

УДК 548.31

МНОГОФАКТОРНЫЙ АНАЛИЗ ТЕМПЕРАТУР ПЛАВЛЕНИЯ ИЗОМЕРОВ ГОМОЛОГИЧЕСКОГО РЯДА $C_xH_{2x-1}NO$ ($x = 3-6$)

Д.Б. Еремин, А.М. Банару

(кафедра физической химии; e-mail: banaru@phys.chem.msu.ru)

Проведен многофакторный анализ массива эмпирических структурных характеристик для изомерных гомологов методом главных компонент. Показана их корреляция с экспериментальными температурами плавления. По структурным данным Кембриджского банка обсужден вклад межмолекулярной агломерации за счет Н-связывания.

Ключевые слова: метод главных компонент, температура плавления, молекулярная структура, водородная связь.

Метод главных компонент [1] как вид многофакторного анализа неявно зависимых величин – перспективный численный метод, который применяется во многих отраслях науки: экономике, социологии, аналитической химии. Его с успехом используют и в структурной химии [2–3] для объяснения и предсказания физико-химических свойств, а также биологической активности. Сам по себе метод главных компонент является центральным инструментом хемометрики [4]. Применимость этого метода для корректного анализа кристаллических структур была неоднократно показана ранее (например, в [5]).

Зависимость температуры плавления от структуры органического соединения неоспорима, но оценка или объяснение получаемых значений температуры возможны лишь при многофакторном анализе молекулярных структурных характеристик, которые желательно дополнить данными о кристаллической структуре. Ввиду изомерного разнообразия и доступности экспериментальных значений температуры плавления объектами нашего изучения стали изомеры гомологического ряда $C_xH_{2x-1}NO$ ($x = 3-6$), однако кристаллоструктурные данные получены к настоящему времени далеко не для всех соединений этого ряда. Общеизвестно, что температура плавления молекулярных кристаллов зависит от многих факторов: величины дипольного момента молекул, их поляризуемости, ван-дер-ваальсова размера и др.

Цель настоящей работы заключалась в обосновании эмпирически установленных значений температуры плавления на основе простых структурных характеристик, не требующих *ab initio* или даже полумпирических расчетов. Общепринятой схемы, объясняющей и предсказывающей температуру

плавления молекулярных кристаллов, до сих пор не существует, и именно это обстоятельство служит стимулом к последующим расчетам.

Обработка данных

Поиск соединений состава $C_xH_{2x-1}NO$ ($x = 3-6$) с известными значениями температуры плавления был осуществлен в интегрированной научной базе данных SPRESI^{web} 2.11 [6]. Были отобраны: 1 соединение состава C_3H_5NO , 6 соединений состава C_4H_7NO , 20 – C_5H_9NO и 36 – $C_6H_{11}NO$. В дальнейшем из них была выбрана модельная группа из 16 соединений, состоящая из двух гомологически связанных изомерных рядов (рис. 1). Для каждого ряда был задан набор структурных параметров, являющихся логическими переменными, т.е. принимающими значения 0 и 1: наличие в молекуле двух N–H-связей, увеличение в ней цикла по отношению к предыдущему гомологу, рост по α -положению, наличие пяти-, и шестичленного цикла, а также резонансное «стягивание» электронной плотности с атома азота. Полученный набор данных (двоичная матрица) был обработан с помощью пакета статистических программ STATISTI^xl [7], включающего в том числе и алгоритм метода главных компонент. Был проведен первичный расчет главных компонент, необходимый для определения вида корреляционных зависимостей температуры плавления от первой компоненты. Соединение 16 вначале было исключено из расчета, так как собственные значения его главных компонент выпадали из общей закономерности распределения (*outlayer*), однако при повторном расчете оно было возвращено (пояснение см. в следующем разделе). На основании первоначальных результатов был уточнен список используемых

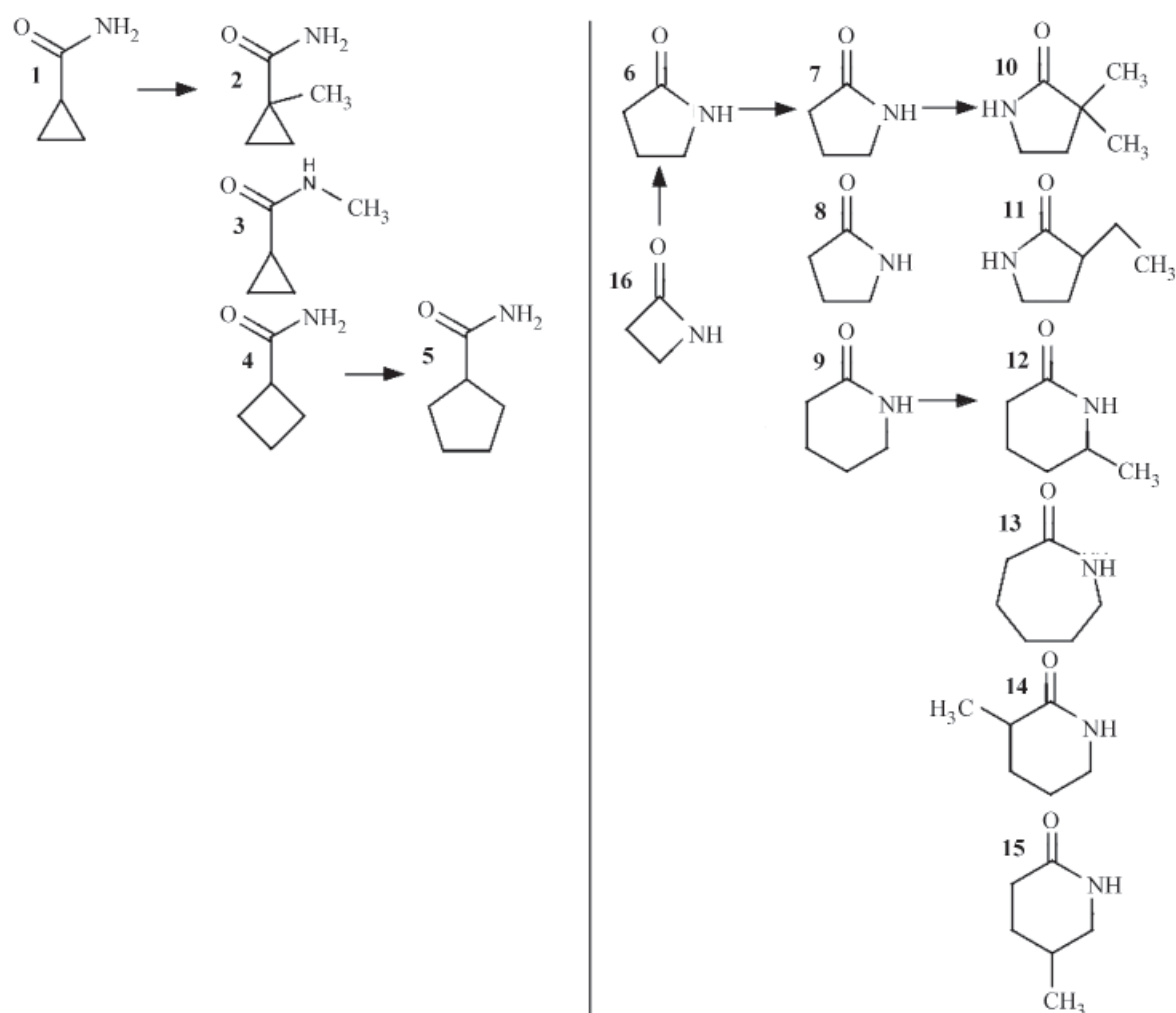


Рис. 1. Соединения в анализе главных компонент структурных характеристик

структурных характеристик и проведен повторный расчет.

В Кембриджском банке (CSD) [8] были найдены данные о пяти кристаллических структурах соединений выбранной группы, и для дополнения структурных переменных были учтены сведения о водородных связях. Так была составлена эмпирическая схема, которая ясно описывает взаимосвязь температуры плавления и структуры (молекулярной и кристаллической).

Обсуждение результатов

При беглом рассмотрении распределение экспериментальных значений температуры плавления и других количественных физико-химических характеристик, приведенных в SPRESI^{web}, не поддается внятому аппроксимированию, именно поэтому в группе отобранных соединений и были заданы структурные критерии, отражаемые двоичным кодом (0 – отсут-

ствие признака, 1 – наличие признака). Из зависимости собственных значений выделенных главных компонент от их числа можно сделать вывод о том, что для ее описания достаточно трех компонент, так как наибольший излом на графике происходит на третьей компоненте (правило «сломанной трости», [1]).

При определении взаимной коррелированности выделенных главных компонент с температурой плавления мы выявили две выраженные группы точек с участками линейности (рис. 2). Ввиду первоначально существенного разброса точек (соединений) на корреляционной диаграмме были проверены коэффициенты корреляции структурных параметров с каждой из трех главных компонент, но так как все они оказались статистически несомы, то их оставили для повторного анализа, но дополнили еще одним эмпирическим параметром – относительной молекулярной массой. Результат расчета при таком рассмотрении показал более отчетливую кластеризацию соединений в две

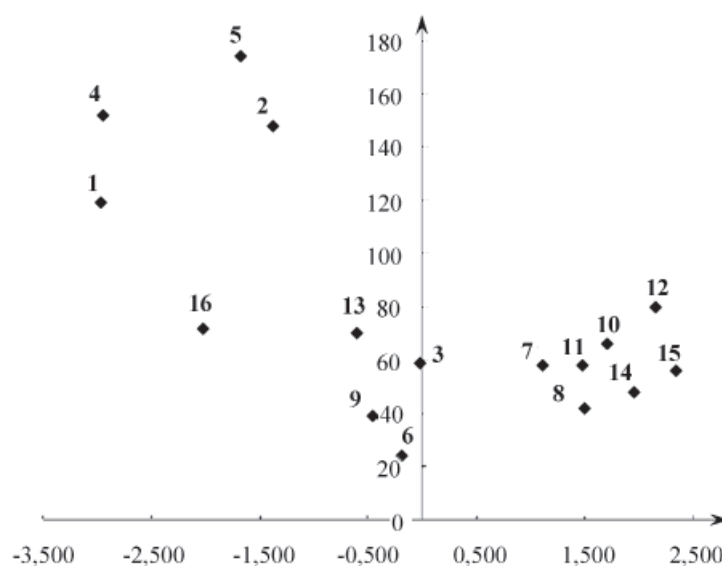


Рис. 2. Диаграмма зависимости температур плавления (°C) от собственного значения первой главной компоненты

группы изомерных соединений, при этом соединение 16 оказалось соответствующим общей закономерности (рис. 2). В табл. 1 приведены нагрузки (парные коэффициенты корреляции) заданных структурных параметров в значения трех главных компонент.

Когда значение нагрузки выше 0,7 по модулю, можно говорить о наличии статистически значимой корреляции между компонентой и выбранной характеристикой. По аналогии с традиционным коэффициентом корреляции Пирсона знак «+» нагрузки означает прямую пропорциональность двух величин, а «-» – обратную. При этом мера пропорциональности так же, как и для корреляции по Пирсону, характеризуется абсолютной величиной нагрузки. Структурные характеристики с меньшим значением нагрузок, на первый взгляд, можно было бы исключить из рассмотрения, однако, как правило, этого делать не рекомендуется при значении хотя бы одной нагрузки по модулю выше 0,2. Это связано с тем, что вследствие высокой чувствительности возможно искажение корреляционной картины между компонентами и исследуемым свойством (температурой плавления) в отсутствие даже слабо коррелированных структурных характеристик. Как видно из табл. 1, для повторного (окончательного) расчета были выбраны только те структурные характеристики, значение нагрузки которых в первую компоненту превышает указанный инфимум коэффициента корреляции, что и дает весьма наглядную зависимость компоненты от температуры. Сами же главные ком-

поненты представляют собой новые комбинированные структурные характеристики, составляющие в многомерном пространстве базис из ортогональных векторов, которые не преобразуются при изменении числа изначально вводимых данных, что является неоспоримым преимуществом данного метода. Численные значения компонент – безразмерные нормированные и центрированные величины, которые дают возможность более определенного их сопоставления с интересующими свойствами. Самое существенное влияние на температуру плавления оказывают такие характеристики, как число атомов водорода при азоте (очевидно, существенно влияющее на число и прочность образуемых им водородных связей), разветвленность цепи, наличие энергетически выгодных пяти- и шестичленных циклов и, безусловно, молекулярная масса, опосредованно связанная с площадью ван-дер-ваальсовой оболочки молекулы.

Анализ кристаллических структур по данным, взятым из Кембриджского банка (CSD) [8], показал, что температура плавления уменьшается при уменьшении размерности реализованного в кристалле мотива водородных связей (табл. 2). То же самое наблюдается и на описанной выше корреляционной диаграмме: чем выше размерность мотива водородных связей в кристаллической структуре, тем более отрицательным становится собственное значение главной компоненты в ходе анализа.

На основании сказанного можно заключить, что, имея кристаллоструктурные данные соединений

Т а б л и ц а 1

Нагрузки характеристик молекулярной структуры в значения трех главных компонент (PCA 1, PCA 2, PCA 3)

Структурный параметр	PCA 1	PCA 2	PCA 3
Гомологичность амиду	0,541	-0,471	0,066
Наличие двух N–H-связей	-0,843	0,159	-0,045
Рост цикла	-0,467	-0,749	0,213
Рост по α -положению	0,247	0,627	0,689
Наличие пяти- или шестичленного цикла	-0,730	0,169	-0,331
Разветвление основной цепи	0,779	0,445	-0,199
«Стягивание» электронной плотности с азота	0,490	-0,002	-0,819
Молекулярная масса	0,795	-0,427	0,214

Т а б л и ц а 2

Мотивы водородных связей в кристаллических структурах

Номер соединения	Рефкод в CSD	Сингония, ПГ, число молекул в электронной ячейке	Структурный мотив и длина водородной связи N...O (Å)
1	CPRPCX10	моноклинная, $P2_1/c$, $Z = 2$	каркас: 2,96; 2,92; 3,00; 2,99
16	FEPNAP	триклинная, $P-1$, $Z = 1$	ленты: 2,97; 3,03
5	BARFEF	моноклинная, $C2/c$, $Z = 1$	слои: 2,86; 2,93
6	NILYAI	триклинная, $P-1$, $Z = 1$	димеры: 2,92
13	CAPLAC	моноклинная, $C2/c$, $Z = 1$	димеры: 2,90

гомологического ряда и результаты проведенного расчета методом главных компонент, можно предсказывать температуру плавления неизвестного соединения, принадлежащего данному ряду, с точностью, позволяющей отличить его от остальных. На рис. 2 выделены участки линейности, которые соответствуют близким по строению соединениям, например, 6–9–13, 8–14–15, 7–11–10–12.

Таким образом, изучение взаимосвязи между температурой плавления и характеристиками кристаллических структур (в частности H-связями) на осно-

ве метода главных компонент может быть полезно в применении к предсказанию еще не установленных значений температуры плавления на основании эмпирических данных кристаллографических исследований. Использование комбинированного метода, включающего в себя анализ главных компонент и анализ кристаллографических данных, позволяет кластеризовать соединения и объяснить тенденции в изменении температуры, что может лечь в основу чрезвычайно простой методики оценки и предсказания неизвестных значений температуры плавления.

СПИСОК ЛИТЕРАТУРЫ

1. Браверман Э.М., Мучник И.Б. Структурные методы обработки эмпирических данных. М., 1983.
2. Bacchi A., Pelizzi G., Nebuloni M., Ferrari P. // J. Med. Chem. 1998. **41**. P. 2319.
3. Сычев С.Н., Сычев К.С., Ланин С.Н. // Вестн. Моск. ун-та. Сер. 2. Химия. 2000. **41**. С. 28.
4. Pomerantsev A.L., Rodionova O.Ye. // J. Chemometrics. 2012. **26**. P. 299.

5. Банару А.М., Словоохотов Ю.Л. // Вестн. Моск. ун-та. Сер. 2. Химия. 2010. 51. С. 435.
6. SPRESI^{web}, InfoChem Gesellschaft für chemische Information mbH, <http://www.spresi.com/>
7. Roberts A., Withers P. StatistiXL, release 1.7. Australia, University of Western Australia, 2007.
8. Allen F.H. // Acta Cryst. 2002. B58. P. 380.

Поступила в редакцию 20.03.13

MULTIVARIATE ANALYSIS OF MELTING POINTS FOR HOMOLOGOUS SERIES $C_xH_{2x-1}NO$ ($x = 3-6$)

D.B. Eremin, A.M. Banaru

(Division of Physical Chemistry)

A multivariate analysis of empirical structure data for isomeric homologues by use of principal component analysis was carried out, and the correlation with experimental melting points was shown. Based on CSD structure data we discuss the impact of molecular H-aggregation.

Key words: principal component analysis, melting point, molecular structure, hydrogen bond.

Сведения об авторах: *Еремин Дмитрий Борисович* – студент химического факультета МГУ (mit.eremin@me.com); *Банару Александр Михайлович* – ст. преподаватель кафедры физической химии химического факультета МГУ, канд. хим. наук (banaru@phys.chem.msu.ru).