоконформационный дескриптор лиганда строится из дескрипторов заданных конформаций применением побитовой операции логического «ИЛИ»:

```
00000001 00000000 ... 00100000 00001000
00000010 00000000 ... 00010000 00000100
00000001 00000000 ... 00001000 00001000
00000100 00000000 ... 00100000 00000100
OR
00000111 00000000 ... 00111000 00001100
```

Полученная в итоге структура данных и является фильтром Блума и, очевидным образом, сохраняет в дескрипторе информацию о различиях конформаций. Простейший алгоритм поиска элемента (дескриптора одной конформации) в фильтре Блума (дескрипторе множества конформаций) основан на поиске единиц в определенных позициях битового массива. Если позиции, на которых стоят единицы в искомом элементе, заняты единицами на тех же позициях фильтра Блума, то структура данных сообщает, что элемент принадлежит множеству. При этом могут возникнуть две ситуации: либо элемент действительно принадлежит к множеству, либо все эти биты оказались установлены по случайности при добавлении других элементов, что и является источником ложноположительных срабатываний (ЛПС) в этой структуре данных (конформации во множестве нет, но структура сообщает, что она есть). Если элемент не найдем в фильтре Блума, значит, он отсутствует в хранимом множестве конформаций (иначе соответствующие ему позиции в фильтре Блума были бы заполнены единицами).

Простейший алгоритм поиска элемента используется для сравнения 3D и 4D дескрипторов, но применять его к дескрипторам лигандов целиком нельзя, поскольку это создаст зависимость от порядка записи атомов. Поэтому следует сравнивать дескрипторы лигандов поатомно. Проблема сопоставления атомов в сравниваемых лигандах сводится к задаче о поиске наибольшего паросочетания в двудольном графе (левая доля – целевой лиганд, правая – тестируемый) и решается при помощи алгоритма Куна со сложностью $O(m^3)$.

Оценка информационной емкости

Информационную емкость дескриптора можно оценить через вероятность ЛПС при поиске элемента в фильтре Блума. Именно высокий процент ЛПС снижает дискриминационную способность дескриптора и вызывает неэффективное «просеивание» БД. Далее будут рассматриваться только такие ЛПС (без учета ЛПС, вызванных негеометрическими характеристиками лигандов).

Вероятность ЛПС при сравнении одноатомных дескрипторов совпадает с известной для фильтра Блума аналогичной величиной [2] и составляет: $(1 - e^{-ln/m})^l$.

Тогда произведение этих вероятностей даст оценку вероятности ЛПС для дескриптора лиганда: $(1 - e^{-ln/m})^{l^2}$ (в данном случае мы пренебрегаем малой вероятностью того, что для разных лигандов совпадут дескрипторы даже двух атомов). Рисунок 1 иллюстрирует зависимость между вероятностью ЛПС и количеством хранимых в дескрипторе одного лиганда конформаций при заданном количестве атомов в лиганде и заданной длине дескриптора.

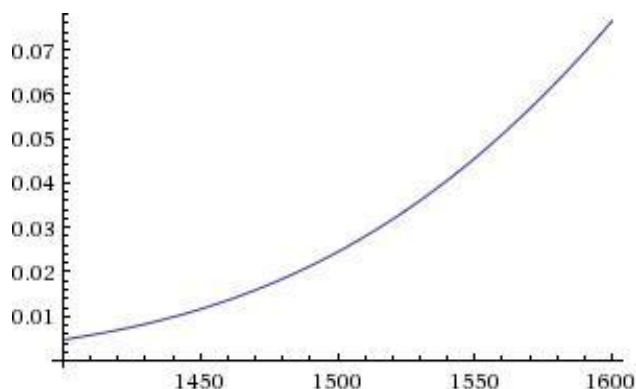


Рис. 1. Зависимость вероятности ложноположительного срабатывания (по оси ординат) от количества конформаций, содержащихся в дескрипторе тестируемого лиганда длиной $m = 1024$ байт (по оси абсцисс), в предположении того, что количество атомов лиганда l не превышает 30

Результаты

Разработанный метод был реализован на языке программирования *Java* в виде двух программных модулей. Модуль *GenerateConformers* генерирует при помощи стороннего свободно-распространяемого программного обеспечения *Balloon* [6] вспомогательную БД, содержащую для каждого из лигандов сканируемой БД ансамбль его конформеров. Модуль *ConformerDescriptor* конструирует дескрипторы тестируемых лигандов при помощи вспомогательной БД, индексирует ими сканируемую БД, строит дескриптор целевого лиганда и проводит первичный виртуальный скрининг проиндексированной БД относительно целевого лиганда.

Тестирование программных модулей проводилось на базе данных размером 5547 лигандов. Результаты работы, полученные для дескрипторов длиной 128 байт, при количестве хранящихся в каждом дескрипторе конформаций, не превышающем 200, соответствуют объявленной оценке информационной емкости дескрипторов, отраженной на Рисунке 2.

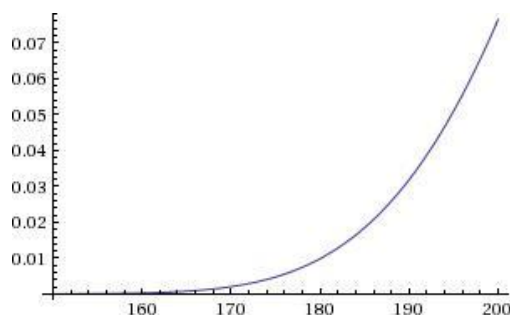


Рис. 2. Зависимость вероятности ложноположительного срабатывания (по оси ординат) от количества конформаций, содержащихся в дескрипторе тестируемого лиганда длиной $m = 128$ байт (по оси абсцисс), в предположении того, что количество атомов лиганда l не превышает 30

Например, для конформационно подвижного лиганда $C_{12}N_3O_9$ (3-[3,5-Bis-(2-carboxy-ethyl)-2,4,6-trioxo[1,3,5]triazinan-1-yl]-propionic acid), количество существенно разных конформеров которого варьируется от 130 до 186, алгоритм находит в тестируемой БД 4 схожих с ним лиганда. При проверке все эти лиганды оказываются различными конформациями $C_{12}N_3O_9$, то есть ложноположительных срабатываний не происходит (действительно, при $l=24$, $n \leq 200$ и $m=128$ байт вероятность ЛПС не превышает 0,48%). Для небольших лигандов малой (например, PO_4 , SO_4) и средней ($C_3H_8O_3$, $C_2H_3O_2$) конформационной подвижности ложноположительных срабатываний также не происходит. Процент отсеивания лигандов из БД составляет от 93%.

К сожалению, для дескрипторов большей длины никаких оценок, кроме временных (Табл. 1), получить не удалось. Это вызвано тем, что задействованное при генерации конформаций программное обеспечение *Balloon* не выдает достаточного большого числа конформаций. Малое же число конформаций при чрезмерно большой длине дескриптора приводит к высокому проценту ложноотрицательных срабатываний. Это происходит из-за того, что точность определения межатомных расстояний повышается с ростом длины дескриптора (одной ячейке большего дескриптора соответствует меньший интервал расстояний) (Рис. 3), но не компенсируется «густым» заполнением дескриптора в областях пиков радиальной функции распределения межатомных расстояний (что является характерным следствием добавления большого числа конформаций).

```
00000001 00000100 00001000
00000000 00000001 00000010 00000001 00000010 00000010
```

Рис. 3. Степень заполнения единицами для дескрипторов разной длины (храняемая конформация – одна и та же). Вверху – 3-байтовый дескриптор, внизу – 6-байтовый

Табл. 1

Длина дескриптора m в байтах, l – кол-во атомов в лиганде	Кол-во конформаций n , которое может храниться в дескрипторе одного лиганда	Время виртуального скрининга БД размером 5547 лигандов
128*/ l	200	~1 секунда
1024*/ l	800	~ 8 секунд
4096*/ l	3200	~ минута

Таблица 1 показывает зависимость количества конформаций n , помещаемого в многоконформационный дескриптор l -атомного лиганда, от размера дескриптора m при постоянной вероятности ложноположительного срабатывания, не превышающей 7,6%, а также время сканирования БД при заданной длине дескриптора.

Заключение

Для обеспечения прогноза о наличии у лиганда конформации, пространственно комплементарной сайту связывания целевого белка, при решении задач виртуального скрининга БД был исследован новый метод построения многоконформационных структурных 4D-QSAR дескрипторов химических соединений. Данный метод позволяет строить дескрипторы высокой информационной емкости, о чем говорит их способность хранить большое число конформаций, не теряя знаний о различиях конформаций и сохраняя низкий процент ложноположительных прогнозов. Также достоинством метода является высокое быстродействие и универсальность дескрипторов – однажды проиндексированная дескрипторами БД может использоваться в задачах виртуального скрининга с различными целевыми лигандами. В частности, потенциально, свойство

универсальности позволяет использовать данный дескриптор в сочетании с методами *de novo* для поиска соответствий между фиктивным лигандом, пространственно комплементарным сайту связывания белка, и реальными химическими соединениями.

К сожалению, проблемы с используемым сторонним программным обеспечением не позволяют точно определить форму зависимости числа ложноотрицательных прогнозов от длины дескриптора и числа хранимых конформаций. Поэтому основное направление развития метода связано с разработкой собственного или поиском и адаптацией существующего генератора ансамбля конформаций, в достаточной мере предоставляющего управление параметрами получения конформаций, в частности шириной диапазона потенциальной энергии, возможностью получения заданного числа конформаций напрямую или через оценку среднеквадратичного отклонения межатомных расстояний.

Список литературы

1. **Bloom B.** Space/Time Tradeoffs in Hash Coding with Allowable Errors // *Communications of the ACM*. 1970. Vol. 13. P. 422-426.
2. **Broder A., Mitzenmacher M.** Network Applications of Bloom Filters: a Survey // *Internet Mathematics*. 2004. № 1 (4). P. 485-509.
3. **Hopfinger A., Wang S., Tokarski J., Jin B., Albuquerque M., Madhav P., Duraiswami C.** Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism // *Journal of the American Chemical Society*. 1997. Vol. 119. P. 10509-10524.
4. **Koes D. R., Camacho C. J.** Pharmer: Efficient and Exact Pharmacophore Search // *Journal of Chemical Information and Modeling*. 2011. Vol. 51. P. 1307-1314.
5. **Stranneheim H., Kaller M., Allander T., Andersson B., Arvestad L., Lundeberg J.** Classification of DNA Sequences Using Bloom Filters // *Bioinformatics*. 2010. № 26 (13). P. 1595-1600.
6. **Vainio M. J., Johnson M. S.** Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm // *Journal of Chemical Information and Modeling*. 2007. Vol. 47. P. 2462-2474.

УДК 536-3

Физико-математические науки

В данной работе исследуются изотермические процессы, протекающие в реальных газах. Проведено компьютерное моделирование уравнения Ван-дер-Ваальса с использованием известных значений постоянных a и b для чистых газов. Модель используется для численных расчетов критических значений температур и изотерм многокомпонентных газовых смесей.

Ключевые слова и фразы: реальные газы; *Mathcad*; уравнение Ван-дер-Ваальса; смеси реальных газов; изотермы реальных газов; углеводороды.

Васильева Лидия Анатольевна, к. ф.-м. н.

Кудряшов Марк Дмитриевич

Баязитов Василий Дмитриевич

Волгоградский государственный архитектурно-строительный университет

bayazitov.vgasu@mail.ru

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ РЕАЛЬНЫХ ИЗОТЕРМ МНОГОКОМПОНЕНТНЫХ ГАЗОВЫХ СМЕСЕЙ[©]

Широкое применение природного газа требует его тщательного изучения. Изучение поведения газа в изотермическом процессе актуально в связи с необходимостью сжижения газа для его транспортировки.

Если для чистых газов условия сжижения известны, то для смесей разных составов требуется расчёт постоянных в уравнении Ван-дер-Ваальса и критических значений температуры и давления. Проведен анализ теоретических моделей для построения изотерм реальных газов и выбрана компьютерная модель изотермических процессов с применением двухконстантного уравнения состояния Ван-дер-Ваальса для реальных газов:

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT$$

Главное достоинство уравнения Ван-дер-Ваальса состоит не только в том, что оно описывает газовую фазу вещества в более широкой области температур и давления, но и в том, что оно описывает картину превращения газа в жидкость при температурах ниже критической.

Видно, что это уравнение фактически является уравнением состояния идеального газа с двумя поправками. Поправка a учитывает силы притяжения между молекулами (давление на стенку уменьшается, так как