

Новосибирский государственный
университет

Лекция №4

Установление первичной структуры нуклеиновых кислот
(секвенирование)

Кечин Андрей Андреевич, к.б.
н.

Новосибирск -
2020

План лекции

- Секвенирование первого поколения (Сэнгера, Максама-Гилберта)
- Секвенирование второго поколения (Illumina, Roche, Applied Biosystems, Ion Torrent)
- Секвенирование третьего поколения (Oxford Nanopore, Pacific Biosciences)

Секвенирование первого поколения 1977 г

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 2, pp. 560-564, February 1977
Biochemistry

A new method for sequencing DNA

(DNA chemistry/dimethyl sulfate cleavage/hydrazine/piperidine)

ALLAN M. MAXAM AND WALTER GILBERT

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138

Contributed by Walter Gilbert, December 9, 1976

ABSTRACT DNA can be sequenced by a chemical procedure that breaks a terminally labeled DNA molecule partially at each repetition of a base. The lengths of the labeled fragments then identify the positions of that base. We describe reactions that cleave DNA preferentially at guanines, at adenines, at cytosines and thymines equally, and at cytosines alone. When the products of these four reactions are resolved by size, by electrophoresis on a polyacrylamide gel, the DNA sequence can be read from the pattern of radioactive bands. The technique will permit sequencing of at least 100 bases from the point of labeling.

We have developed a new technique for sequencing DNA molecules. The procedure determines the nucleotide sequence of a terminally labeled DNA molecule by breaking it at adenine, guanine, cytosine, or thymine with chemical agents.

THE SPECIFIC CHEMISTRY

A Guanine/Adenine Cleavage (2). Dimethyl sulfate methylates the guanines in DNA at the N7 position and the adenines at the N3 (3). The glycosidic bond of a methylated purine is unstable (3, 4) and breaks easily on heating at neutral pH, leaving the sugar free. Treatment with 0.1 M alkali at 90° then will cleave the sugar from the neighboring phosphate groups. When the resulting end-labeled fragments are resolved on a polyacrylamide gel, the autoradiograph contains a pattern of dark and light bands. The dark bands arise from breakage at guanines, which methylate 5-fold faster than adenines (3).

This strong guanine/weak adenine pattern contains almost half the information necessary for sequencing; however, ambiguities can arise in the interpretation of this pattern because

Proc. Natl. Acad. Sci. USA
Vol. 74, No. 12, pp. 5463-5467, December 1977
Biochemistry

DNA sequencing with chain-terminating inhibitors

(DNA polymerase/nucleotide sequences/bacteriophage ϕ X174)

F. SANGER, S. NICKLEN, AND A. R. COULSON

Medical Research Council Laboratory of Molecular Biology, Cambridge CB2 2QH, England

Contributed by F. Sanger, October 3, 1977

ABSTRACT A new method for determining nucleotide sequences in DNA is described. It is similar to the "plus and minus" method [Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* 94, 441-448] but makes use of the 2',3'-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which act as specific chain-terminating inhibitors of DNA polymerase. The technique has been applied to the DNA of bacteriophage ϕ X174 and is more rapid and more accurate than either the plus or the minus method.

The "plus and minus" method (1) is a relatively rapid and simple technique that has made possible the determination of the sequence of the genome of bacteriophage ϕ X174 (2). It depends on the use of DNA polymerase to transcribe specific regions of the DNA under controlled conditions. Although the method is considerably more rapid and simple than other available techniques, neither the "plus" nor the "minus" method is completely accurate, and in order to establish a sequence both must be used together, and sometimes confirmatory data are necessary. W. M. Barnes (*J. Mol. Biol.*, in press) has recently developed a third method, involving ribo-substi-

a stereoisomer of ribose in which the 3'-hydroxyl group is oriented in *trans* position with respect to the 2'-hydroxyl group. The arabinosyl (ara) nucleotides act as chain terminating inhibitors of *Escherichia coli* DNA polymerase I in a manner comparable to ddT (4), although synthesized chains ending in 3' araC can be further extended by some mammalian DNA polymerases (5). In order to obtain a suitable pattern of bands from which an extensive sequence can be read it is necessary to have a ratio of terminating triphosphate to normal triphosphate such that only partial incorporation of the terminator occurs. For the dideoxy derivatives this ratio is about 100, and for the arabinosyl derivatives about 5000.

METHODS

Preparation of the Triphosphate Analogues. The preparation of ddTTP has been described (6, 7), and the material is now commercially available. ddA has been prepared by McCarthy *et al.* (8). We essentially followed their procedure and used the methods of Tener (9) and of Hoard and Ott (10)

<https://www.pnas.org/content/pnas/74/2/560.full.pdf>

Февраль

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC431765/>

Декабрь

Метод Максама-Гилберта

CTCAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGG
GCTCCGGGGACACTTTGCGTTCGGGC



P^*O_4 CTCAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGC
TGGGCTCCGGGGACACTTTGCGTTCGGGC

P^*O_4 CTCAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGC
TGGGCTCCGGGGACACTTTGCGTTCGGGC

Метод Максама-Гилберта

A+G G T+C C



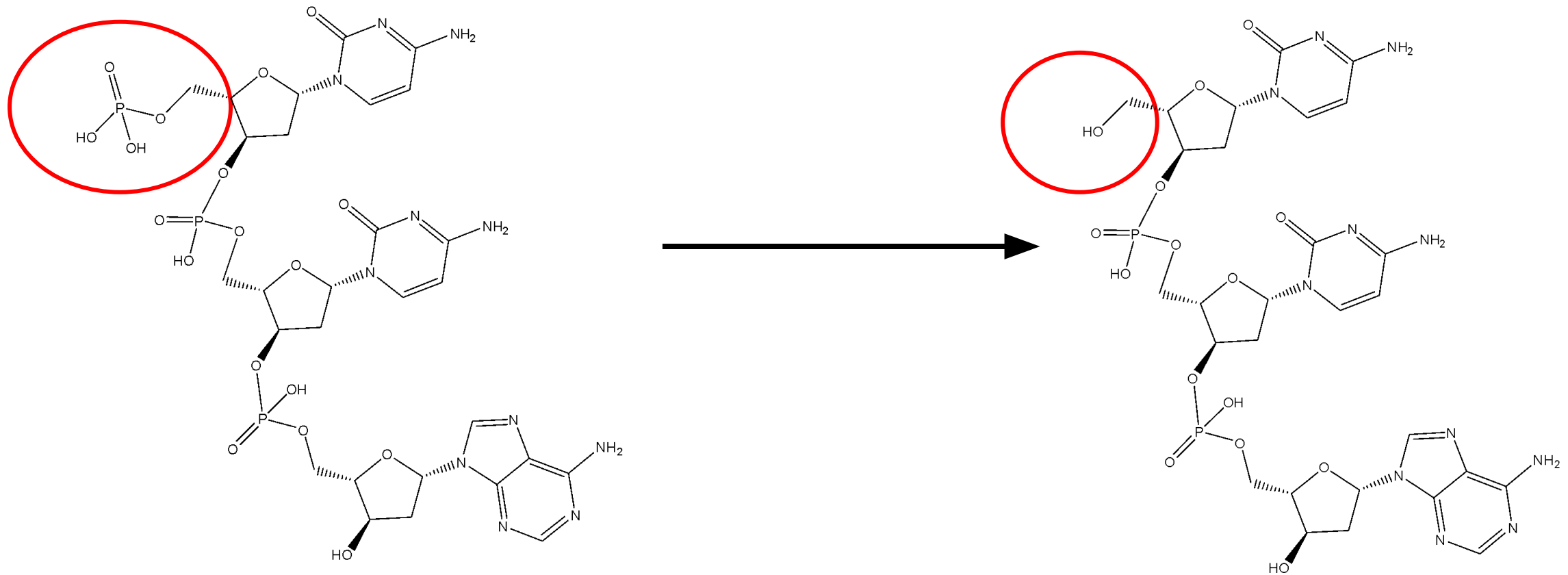
P^*O_4 CTCAAAAGTCTAGAGCCACCGTCCA
GGGAGCAGGTAGCTGCTGGGGCTCCGGGG
ACACTTTGCGTTCGGGC

Метод Максама-Гилберта (1980 г.)

- Удалить немеченный фосфат с 5'-конца молекулы с помощью фосфатазы (она же – фосфомоноэстераза)
- Ввести меченный фосфат на 5'-конец с помощью полинуклеотидкиназы и АТФ с меченым γ -фосфатом
- В четырех разных реакциях (для G, для A+G, для T+C и для C) провести расщепление одной фосфодиэфирной связи на каждую молекулу по одному случайному соответствующему нуклеотиду
- С помощью электрофореза в полиакриламидном геле разделить получившиеся фрагменты

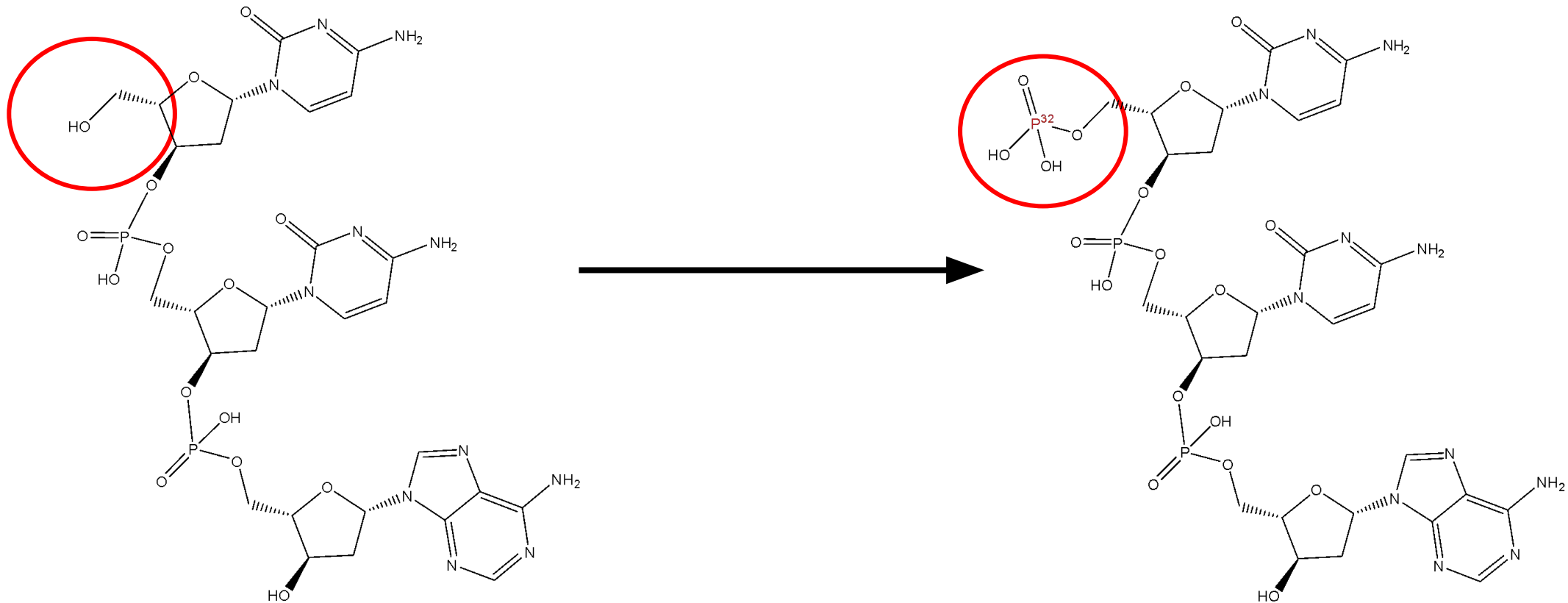
Метод Максама-Гилберта (1980 г.)

- Удалить немеченный фосфат с 5'-конца молекулы с помощью фосфатазы (она же – фосфомоноэстераза)



Метод Максама-Гилберта (1980 г.)

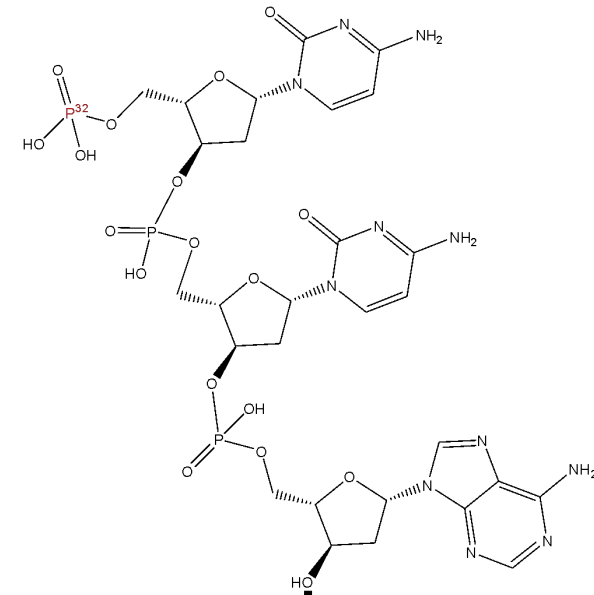
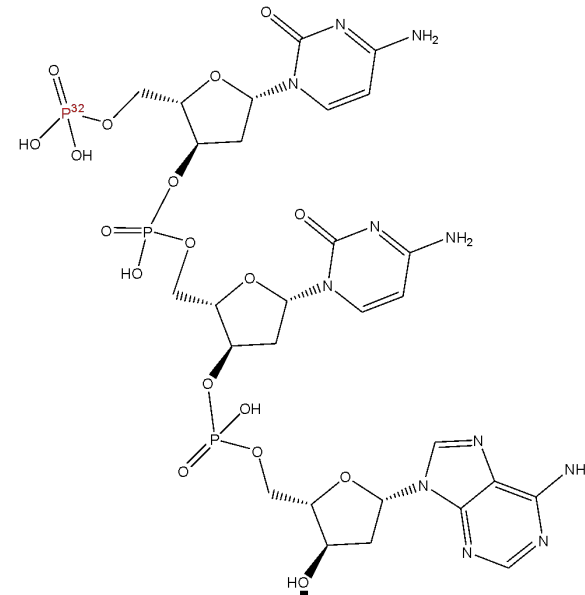
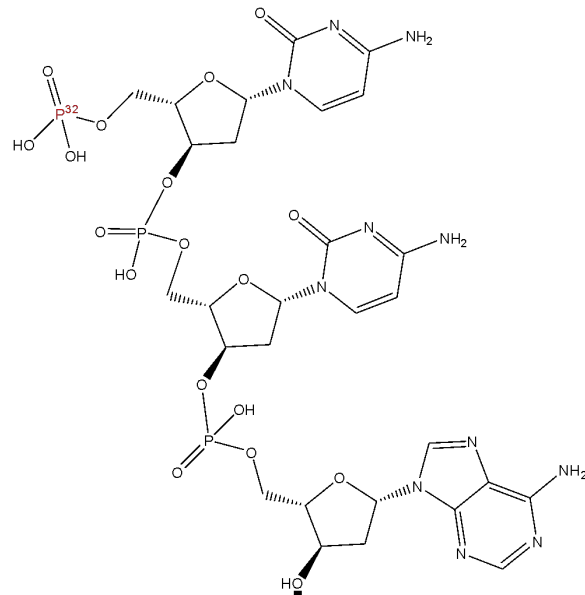
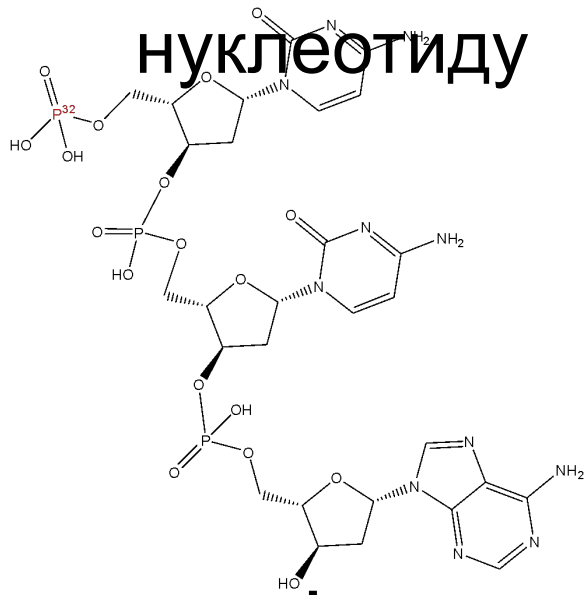
- Ввести меченный фосфат на 5'-конец с помощью полинуклеотидкиназы и АТФ с меченым γ -фосфатом



Метод Максама-Гилберта (1980 г.)

- В четырех разных реакциях (для G, для A+G, для T+C и для C) провести расщепление одной фосфодиэфирной связи на каждую молекулу по одному случайному соответствующему

нуклеотиду



Разрезаем после:

Метод Максама-Гилберта (1980 г.)

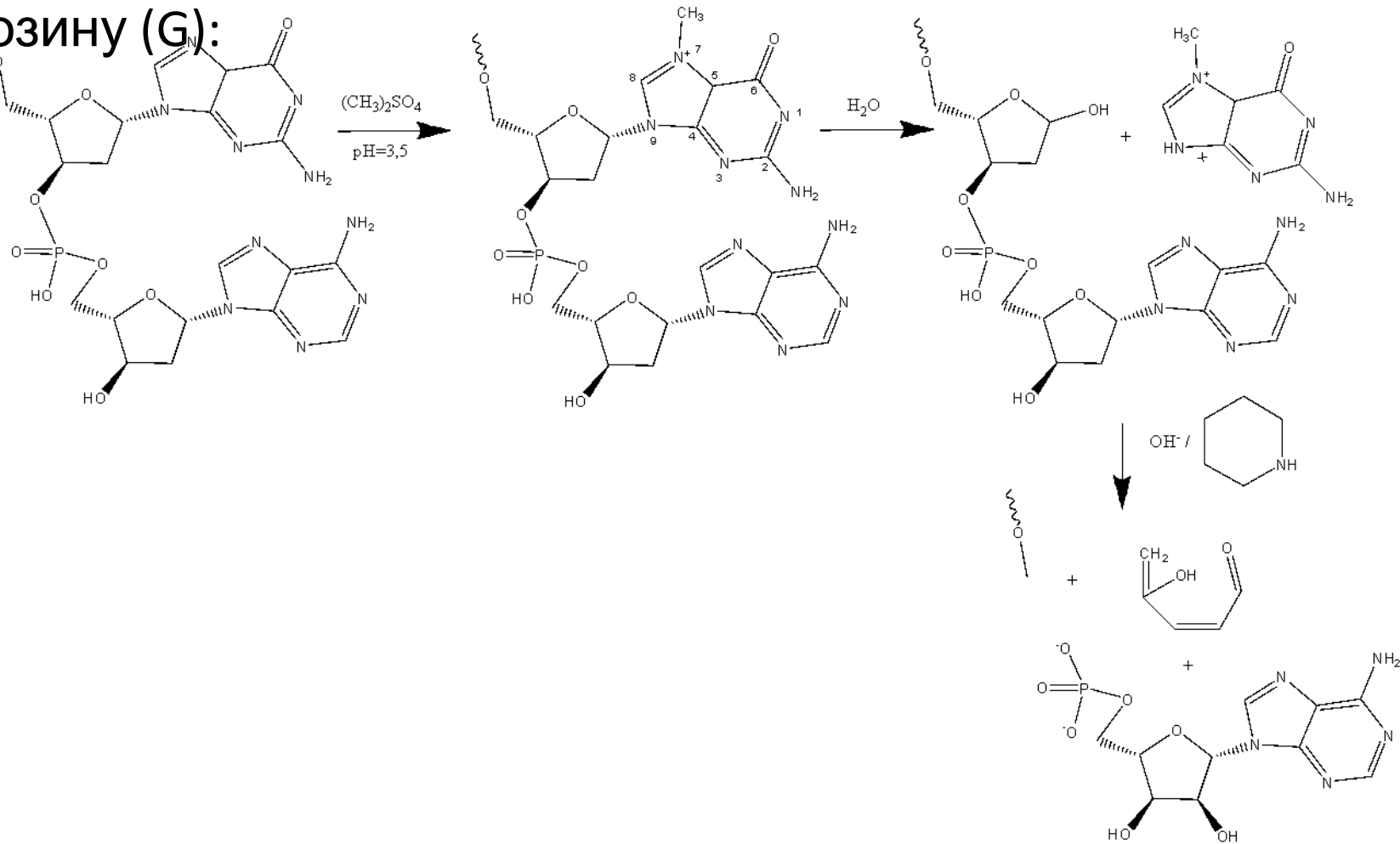
P^*O_4 **CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTTAGCTGCT**
GGGCTCCGGGGACACTTTGCGTTCCGGGC

↓ разрезаем после C

P^*O_4
 P^*O_4 **CT**
 P^*O_4 **CTCAAAAGT**
 P^*O_4 **CTCAAAAGTCTAGAG**
 P^*O_4 **CTCAAAAGTCTAGAGC**
 P^*O_4 **CTCAAAAGTCTAGAGCCA**
 P^*O_4 **CTCAAAAGTCTAGAGCCAC**
 P^*O_4 **CTCAAAAGTCTAGAGCCACCGT**
 P^*O_4 **CTCAAAAGTCTAGAGCCACCGTC**
 P^*O_4 **CTCAAAAGTCTAGAGCCACCGTCCAGGGAG**
 P^*O_4 **CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCA**
 P^*O_4 **CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTTAG**
 P^*O_4 **CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTTAGCTG** и т.д.

Метод Максама-Гилберта (1980 г.)

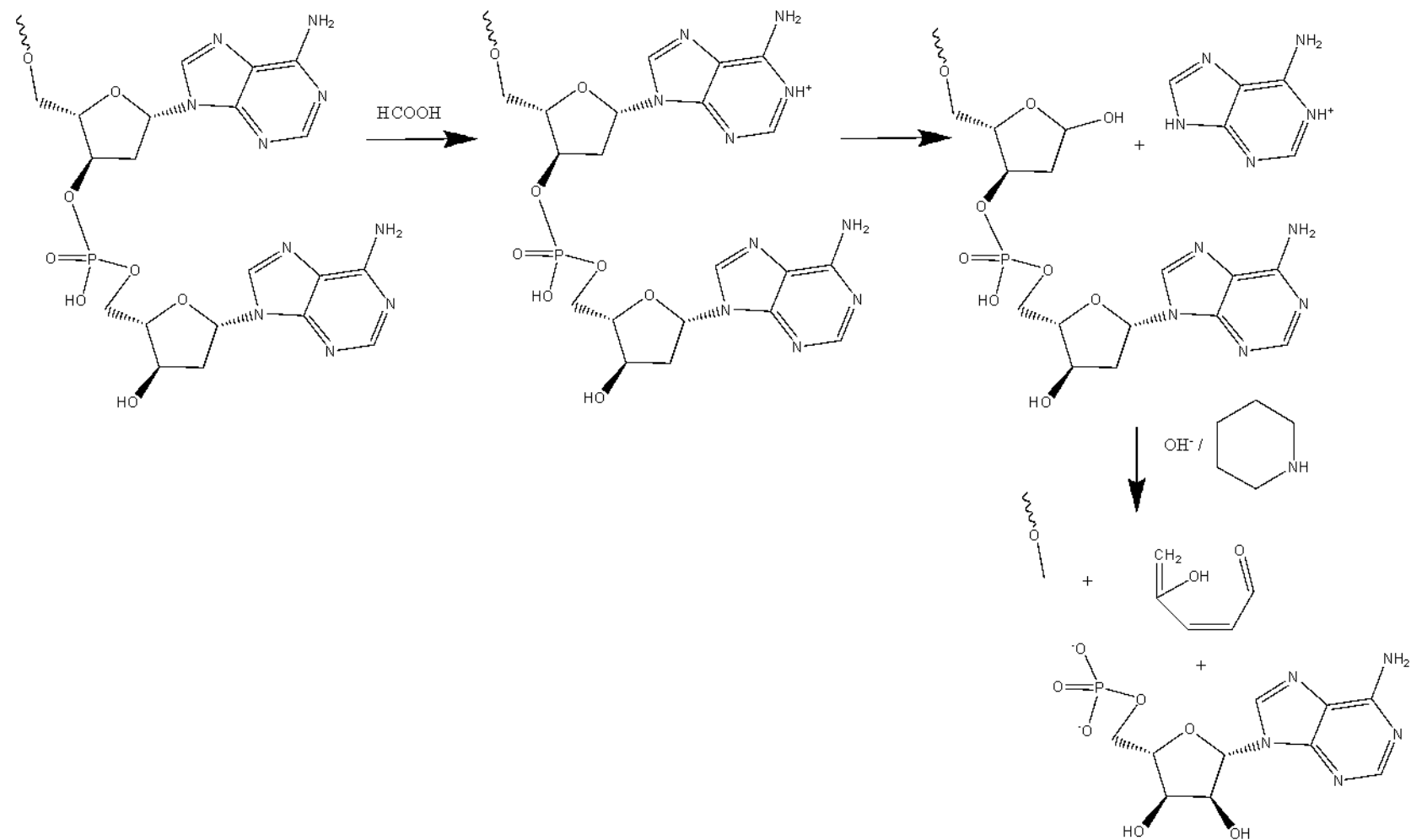
Как разрезать фосфодиэфирную связь по гуанозину (G):



Метод Максама-Гилберта (1980 г.)

Как разрезать фосфодиэфирную связь по аденозинам и гуанозинам

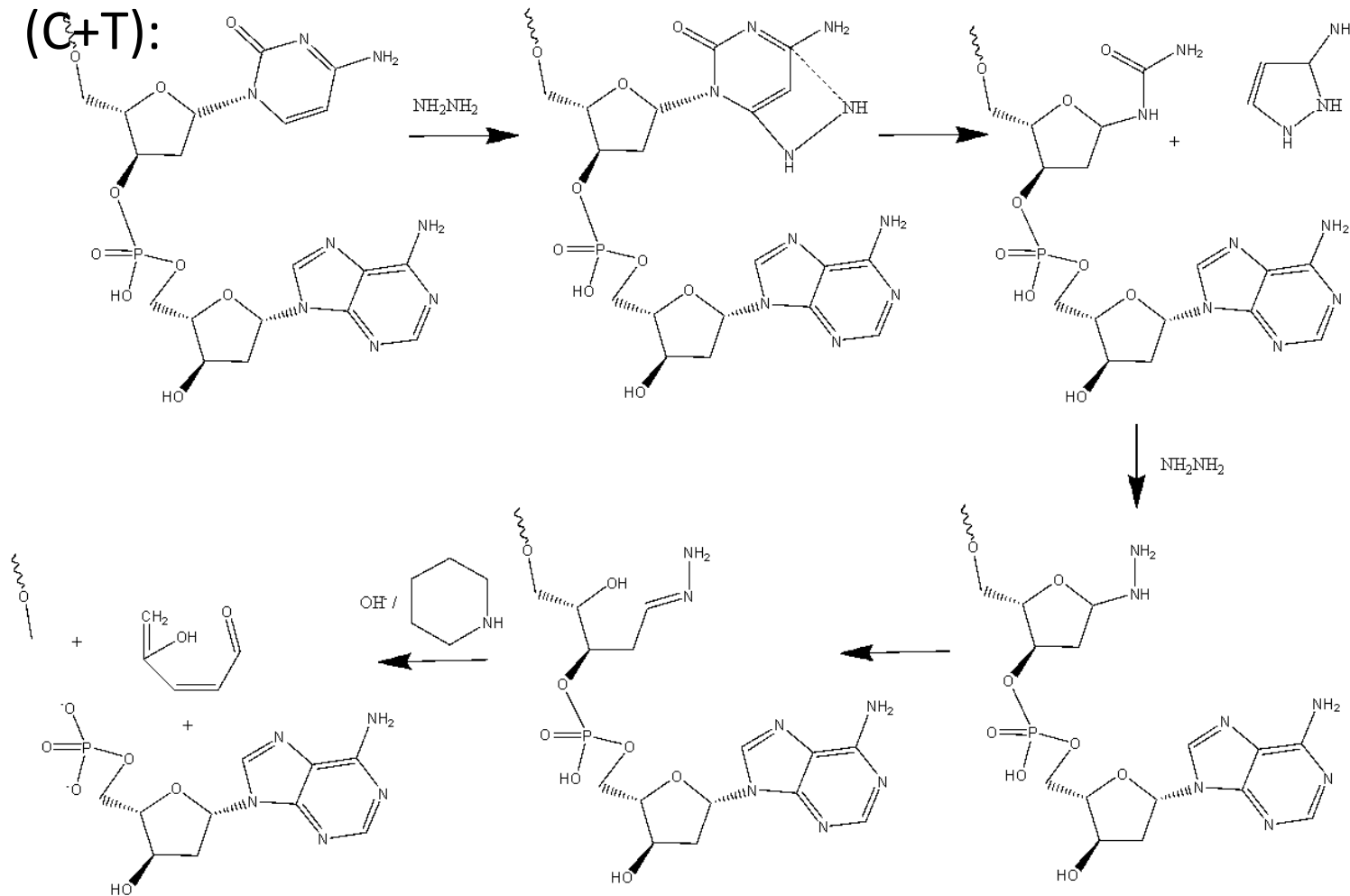
(A+G):



Метод Максама-Гилберта (1980 г.)

Как разрезать фосфодиэфирную связь по цитидинам и тимидинам

(C+T):

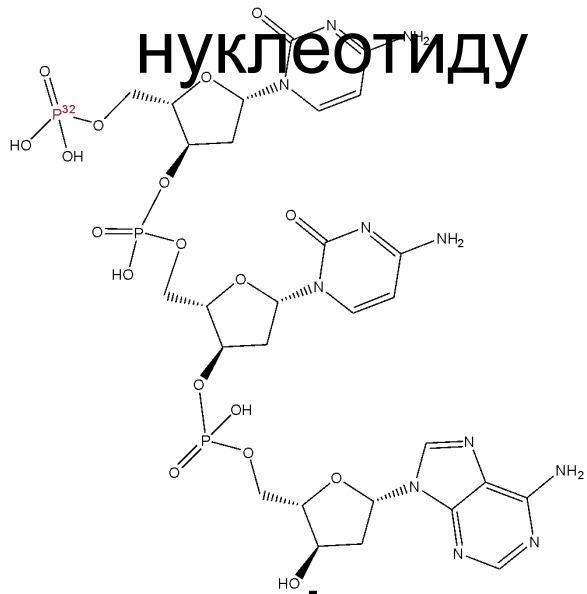


В щелочных условиях (0,1 М КОН) или при высокой соли (1 М NaCl) реакция проходит только для дезоксицитидинов

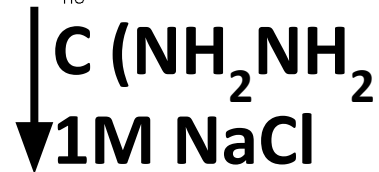
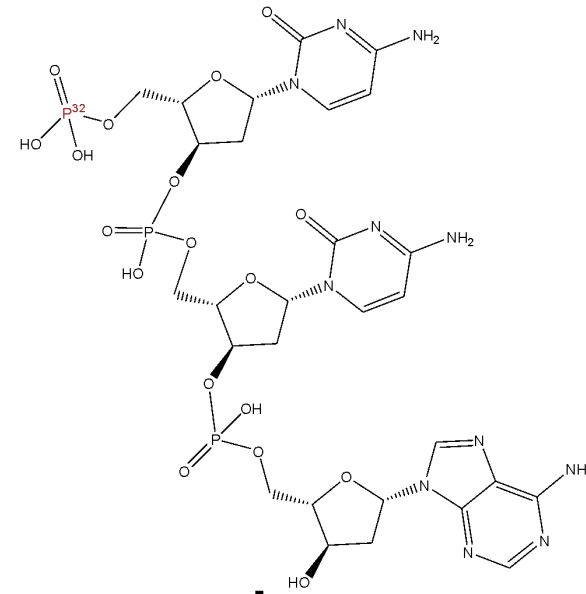
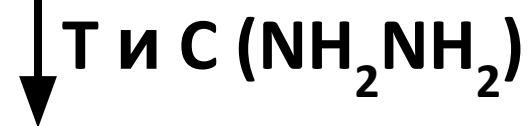
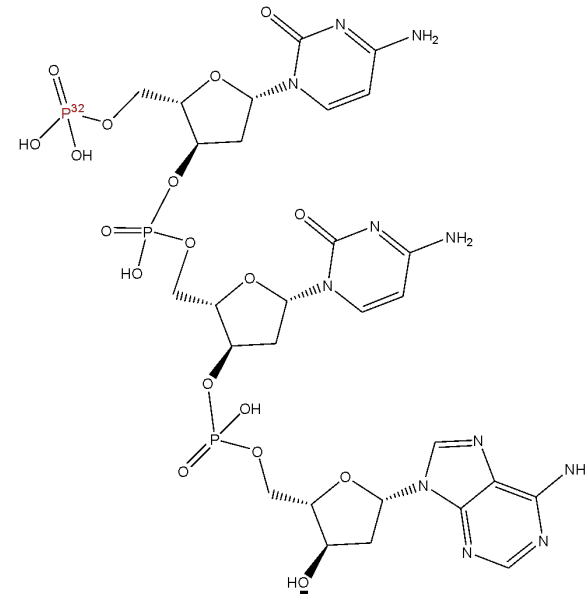
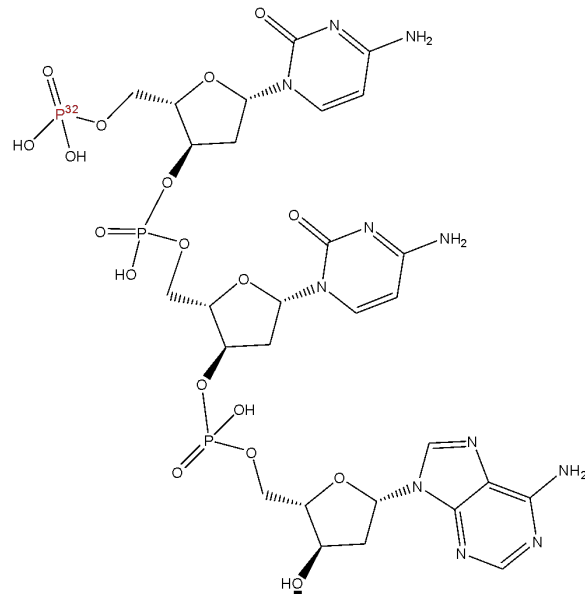
Метод Максама-Гилберта (1980 г.)

- В четырех разных реакциях (для G, для A+G, для T+C и для C) провести расщепление одной фосфодиэфирной связи на каждую молекулу по одному случайному соответствующему

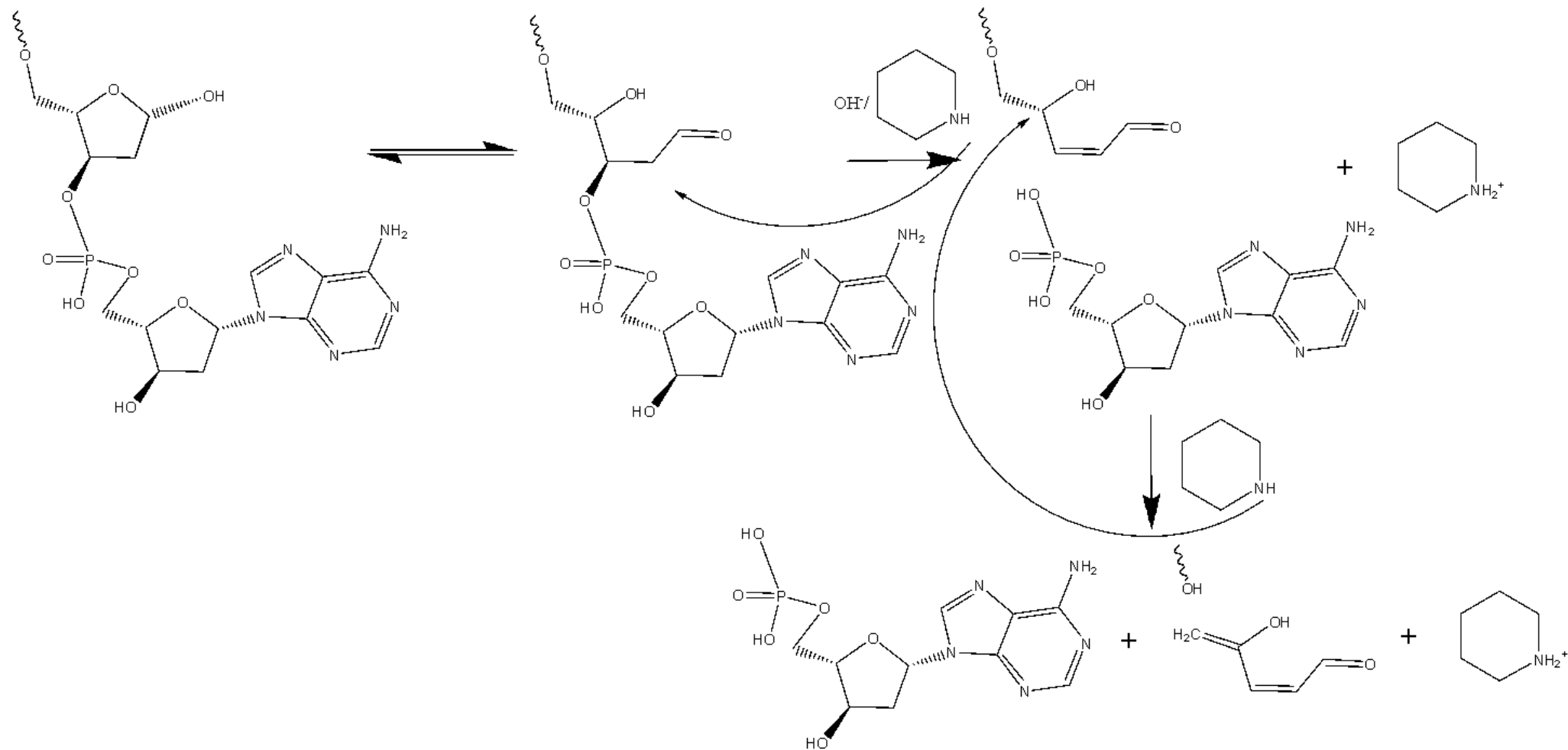
нуклеотиду



Разрезаем после:



Метод Максама-Гилберта



Метод Максама-Гилберта (1980 г.)

P^*O_4 CTC AAA AGT CTAGAGCCACCGTCCAGGGAGCAGGGTAGCTGCT
GGGCTCCGGGGACACTTTGCGTTTCGGGC

↓ по C

P^*O_4
 P^*O_4 CT
 P^*O_4 CTC AAA AGT
 P^*O_4 CTC AAA AGT CTAGAG
 P^*O_4 CTC AAA AGT CTAGAGC
 P^*O_4 CTC AAA AGT CTAGAGCCA
 P^*O_4 CTC AAA AGT CTAGAGCCAC и т.
д.

↓ по G

P^*O_4 CTC AAAA
 P^*O_4 CTC AAAA AGTCTA
 P^*O_4 CTC AAAA AGTCTAGA
 P^*O_4 CTC AAAA AGTCTAGAGCCACC
 P^*O_4 CTC AAAA AGTCTAGAGCCACCGTCCA
 P^*O_4 CTC AAAA AGTCTAGAGCCACCGTCCAG
 P^*O_4 CTC AAAA AGTCTAGAGCCACCGTCCAGGGA и т.д.

↓ по A+G

P^*O_4 CTC
 P^*O_4 CTC A
 P^*O_4 CTC AA
 P^*O_4 CTC AAA
 P^*O_4 CTC AAAA AGTCT
 P^*O_4 CTC AAAA AGTCTAG
 P^*O_4 CTC AAAA AGTCTAGAGCC и т.д.

↓ по T+C

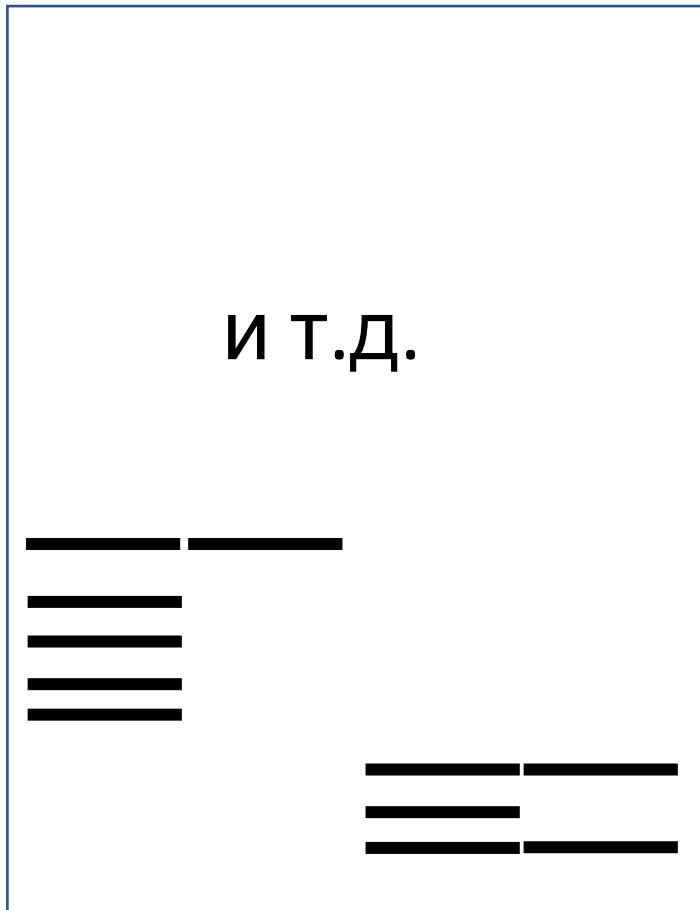
P^*O_4 C
 P^*O_4 CTC AAAA AG
 P^*O_4 CTC AAAA AGTC
 P^*O_4 CTC AAAA AGTCTAGAGCCACCG и т.д.



Электрофорез в
ПААГ

Метод Максама-Гилберта

A+G G T+C C



P^*O_4 CTCAAAAGTCTAGAGCCACCGTCCA
GGGAGCAGGTAGCTGCTGGGGCTCCGGGG
ACACTTTGCGTTCGGGC

Метод Сэнгера

5' -GAGTTTTTCAGATCTCGGT-3' -OH

|||||

3' -CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGC -5'

+ фрагмент Клёнова
+ dATP + dTTP + dCTP + dGTP

+ddATP*

+ddTTP*

+ddCTP*

+ddGTP*

GAGTTTTTCAGATCTCGGTGGC**A**-H
GAGTTTTTCAGATCTCGGTGGCAGGTCCCTCGTCCA-H
GAGTTTTTCAGATCTCGGTGGCAGGTCCCTCGTCCATCGA-H
и т.д.

GAGTTTTTCAGATCTCGGTGGCAGGT**H**-H
GAGTTTTTCAGATCTCGGTGGCAGGTCCCT-H
и т.д.

GAGTTTTTCAGATCTCGGTGGC**H**-H
GAGTTTTTCAGATCTCGGTGGCAGGT**C**-H
и т.д.

GAGTTTTTCAGATCTCGGT**G**-H
GAGTTTTTCAGATCTCGGTGG**H**-H
GAGTTTTTCAGATCTCGGTGGCAG**H**-H
и т.д.

Электрофорез в
ПААГ

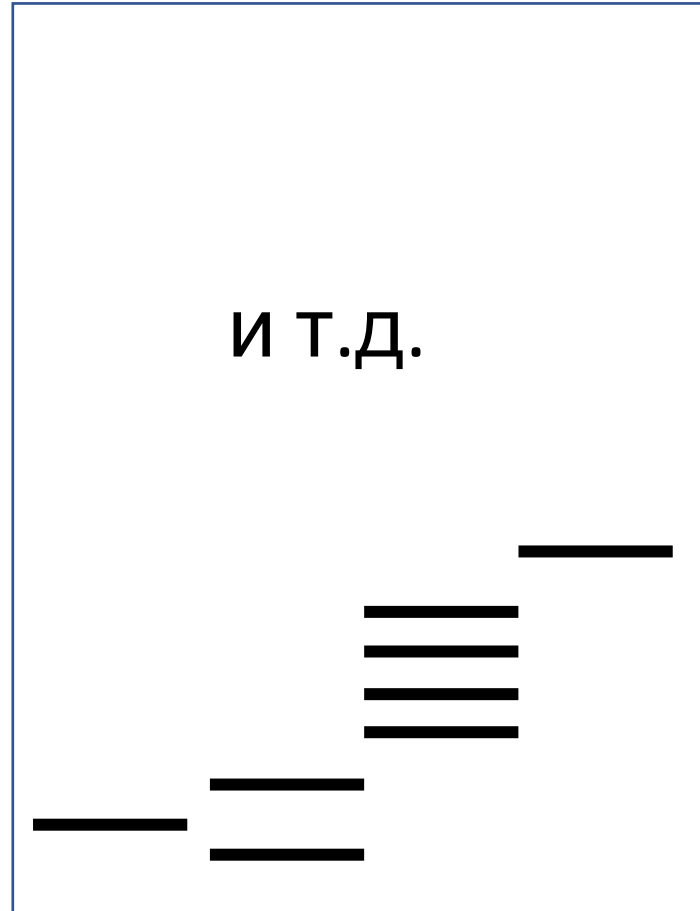
Метод Сэнгера

A+G G T+C C



Максама-Гилберта

A G T C



Сэнгер

GAGTTTTTCAGATCTCGGTGGCAGGTCCCTCG
TCCATCGACGACCCGAGGCCCTGTGAAACG
CAAGCCCG

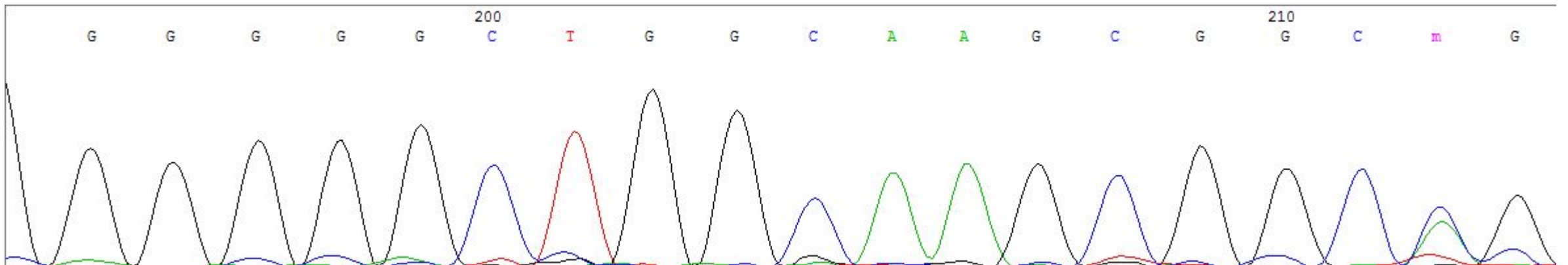
Метод Сэнгера vs Максама-Гилберта

- Требуется большого количества исходного материала
- Требуется знания о районе для праймера
- + Нет высокотоксичных соединений
- + Возможность автоматизации

- + Возможно секвенирование сразу после выделения ДНК
- + Можно секвенировать любую последовательность
- ДМС и гидразин – сильные канцерогены
- Трудно автоматизировать

Метод Сэнгера. Усовершенствования

- Радиоактивное мечение заменено на флуорофоры
- Каждый из типов нуклеотидов мечен отдельным флуорофором
- Разделение фрагментов и обработка результатов электрофореза проводятся автоматически на приборе



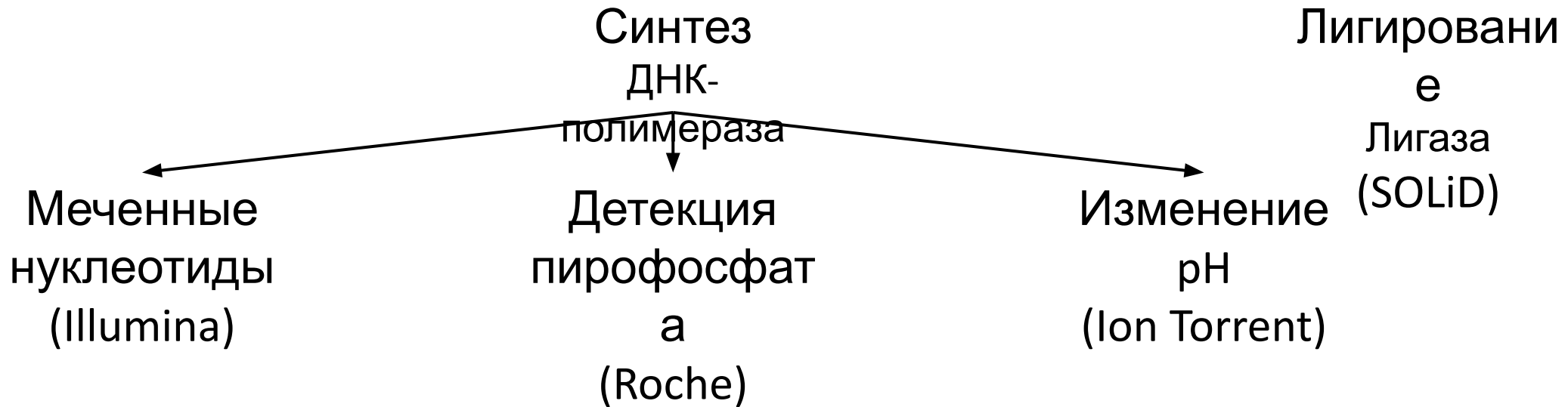
Секвенирование второго (следующего) поколения – Next generation sequencing

- Появление ПЦР в 1985 году
- Появление термостабильной ДНК-полимеразы в 1986 году
- Развитие микрочиповых исследований с 1995 года
- Крупные вложения в технологии секвенирования в процессе секвенирования генома человека в начале 2000-х

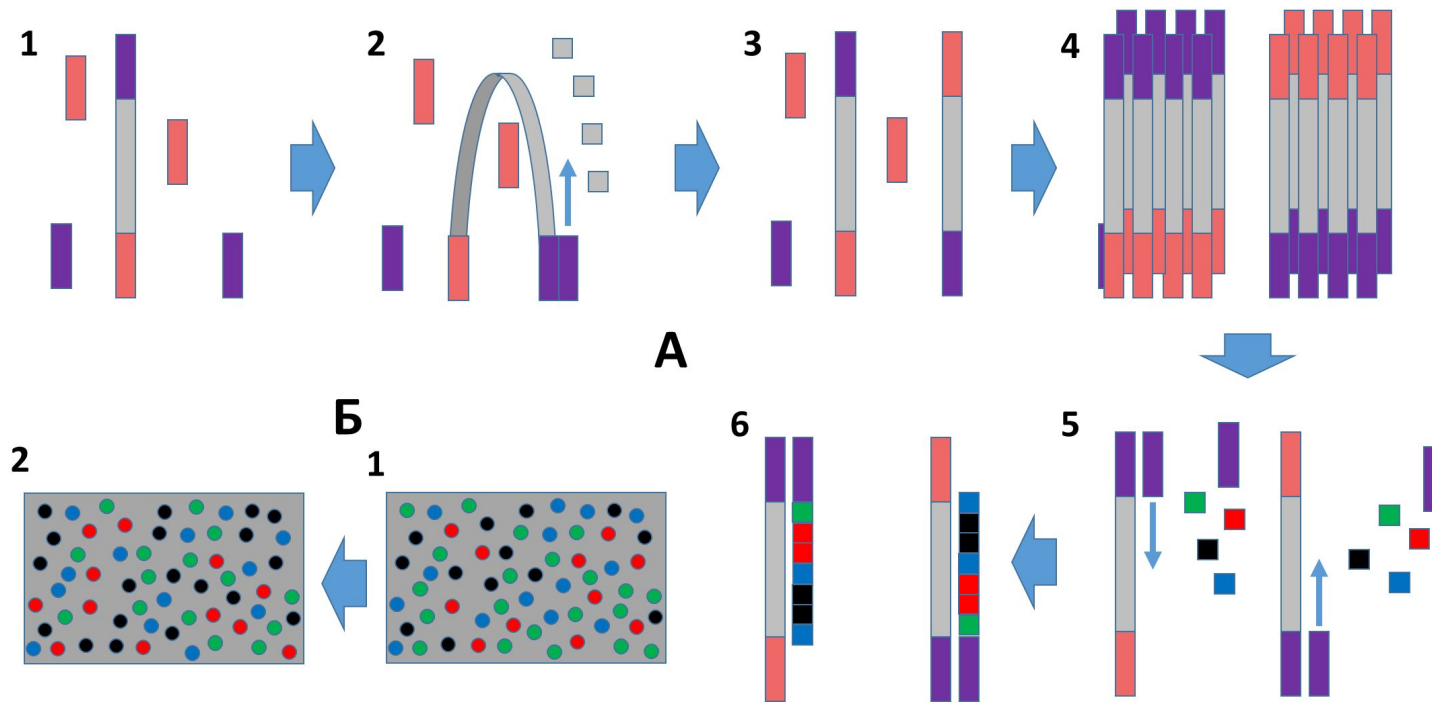
Отличия NGS от «обычного» секвенирования

- Не требуется проведение электрофореза
- Чтение нуклеотидов осуществляется в процессе секвенирования
- Секвенируются одновременно сотни тысяч, миллионы или миллиарды последовательностей (против 8–96 в реакции Сэнгера)
- Реакции секвенирования проходят на поверхности или в лунках

Варианты подходов в NGS (второе поколение)

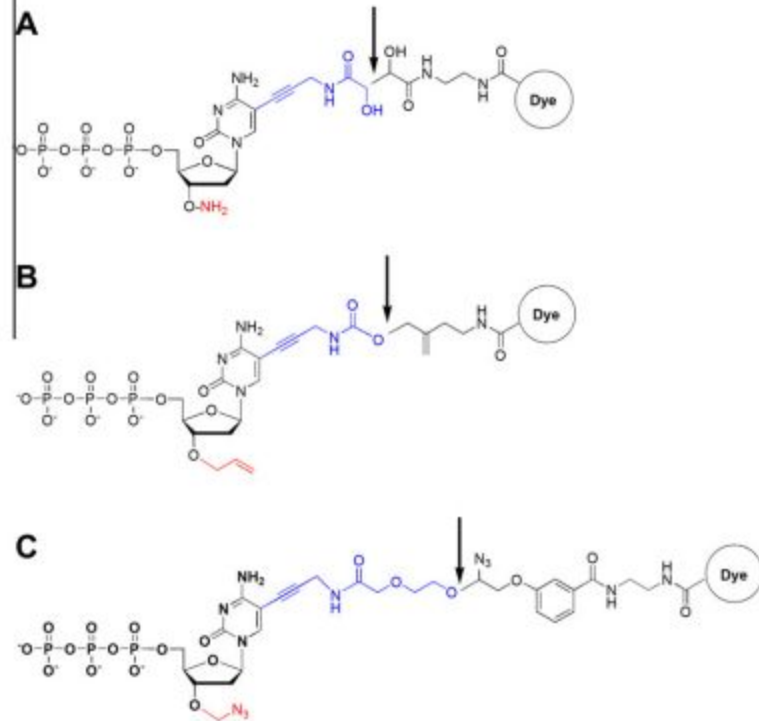


Варианты подходов в NGS (второе поколение): технология Illumina (Solexa)

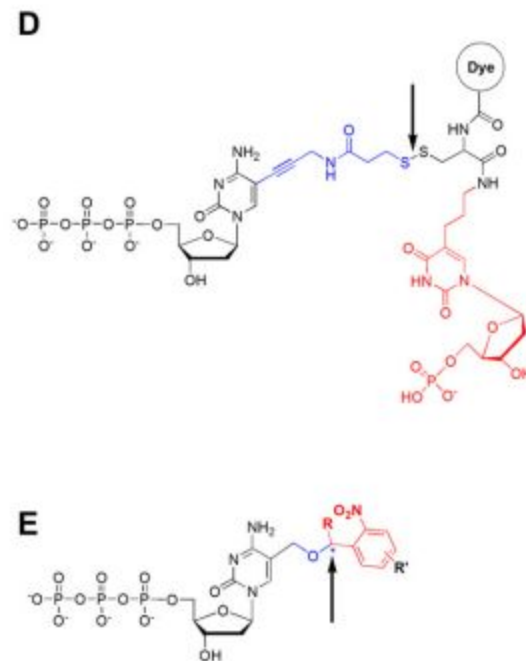


Варианты подходов в NGS (второе поколение): технология Illumina (Solexa)

3'-blocked reversible terminators

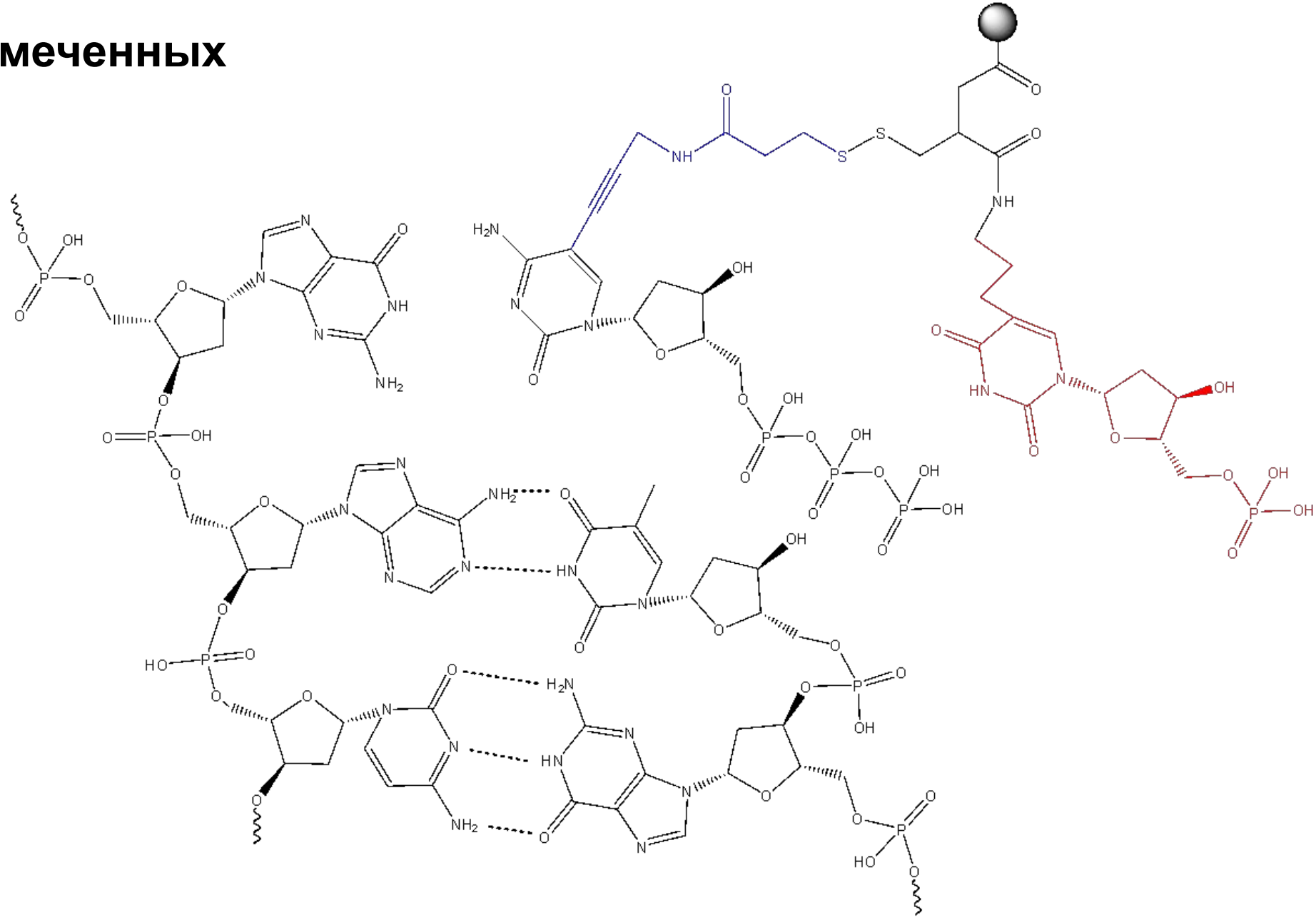
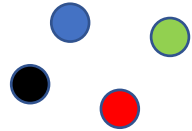


3'-unblocked reversible terminators

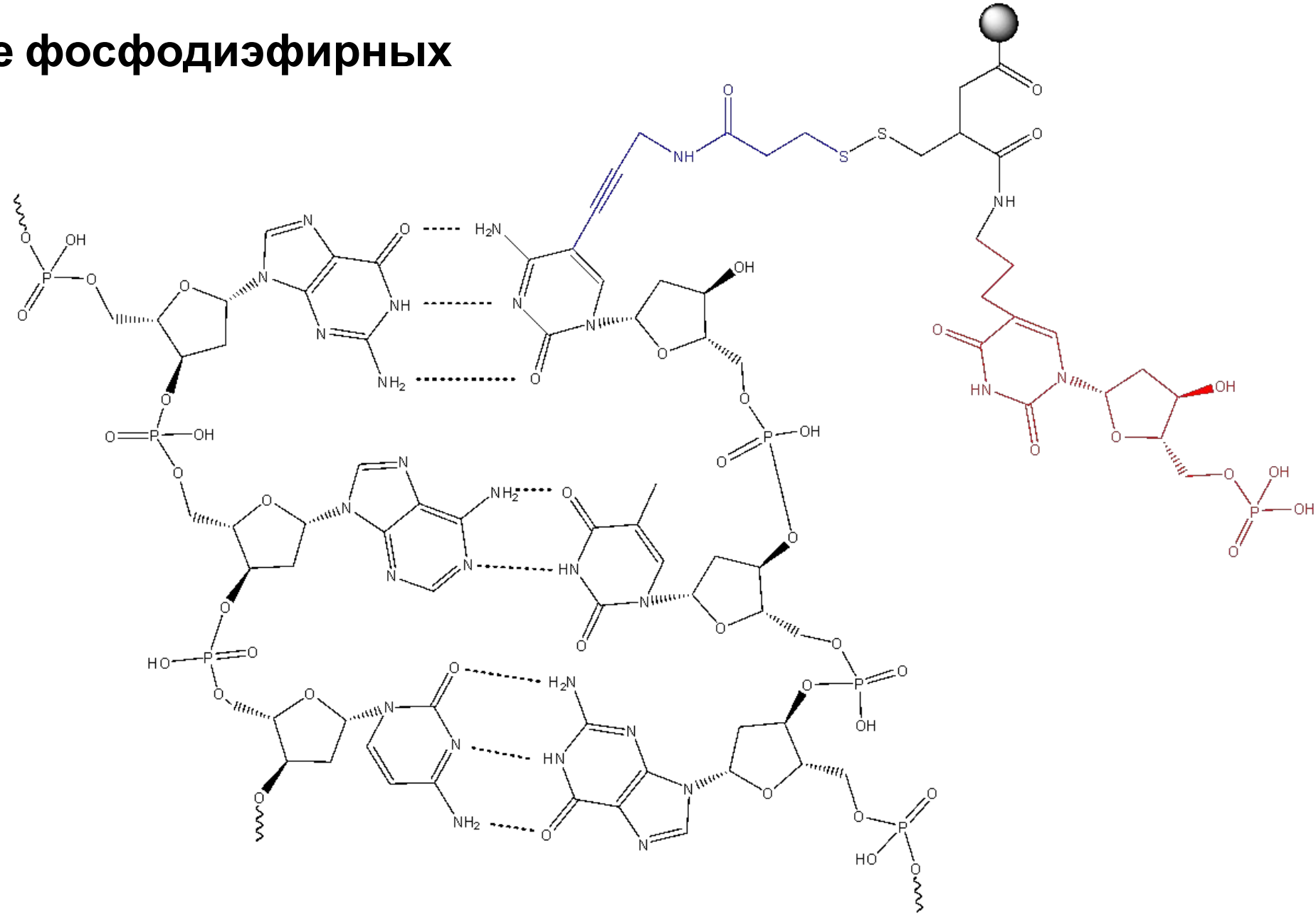
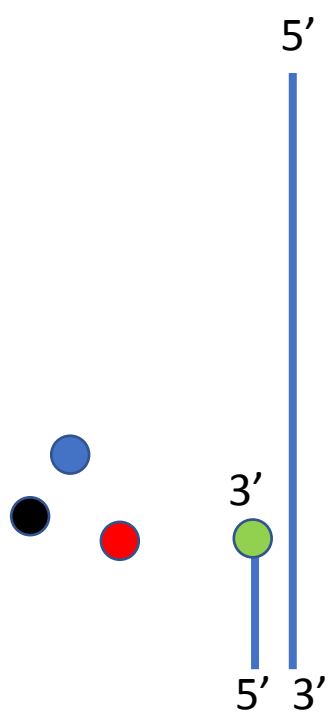


Красным – терминирующая группа.
Могут быть 3'-блокирующие (слева) и 3'-неблокирующие (справа)
Синее – неудаляемый заместитель

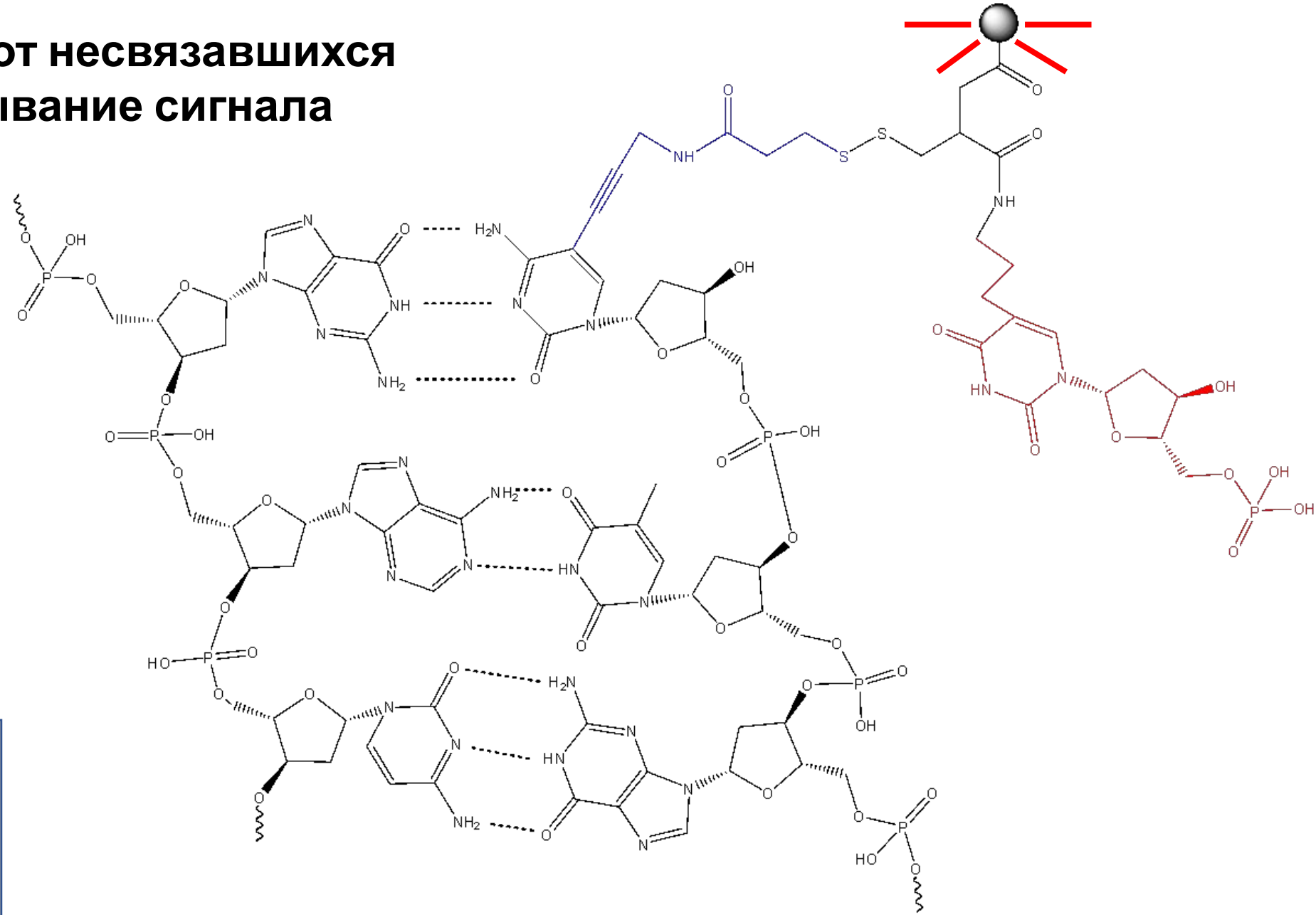
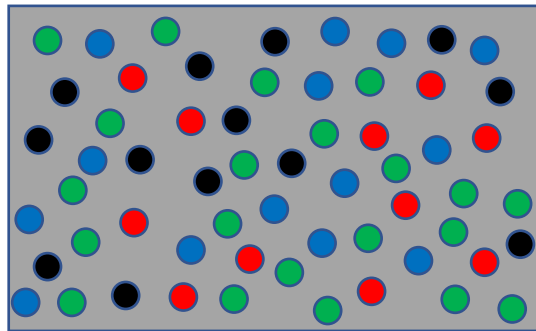
Этап 1 – Добавление меченных дНТФ



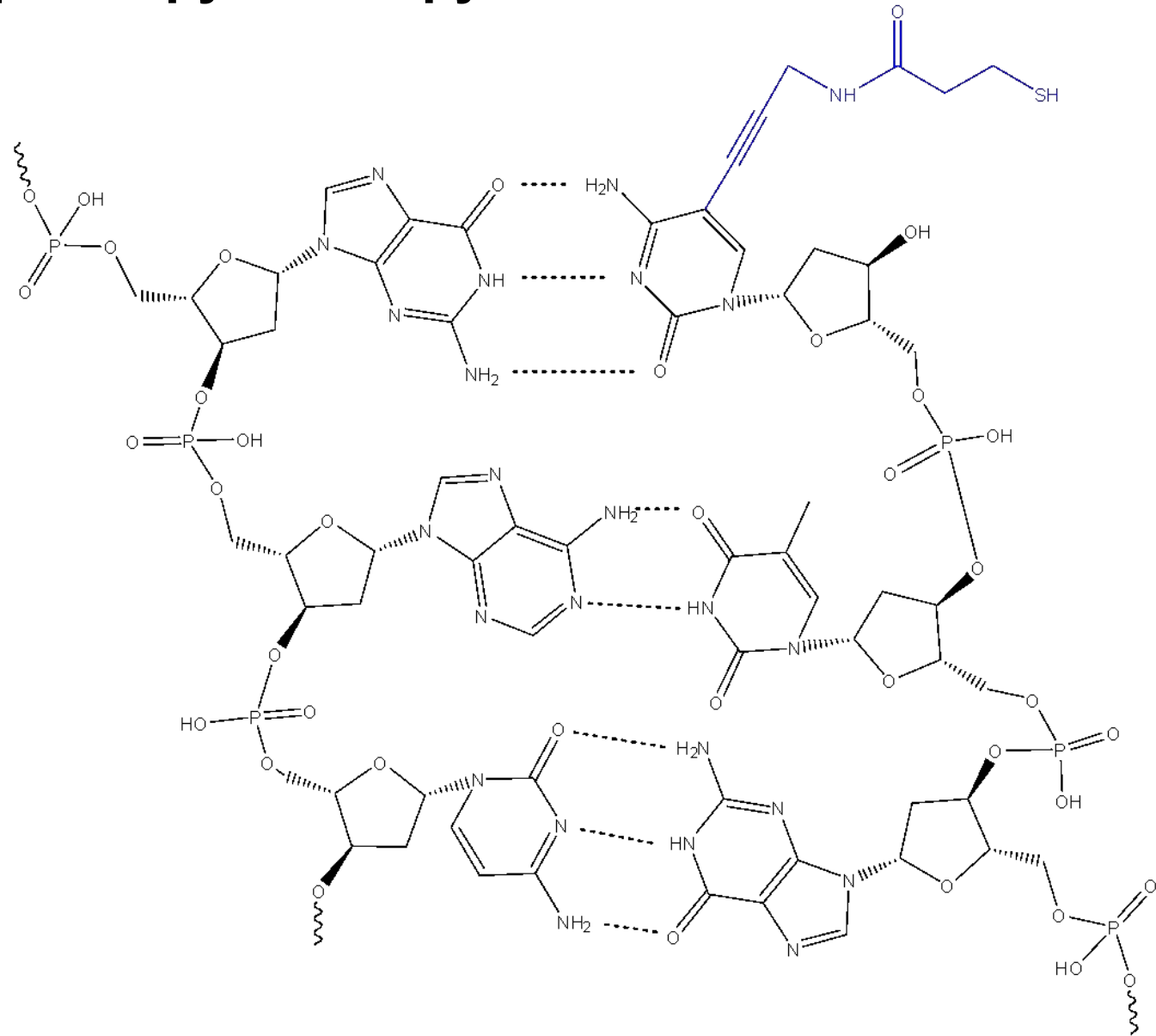
Этап 2 – Образование фосфодиэфирных связей



Этап 3 – Отмывание от несвязавшихся нуклеотидов и считывание сигнала

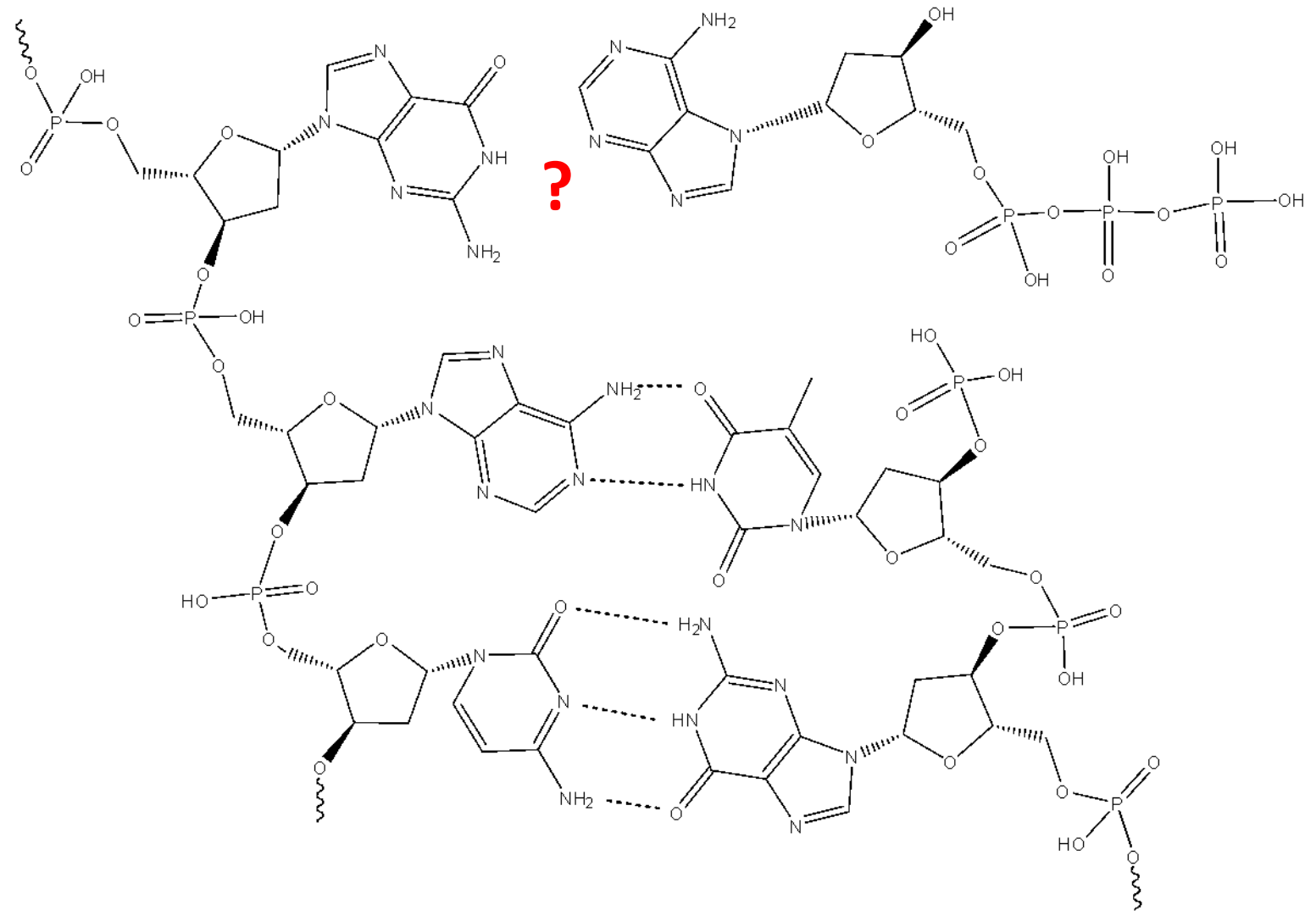


Этап 4 – Удаление терминирующей группы



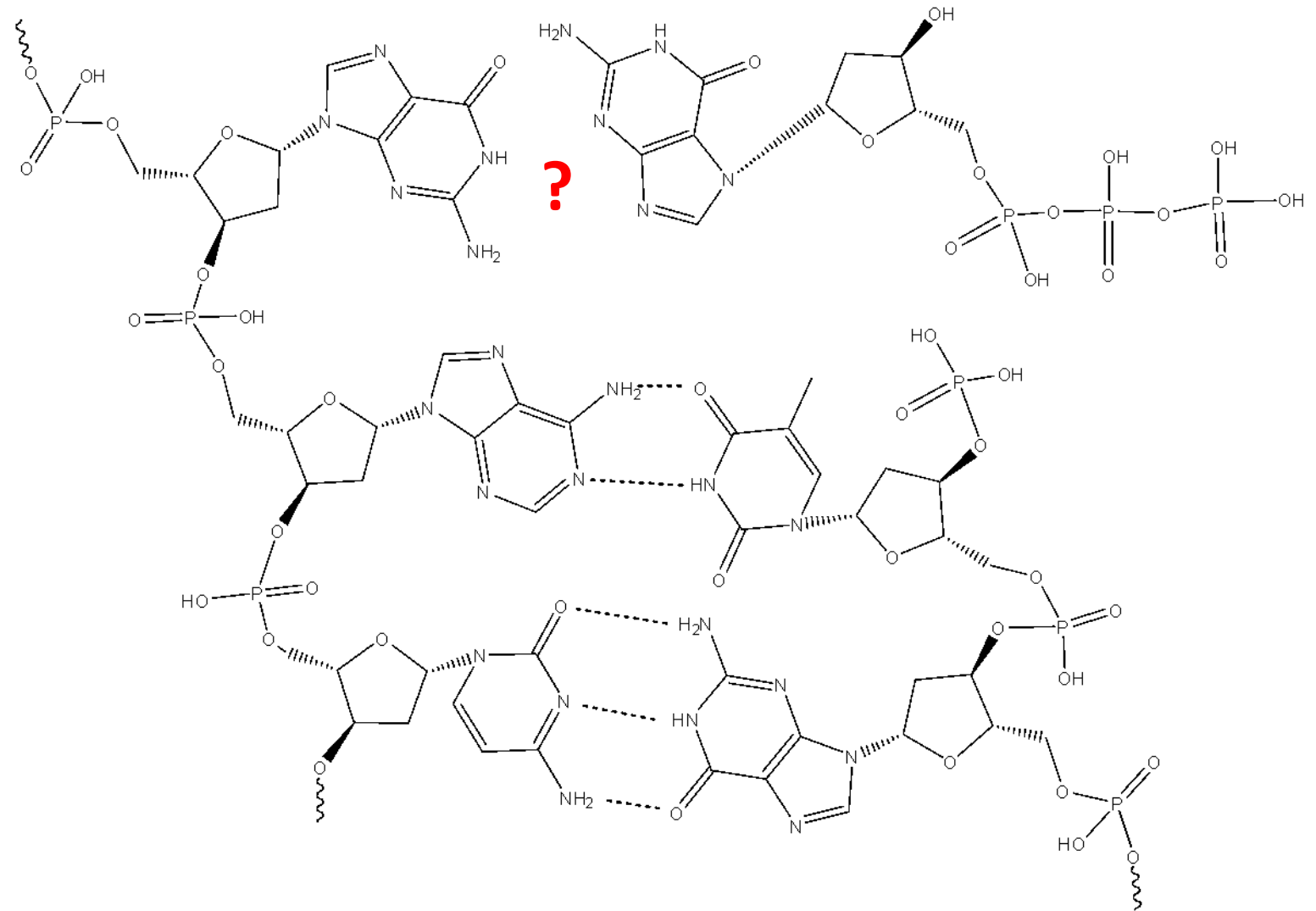
Варианты подходов в NGS (второе поколение): технология 454 (Roche)

Каждый нуклеотид добавляется поочередно, и в каждой точке проверяется, выделился ли пирофосфат



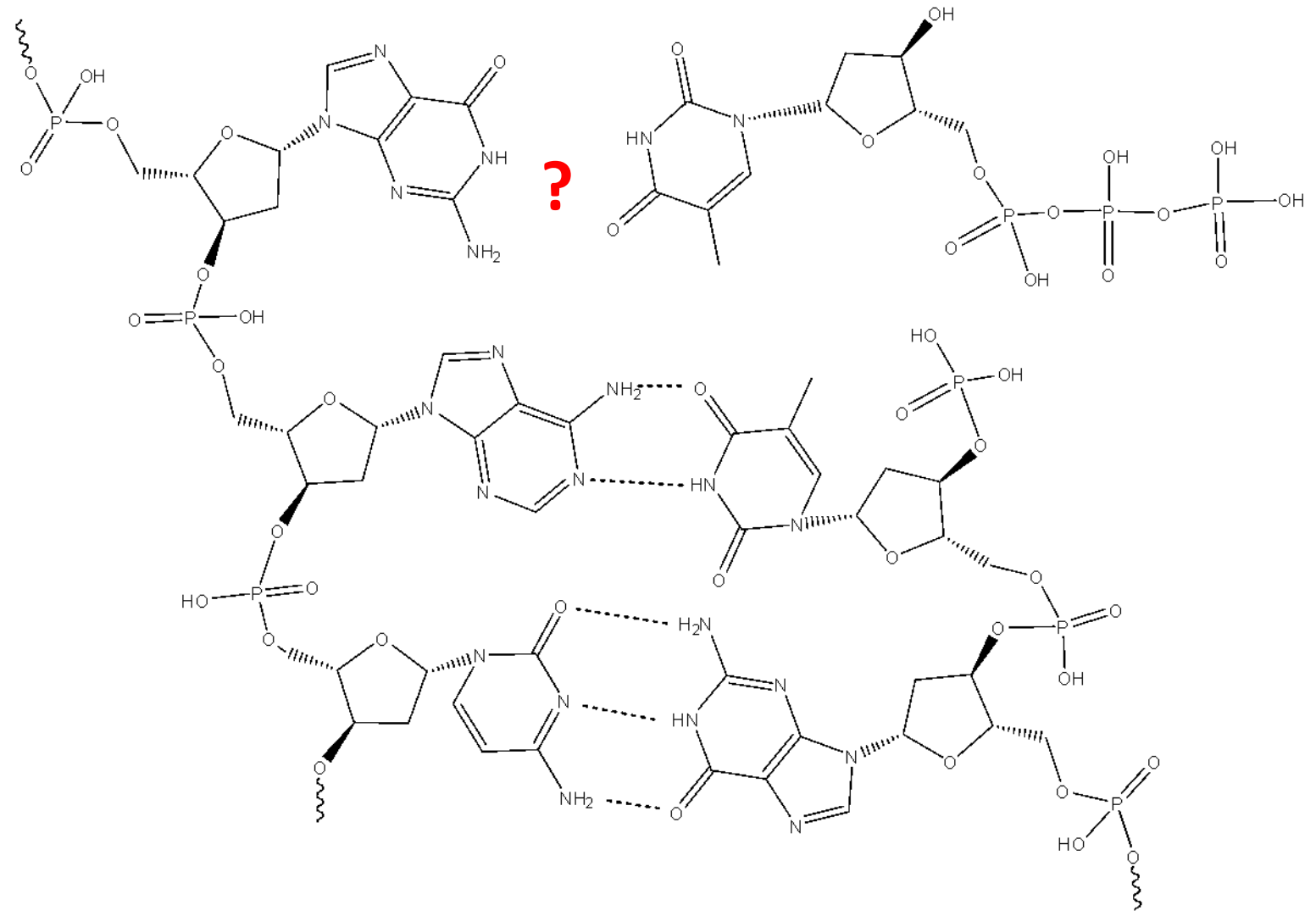
Варианты подходов в NGS (второе поколение): технология 454 (Roche)

Каждый нуклеотид добавляется поочередно, и в каждой точке проверяется, выделился ли пирофосфат



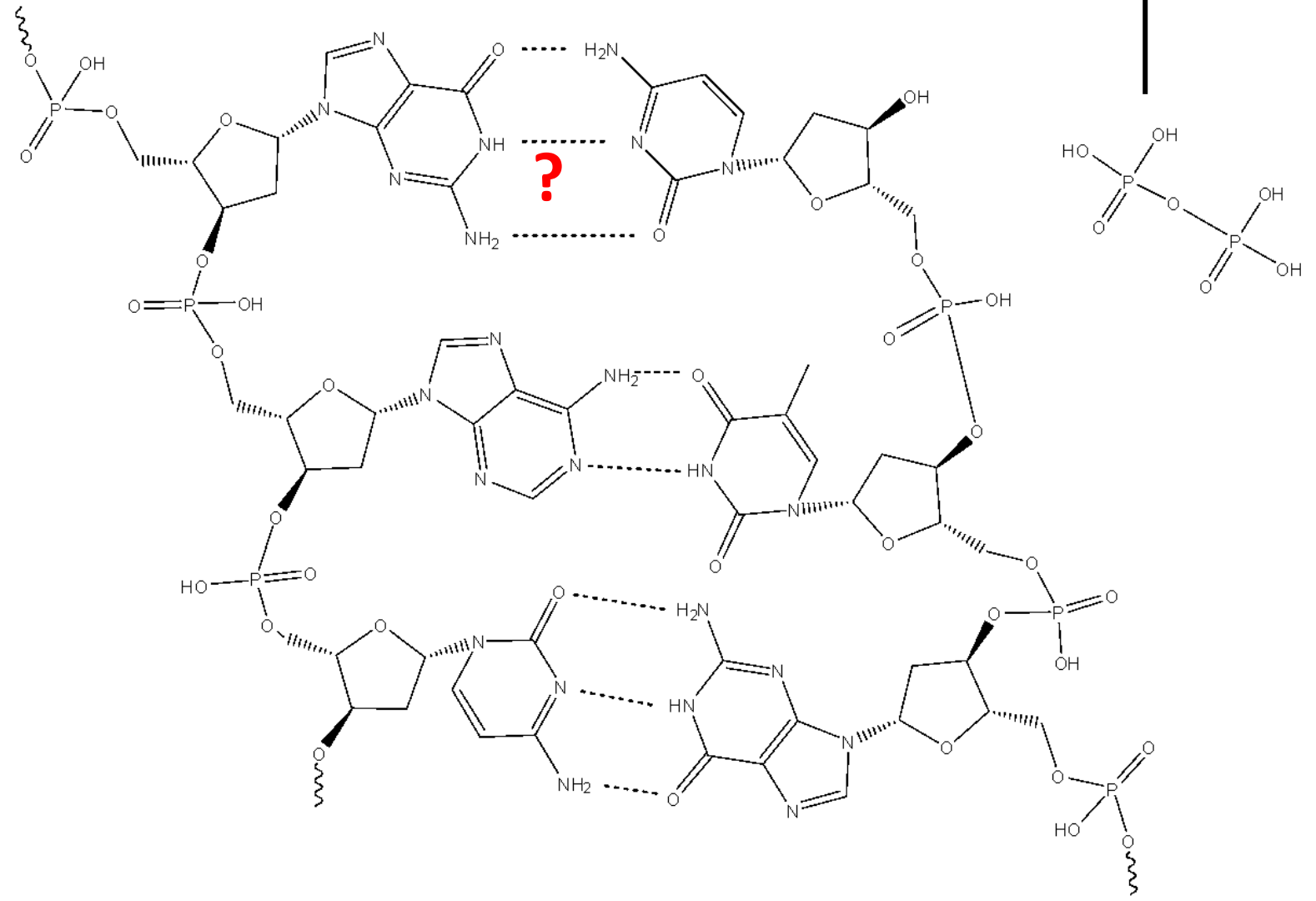
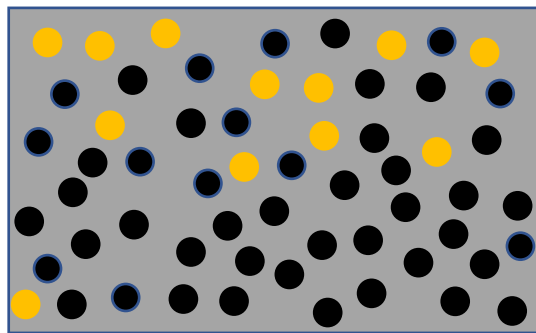
Варианты подходов в NGS (второе поколение): технология 454 (Roche)

Каждый нуклеотид добавляется поочередно, и в каждой точке проверяется, выделился ли пирофосфат



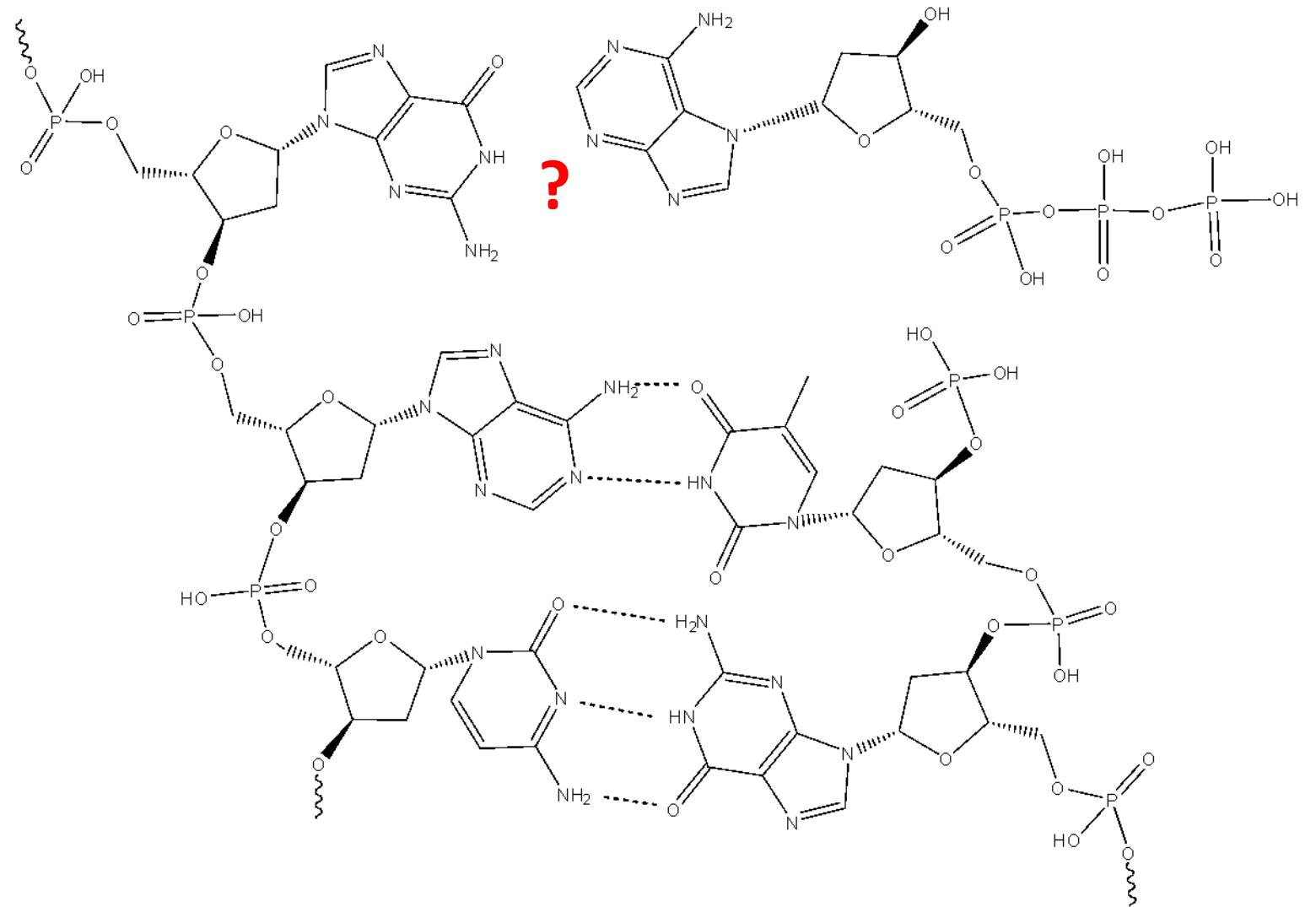
Варианты подходов в NGS (второе поколение): технология 454 (Roche)

Каждый нуклеотид добавляется поочередно, и в каждой точке проверяется, выделился ли пирофосфат



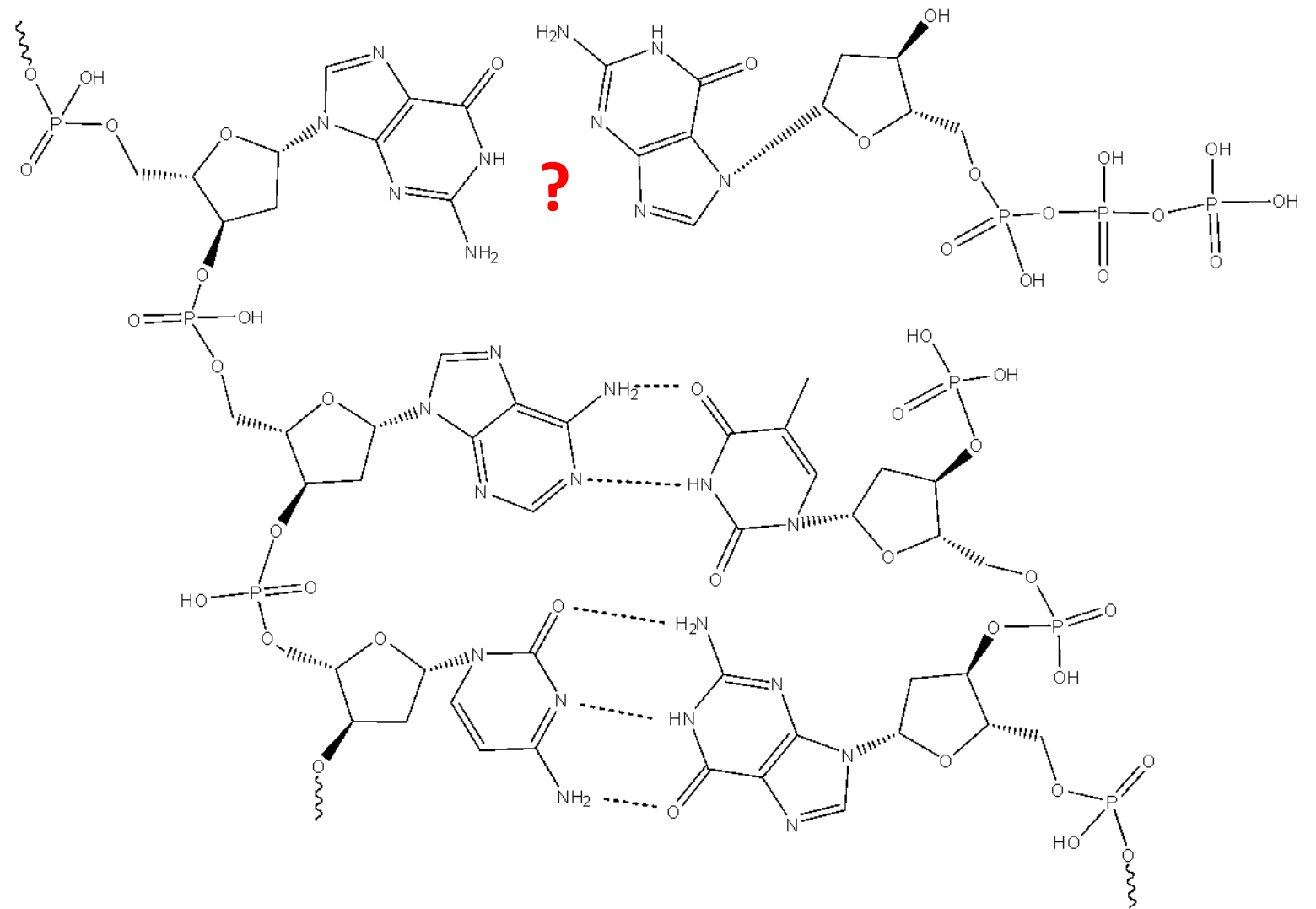
Варианты подходов в NGS (второе поколение): технология Ion Torrent (ThermoFisher)

Каждый нуклеотид добавляется поочередно, и в каждой точке проверяется, изменился ли pH



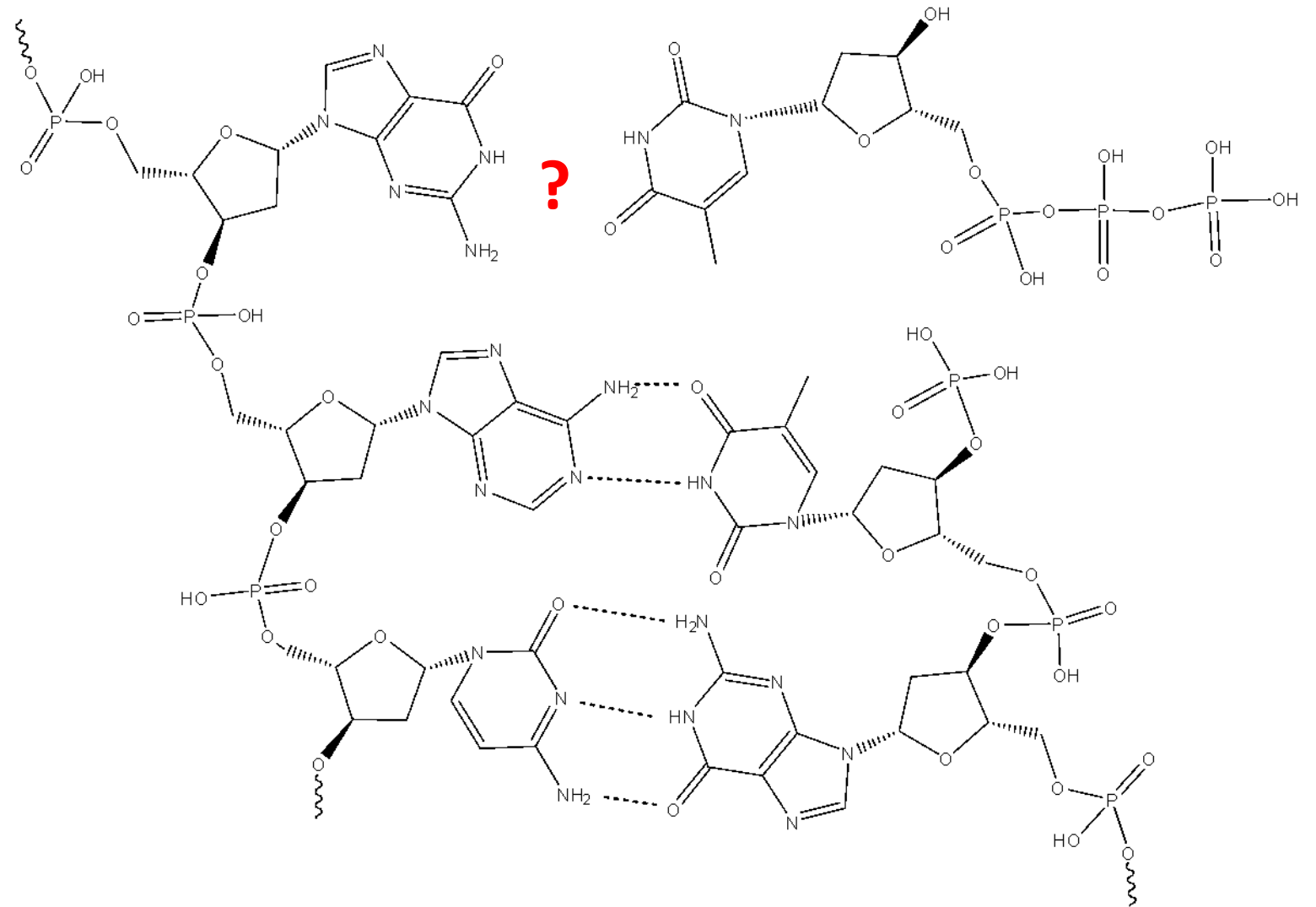
Варианты подходов в NGS (второе поколение): технология Ion Torrent (ThermoFisher)

Каждый нуклеотид добавляется поочередно, и в каждой точке проверяется, изменился ли pH



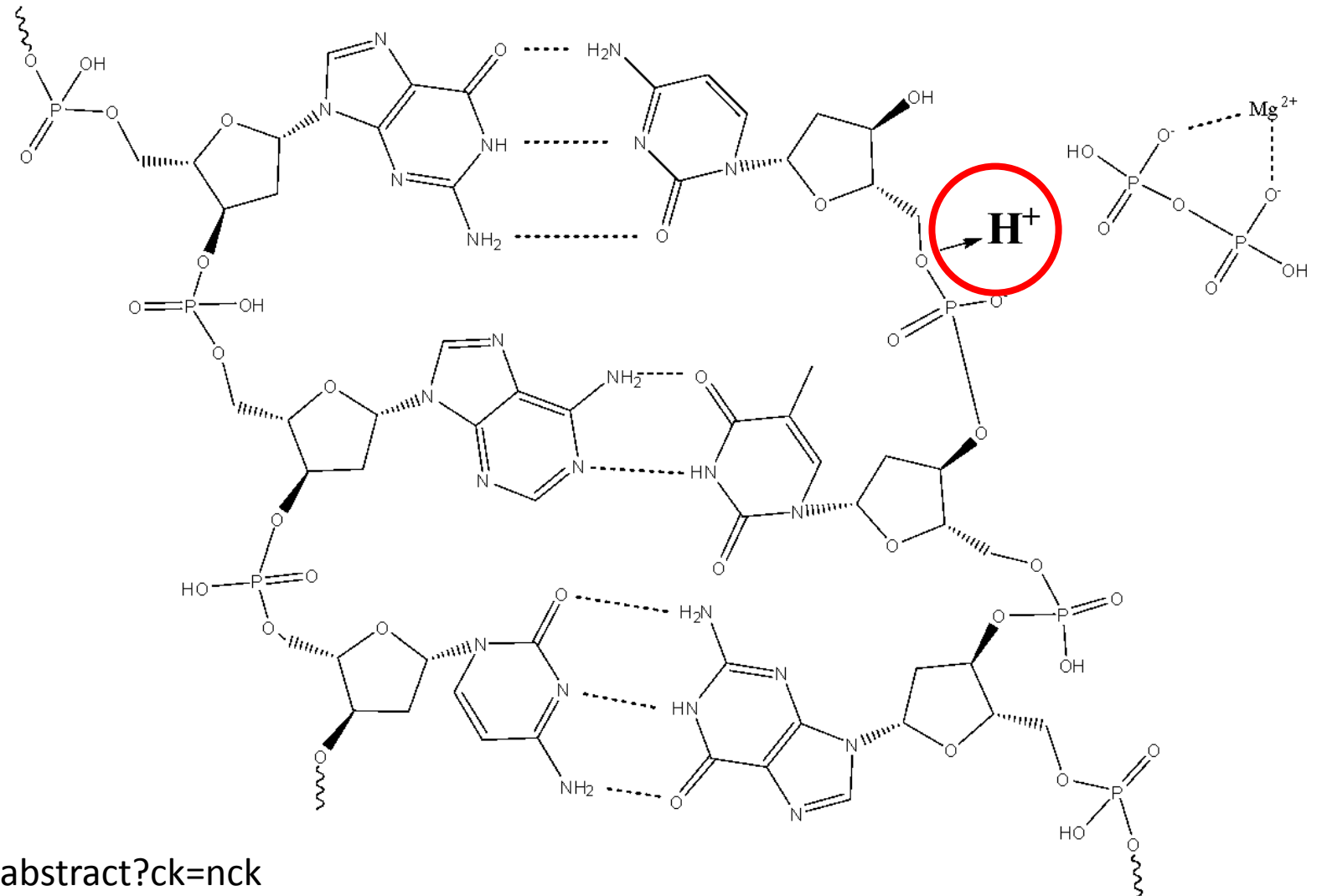
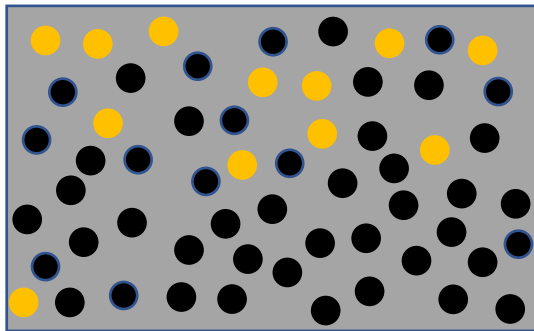
Варианты подходов в NGS (второе поколение): технология Ion Torrent (ThermoFisher)

Каждый нуклеотид добавляется поочередно, и в каждой точке проверяется, изменился ли pH

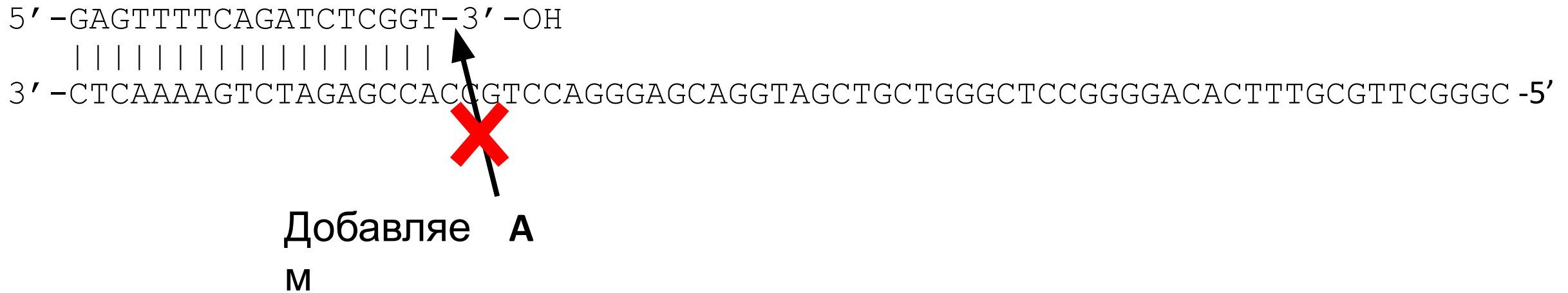


Варианты подходов в NGS (второе поколение): технология Ion Torrent (ThermoFisher)

Каждый нуклеотид добавляется поочередно, и в каждой точке проверяется, изменился ли pH



Варианты подходов в NGS (второе поколение): основная проблема пиросеквенирования и Ion Torrent (ThermoFisher)



Варианты подходов в NGS (второе поколение): основная проблема пиросеквенирования и Ion Torrent (ThermoFisher)

5' -GAGTTTTTCAGATCTCGGT-3' -OH

|||||

3' -CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGC -5'

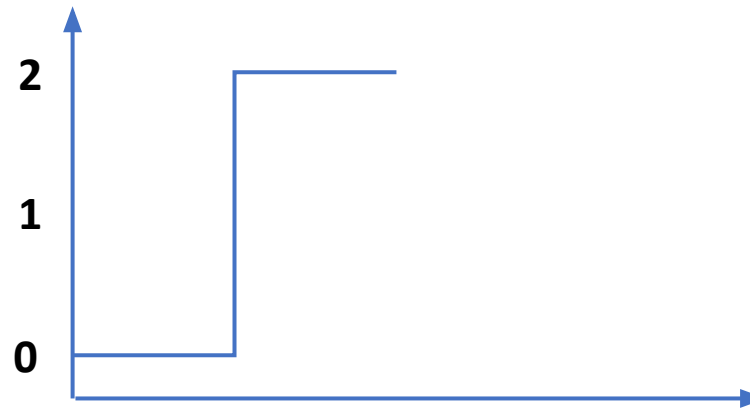
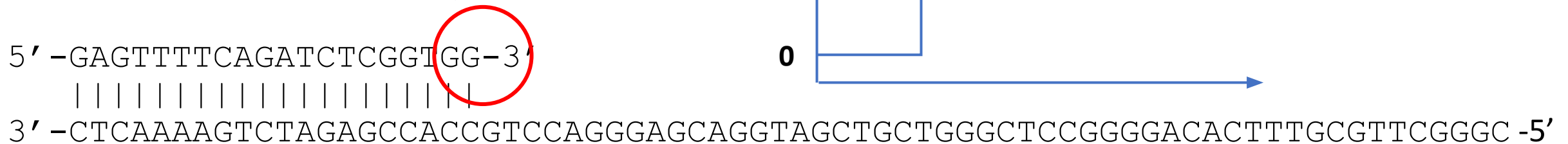


Добавляе Т
М

Варианты подходов в NGS (второе поколение): основная проблема пиросеквенирования и Ion Torrent (ThermoFisher)



Добавляе **G**
M



Варианты подходов в NGS (второе поколение): основная проблема пиросеквенирования и Ion Torrent (ThermoFisher)

5' -GAGTTTTTCAGATCTCGGTGG-3' -OH

|||||

3' -CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGC -5'

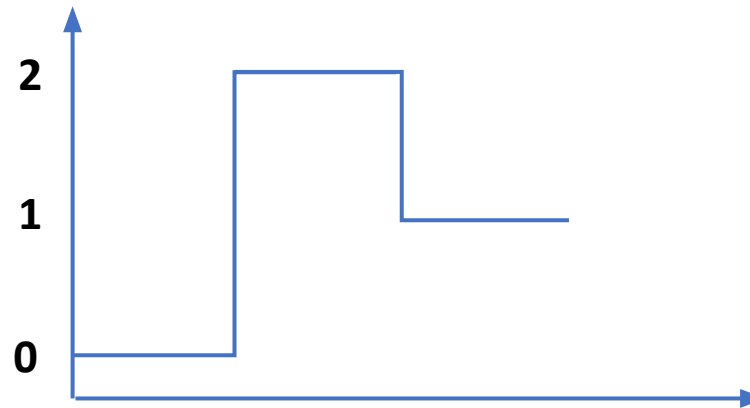
Добавляе С
М



5' -GAGTTTTTCAGATCTCGGTGGC-3'

|||||

3' -CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGC -5'

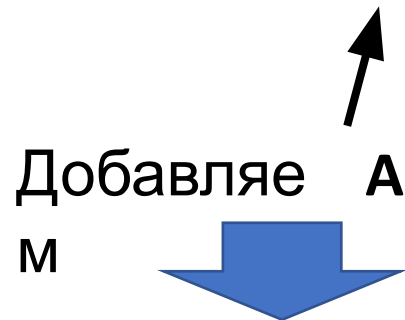


Варианты подходов в NGS (второе поколение): основная проблема пиросеквенирования и Ion Torrent (ThermoFisher)

5' -GAGTTTTTCAGATCTCGGTGGC-3' -OH

|||||

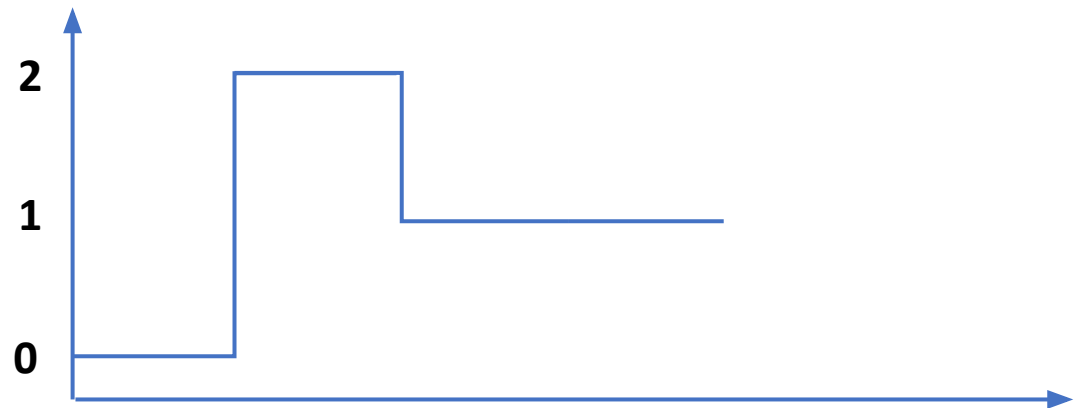
3' -CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGC -5'



5' -GAGTTTTTCAGATCTCGGTGGC**CA**-3'

|||||

3' -CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGTTCGGGC -5'



Варианты подходов в NGS (второе поколение)



Варианты подходов в NGS (третье поколение)

- Не требует фрагментации исследуемой молекулы, то есть выделенная ДНК (или даже РНК!) почти сразу загружается в прибор
- Ну очень длинные прочтения (до 1 миллиона п.о.)
- Проходит в режиме реального времени (нет остановки на добавление новых нуклеотидов)
- Множество ошибок в определенных нуклеотидах (точность 85–95 %)

Варианты подходов в NGS (третье поколение)

Oxford
Nanopore
MinION



Прибор – 1000 \$
Запуск 500–900 \$
Основан на поре с
геликазой



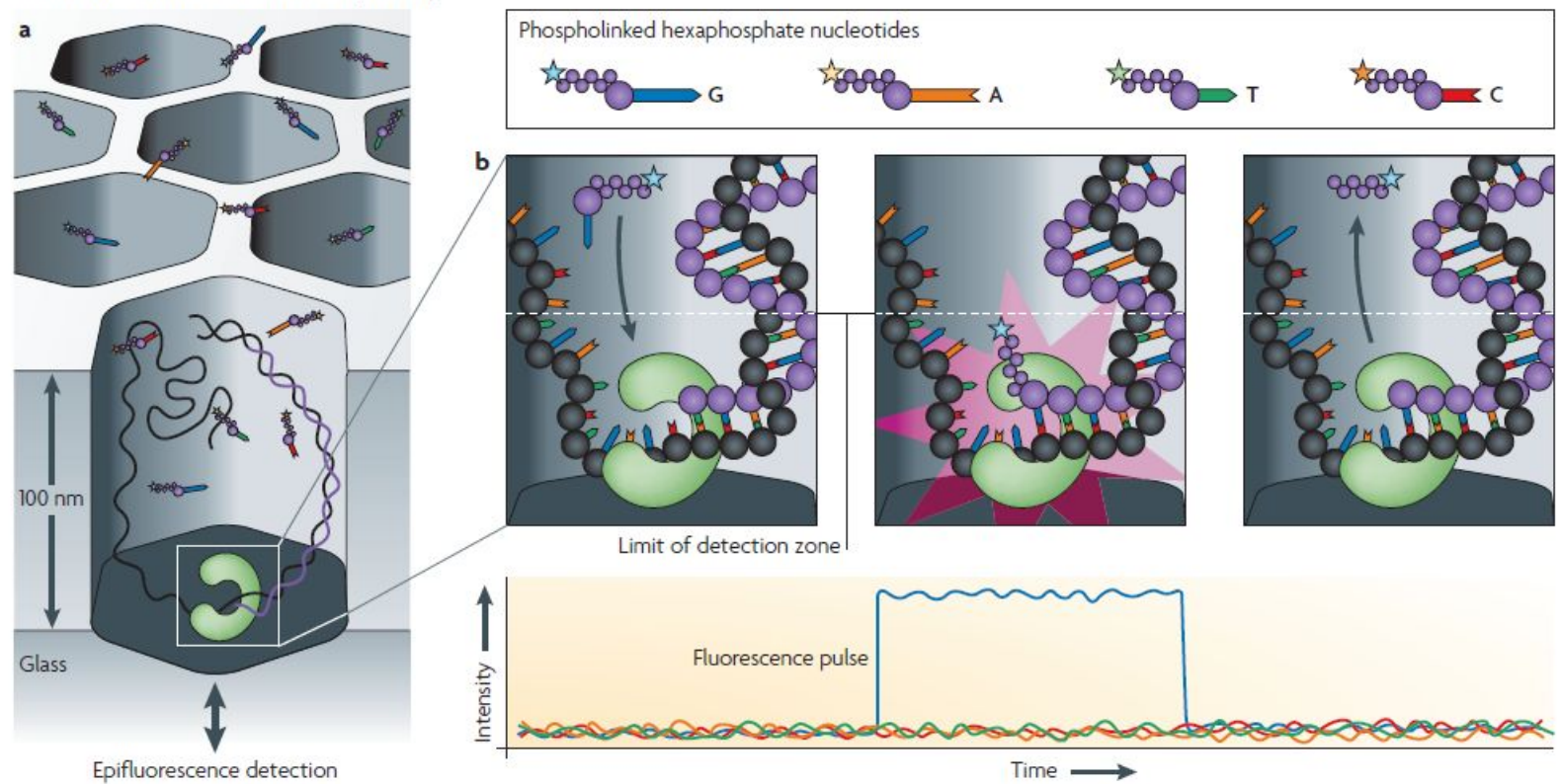
Pacific
Biosciences
Sequel

Прибор – 350'000 \$
Запуск – 850 \$
Основан на ДНК-полимеразе
и конфокальной микроскопии

Варианты подходов в NGS (третье поколение): PacBio

- В каждой лунке закреплена молекула ДНК-полимеразы
- Снизу в лунку поступает свет
- Считывание сигнала происходит только около полимеразы
- К γ -фосфату каждого нуклеотида пришит свой флуорофор
- Время присоединения нуклеотида – мс, время диффузии остальных

Pacific Biosciences — Real-time sequencing



Варианты подходов в NGS (третье поколение): Oxford Nanopore

- На мембране закреплено множество пор, с которыми связана геликаза, разворачивающая ДНК
- Из-за разного размера азотистых оснований разные нуклеотиды по-разному «затыкают» пору, не давая проходить через нее ионам, что меняет силу тока ионов
- Через пору может протягиваться не только ДНК, но и РНК, что позволяет использовать ее напрямую, без этапа обратной транскрипции
- Могут быть различены модифицированные и

