

МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ МЕДИЦИНСКИЙ УНИВЕРСИТЕТ  
КАФЕДРА ОБЩЕЙ ХИМИИ  
КАФЕДРА БИОЛОГИИ

**А. В. БУТВИЛОВСКИЙ, Е. А. ЧЕРНОУС**

# **МЕТОДЫ ОПРЕДЕЛЕНИЯ ХАРАКТЕРА АМИНОКИСЛОТНЫХ ЗАМЕН**

Учебно-методическое пособие



Минск БГМУ 2008

УДК 577.112.3 (075.8)  
ББК 28.072 я73  
Б 93

Утверждено Научно-методическим советом университета в качестве  
учебно-методического пособия 24.10.07 г., протокол № 2

А в т о р ы: А. В. Бутвиловский (гл. 1, 2); Е. А. Черноус (гл. 1)

Р е ц е н з е н т ы: лауреат Государственной премии Республики Беларусь,  
д-р биол. наук, проф. Е. В. Барковский; канд. мед. наук, доц. О. Н. Ринейская

**Бутвиловский, А. В.**

Б 93 Методы определения характера аминокислотных замен : учеб.-метод. пособие /  
А. В. Бутвиловский, Е. А. Черноус. – Минск : БГМУ, 2008. – 28 с.

ISBN 978–985–462–817–2.

Рассмотрены наиболее часто используемые в эволюционной биохимии и молекулярной  
эволюции методы определения характера аминокислотных замен. Приведен пример применения  
данных методов для решения прикладных задач.

Предназначено для студентов 1-го курса всех факультетов, слушателей факультативного кур-  
са «Молекулярная эволюция», а также может быть использовано студентами старших курсов  
и всеми, кто интересуется вопросами данной отрасли науки.

УДК 577.112.3 (075.8)  
ББК 28.072 я73

Учебное издание

**Бутвиловский Александр Валерьевич**  
**Черноус Евгений Анатольевич**

## **МЕТОДЫ ОПРЕДЕЛЕНИЯ ХАРАКТЕРА АМИНОКИСЛОТНЫХ ЗАМЕН**

Учебно-методическое пособие

Ответственный за выпуск Е. В. Барковский  
Редактор О. В. Иванова  
Компьютерная верстка Н. М. Федорцовой

Подписано в печать 25.10.07. Формат 60×84/16. Бумага писчая «КюмЛюкс».

Печать офсетная. Гарнитура «Times».

Усл. печ. л. 1,63. Уч.-изд. л. 1,14. Тираж 60 экз. Заказ 221.

Издатель и полиграфическое исполнение –

Белорусский государственный медицинский университет.

ЛИ № 02330/0133420 от 14.10.2004; ЛП № 02330/0131503 от 27.08.2004.

220030, г. Минск, Ленинградская, 6.

ISBN 978–985–462–817–2

© Оформление. Белорусский государственный  
медицинский университет, 2008

## Введение

Анализ и сравнение аминокислотных последовательностей белков является составной частью многих исследований в области молекулярной биологии, биохимии, молекулярной эволюции и других отраслях науки.

Достаточно часто требуется качественно и количественно оценить сходство (и/или различие) сравниваемых аминокислотных последовательностей. Для этого изучаются их аминокислотный состав, степень различий первичной структуры (эволюционные дистанции), вторичная структура, характер наблюдаемых аминокислотных замен и другие показатели.

В Республике Беларусь отсутствует обобщенная информация по методам определения характера аминокислотных замен (аминокислотным индексам). Вместе с тем существует большое количество методов этой группы, что затрудняет выбор исследователей.

Первая попытка систематизации и классификации данных методов была предпринята в 1988 г.: К. Накай и соавт. проанализировали 222 аминокислотных индекса и разделили их на 4 группы, основанных на анализе:

- 1) поворотов и  $\alpha$ -спиралей;
- 2)  $\beta$ -структур;
- 3) гидрофобности;
- 4) физико-химических свойств.

Данная классификация была дополнена в 1996 г. К. Томии и М. Канехиса в результате анализа ими большего числа аминокислотных индексов (402) и 42 матриц аминокислотных замен.

В 2000 г. с целью систематизации накопленных данных Ш. Кавашима и М. Канехиса разработали базу индексов AAindex ([www.genome.jp/aaindex](http://www.genome.jp/aaindex)), позволяющих оценить характер аминокислотных замен. Данная база (версия 9.1) состоит из 3 разделов:

1. AAindex1 содержит 544 аминокислотных индекса, числовые значения которых характеризуют различные физико-химические и биохимические свойства взаимозаменяемых аминокислот. При выборе любого индекса база данных дает справочную информацию: краткое описание методики его расчета, рекомендованную литературу, а также ссылки на индексы в пределах группы (формируется по принципу высокой межиндексной корреляции ( $r \geq 0,80$ )).

2. AAindex2 содержит 94 матрицы аминокислотных замен в формате  $20 \times 20$ .

3. AAindex3 включает 47 матриц взаимодействий боковых радикалов аминокислот в белках.

Наличие регулярно обновляемой базы данных позволяет оперативно получать всю необходимую информацию и знакомиться с новыми методами этой группы.

В данном издании<sup>1</sup> внимание сконцентрировано на наиболее часто используемых методах определения характера аминокислотных замен.

## Глава 1

---

<sup>1</sup> Подготовлено при финансовой поддержке ЕНЧР И (грант № B06M-060 от 01.04.06 г.)

# Характеристика наиболее часто используемых методов определения характера аминокислотных замен

## 1.1. МЕТОД П. СНИТА

В 1966 г. П. Снит предложил метод изучения характера аминокислотных замен, в основе которого лежит анализ всех доступных для того времени физико-химических свойств аминокислот. При этом учитывалось любое свойство аминокислоты, которое могло повлиять на биологическую активность белка. Например, замена водорода на ОН-группу в пара-положении молекулы фенилаланина приводит к образованию тирозина (формулы аминокислот приведены в прил.). Данная замена имеет множество эффектов, каждый из которых может повлиять на биологическую активность (отрицательный заряд при высоком рН, большой объем бокового радикала и т. д.). Поэтому каждый из данных эффектов должен учитываться. Вместе с тем достаточно сложно установить степень влияния каждого эффекта на биологическую активность белка, поэтому априори был принят равный вклад каждого эффекта в изменение биологической активности. Проанализированные П. Снитом физико-химические свойства аминокислот перечислены в табл. 1. Если свойство может быть оценено количественно, то оно дополнительно подразделяется на классы.

Таблица 1

**Физико-химические свойства аминокислот, проанализированные П. Снитом**

Классы	Свойства
1–5	Присутствие соответственно: $\beta$ -CH <sub>2</sub> ; $\gamma$ -CH <sub>2</sub> ; $\delta$ -CH <sub>2</sub> (пролин учтен); $\varepsilon$ -CH <sub>2</sub> и $\alpha$ -CH <sub>3</sub> групп
6–10	Присутствие соответственно: $\omega$ -SH; $\omega$ -COOH; $\omega$ -NH <sub>2</sub> ; $\omega$ -CONH <sub>2</sub> и -ОН групп в боковой цепи
11–15	Присутствие соответственно: бензольного кольца (триптофан учтен); разветвление в боковом радикале; второй CH <sub>3</sub> -группы; только два атома Н в конце боковой цепи (пролин учтен) и -S-CH <sub>3</sub> группы
16–20	Присутствие соответственно: гуанидильной группы; $\alpha$ -NH <sub>2</sub> ; $\alpha$ -NH группы в цикле; группы $\delta$ -NH в цикле; пиридинового атома N в цикле
21–25	Присутствие соответственно: -CH=N-; ядра индола; ядра имидазола; C=O группы в боковой цепи; и конфигурации у $\alpha$ -C, потенциально изменяющей направление пептидной цепи (включен только пролин)
26–30	Присутствие соответственно: атома серы; первичной алифатической -ОН группы; вторичной алифатической -ОН группы; фенольной -ОН группы; способности образовывать дисульфидные мостики
31–35	Присутствие соответственно: -NH группы в ядре имидазола; -NH группы в ядре индола; -SCH <sub>3</sub> группы; второго хирального центра; N=CR-NH группы

Классы	Свойства
36–40	Присутствия соответственно: изопропильного радикала; ароматичности; сильной ароматичности; терминальный положительный заряд; отрицательный заряд при $pH > 7$ (тирозин учтен)
41	Присутствие пирролидинового кольца
42–53	Молекулярные массы боковой цепи (приблизительно), рассчитанные в 12 аддитивных шагах (сера учтена как эквивалент двух атомов углерода, азота или кислорода)
54–56	Присутствие соответственно: 5-, 6- и 9-членной плоскостной циклической системы
57–64	$pK$ в изоэлектрической точке, рассчитанной аддитивно ступенчато для $pH = 1$
65–68	Логарифм растворимости в воде L-изомера в мг/100 мл, рассчитанный аддитивно
69–70	Удельное оптическое вращение в 5 N-HCl, $[\alpha]_D$ от 0 до $-25$ , и менее $-25$ соответственно
71–72	Удельное оптическое вращение в 5 N-HCl, $[\alpha]$ от 0 до $+25$ соответственно (значения для глутамина и триптофана с водой как растворителем, а для аспарагина с 3-4 N-HCl)
73–74	Образование водородной связи боковой цепи (ионный тип), сильный донор и сильный акцептор соответственно
75–76	Образование водородной связи боковой цепи (нейтральный тип), сильный донор и сильный акцептор соответственно
77–78	Способность гидратироваться, умеренная и сильная соответственно
79	Разрушители водных структур (хорошо гидратирующиеся)
80–82	Мобильные электроны: немного; умеренно и много соответственно (рассчитано аддитивно)
83–85	Температурная и временная стабильность: умеренная; высокая и очень высокая соответственно (рассчитана аддитивно)
86–89	$R_F$ в фенольно-водной хроматографии на бумаге ступенчато 0–2 (рассчитано аддитивно)
90–93	$R_F$ в толуол-пиридин-гликольхлоргидрин (хроматографии на бумаге производной DNP) ступенчато 0–2 (рассчитано аддитивно: для лизина ди-DNP производное)
94–97	Цвет нингидрина после коллидин-лутидин хроматографии и нагревания 5 мин в $100\text{ }^\circ\text{C}$ : пурпурный; розовый; коричневый и желтый соответственно.
98	Разветвление в конце боковой цепи
99–101	Число замещений у $\beta$ -углеродного атома, соответственно 1, 2 или 3 (рассчитано аддитивно)
102–111	Среднее число электронных пар на боковой цепи, не участвующих в сопряжении (рассчитано аддитивно)
112–115	Количество связей в боковой цепи, позволяющих вращение (рассчитано аддитивно)
116–117	Ионный объем в пределах колец: небольшой или умеренный (рассчитано аддитивно)
118–124	Максимальный момент инерции для вращения в $\alpha$ - $\beta$ связи (рассчитано аддитивно в семи приблизительных шагах)

Классы	Свойства
125–131	Максимальный момент инерции для вращения в $\beta$ – $\gamma$ связи (рассчитано аддитивно в семи приближительных шагах)
132–134	Максимальный момент инерции для вращения в $\gamma$ – $\delta$ связи (рассчитанный аддитивно в трех приближительных шагах)

Коэффициент П. Снита ( $\phi$ ), характеризующий степень подобия аминокислот, рассчитывается по следующей формуле:

$$\phi = \frac{(ad - bc)}{\sqrt{(a+c)(b+d)(a+b)(c+d)}}$$

где  $a$  — это количество признаков в табл. 1 одинаковое для аминокислот  $i$  и  $j$ ;  $b$  и  $c$  — число признаков, свойственных аминокислоте  $i$ , но не  $j$ ;  $d$  — число признаков, не свойственных ни одной из аминокислот.

Помимо данного коэффициента, П. Снит также предложил использовать показатель различия аминокислот  $D$ , который рассчитывается по формуле:

$$D = \frac{100(b+c)}{(a+b+c+d)}$$

В табл. 2 приведены значения коэффициента Снита и различия  $D$  для всех возможных пар аминокислот.

Аминокислоты обозначены по однобуквенной системе: А — аланин; С — цистеин; D — аспарагиновая кислота; Е — глутаминовая кислота; F — фенилаланин; G — глицин; Н — гистидин; I — изолейцин; К — лизин; L — лейцин; М — метионин; N — аспарагин; P — пролин; Q — глутамин; R — аргинин; S — серин; T — треонин; V — валин; W — триптофан; Y — тирозин.

При использовании данного метода замена считается консервативной, если коэффициент Снита больше 0,416 (для одношаговых замен), в обратном случае замена считается радикальной.

Таблица 2

Коэффициенты П. Снита и показатели различия аминокислот *D* (выделены жирным шрифтом)

	<b>L</b>	<b>I</b>	<b>V</b>	<b>G</b>	<b>A</b>	<b>P</b>	<b>Q</b>	<b>N</b>	<b>M</b>	<b>T</b>	<b>S</b>	<b>C</b>	<b>E</b>	<b>D</b>	<b>K</b>	<b>R</b>	<b>Y</b>	<b>F</b>	<b>W</b>	<b>H</b>
<b>L</b>		0,889	0,785	0,380	0,643	0,432	0,501	0,506	0,515	0,432	0,411	0,398	0,333	0,387	0,492	0,360	0,347	0,570	0,368	0,450
<b>I</b>	<b>5</b>		0,843	0,371	0,588	0,419	0,453	0,456	0,494	0,493	0,360	0,348	0,366	0,338	0,477	0,342	0,266	0,487	0,287	0,368
<b>V</b>	<b>9</b>	<b>7</b>		0,437	0,675	0,473	0,416	0,395	0,465	0,551	0,439	0,430	0,239	0,279	0,419	0,307	0,199	0,380	0,195	0,300
<b>G</b>	<b>24</b>	<b>25</b>	<b>19</b>		0,659	0,499	0,163	0,190	0,149	0,396	0,323	0,290	0,049	0,015	0,309	0,149	0,163	0,259	0,138	0,183
<b>A</b>	<b>15</b>	<b>17</b>	<b>12</b>	<b>9</b>		0,533	0,356	0,280	0,421	0,417	0,477	0,578	0,159	0,156	0,426	0,315	0,214	0,356	0,224	0,320
<b>P</b>	<b>23</b>	<b>24</b>	<b>20</b>	<b>17</b>	<b>16</b>		0,168	0,172	0,265	0,330	0,321	0,318	0,003	0,015	0,295	0,155	0,179	0,282	0,211	0,172
<b>Q</b>	<b>22</b>	<b>24</b>	<b>25</b>	<b>32</b>	<b>26</b>	<b>33</b>		0,589	0,699	0,340	0,501	0,482	0,685	0,492	0,545	0,561	0,368	0,459	0,353	0,406
<b>N</b>	<b>20</b>	<b>23</b>	<b>23</b>	<b>26</b>	<b>25</b>	<b>31</b>	<b>10</b>		0,518	0,488	0,581	0,485	0,578	0,637	0,401	0,427	0,391	0,340	0,316	0,459
<b>M</b>	<b>20</b>	<b>22</b>	<b>23</b>	<b>34</b>	<b>25</b>	<b>31</b>	<b>13</b>	<b>21</b>		0,409	0,480	0,612	0,402	0,292	0,482	0,522	0,307	0,465	0,355	0,345
<b>T</b>	<b>23</b>	<b>21</b>	<b>17</b>	<b>20</b>	<b>20</b>	<b>25</b>	<b>24</b>	<b>19</b>	<b>25</b>		0,668	0,485	0,218	0,254	0,224	0,258	0,285	0,254	0,176	0,208
<b>S</b>	<b>23</b>	<b>25</b>	<b>20</b>	<b>19</b>	<b>16</b>	<b>24</b>	<b>21</b>	<b>15</b>	<b>22</b>	<b>12</b>		0,613	0,312	0,330	0,285	0,317	0,354	0,380	0,243	0,342
<b>C</b>	<b>24</b>	<b>26</b>	<b>21</b>	<b>21</b>	<b>13</b>	<b>25</b>	<b>22</b>	<b>19</b>	<b>17</b>	<b>19</b>	<b>13</b>		0,221	0,243	0,269	0,324	0,223	0,289	0,188	0,288
<b>E</b>	<b>30</b>	<b>31</b>	<b>31</b>	<b>37</b>	<b>34</b>	<b>43</b>	<b>14</b>	<b>19</b>	<b>26</b>	<b>34</b>	<b>29</b>	<b>33</b>		0,840	0,435	0,382	0,261	0,219	0,086	0,201
<b>D</b>	<b>25</b>	<b>28</b>	<b>28</b>	<b>33</b>	<b>30</b>	<b>40</b>	<b>22</b>	<b>14</b>	<b>31</b>	<b>29</b>	<b>25</b>	<b>28</b>	<b>7</b>		0,248	0,236	0,287	0,172	0,028	0,200
<b>K</b>	<b>23</b>	<b>24</b>	<b>26</b>	<b>31</b>	<b>26</b>	<b>31</b>	<b>21</b>	<b>27</b>	<b>24</b>	<b>34</b>	<b>31</b>	<b>32</b>	<b>26</b>	<b>34</b>		0,733	0,285	0,381	0,297	0,421
<b>R</b>	<b>33</b>	<b>34</b>	<b>36</b>	<b>43</b>	<b>37</b>	<b>43</b>	<b>23</b>	<b>31</b>	<b>28</b>	<b>38</b>	<b>37</b>	<b>36</b>	<b>31</b>	<b>39</b>	<b>14</b>		0,407	0,339	0,288	0,396
<b>Y</b>	<b>30</b>	<b>34</b>	<b>36</b>	<b>36</b>	<b>34</b>	<b>37</b>	<b>29</b>	<b>28</b>	<b>32</b>	<b>32</b>	<b>29</b>	<b>34</b>	<b>34</b>	<b>34</b>	<b>34</b>	<b>36</b>		0,729	0,565	0,504
<b>F</b>	<b>19</b>	<b>22</b>	<b>26</b>	<b>29</b>	<b>26</b>	<b>27</b>	<b>24</b>	<b>24</b>	<b>24</b>	<b>28</b>	<b>25</b>	<b>29</b>	<b>35</b>	<b>35</b>	<b>28</b>	<b>34</b>	<b>13</b>		0,741	0,605
<b>W</b>	<b>30</b>	<b>34</b>	<b>37</b>	<b>39</b>	<b>36</b>	<b>37</b>	<b>31</b>	<b>32</b>	<b>31</b>	<b>38</b>	<b>35</b>	<b>37</b>	<b>43</b>	<b>45</b>	<b>34</b>	<b>36</b>	<b>21</b>	<b>13</b>		0,484
<b>H</b>	<b>25</b>	<b>28</b>	<b>31</b>	<b>34</b>	<b>29</b>	<b>36</b>	<b>27</b>	<b>24</b>	<b>30</b>	<b>34</b>	<b>28</b>	<b>31</b>	<b>27</b>	<b>35</b>	<b>27</b>	<b>31</b>	<b>23</b>	<b>18</b>	<b>25</b>	

## 1.2. МЕТОД М. ВОЛЬКЕНШТЕЙНА

Одним из свойств, определяющих конформацию белка, является гидрофобность составляющих его аминокислот. Пространственная структура белка во многом определяет его активность, а, следовательно, и биологическую значимость. Проведенный в 1970-е гг. М. Волькенштейном анализ 790 эволюционных замещений в ряде белков показал, что наиболее многочисленны и предпочтительны замены с минимальными изменениями гидрофобности аминокислотного остатка. Таким образом, чем меньше разность гидрофобностей двух сравниваемых аминокислот, тем выше вероятность мутационной замены одной аминокислоты на другую.

В качестве показателя взаимозаменяемости аминокислоты  $i$  на  $j$  М. Волькенштейн предложил использовать показатель разности их гидрофобностей  $\Delta H_{ij} = H_i - H_j$  (гидрофобность определяется как разность свободных энергий растворения аминокислоты в воде и этаноле). Традиционно пользуются шкалой гидрофобностей аминокислот, установленной Ч. Танфордом.

При произвольных заменах аминокислот среднее изменение гидрофобности составляет 1,28 ккал/моль. Замена одной аминокислоты на другую считается консервативной в том случае, если разность их гидрофобностей  $\Delta H < 1,28$  ккал/моль (для замен в целом) и  $\Delta H < 1,22$  ккал/моль (для одношаговых замен — обусловлены заменой одного нуклеотида в кодоне ДНК), в противном случае она считается радикальной.

В табл. 3 представлены разности гидрофобностей ( $\Delta H$ ) для 75 одношаговых аминокислотных замен.

Таблица 3

Разность гидрофобностей для 75 одношаговых аминокислотных замен

Замены	$\Delta H$	Замены	$\Delta H$	Замены	$\Delta H$	Замены	$\Delta H$	Замены	$\Delta H$
P-S	2,56	S-Y	2,83	T-I	2,53	E-D	0,01	I-N	2,96
P-T	2,16	S-T	0,40	T-A	0,19	H-R	0,67	V-F	0,96
P-A	1,97	S-I	2,93	Z-Q	2,32	H-D	0,86	V-A	1,06
P-L	0,18	S-F	2,61	L-H	1,02	H-N	1,39	V-L	0,73
P-Q	2,50	S-A	0,59	L-M	1,12	M-R	0,57	T-K	1,06
P-H	1,20	S-L	2,38	L-R	1,69	R-C	0,08	T-M	0,86
P-R	1,87	S-R	0,69	L-W	0,58	R-W	2,27	T-R	0,29
G-S	0,04	S-N	0,03	K-Q	1,40	D-N	0,53	T-N	0,44
G-V	1,69	S-C	0,61	K-E	0,95	C-W	2,35	A-E	0,08
G-A	0,63	S-W	2,96	K-M	0,20	I-V	1,28	A-D	0,09
G-E	0,55	Y-F	0,22	K-R	0,77	I-F	0,32	V-E	1,14
G-R	0,73	Y-H	1,47	K-N	1,49	I-L	0,55	V-M	0,39
G-D	0,54	Y-D	2,33	Q-E	0,45	I-K	1,47	V-D	1,15
G-C	0,65	Y-N	2,86	Q-H	1,30	I-M	1,67	F-L	0,23
G-W	3,00	Y-C	2,22	Q-R	0,63	Y-R	2,24	F-C	2,00



### 1.3. МЕТОД А. БАЧИНСКОГО

Известно, что молекулярная эволюция белков сопровождается изменением их первичной структуры. В одном случае замена аминокислоты  $i$  на  $j$  не изменит активности и функции белка, в другом приведет к инактивации фермента или его дисфункции. Для оценки степени взаимозаменяемости одной аминокислоты на другую в 1976 г. А. Бачинский предложил показатель «функциональной близости аминокислот» (ФБА).

Под ФБА подразумевается способность аминокислот заменять друг друга в белках с полным или частичным сохранением их активности. Величины ФБА были получены на основе анализа числа замен аминокислот в семействах изофункциональных белков (выполняющих сходные функции у организмов разных видов). Фиксация замены аминокислоты  $i$  на  $j$  в определенной позиции белка возможна лишь в том случае, если его активность сохранится на приемлемом для обеспечения жизнедеятельности организма уровне. Следовательно, можно предположить, что данные аминокислоты взаимозаменяемы, а замена функционально синонимична.

Чем чаще в определенной позиции встречается замена аминокислоты  $i$  на аминокислоту  $j$  в семействах изофункциональных белков, тем выше степень их взаимозаменяемости. Таким образом, проанализировав большую выборку изофункциональных белков и их семейств, можно получить относительные величины, отражающие число фиксированных замен аминокислот  $i$  на  $j$ . Безусловно, эти величины характеризуют взаимозаменяемость аминокислот.

Для 28 семейств изофункциональных белков, взятых из атласа М. Дэйхофф, А. Бачинским определены величины  $N_{ij}$ , отражающие совокупное число сайтов в их последовательностях, в которых у одних представителей находится аминокислота  $i$ , а у других  $j$ . Нет сомнений, что различная частота встречаемости аминокислот в белках ( $N_i$  и  $N_j$ ) будет влиять на показатель взаимозаменяемости аминокислот  $N_{ij}$ . Для исключения этого влияния А. Бачинский предложил следующую формулу для расчета ФБА:

$$\alpha_{ij} = \frac{2N_{ij}}{N_i + N_j},$$

где  $\alpha_{ij}$  (ФБА) — средняя доля функционально синонимичных замен аминокислоты  $i$  на аминокислоту  $j$ ;  $N_{ij}$  — взаимозаменяемость аминокислот  $i$  и  $j$ ;  $N_i$  и  $N_j$  — частоты встречаемости аминокислот  $i$  и  $j$  в последовательности белка. ФБА и матрицы взаимозаменяемости аминокислот представлены в табл. 4.

Замена одной аминокислоты на другую считается консервативной, если значение ФБА больше 12,4 (для одношаговых замен), при меньшем значении она считается радикальной.

Таблица 4

## Матрицы взаимозаменяемости и функциональной близости аминокислот (выделена жирным шрифтом)

Аминокислота	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	NS
<b>A</b>		<b>3</b>	<b>17</b>	<b>18</b>	<b>8</b>	<b>29</b>	<b>7</b>	<b>10</b>	<b>13</b>	<b>9</b>	<b>6</b>	<b>16</b>	<b>12</b>	<b>12</b>	<b>6</b>	<b>29</b>	<b>27</b>	<b>17</b>	<b>4</b>	<b>4</b>	<b>1</b>
<b>C</b>	9		<b>1</b>	<b>2</b>	<b>1</b>	<b>4</b>	<b>2</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>4</b>	<b>6</b>	<b>0</b>	<b>1</b>	<b>0</b>
<b>D</b>	61	1		<b>31</b>	<b>0</b>	<b>18</b>	<b>6</b>	<b>3</b>	<b>14</b>	<b>2</b>	<b>2</b>	<b>27</b>	<b>8</b>	<b>12</b>	<b>6</b>	<b>18</b>	<b>14</b>	<b>5</b>	<b>0</b>	<b>5</b>	<b>1</b>
<b>E</b>	66	4	86		<b>1</b>	<b>14</b>	<b>4</b>	<b>5</b>	<b>17</b>	<b>6</b>	<b>3</b>	<b>14</b>	<b>7</b>	<b>25</b>	<b>8</b>	<b>13</b>	<b>17</b>	<b>9</b>	<b>0</b>	<b>5</b>	<b>2</b>
<b>F</b>	26	1	1	3		<b>6</b>	<b>9</b>	<b>10</b>	<b>1</b>	<b>16</b>	<b>5</b>	<b>4</b>	<b>3</b>	<b>2</b>	<b>5</b>	<b>4</b>	<b>4</b>	<b>8</b>	<b>13</b>	<b>28</b>	<b>2</b>
<b>G</b>	99	6	44	35	12		<b>4</b>	<b>5</b>	<b>11</b>	<b>5</b>	<b>2</b>	<b>20</b>	<b>11</b>	<b>12</b>	<b>8</b>	<b>25</b>	<b>15</b>	<b>13</b>	<b>2</b>	<b>5</b>	<b>2</b>
<b>H</b>	20	2	13	8	13	8		<b>4</b>	<b>7</b>	<b>5</b>	<b>2</b>	<b>14</b>	<b>3</b>	<b>12</b>	<b>7</b>	<b>5</b>	<b>6</b>	<b>5</b>	<b>6</b>	<b>9</b>	<b>0</b>
<b>I</b>	35	2	7	14	19	11	7		<b>5</b>	<b>24</b>	<b>10</b>	<b>3</b>	<b>4</b>	<b>8</b>	<b>7</b>	<b>7</b>	<b>12</b>	<b>35</b>	<b>4</b>	<b>11</b>	<b>1</b>
<b>K</b>	50	1	39	51	3	31	16	13		<b>6</b>	<b>5</b>	<b>17</b>	<b>9</b>	<b>18</b>	<b>25</b>	<b>16</b>	<b>19</b>	<b>10</b>	<b>0</b>	<b>5</b>	<b>4</b>
<b>L</b>	37	3	7	20	40	15	11	67	19		<b>15</b>	<b>6</b>	<b>4</b>	<b>10</b>	<b>5</b>	<b>7</b>	<b>9</b>	<b>23</b>	<b>4</b>	<b>7</b>	<b>1</b>
<b>M</b>	17	2	4	6	7	3	2	16	11	32		<b>3</b>	<b>1</b>	<b>4</b>	<b>6</b>	<b>5</b>	<b>7</b>	<b>11</b>	<b>4</b>	<b>5</b>	<b>0</b>
<b>N</b>	60	2	75	41	10	52	29	8	50	20	5		<b>6</b>	<b>15</b>	<b>11</b>	<b>25</b>	<b>17</b>	<b>7</b>	<b>2</b>	<b>7</b>	<b>1</b>
<b>P</b>	39	1	17	17	5	23	5	8	23	11	1	15		<b>10</b>	<b>5</b>	<b>12</b>	<b>9</b>	<b>9</b>	<b>1</b>	<b>3</b>	<b>3</b>
<b>Q</b>	41	3	29	63	3	27	23	19	49	27	7	39	21		<b>15</b>	<b>13</b>	<b>15</b>	<b>12</b>	<b>3</b>	<b>4</b>	<b>0</b>
<b>R</b>	21	5	15	20	9	18	13	16	66	14	9	27	10	35		<b>12</b>	<b>13</b>	<b>7</b>	<b>3</b>	<b>8</b>	<b>2</b>
<b>S</b>	129	12	63	48	13	84	15	24	58	25	13	32	38	43	40		<b>31</b>	<b>10</b>	<b>4</b>	<b>7</b>	<b>3</b>
<b>T</b>	109	9	44	52	10	43	15	36	62	29	16	54	25	44	36	122		<b>16</b>	<b>2</b>	<b>6</b>	<b>2</b>
<b>V</b>	69	14	14	27	20	37	11	99	32	78	25	21	23	35	19	40	55		<b>2</b>	<b>10</b>	<b>0</b>
<b>W</b>	9	0	0	0	13	3	6	5	0	8	3	3	1	4	4	9	4	3		<b>6</b>	<b>0</b>
<b>Y</b>	12	1	10	10	42	10	13	21	12	17	6	14	5	7	14	20	15	23	6		<b>0</b>
<b>NS</b>	2	0	2	3	2	2	0	1	6	1	0	1	3	0	3	7	3	0	0	0	

Примечание. NS —нонсенс-кодон.

#### 1.4. МЕТОД Р. ГРЭНТСЕМА И ЕГО МОДИФИКАЦИЯ

Согласно представлениям Р. Грэнтсема (1974), основными детерминантами эволюции белков являются физико-химические силы. Р. Грэнтсем считает, что эволюционное различие между аминокислотами, определяемое их относительными частотами замены (relative substitution frequency, RSF), зависит от физико-химических свойств их боковых радикалов. Другими словами, вероятность замены одной аминокислоты на другую тем выше, чем ближе их физико-химические свойства. Наиболее сильно с RSF коррелируют несколько свойств боковой цепи аминокислот: состав, полярность и объем ( $r = 0,49, 0,47$  и  $0,37$  соответственно). Состав ( $c$ ) определяется как отношение атомных масс неуглеродных элементов в концевых группах или кольцах к атомным массам углеродов, входящих в боковую цепь. Это довольно простой способ отражения различий состава между аминокислотами. Например, боковая цепь серина —  $\text{CH}_2\text{OH}$ ,  $c = 17/12$  (1,42), лизина —  $\text{CCCCNH}_2$ ,  $c = 16/48$  (0,33). Различие состава этих аминокислот можно вычислить по формуле:  $c_{\text{сер}} - c_{\text{лиз}} = 1,42 - 0,33 = 1,09$ . Состав, полярность ( $p$ ) и объем аминокислот ( $v$ ) представлены в табл. 5.

Таблица 5

Состав, полярность и объем аминокислот

Аминокислота	Свойства аминокислот			Аминокислота	Свойства аминокислот		
	$c$	$p$	$v$		$c$	$p$	$v$
S	1,42	9,2	32	Y	0,20	6,2	136
R	0,65	10,5	124	C	2,75	5,5	55
L	0	4,9	111	H	0,58	10,4	96
P	0,39	8,0	32,5	Q	0,89	10,5	85
T	0,71	8,6	61	N	1,33	11,6	56
A	0	8,1	31	K	0,33	11,3	119
V	0	5,9	84	D	1,38	13,0	54
G	0,74	9,0	3	E	0,92	12,3	83
I	0	5,2	111	M	0	5,7	105
F	0	5,2	132	W	0,13	5,4	170

Для объединения в единой формуле сильно отличающихся значений по составу, полярности и объему боковых радикалов аминокислот Р. Грэнтсем предложил поправочные коэффициенты  $\alpha$ ,  $\beta$  и  $\gamma$ , равные  $\alpha = (1/D_c)^2 = 1,833$ ;  $\beta = (1/D_p)^2 = 0,01018$ ;  $\gamma = (1/D_v)^2 = 0,000399$ , а  $D_c = \sum[(c_i - c_j)^2]^{1/2}/190 = 0,739$ ;  $D_p = \sum[(p_i - p_j)^2]^{1/2}/190 = 3,134$ ;  $D_v = \sum[(v_i - v_j)^2]^{1/2}/190 = 50,06$ .

Предложенная в 1974 г. Р. Грэнтсемом формула расчета различий ( $D_{ij}$ ) между любыми аминокислотами  $i$  и  $j$  объединяет в себе перечисленные выше параметры взаимозаменяемых аминокислот (состав, полярность и объем):

$$D_{ij} = [\alpha (c_i - c_j)^2 + \beta (p_i - p_j)^2 + \gamma (v_i - v_j)^2]^{1/2},$$

где  $c_i$  и  $c_j$  — состав,  $p_i$  и  $p_j$  — полярность,  $v_i$  и  $v_j$  — объем аминокислот  $i$  и  $j$  соответственно. Использование коэффициентов  $\alpha$ ,  $\beta$  и  $\gamma$  в приведенной выше формуле обеспечивает сильную корреляцию значений  $D$  с логарифмом RSF ( $r = 0,72$ ). Вычисление физико-химической дистанции Р. Грантсема (Grantham distance, GD) проводится по формуле:

$$GD_{ij} = 50,723D_{ij}.$$

Значения физико-химических дистанций представлены в табл. 6. В данном методе аминокислотные замены, для которых значение физико-химической дистанции меньше среднего (100,0), считаются консервативными, в обратном случае — радикальными.

Для удобства анализа М. Джонсон предложил использовать относительные (модифицированные) значения физико-химической дистанции (Grantham distance modified,  $GD_M$ ), вычисленные по формуле:

$$GD_M = \frac{215 - GD_{ij}}{215} 100\%,$$

где 215 — максимальное значение физико-химической дистанции, соответствующее замене триптофана на цистеин. Модифицированные физико-химические дистанции показаны в табл. 7. При использовании модифицированных физико-химических дистанций замены аминокислот считаются консервативными при значении  $GD_M$  выше среднего (более 52,3 для замен в целом, более 57,9 для одношаговых замен), в обратном случае — радикальными.

Таблица 6

## Физико-химические дистанции

<b>R</b>	<b>L</b>	<b>P</b>	<b>T</b>	<b>A</b>	<b>V</b>	<b>G</b>	<b>I</b>	<b>F</b>	<b>Y</b>	<b>C</b>	<b>H</b>	<b>Q</b>	<b>N</b>	<b>K</b>	<b>D</b>	<b>E</b>	<b>M</b>	<b>W</b>	<b>Аминокислота</b>
110	145	74	58	99	124	56	142	155	144	112	89	68	46	121	65	80	135	177	<b>S</b>
	102	103	71	112	96	125	97	97	77	180	29	43	86	26	96	54	91	101	<b>R</b>
		98	92	96	32	138	5	22	36	198	99	113	153	107	172	138	15	61	<b>L</b>
			38	27	68	42	95	114	110	169	70	76	91	103	108	9	87	147	<b>P</b>
				58	69	59	89	103	92	149	47	42	65	78	85	65	81	128	<b>T</b>
					64	60	94	113	112	195	86	91	111	106	126	107	84	148	<b>A</b>
						109	29	50	55	192	84	96	133	97	152	121	21	88	<b>V</b>
							135	153	147	159	98	87	80	127	94	98	127	184	<b>G</b>
								21	33	198	94	109	149	102	168	134	10	61	<b>I</b>
									22	205	100	116	158	102	177	140	28	40	<b>F</b>
										194	83	99	143	85	160	122	36	37	<b>Y</b>
											174	154	139	202	154	170	196	215	<b>C</b>
												24	68	32	81	40	87	115	<b>H</b>
													46	53	61	29	101	130	<b>Q</b>
														94	23	42	142	174	<b>N</b>
															101	56	95	110	<b>K</b>
																45	160	181	<b>D</b>
																	126	152	<b>E</b>
																		67	<b>M</b>

Таблица 7

## Модифицированные физико-химические дистанции

Амино-кислота	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	100	9	41	50	45	73	59	54	50	54	59	45	86	54	45	54	73	68	32	45
C		100	27	23	4	27	18	9	4	9	9	36	23	27	18	45	32	9	0	9
D			100	77	18	54	59	23	50	18	27	86	50	68	54	68	59	27	14	27
E				100	36	54	82	36	73	36	41	77	54	86	73	64	68	41	27	41
F					100	27	54	86	50	86	86	27	45	45	54	27	50	77	82	86
G						100	54	36	41	36	41	64	77	59	41	73	73	50	14	32
H							100	54	82	54	59	68	64	86	86	59	77	59	45	59
I								100	50	95	95	32	54	50	54	32	59	86	68	82
K									100	50	54	54	50	73	86	41	64	54	50	59
L										100	91	27	54	45	50	32	54	82	68	82
M											100	32	59	50	54	36	59	86	68	82
N												100	54	77	59	77	68	36	18	32
P													100	64	50	64	82	68	32	50
Q														100	77	68	77	54	41	54
R															100	50	64	54	50	64
S																100	73	41	18	32
T																	100	68	41	54
V																		100	59	73
W																			100	82
Y																				100

## 1.5. МЕТОДЫ ДЖ. ЖАНГА

Методы Дж. Жанга, предложенные в 2000 г., основаны на двух классификациях аминокислот: 1) по заряду; 2) полярности и объему. Согласно первой классификации, все аминокислоты подразделяются на три группы:

- 1) положительно заряженные (R, H, K);
- 2) отрицательно заряженные (D, E);
- 3) незаряженные (A, N, C, Q, G, I, L, M, F, P, S, T, W, Y, V).

В соответствии со второй классификацией, все аминокислоты делятся на 6 групп:

- 1) специальные (C);
- 2) нейтральные и малые (A, G, P, S, T);
- 3) полярные и относительно малые (N, D, Q, E);
- 4) полярные и относительно большие (R, H, K);
- 5) неполярные и относительно малые (I, L, M, V);
- 6) неполярные и относительно большие (F, W, Y).

Если аминокислотная замена происходит в пределах одной группы, то она считается консервативной, в обратном случае — радикальной.

## 1.6. УНИВЕРСАЛЬНЫЙ ЭВОЛЮЦИОННЫЙ ИНДЕКС Х. ТАНГА

В 2004 г. Х. Танг и соавт. предложили новый метод изучения характера аминокислотных замен, разработанный на основании анализа белок-кодирующих участков 4383 генов четырех групп организмов: приматов (человека и макаки, 280 генов), грызунов (мыши и крысы, 1306 генов), дрозофил (*D. melanogaster* и *D. simulans*, 309 генов) и дрожжей (*S. cerevisiae* и *S. paradoxus*, 2488 генов).

Первоначально авторами для каждой из 75 одношаговых аминокислотных замен вычислен некорректированный эволюционный индекс.

Некорректированный эволюционный индекс ( $EI_{ij}^{HK}$ ) — это соотношение количества наблюдаемых и теоретически ожидаемых аминокислотных замен. Данный индекс вычисляется по формуле:

$$EI_{ij}^{HK} = \frac{\text{Obs}(i \leftrightarrow j)}{\text{Exp}(i \leftrightarrow j)},$$

где  $\text{Obs}(i \leftrightarrow j)$  — число наблюдаемых взаимных замен аминокислоты  $i$  на аминокислоту  $j$ ;  $\text{Exp}(i \leftrightarrow j)$  — число теоретически ожидаемых замен аминокислоты  $i$  на аминокислоту  $j$ .

Вычисление числа наблюдаемых аминокислотных замен проводится следующим образом. Все наблюдаемые в сравниваемых последовательностях генов замены нуклеотидов заносятся в матрицу смысловых кодонов (размером  $61 \times 61$ ). Двухшаговые замены учитываются как две одношаговые замены, а трехшаговые замены не анализируются. Далее матрица ко-

донов превращается в матрицу аминокислот  $20 \times 20$  путем суммирования всех значений в пределах серии кодонов.

Вычисление числа теоретически ожидаемых аминокислотных замен основано на том, что для каждого из смысловых кодонов возможны 9 видов одношаговых мутаций (от 1 до 3 синонимичных и от 6 до 9 — несинонимичных, за исключением шестикратно вырожденных кодонов, соответствующих лейцину и аргинину).

Рассмотрим данные виды мутаций для кодона ТТТ, соответствующего фенилаланину. По первому положению данного кодона возможны мутации ТТТ (F) → ЦТТ (L), АТТ (I), ГТТ (V), по второму — ТТТ (F) → ТЦТ (S), ТАТ (Y), ТГТ (C), по третьему — ТТТ (F) → ТТЦ (F), ТТА (L) и ТТГ (L). Обобщенная картина синонимичных и несинонимичных замен по данному кодону выглядит так: переходы F → F, S осуществляются путем транзиций, переход F → L — транзиции и двух трансверсий, переходы F → I, V, Y, C — путем трансверсий. Аналогичным образом определяются картины для других кодонов. Вычисленные вероятности корректируются с учетом числа кодонов, по которым сравниваются последовательности организмов двух видов, и заносятся в матрицу  $61 \times 61$ . Такой подход позволяет учитывать, наряду с аминокислотным составом, и использование синонимичных кодонов.

Общее количество синонимичных замен в матрице устанавливается равным числу наблюдаемых синонимичных замен, а затем проводится калибровка количества каждого типа несинонимичных замен. Полученная матрица кодонов аналогично переводится в матрицу аминокислот. Важно, что данные матрицы являются симметричными, поскольку в них нет направленности аминокислотных замен из-за отсутствия анцестральных последовательностей.

Вычисление скорректированного эволюционного индекса ( $EI_{ij}^K$ ) осуществляется по формуле:

$$EI_{ij}^K = \frac{EI_{ij}^{HK}}{f_c},$$

где  $f_c$  — фактор коррекции  $c$ , равный отношению некорректированных  $d_N/d_S$  к скорректированным (для всей группы данных)  $d_N/d_S$ .

Универсальный эволюционный индекс ( $U_i$ ) — это мера взаимозаменяемости аминокислот, вычисленная на основании полученных по большому числу генов эволюционных индексов для грызунов и дрожжей:

$$U_{ij} = \frac{R_{(r)} \cdot EI_{ij(r)} + R_{(y)} \cdot EI_{ij(y)}}{R_{(r)}^2 + R_{(y)}^2},$$

где  $R_{(r)}$  — среднее скорректированное соотношение  $d_N/d_S$  для грызунов;  $EI_{ij(r)}$  — эмпирический эволюционный индекс взаимной замены аминокислот  $i$  и  $j$  для грызунов;  $R_{(y)}$  — среднее скорректированное соотношение



$d_N/d_S$  для дрожжей;  $EI_{ij(v)}$  — эмпирический эволюционный индекс взаимной замены аминокислот  $i$  и  $j$  для дрожжей. Показатели  $U$  калиброваны таким образом, что среднее значение равно 1 (табл. 8).

Таблица 8

Показатели универсального эволюционного индекса для элементарных аминокислотных замен

Замена	Показатель $U$	Замена	Показатель $U$	Замена	Показатель $U$	Замена	Показатель $U$	Замена	Показатель $U$
S–T	2,490	Q–P	1,377	E–A	0,906	L–H	0,560	R–L	0,414
V–I	2,415	S–G	1,360	S–C	0,852	K–M	0,559	G–C	0,414
S–A	2,380	Q–H	1,351	R–S	0,850	R–P	0,559	P–L	0,388
N–S	2,053	V–L	1,329	R–T	0,827	E–G	0,553	R–C	0,382
D–E	2,033	R–H	1,317	I–M	0,827	V–F	0,548	N–Y	0,378
I–L	1,726	A–P	1,288	Q–L	0,805	E–K	0,548	S–W	0,375
N–T	1,695	K–N	1,075	L–W	0,793	D–G	0,548	S–F	0,365
Y–F	1,649	R–Q	1,045	P–H	0,784	I–F	0,545	D–V	0,361
E–Q	1,634	S–P	1,039	T–I	0,750	S–I	0,540	C–F	0,321
L–M	1,601	A–V	1,017	L–F	0,732	G–V	0,539	N–I	0,321
T–A	1,587	D–N	1,015	S–L	0,725	R–G	0,534	C–W	0,271
R–K	1,583	T–M	1,007	K–I	0,688	E–V	0,506	C–Y	0,268
K–Q	1,466	T–P	1,001	H–Y	0,665	S–Y	0,503	R–W	0,263
N–H	1,382	K–T	0,989	D–A	0,657	R–I	0,490	G–W	0,242
G–A	1,379	V–M	0,986	D–H	0,560	R–M	0,470	D–Y	0,241

Чем больше значение  $U$ , тем чаще происходят взаимные замены аминокислот данной пары. Так, максимальное значение  $U$  свойственно паре S–T, а минимальное — паре D–Y, что свидетельствует об их соответственно наибольшей и наименьшей (по сравнению с другими элементарными аминокислотными заменами) взаимозаменяемости в ходе эволюции.

Таким образом, универсальный эволюционный индекс является эмпирической системой классификации аминокислотных замен по степени их эволюционной изменчивости. Аминокислоты часто взаимозаменяемых пар, как правило, более сходны по своим физико-химическим свойствам. Установлена корреляция значений эмпирических значений эволюционного индекса с другими методами изучения характера аминокислотных замен: с дистанцией Грэнтсема ( $r$  от  $-0,67$  до  $-0,56$ ) и дистанцией РАМ ( $r = 0,67$ ). Корреляционные связи средней силы объясняются тем, что дистанции Грэнтсема и РАМ учитывают лишь аминокислотные замены (первая — теоретически, вторая — эмпирически) без учета соответствующих им кодонов.

Для любого большого набора данных можно вычислить расчетное значение  $EI$  ( $EI_{\text{расч.}}$ ):

$$EI_{\text{расч.}} = U \cdot R = \frac{U d_N}{d_S}.$$

Для проверки корректности данной формулы проведен корреляционный анализ между эмпирическими и расчетными значениями эволюционного индекса. Авторами установлена сильная корреляция данных показателей не только для грызунов и дрожжей ( $r = 0,99$  и  $r = 0,96$  соответственно; вычисленные по их генам значения  $EI$  использовались для расчета универсального эволюционного индекса), но для приматов и дрозофил ( $r = 0,84$  и  $r = 0,84$ ). Применение данной формулы рекомендовано авторами для набора данных с более чем 20 000 аминокислотных замен или, по меньшей мере, более чем 2500 замен, что и является основным ограничением широкого использования метода.

Таким образом, метод Х. Танга может быть охарактеризован как эмпирический, основанный на анализе эволюционных изменений кодонов, применимый для близкородственных видов, и универсальный (для различных таксономических групп организмов).

При апробации данного метода установлено, что 10 пар аминокислот с наиболее высокими значениями составляют 25–30 % от общего числа элементарных замен. Преобладание скорости данных замен ( $d_h$ ) над таковой синонимичных замен рекомендовано в качестве критерия наличия положительного отбора.

Потенциальными областями применения данного метода являются оценка функциональных эффектов аминокислотных замен в белках и создание эволюционной классификации аминокислот.

## Глава 2

### Использование методов определения характера аминокислотных замен для решения прикладных задач

Проанализированы нуклеотидные последовательности мРНК и аминокислотные последовательности алкогольдегидрогеназ (АДГ) классов 1–4 мыши и человека. Выравнивание последовательностей произведено с помощью программ Clustal W DNA и Clustal W Protein. Определены количество ( $n_d$ ) и вид наблюдаемых замещений аминокислот. Характер аминокислотных замен (консервативный или радикальный) определен при помощи коэффициента П. Снита ( $\phi$ ), показателя ФБА и модифицированной физико-химической дистанции Р. Грэнтсема.

Известно, что аминокислотная замена может быть обусловлена замещением одного (одношаговая замена), двух (двухшаговая замена) и трех нуклеотидов (трехшаговая замена) в кодоне. Количество наблюдаемых одно- ( $n_{dI}$ ), двух- ( $n_{dII}$ ) и трехшаговых ( $n_{dIII}$ ) аминокислотных замен в последовательностях алкогольдегидрогеназ классов 1–4 мыши и человека приведено в табл. 9.

Таблица 9

**Количество одно-, двух- и трехшаговых аминокислотных замен  
в последовательностях алкогольдегидрогеназ классов 1–4 мыши и человека**

Класс АДГ	1А	1В	1С	2	3	4	Всего
$n_{dI}$	33	33	30	61	20	31	208
$n_{dII}$	24	25	22	35	6	8	120
$n_{dIII}$	6	5	4	9	1	2	27
<b>Общее, <math>n_d</math></b>	63	63	56	105	27	41	355

Из приведенных данных видно, что большая часть замен ( $58,6 \pm 2,61$  %) в изучаемых аминокислотных последовательностях АДГ является одношаговыми, что согласуется с данными других исследователей. Согласно таблице генетического кода, из 190 всевозможных пар взаимозаменяемых аминокислотных остатков 75 ( $39,5 \pm 3,55$  %) связаны одним мутационным шагом. По этим причинам анализ характера аминокислотных замещений следует начинать с изучения характера одношаговых аминокислотных замен.

Известно, что чем меньше значения коэффициента П. Снита, показателя ФБА и модифицированной дистанции Р. Грэнтсема для взаимозаменяемых остатков, тем больше вероятность изменения биологической функции белка. Средние значения данных показателей, а также процент консервативных одношаговых аминокислотных замен в последовательностях АДГ классов 1–4 мыши и человека представлены в табл. 10.

**Показатели консервативности для одношаговых аминокислотных замен  
в последовательностях АДГ классов 1–4 мыши и человека**

Класс АДГ	1А	1В	1С	2	3	4	Среднее
$\phi \pm SE$	0,61 ± 0,03 (90,9 %* <sup>4</sup> )	0,61 ± 0,03 (87,9 %)	0,64 ± 0,03 (93,3 %* <sup>4</sup> )	0,62 ± 0,03 (82,0 %)	0,66 ± 0,04 (95,0 %* <sup>4</sup> )	0,57 ± 0,04 (67,7 %* <sup>1А,1С,3</sup> )	0,61 ± 0,01 (85,1 %)
ФБА ± SE	23,91 ± 1,53 (87,9 %)	24,03 ± 1,56 (81,8 %)	25,83 ± 1,53 (90,0 %)	22,75 ± 1,15 (80,3 %)	24,25 ± 1,59 (90,0 %)	21,77 ± 1,73 (83,9 %)	23,58 ± 0,61 (84,6 %)
GD <sub>М</sub> ± SE	72,58 ± 2,27 (90,9 %* <sup>4</sup> )	74,48 ± 2,12 (84,8 %)	74,53 ± 2,07 (86,7 %)	74,30 ± 1,86 (86,9 %)	80,00 ± 2,63* <sup>4</sup> (90,0 %)	70,58 ± 3,24* <sup>3</sup> (71,0 %* <sup>1А</sup> )	73,92 ± 0,96 (85,1 %)

*Примечание.* В скобках указан процент консервативных замен; знаком \*<sup>(1А, 2А...)</sup> обозначены статистически достоверные ( $p < 0,05$ ) различия с соответствующим классом АДГ.

Для изученных аминокислотных последовательностей АДГ средние значения коэффициента П. Снита и модифицированной дистанции Р. Грэнтсема незначительно варьируют, максимальные значения характерны для АДГ третьего класса ( $0,66 \pm 0,04$  и  $80,00 \pm 2,63$  соответственно), а минимальные ( $0,57 \pm 0,04$  и  $70,58 \pm 3,24$  соответственно) — для АДГ класса 4. Наименьшее значение показателя ФБА также отмечается у АДГ класса 4 ( $21,77 \pm 1,73$ ), а наибольшее — у АДГ класса 1С ( $25,83 \pm 1,53$ ).

Наиболее высокий процент консервативных замен в соответствии с коэффициентом П. Снита характерен для АДГ классов 1С и 3 (93,3 % и 95,0 % соответственно), в соответствии с ФБА — для классов 1С и 3 (по 90,0 %), в соответствии с модифицированной дистанцией Р. Грэнтсема — для классов 1А и 3 (90,9 % и 90,0 % соответственно).

Наименьший процент консервативных замен в соответствии с коэффициентом П. Снита и модифицированной дистанцией Р. Грэнтсема наблюдается у АДГ класса 4 (67,7 % и 71,0 %), а в соответствии с показателем ФБА — у АДГ класса 2 (80,3 %). Для получения более корректных результатов следует рассчитать средний процент консервативных замен по трем показателям, так как коэффициент П. Снита и модифицированная дистанция Р. Грэнтсема учитывают лишь физико-химические свойства взаимозаменяемых аминокислот, а показатель А. Бачинского — лишь их биохимические свойства. Так, для АДГ класса 1А средний процент консервативных замен равен 89,9 %, для класса 1В — 84,8 %, класса 1С — 90,0 %, класса 2 — 83,1 %, класса 3 — 91,7 % и, наконец, для класса 4 — 74,2 %.

Таким образом, замены в аминокислотных последовательностях алкогольдегидрогеназ класса 4 наименее, а классов 1А, 1С и 3 — наибо-

лее консервативны. Эволюционная консервативность АДГЗ обусловлена универсальностью катализа глутатионзависимого окисления формальдегида, лежащего в основе многих процессов переноса одноуглеродных остатков.

В зависимости от положения нуклеотида в кодоне, в котором произошла мутация, одношаговые аминокислотные замены подразделяются на 3 группы (обусловленные заменой в первом ( $n_1$ ), втором ( $n_2$ ) и третьем положениях кодона ( $n_3$ )). Количество одношаговых аминокислотных замен в последовательностях АДГ классов 1–4 мыши и человека, обусловленных заменами в первом ( $n_{d11}$ ), втором ( $n_{d12}$ ) и третьем ( $n_{d13}$ ) положениях кодона показано в табл. 11. Большая часть одношаговых замен обусловлена изменением нуклеотида в первом положении кодона (54,8 %), меньшая — во втором (23,6 %), а наименьшая — в третьем (21,6 %).

Таблица 11

**Количество одношаговых аминокислотных замен  
в последовательностях АДГ классов 1–4 мыши и человека, обусловленных  
заменами в первом, втором и третьем положениях кодона**

Класс АДГ	1А	1В	1С	2	3	4	Всего
$n_{d11}$	18	18	17	34	12	15	114
$n_{d12}$	8	7	4	16	4	10	49
$n_{d13}$	7	8	9	11	4	6	45

Средние значения коэффициента П. Снита, показателя ФБА и модифицированной дистанции Р. Грэнтсема для одношаговых аминокислотных замен в последовательностях АДГ классов 1–4 мыши и человека в зависимости от положения кодона, в котором произошла замена, представлены в табл. 12.

Таблица 12

**Средние значения показателей консервативности  
для одношаговых замен (по отдельным положениям в кодоне)  
в последовательностях АДГ классов 1–4 мыши и человека**

Показатель	Первое положение кодона	Второе положение кодона	Третье положение кодона
$\phi \pm SE$	$0,62 \pm 0,02$ (89,5 % <sup>*3</sup> )	$0,56 \pm 0,03$ <sup>*3</sup> (81,6 %)	$0,67 \pm 0,03$ <sup>*2</sup> (77,8 % <sup>*1</sup> )
ФБА $\pm SE$	$25,04 \pm 0,81$ <sup>*2</sup> (90,4 % <sup>*3</sup> )	$20,53 \pm 1,16$ <sup>*1</sup> (79,6 %)	$23,20 \pm 1,35$ (75,6 % <sup>*1</sup> )
GD <sub>M</sub> $\pm SE$	$74,67 \pm 1,36$ (86,0 %)	$71,67 \pm 1,83$ (87,8 %)	$75,22 \pm 1,99$ (80,0 %)

*Примечание.* В скобках указан процент консервативных замен; знаком <sup>\*(1, 2...)</sup> обозначены статистически достоверные ( $p < 0,05$ ) различия с соответствующим положением кодона.

Установлено, что наибольшие значения коэффициента П. Снита и модифицированной дистанции Р. Грэнтсема характерны для замен в третьем положении нуклеотида в кодоне, меньшие — для первого,

а наименьшие — для второго. Максимальные значения ФБА наблюдаются при заменах в первом положении кодона, меньшие — в третьем, а минимальные — во втором.

Средний процент консервативных замен по трем показателям для первого положения нуклеотида в кодоне составляет 88,6 %, для второго — 83,0 % и для третьего — 77,8 %. Исходя из этого следует, что замены нуклеотидов в первом положении сопровождаются заменами на такие аминокислоты, физико-химические и функциональные свойства которых отличаются от свойств замещаемых кислот сильнее, чем в случае замен по второму и третьему положению, что соответствует теории нейтральности. Можно также предположить, что именно высокая частота радикальных аминокислотных замен, связанных с мутациями в третьем положении нуклеотида в кодоне, обуславливает низкую скорость несинонимичных замен в этом положении.

Перейдем к сравнению характера одно- и многошаговых аминокислотных замен. Для этого проанализируем средние значения модифицированной дистанции Р. Грэнтсема и процент консервативных замен в последовательностях АДГ классов 1–4 мыши и человека (табл. 13). При сравнении последовательностей АДГ класса 3 мыши и человека обнаружена лишь одна трехшаговая замена, поэтому значение модифицированной дистанции Р. Грэнтсема для нее ( $9,00 \pm n/calc$ ) не учитывалось при анализе.

Таблица 13

**Средние значения модифицированной дистанции Р. Грэнтсема  
для одно- и многошаговых аминокислотных замен  
в последовательностях АДГ классов 1–4 мыши и человека**

Класс АДГ	1А	1В	1С	2	3	4
Одношаговые замены	72,58 ± 2,27	74,48 ± 2,12	74,53 ± 2,07	74,30 ± 1,86	80,00 ± 2,63* <sup>4</sup>	70,58 ± 3,24* <sup>3</sup>
Двухшаговые замены	66,17 ± 4,05	68,44 ± 3,75	65,64 ± 3,96	66,00 ± 3,51	60,50 ± 8,55	68,00 ± 6,08
Трехшаговые замены	41,67 ± 6,98*	44,40 ± 8,0	38,50 ± 6,06*	52,33 ± 3,9	9,00 ± n/calc	63,50 ± 6,36* <sup>1А, 1С</sup>

*Примечание.* Знаком <sup>(1, 2...)</sup> обозначены статистически достоверные ( $p < 0,05$ ) различия с соответствующим классом АДГ.

Для одношаговых замен в изучаемых аминокислотных последовательностях АДГ средние значения модифицированной дистанции Р. Грэнтсема незначительно варьируют, максимальные значения характерны для АДГ третьего класса ( $80,00 \pm 2,63$ ), а минимальные ( $70,58 \pm 3,24$ ) — для АДГ четвертого класса. Наименьшее значение  $GD_M$  для двухшаговых замен отмечается у АДГ третьего класса ( $60,50 \pm 8,55$ ), а трехшаговых — АДГ класса 1С ( $38,50 \pm 6,06$ ). Наибольшее значение модифицированной дистанции Р. Грэнтсема по двухшаговым заменам характерно для АДГ 1В

и 4 ( $68,44 \pm 3,75$  и  $68,00 \pm 6,08$  соответственно), а по трехшаговым — для АДГ класса 4 ( $63,50 \pm 6,36$ ).

Из приведенных в табл. 13 данных видно, что для всех классов АДГ наибольшие значения модифицированной дистанции Р. Грэнтсема характерны для одношаговых замен, меньшие — для двухшаговых и наименьшие — для трехшаговых.

Среднее значение  $GD_M$  для одношаговых замен, наблюдаемых во всех изучаемых алкогольдегидрогеназах, составило  $73,92 \pm 0,96$ , для двухшаговых —  $66,33 \pm 1,72$  и для трехшаговых —  $45,67 \pm 3,05$  (различия статистически достоверны,  $p < 0,01$ ). Между значениями физико-химической дистанции и частотой замен в зависимости от их вида (одно-, двух- или трехшаговые) обнаружена четкая связь ( $r = 0,98$ ).

Определим, какой вид замен нуклеотидов определяет более консервативный и более радикальный характер аминокислотных замен. Для этого вычислим средние значения коэффициента П. Снита, показателя ФБА и модифицированной дистанции Р. Грэнтсема и процент консервативных замен для одношаговых замен аминокислот в последовательностях алкогольдегидрогеназ классов 1–4 мыши и человека (табл. 14).

Таблица 14

**Значения показателей консервативности одношаговых аминокислотных замен, наблюдаемых в последовательностях алкогольдегидрогеназ классов 1–4 мыши и человека, в зависимости от вида произошедшей замены нуклеотида**

Показатель	A↔Г n=63	Ц↔У n=15	A↔Ц n=29	Г↔Ц n=37	Г↔У n=33	A↔У n=31
$\phi \pm SE$	$0,66 \pm 0,03$ (87,3 %)	$0,51 \pm 0,04$ (73,3 %)	$0,54 \pm 0,03$ (86,2 %)	$0,65 \pm 0,03$ (81,1 %)	$0,55 \pm 0,03$ (87,9 %)	$0,67 \pm 0,03$ (87,1 %)
ФБА ± SE	$27,44 \pm 1,00$ (93,7 %)	$12,93 \pm 1,13$ (40,0 %)	$17,31 \pm 0,92$ (96,6 %)	$24,14 \pm 1,24$ (86,5 %)	$23,09 \pm 1,73$ (72,7 %)	$26,61 \pm 1,40$ (83,9 %)
$GD_M \pm SE$	$81,94 \pm 0,97$ (96,8 %)	$63,27 \pm 4,08$ (93,3 %)	$73,55 \pm 2,36$ (89,7 %)	$72,78 \pm 2,03$ (83,8 %)	$66,55 \pm 3,18$ (48,5 %)	$73,42 \pm 2,29$ (87,1 %)

*Примечание.* В скобках указан процент консервативных замен.

Установлено, что минимальные средние значения всех трех показателей консервативности отмечаются для транзиции Ц↔У (различия статистически достоверны по сравнению со средними значениями трех показателей для замен А↔Г, Г↔Ц, А↔У с  $p < 0,05$ ). Наибольшая величина коэффициента П. Снита характерна для замен А↔Г, Г↔Ц и А↔У, а показателя ФБА — для замен А↔Г и А↔У. Максимальное значение модифицированной дистанции Р. Грэнтсема наблюдается для транзиции А↔Г, что достоверно выше ( $p < 0,01$ ) таковых для всех остальных замен. Таким образом, судя по средним значениям показателей консервативности наиболее благоприятными являются замены аминокислот, обусловленные мутациями А↔Г, а наименее благоприятными — Ц↔У. В этом контексте

легко понять преобладание числа транзиций  $A \leftrightarrow G$  над числом транзиций  $C \leftrightarrow U$  более, чем в 4 раза.

Средний процент консервативных замен аминокислот по трем показателям для транзиций  $A \leftrightarrow G$  составил 96,2 %, для  $C \leftrightarrow U$  — 68,9 %, для трансверсий  $A \leftrightarrow C$  — 90,8 %,  $G \leftrightarrow C$  — 83,8 %,  $G \leftrightarrow U$  — 69,8 % и  $A \leftrightarrow U$  — 86,0 %. Эти данные наводят на мысль о том, что существует определенная взаимосвязь между частотой закрепленных мутаций нуклеотидов и процентом консервативных аминокислотных замен. Так, в изучаемых последовательностях АДГ наблюдается большое количество ( $n = 63$ ) вызванных транзициями  $A \leftrightarrow G$  замен аминокислот, носящих высоко консервативный характер (92,6 % замен), и в то же время отмечается малое количество ( $n = 15$ ) замен, связанных с транзициями  $C \leftrightarrow U$  и характеризующихся более низкой долей консервативных замен (68,9 %).

Таким образом, с использованием методов определения характера аминокислотных замен установлено, что в аминокислотных последовательностях алкогольдегидрогеназ мыши и человека классов 1–4 преобладают одношаговые замены; аминокислотные замены в последовательностях алкогольдегидрогеназ класса 4 наименее, а классов 1С и 3 — наиболее консервативны; наибольший процент консервативных замен наблюдается в первом положении кодона, меньший — во втором и наименьший — в третьем; между долей консервативных замен и частотой замен в зависимости от положения нуклеотида в кодоне существует сильная положительная корреляционная связь.

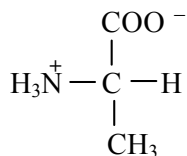


## Литература

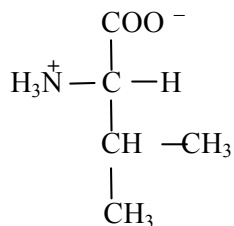
1. *Бутвиловский, А. В.* Алкогольдегидрогеназы хордовых животных : монография / А. В. Бутвиловский, Е. В. Барковский, В. Э. Бутвиловский ; под общ. ред. Е. В. Барковского. Минск : БГМУ, 2007. 144 с.
2. *Бачинский, А. Г.* Структура и помехоустойчивость генетического кода / А. Г. Бачинский // Журн. общей биол. 1976. Т. 37. С. 163–173.
3. *Волькенштейн, М.* Общая биофизика / М. Волькенштейн. М. : Наука, 1978. 590 с.
4. *Grantham, R.* Amino acid difference formula to help explain protein evolution / R. Grantham // Science. 1974. Vol. 185 (4154). P. 862–864.
5. *McClellan, D. A.* Estimating the influence of selection on variable amino acid sites of the cytochrome b protein functional domains / D. A. McClellan, K. G. McCracken // Mol. Biol. Evol. 2001. Vol. 18. P. 917–925.
6. *Dagan, T.* Ratios of radical to conservative amino acid replacement are affected by mutational and compositional factors and may not be indicative of positive Darwinian selection / T. Dagan, Y. Talmor, D. Graur // Mol. Biol. Evol. 2002. Vol. 19 (7). P. 1022–1025.
7. *Sneath, P. H. A.* Relations between chemical structure and biological activity in peptides / P. H. A. Sneath // J. Theoret. Biol. 1966. Vol. 12 (2). P. 157–195.
8. *Tang, H.* A new method for estimating nonsynonymous substitutions and its applications to detecting positive selection / H. Tang, C. I. Wu // Mol. Biol. Evol. 2006. Vol. 23(2). P. 372–379.
9. *Universal evolutionary index for amino acid changes* / H. Tang [et al.] // Mol. Biol. Evol. 2004. Vol. 21 (8). P. 1548–1556.
10. *Zhang, J.* Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes / J. Zhang // J. Mol. Evol. 2000. Vol. 50 (1). P. 56–68.

## Приложение

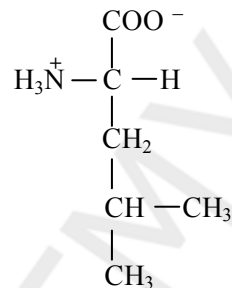
### Формулы и обозначения аминокислот



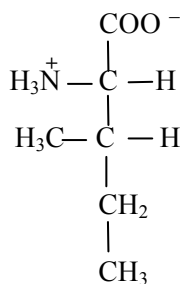
**Л-аланин, А, Ала**



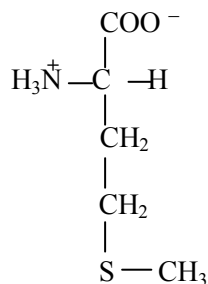
**Л-валин, V, Вал**



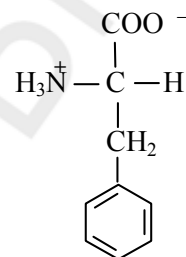
**Л-лейцин, L, Лей**



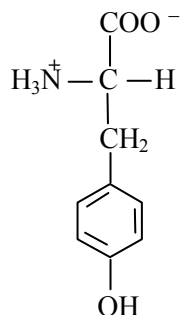
**Л-изолейцин, I, Иле**



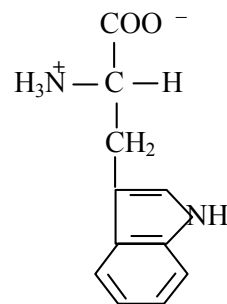
**Л-метионин, M, Мет**



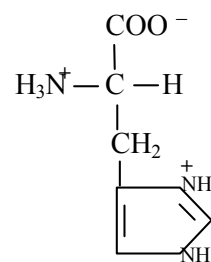
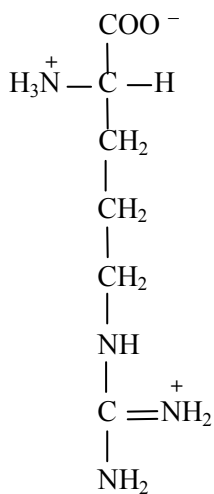
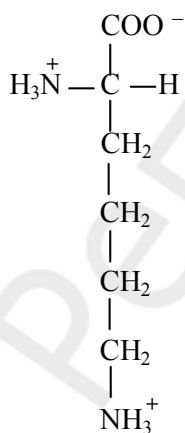
**Л-фенилаланин, F, Фен**

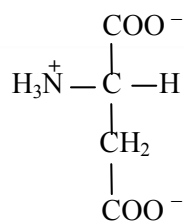
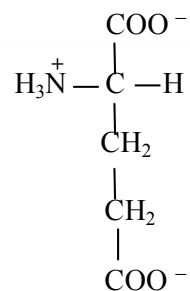
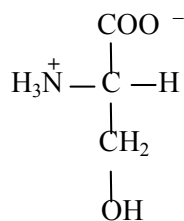
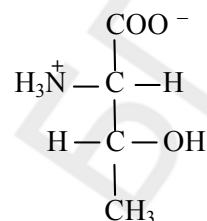
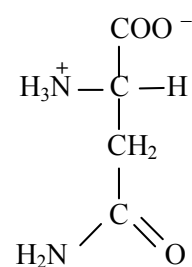
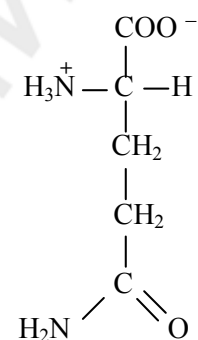
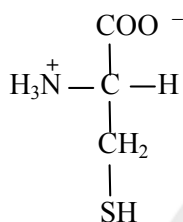
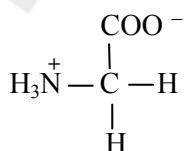
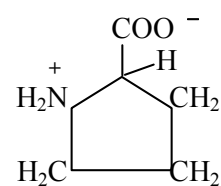


**Л-тирозин, Y, Тир**



**Л-триптофан, W, Три**



**L-лизин, К, Лиз****L-аргинин, R, Арг****L-гистидин, H, Гис****L-аспарагиновая кислота, D, Асп    L-глутаминовая кислота, E, Глу****L-серин, S, Сер****L-треонин, T, Тре****L-аспарагин, N, Аспн****L-глутамин, Q, Глн****L-цистеин, C, Цис****Глицин, G, Гли****L-пролин, P, Про**

## Оглавление

Введение .....	3
Глава 1. Характеристика наиболее часто используемых методов определения характера аминокислотных замен .....	4
1.1. Метод П. Снита .....	8
1.2. Метод М. Волькенштейна .....	9
1.3. Метод А. Бачинского .....	11
1.4. Метод Р. Грэнтсема и его модификация .....	15
1.5. Методы Дж. Жанга .....	15
1.6. Универсальный эволюционный индекс Х. Танга .....	17
Глава 2. Использование методов определения характера аминокислотных замен для решения прикладных задач .....	19
Литература .....	25
Приложение .....	26