

УДК 577.113

О.С. Завьялова, доцент, канд. физ.-мат. наук,

Н.В. Петренко, доцент, канд. физ.-мат. наук,

П.А. Сухонос, студент

Севастопольский национальный технический университет

ул. Университетская, 33, г. Севастополь, Украина, 99053

E-mail: bprc@ukr.net

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭКСПЕРИМЕНТАЛЬНЫХ ЗНАЧЕНИЙ ХИМИЧЕСКИХ СДВИГОВ ПРОТОНОВ ДЕЗОКСИОЛИГОНУКЛЕОТИДОВ РАЗЛИЧНОГО СОСТАВА И ДЛИНЫ НА ОСНОВЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ

Методом главных компонент проведен анализ экспериментальных значений химических сдвигов необменивающихся протонов самокомплементарных дезоксиолигонуклеотидов различного состава и последовательности оснований в цепи. Выявлены признаки, наиболее существенно влияющие на магнитное экранирование протонов рассмотренных коротких фрагментов ДНК.

Ключевые слова: *дезоксиолигонуклеотид, химический сдвиг, триплет, главные компоненты.*

Введение. ЯМР-спектроскопия является одним из эффективных экспериментальных методов исследования молекулярных систем в растворе. Она дает возможность получить информацию о пространственной организации и динамических особенностях исследуемых соединений на атомном уровне. Химический сдвиг, как один из экспериментальных параметров ЯМР, непосредственно связан с электронным и магнитным окружением молекул и позволяет изучить конформационные особенности соединений непосредственно в растворе.

Целью настоящей работы являлось проведение сравнительного анализа резонансных сигналов протонов самокомплементарных дезоксиолигонуклеотидов, отличающихся длиной и чередованием оснований в цепи. Тетрануклеотиды, рассмотренные в работе, представлены следующими пятью последовательностями: 5'-d(ACGT), 5'-d(AGCT), 5'-d(GCGC), 5'-d(CGCG), 5'-d(TGCA), гексануклеотиды – 5'-d(CGTAACG), 5'-d(CGCGCG), 5'-d(TACGTA), октамер – 5'-(GACATGTC) [1-3]. Каждый из дезоксиолигонуклеотидов может быть разбит на триплеты, отличающиеся составом и локализацией, что позволяет проанализировать особенности магнитного экранирования протонов центрального в триплете основания и выявить структуры, обладающие отклонениями от стандартной геометрии двойной спирали, в частности, это касается одноцепочечных и шпильчатых структур ДНК.

Методы исследования. В качестве метода статистической обработки данных использовался Метод главных компонент (Principal Components Analysis PCA) [4], реализованный в программной среде Statistica 6.0. PCA относится к методам снижения размерности многомерного пространства. Он позволяет без существенной потери информации перейти от первоначальной системы большого числа наблюдаемых взаимосвязанных факторов к системе существенно меньшего числа скрытых (ненаблюдаемых) факторов, определяющих вариацию первоначальных признаков. Известно, что статистические методы обработки дают надежные результаты в случае использования достаточно больших массивов данных. Поэтому, для повышения эффективности численного эксперимента в исходную базу данных были дополнительно введены приведенные в [5] значения химических сдвигов соответствующих протонов азотистых оснований, из триплетов того же состава, что и рассмотренные в данной работе. Это обоснованно тем, что, как ранее было показано в работе [6], данные С. Altona [5] имеют хорошую корреляцию с экспериментальными значениями химических сдвигов протонов самокомплементарных последовательностей, изученных в нашей лаборатории [1-3], и следовательно могут быть использованы совместно для повышения достоверности численного эксперимента.

Результаты и обсуждение. На рисунке 1, в качестве примера, приведено разбиение на триплеты двух гексамерных последовательностей.



Рисунок 1 – Разбиение самокомплементарных гексамеров на триплеты

В зависимости от длины последовательности триплеты одного состава могут быть как терминальными, так и внутренними; терминальные триплеты, в свою очередь, могут принадлежать как к 5'-, так и к 3'- концу цепи; центральное основание может быть окружено двумя пуринами, пиримидинами или какой-либо чередующейся комбинацией оснований. Все эти «признаки» могут быть использованы для выявления особенностей экранирования протонов азотистых оснований методом главных компонент. В таблице 1, в качестве примера, приведены значения химических сдвигов протонов азотистых оснований и сахарных остатков центральных нуклеотидов, входящих в триплеты, самокомплементарных дезоксигексануклеотидов [2].

Для проведения первого численного эксперимента методом главных компонент был использован массив данных, включающий химические сдвиги протонов азотистых оснований всех триплетных последовательностей рассмотренных в работе. На рисунке 2 видны три компактные группы, в каждой из которых точки практически сливаются, что означает положительную корреляцию экспериментальных данных. Действительно, область «1» соответствует протонам центрального аденина в триплете, область «2» – гуанина, в области «3» представлены протоны центрального пиримидина: тимина или цитозина.

Таблица 1 – Химические сдвиги протонов (млн⁻¹) центральных нуклеотидов, входящих в состав триплетов самокомплементарных дезоксигексануклеотидов 5'-C₁G₂T₃A₄C₅G₆¹⁾, 5'-C₁G₂C₃G₄C₅G₆²⁾, 5'-T₁A₂C₃G₄T₅A₆³⁾

	H8/H6	H2/H5	H1'	H2'	H2''	H3'	H4'	H5'	H5''
YAY									
5'-TAC	8,40(4)	7,64(4)	6,26(3)	2,81(3)	2,92(3)	5,02(2)	4,42(2)	4,16(3)	4,09(3)
5'-T ₁ A ₂ C ₃ ³⁾	8,41	8,08	6,69	2,91	2,91	5,00	4,40	4,07	4,00
TAC	8,28(2)	7,46(3)	6,19(2)	2,71(2)	2,86(2)	5,02(2)	4,42(2)	4,16(3)	4,09(3)
T ₃ A ₄ C ₅ ¹⁾	8,35	7,62	6,26	2,75	2,88	5,06	4,45	4,18	4,05
YGY									
5'-CGT	8,00(4)	-	6,04(3)	2,71(3)	2,84(3)	4,98(2)	4,38(2)	4,14(6)	4,10(6)
5'-C ₁ G ₂ T ₃ ¹⁾	8,02	-	6,00	2,72	2,78	5,00	4,38	4,12	4,02
CGT	7,88(2)	-	5,97(2)	2,61(2)	2,78(2)	4,98(2)	4,38(2)	4,14(6)	4,10(6)
C ₃ G ₄ T ₅ ³⁾	7,92	-	5,97	2,65	2,61	4,93	4,38	4,13	4,13
5'-CGC	7,99(4)	-	5,95(3)	2,73(3)	2,78(3)	4,96(2)	4,34(2)	4,09(2)	4,00(3)
5'-C ₁ G ₂ C ₃ ²⁾	8,00	-	5,94	2,74	2,77	5,01	4,38	4,12	3,95
CGC	7,87(2)	-	5,88(2)	2,63(2)	2,72(2)	4,96(2)	4,34(2)	4,09(2)	4,00(3)
C ₃ G ₄ C ₅ ²⁾	7,94	-	5,92	2,68	2,73	4,98	4,38	4,15	3,95
RCR									
ACG-3'	7,19(3)	5,33(4)	5,64(3)	1,85(3)	2,28(2)	4,80(2)	4,13(2)	4,13(3)	4,14(3)
A ₄ C ₅ G ₆ ¹⁾ -3'	7,31	5,39	5,64	1,87	2,29	4,80	4,16	4,16	4,15
ACG	7,19(3)	5,24(3)	5,57(3)	1,85(3)	2,28(2)	4,80(2)	4,13(2)	4,12(2)	4,10(3)
A ₂ C ₃ G ₄ ³⁾	7,37	5,47	5,81	1,97	2,42	4,82	4,21	4,13	4,13
GCG	7,26(2)	5,30(2)	5,67(2)	1,89(3)	2,32(2)	4,80(2)	4,15(2)	4,14(4)	4,10(4)
G ₂ C ₃ G ₄ ²⁾	7,39	5,45	5,77	2,08	2,44	4,90	4,22	4,02	4,02
GCG-3'	7,26(2)	5,39(4)	5,74(3)	1,89(3)	2,32(2)	4,80(2)	4,15(2)	4,14(4)	4,10(4)
G ₄ C ₅ G ₆ ²⁾ -3'	7,35	5,48	5,78	1,93	2,37	4,84	4,17	4,12	4,12
RTR									
GTA	7,24(2)	-	5,72(4)	2,07(4)	2,43(5)	4,88(3)	4,21(3)	4,20(9)	4,11(9)
G ₂ T ₃ A ₄ ¹⁾	7,32	-	5,67	2,16	2,46	4,19	4,22	-	-
GTA-3'	7,24(2)	-	5,79(4)	2,07(4)	2,43(5)	4,88(3)	4,21(3)	4,20(9)	4,11(9)
G ₄ T ₅ A ₆ ³⁾ -3'	7,23	-	5,94	1,98	2,19	4,66	4,12	3,70	3,45

^{*)} жирным шрифтом представлены данные из работы [5], где в скобках указаны величины погрешностей химических сдвигов в второй цифре после запятой

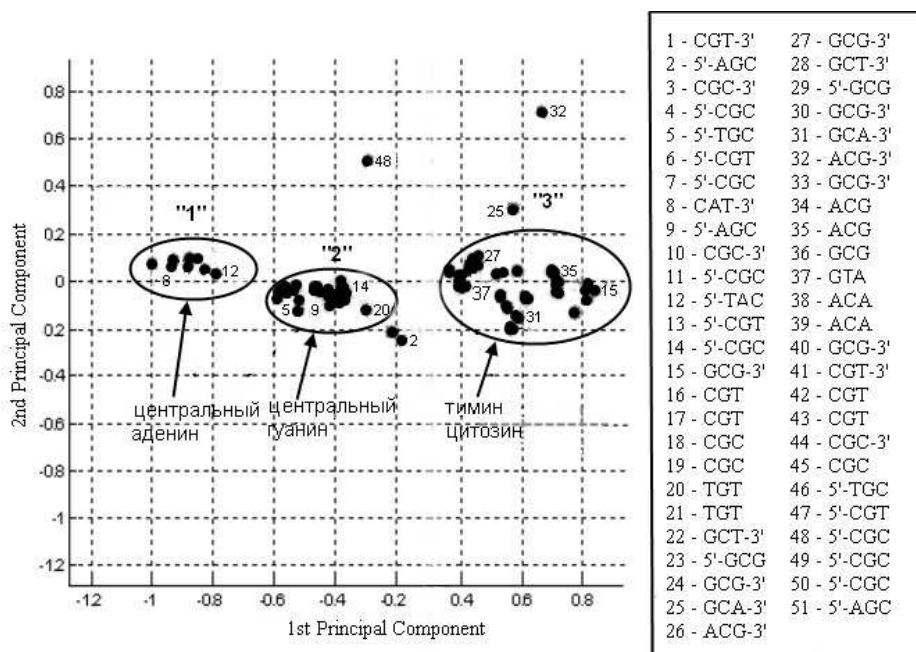


Рисунок 2 – Диаграмма главных компонент PC1-PC2 для триплетов различного состава

Интересным является тот факт, что химические сдвиги протонов пиримидинов более тесно связаны между собой, образуя одну область, в то время как пурины дают две группы – для А и Г. Кроме того, области «1, 2» и «3» располагаются вдоль оси PC1 по разные стороны от точки начала отсчета (0,0), что в рамках метода PCA свидетельствует о высокой отрицательной корреляции присущих им свойств. Это можно объяснить тем, что химические сдвиги пуриновых оснований лежат в более слабом поле, чем пиримидиновых, что согласуется с данными работы [5]. Так как абсолютное большинство точек расположено вдоль оси PC1 – это означает, что первая главная компонента является определяющей в таком наборе данных и характеризует принадлежность центрального основания к пуринам и пиримидинам.

Во втором численном эксперименте использовался массив данных, содержащий триплеты с центральным цитозином и гуанином. При этом наблюдаемая картина распределения по компонентам существенно изменилась. Рассмотрим график счетов, который показывает, как расположились «образцы» в новом пространстве. На графике младших счетов PC1-PC2, показанном на рисунке 3, видно несколько отдельных групп с тесной внутренней корреляцией. В таком эксперименте принадлежность центрального основания к пуринам-пиримидинам теряет основную нагрузку и определяющим становится другой признак. На рисунке 3 видно, что в отличие от первого эксперимента данные распределены по всей плоскости компонент PC1, PC2. Это означает, что в таком массиве данных оба признака, отвечающие за первую и вторую компоненты, достаточно существенны. Причем, относительно оси PC1 группы расположены практически вертикально, что говорит о наличии преобладающего свойства, характеризующего ось этой компоненты. Если проанализировать график счетов совместно с исходным массивом данных, то можно выяснить, что таким свойством является длина последовательности, из которой взяты триплеты.

В экспериментальной базе данных представлены триплеты, входящие в состав последовательностей различной длины, и, хотя среди них встречаются триплеты одинакового состава, левая вертикальная группа на рисунке 3 представлена триплетами, выделенными из тетрануклеотидов, средняя – из гексануклеотидов, и правая – из октамера. Так, например, триплеты состава GCG, CGC входят в «тетра-область» под номерами 29, 11, а в «гекса-область» – под номерами 37, 7. Это подтверждает смысл первой компоненты (PC1)- разделение по длине нуклеотидной последовательности. Что касается второй компоненты, она отвечает за местоположение триплета в цепи, а именно – примыкание к 5'-, 3'- концу, либо внутреннее положение. Причем, в нижней части графика, на рисунке 3, располагаются триплеты, примыкающие к 3'-концу, выше по оси PC2 – внутренние триплеты, а затем – триплеты соответствующие 5'-концу. Так, например, триплет состава CGC, примыкающий к 3'- концу, расположен в левом нижнем углу диаграммы под номером 10, являющийся центральным практически в центре диаграммы под номером 19, а принадлежащий к 5'- концу – в левом верхнем углу диаграммы под номером 11.

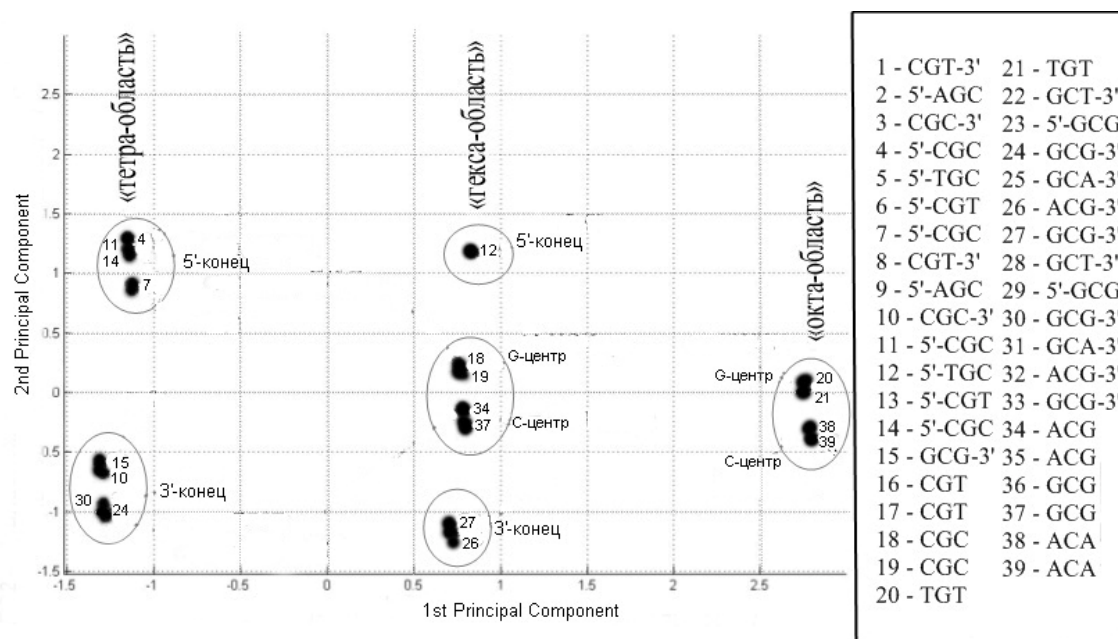


Рисунок 3 – Диаграмма главных компонент PC1-PC2 для триплетов различного состава с центральным цитозином и гуанином

Необходимо отметить тот факт, что в пространстве PC1-PC2, переменные, характеризующие химические сдвиги областей «2» и «4» для нагрузки второй компоненты, близки к нулю – это значит, что значения химических сдвигов оснований внутренних триплетов претерпевают значительно меньшие изменения в сравнении с концевыми, что можно объяснить эффектом плавления концевых пар. На рисунке 4 показано пространство компонент PC1-PC3. Вдоль оси первой компоненты PC1 группы триплетов расположены строго по принадлежности к последовательностям различной длины и образуют «тетра-, гекса- и окта-области».

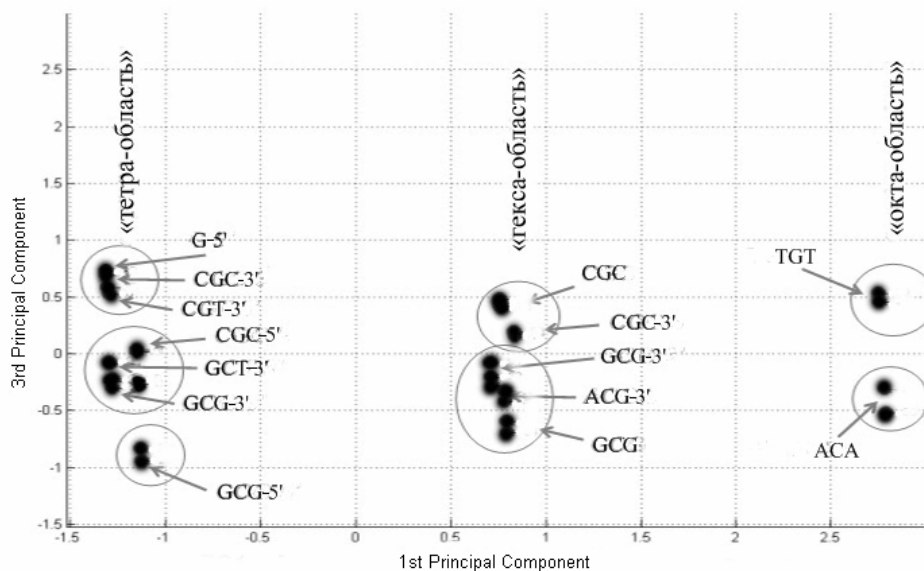


Рисунок 4 – Диаграмма главных компонент PC1-PC3 для триплетов различного состава с центральным цитозином и гуанином

Рассмотрим диаграмму старших счетов PC2-PC3, приведенную на рисунке 5.

Во-первых, она подтверждает нагрузку второй главной компоненты PC2 – местоположение триплета в цепи, во-вторых, из графика видно, что направление оси третьей компоненты PC3 отвечает за принадлежность центрального основания к пуринам или пиримидинам. Верхняя полуплоскость по оси PC3 соответствует центральному гуанину в триplete, а нижняя – центральному цитозину.

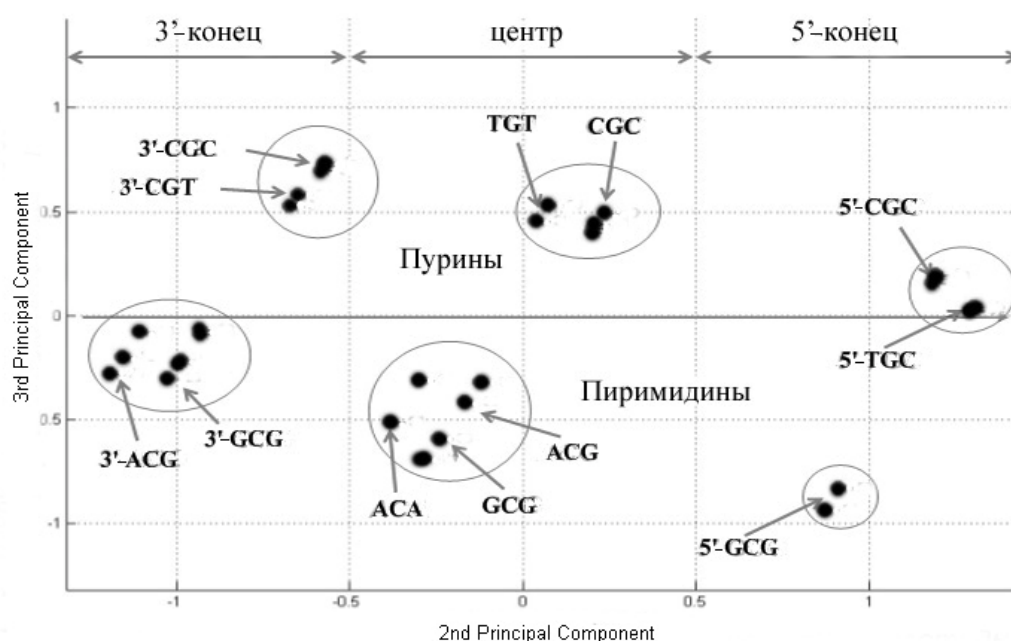


Рисунок 5 – Диаграмма главных компонент PC2-PC3 для триплетов различного состава с центральным цитозином и гуанином

Последующее выделение «более старших» компонент PC4, PC5 и т.д. в данном массиве данных не позволило соотнести с ними каких-либо существенных признаков. Матрица нагрузок главных компонент на исходные признаки, соответствующая данному численному эксперименту, приведена в таблице 2. Из таблицы видно, что на первые три главные компоненты приходится порядка 99% суммарной вариации данных. Это означает, что из всех признаков, которыми обладает использованная экспериментальная база данных, наиболее существенными, влияющими на значения экранирования протонов, оказываются следующие: длина нуклеотидной последовательности (PC1), местоположение триплета в нуклеотидной цепи (PC2), его принадлежность к пуринам или пиримидинам (PC3). При этом наибольшую нагрузку на главные компоненты несет такой признак, как длина олигонуклеотидной последовательности. Это, по всей видимости, связано с эффектом плавления концевых нуклеотидных пар, который в случае коротких последовательностей дает более существенный вклад в экранирование по сравнению с последовательностями большей длины.

Таблица 2 – Матрица нагрузок главных компонент на исходные признаки

Главная компонента PC _j	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Вклад j-главной компоненты, %	64,234	28,042	6,621	0,621	0,281	0,095	0,041	0,031	0,022	0,012
Суммарный вклад, %	64,234	92,276	98,897	99,518	99,799	99,894	99,935	99,966	99,988	100

Следует подчеркнуть тот факт, что влияние «ближайших соседей» (руг-руг или руг-руг окружение) в триплете на экранирование протонов, оказалось значительно меньшим, чем предполагалось первоначально. По значимости этот признак не вошел в первую тройку PC1-PC3.

Вывод. Таким образом, на основе исследований, проведенных выше, могут быть сделаны вполне определенные выводы об особенностях конформационных состояний молекулы ДНК, проведен анализ влияния плавления концевых нуклеотидных пар в коротких фрагментах ДНК на геометрию двойной спирали. Следует отметить, что для более детального выявления закономерностей экранирования протонов ароматических оснований и дезоксирибозы, необходимо рассмотрение более широкого спектра триплетных последовательностей, принадлежащих как к само-, так и несамокомплементарным олигонуклеотидам различного состава и последовательности оснований в

цепи. Представленная в настоящей работе методика статистического анализа протонных химических сдвигов методом главных компонент может быть использована в дальнейшем для решения различных задач структурного анализа в молекулярной биофизике, в частности, для выявления структур, обладающих отклонениями от стандартной геометрии двойной спирали ДНК.

Библиографический список использованной литературы

1. Веселков А.Н. 1М- и 2М-1Н-ЯМР-исследование самоассоциации дезокситетрануклеозид-трифосфатов различной последовательности оснований в водном растворе / А.Н. Веселков, Д. Дэвис, Л. Н. Дымант, Х. Паркес, Д. Шипп // Біополімери і клітина. — 1991. — Т. 7. — № 5. — С. 15–22.
2. Веселков А.Н. ЯМР-исследование самоассоциации молекул дезоксигексануклеотидов 5'-d(CpGpTpApCpG) и 5'-d(CpGpCpGpCpG) в водном растворе / А.Н. Веселков., В.И. Пахомов, Р. Дж. Итон, Д. Дэвис // Биофизика. — 2000. — Т. 45. — С. 20–26.
3. Веселков А.Н. Анализ взаимодействия бромистого этидия с октамером ДНК 5'-d(GpApCpApTpGpTpC) в водном растворе по данным ¹H ЯМР спектроскопии / А.Н. Веселков, Л.Н. Дымант, В.И. Пахомов [и др.] // Журн. структур. химии. — 1999. — Т. 40. — С. 265–275.
4. Эсбенсен К. Анализ многомерных данных / К. Эсбенсен; сокр. пер. с англ. под ред. О. Родионовой. — М: Изд-во ИПХФ РАН, 2005. — 204с.
5. Altona C. Double-helical DNA ¹H chemical shifts: an accurate and balanced predictive empirical scheme / C. Altona, D.H. Faber, A.J.A. Westra Hoekzema // Magn. Reson. Chem. — 2000. — V. 38. — P. 95–107.
6. Барановский С.Ф. Особенности вторичной структуры коротких фрагментов ДНК по экспериментальным значениям протонных химических сдвигов / С.Ф. Барановский, О.С. Завьялова, А.И. Печенкина, О.В. Рогова // Вестник СевНТУ. Сер. Физика и математика: сб. науч. тр. — Севастополь, 2004. — Вып. 33. — С. 62–68.

Поступила в редакцию 24.09.2010 г.

Зав'ялова О.С., Петренко Н.В., Сухоніс П.О. Порівняльний аналіз експериментальних значень хімічних зсувів протонів дезоксиолігонуклеотидів різного складу та довжини на основі Методу головних компонент

Анотація. Проведено порівняльний аналіз експериментальних значень хімічних зсувів протонів самокомплементарних дезоксиолігонуклеотидів різного складу та послідовності основ у нуклеотидному ланцюгу, методом головних компонентів. Виявлені найвпливовіші ознаки на екранування протонів розглянутих коротких фрагментів ДНК.

Ключові слова: дезоксиолігонуклеотід, хімічний зсув, триплет, головні компоненти.

Zavyalova O.S., Petrenko N.V., Sukhonos P.A. Comparative analysis of experimental values of protons chemical shifts of oligonucleotides of various structure and length based on the Principal Components Analysis method

The analysis of experimental values of chemical shifts of non-exchanging protons of deoxyoligonucleotides of various structure and sequence is carried out using the method of Principal Components Analysis. Most important factors influencing the shielding of protons of the considered short fragments of DNA are revealed.

Keywords: deoxyoligonucleotide, chemical shift, triplet, the principal components.