Assignment 0

During the last weeks the Coronavirus spread worldwide. To control the pandemic it is of importance to be able to screen populations for putative infections. This requires the development of virus detection kits. This is a typical bioanalytical problem.

A. How would you develop such a test? Consider how to "measure" the virus – which analytical method would you suggest? Which specification should be fulfilled?

The questions are not intended to examine your knowledge – there are no wrong answers. You are not expected to search literature of other sources for solution. Come up with your own ideas, take the role of a scientist that approaches a novel problem.

Assignment 1

During the past weeks many laboratories worked on the development on test kits for the detection of the Coronavirus. The World Health Organisation (WHO) aims to review this activities to ensure rapid international distribution of the obtained information. The WHO provides the following overview over test kits developed by national laboratories. The WHO points out, that these tests are not validated by the WHO and may not correspond to the standards of the WHO.

World Health Organization Health Topics Countries Newsroom Emergencies About Us

Coronavirus disease 2019
Technical guidance
National laboratories
Early investigations protocols
Case management
Surveillance, rapid response teams, and case investigation
Infection prevention and control
Points of entry and mass gatherings
Naming the coronavirus disease (COVID-19) and the virus that causes it

Summary table of available protocols

Country	Institute	Gene targets
China	China CDC	ORF1ab and N
Germany	Charité	RdRP, E, N
Hong Kong SAR	HKU	ORF1b-nsp14, N
Japan	National Institute of Infectious Diseases, Department of Virology III	Pancorona and multiple targets, Spike protein
Thailand	National Institute of Health	N
US	US CDC	Three targets in N gene
France	Institut Pasteur, Paris	Two targets in RdRP

China CDC Primers and probes for detection 2019-nCoV (24 January 2020)

Diagnostic detection of Wuhan coronavirus 2019 by real-time RT-PCR – Charité, Berlin Germany (17 January 2020)

Detection of 2019 novel coronavirus (2019-nCoV) in suspected human cases by RT-PCR – Hong Kong University (23 January 2020)

PCR and sequencing protocol for 2019-nCoV - Department of Medical Sciences, Ministry of Public Health, Thailand (Updated 28 January 2020)

PCR and sequencing protocols for 2019-nCoV- National Institute of Infectious Diseases Japan (24 January 2020)

US CDC Real-Time RT-PCR Panel for Detection 2019-Novel Coronavirus (28 January 2020)

US CDC panel primer and probes– U.S. CDC, USA (28 January 2020)

Real-time RT-PCR assays for the detection of SARS-CoV-2 Institut Pasteur, Paris (2 March 2020)

Disclaimer: *The order on the list is by country of the hosting institution and does not imply any preference of WHO. Neither the names of vendors or manufacturers included in the protocols are preferred/endorsed by WHO. The protocols have not yet been validated through a WHO process.*

Assumptions: *Most procedures assume a basic familiarity with PCR/RT-PCR assays.*

Safety Information: *Specimen processing should be performed in accordance with pertaining national biological safety regulations and following the recommended WHO guidelines on biosafety and biosecurity.*

Protocol use limitations: *Optional clinical specimens for testing has not yet been validated.*

Scientist at The Charité (Germany) were one of the earliest to release the description of a test kit (see appendix).

As common for a scientific publication information in presented in a very compressed manner and the authors assume a certain level of background knowledge. Reading the document for the first time you may not understand everything directly. Do not become discourage! Try to identify familiar terms and techniques to get an overview. Some seemingly complex stuff is in reality rather simple. In fact, the described method can be executed in any laboratory equipped for work in the field of molecular biology – after a short one day instruction also by you.

A. Which method is used by The Charité to detect the Coronavirus? Describe the method. You may provide also schematic drawings.

B. Welches Verfahren setzt das Charité ein, um den Coroanvirus nachzuweisen? Try to become familiar with the document. With extra information is provided by The Charité? Why is this information provided?

C. What do you not understand? Which information are you lacking? Try to formulate concrete questions.

Berlin, Jan 17th, 2020

Diagnostic detection of 2019-nCoV by real-time RT-PCR

-Protocol and preliminary evaluation as of Jan 17, 2020-

Victor Corman, Tobias Bleicker, Sebastian Brünink, Christian Drosten
Charité Virology, Berlin, Germany

Olfert Landt, Tib-Molbiol, Berlin, Germany

Marion Koopmans
Erasmus MC, Rotterdam, The Netherlands

Maria Zambon
Public Health England, London

Additional advice by Malik Peiris, University of Hong Kong

Users looking for a workflow protocol consult the last three pages of this document

Contact: christian.drosten@charite.de
https://virologie-ccm.charite.de/en/

Positive control material is available from Charité, Berlin, via EVAg
(https://www.european-virus-archive.com/).

This is document Version 2.
Changes against Version 1 (Jan 13, 2019): Workflow protocols included, N gene assay removed, data for single probe versions of RdRp assay added; information on availability of controls updated.

We acknowledge the originators of sequences in GISAID (www.gisaid.org): National Institute for Viral Disease Control and Prevention, China, Institute of Pathogen Biology, Chinese Academy of Medical Sciences, Peking Union Medical College, China, and Wuhan Jinyintan Hospital Wuhan Institute of Virology, Chinese Academy of Sciences, China). We acknowledge Professor Yong-Zhen Zhang, Shanghai Public Health Clinical Center & School of Public Health, Fudan University, Shanghai, China for release of another sequence (MN908947).

We use the term “SARS-related Coronavirus” to include the SARS virus as well as the clade of betacoronaviruses known to be associated with (mainly) rhinolophid bats across the Palearctic. The latest taxonomy classifies these viruses in a subgenus termed *Sarbecovirus*.

Berlin, Jan 17th, 2020

Background

We used known SARS- and SARS-related coronaviruses (bat viruses from our own studies as well as literature sources) to generate a non-redundant alignment (excerpts shown in Annex). We designed candidate diagnostic RT-PCR assays before release of the first sequence of 2019-nCoV. Upon sequence release, the following assays were selected based on their matching to 2019-nCoV as per inspection of the sequence alignment and initial evaluation (Figures 1 and 2).

All assays can use SARS-CoV genomic RNA as positive control. Synthetic control RNA for 2019-nCoV E gene assay is available via EVAg. Synthetic control for 2019-nCoV RdRp is expected to be available via EVAg from Jan 21st onward.

First line screening assay: E gene assay
Confirmatory assay: RdRp gene assay

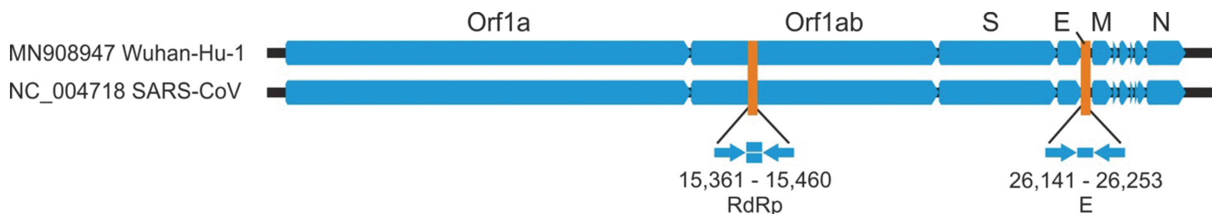


Figure 1 relative positions of amplicon targets on SARS-CoV an 2019-nCoV genome. ORF: open reading frame; RdRp: RNA-dependent RNA polymerase. Numbers below amplicon are genome positions according to SARS-CoV, NC_004718.

Materials and assay formulation

Clinical samples and CoV cell culture supernatants

Respiratory samples were obtained during 2019 from patients hospitalized at Charité medical center and tested by the NxTAG® Respiratory Pathogen Panel (Luminex) or in cases of MERS-CoV by the MERS-CoV upE assay as published before (1).

Cell culture supernatants from typed coronaviruses were available at our research and clinical laboratories. The typed avian influenza virus RNA (H5N1) was obtained from the German Society for Promotion of Quality Assurance in Medical Laboratories (INSTAND) proficiency testing panels. RNA was extracted from clinical samples by using the MagNA Pure 96 system (Roche) and from cell culture supernatants by the viral RNA mini kit (Qiagen).

Assay design

For oligonucleotide design and in-silico evaluation we downloaded all complete and partial (if >400 nucleotides) SARS-related virus sequences available at GenBank by January 1st, 2020. The list (n=729 entries) was manually checked and artificial sequences (lab-derived, synthetic etc.), as well as sequence duplicates removed, resulting in a final list of 375

sequences. These sequences were aligned and the alignment used for assay design. The alignment was later complemented by sequences released from the Wuhan cluster. All presently release sequences match the amplicons (Figure 2). An overview of oligonucleotide binding sites in all unique sequences of bat-associated SARS-related viruses is shown in the appendix.

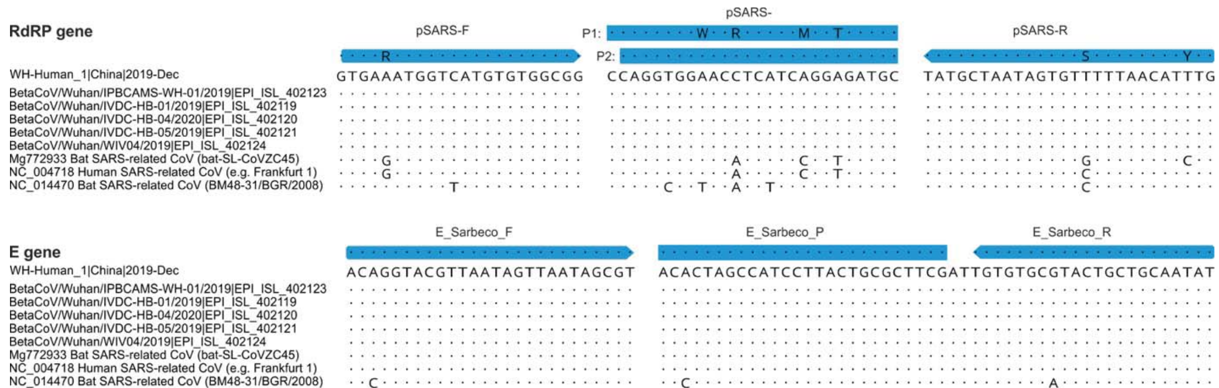


Figure 2 Partial alignments of oligonucleotide binding regions. Panels show six available sequences of 2019-nCoV, aligned to the corresponding partial sequences of SARS-CoV strain Frankfurt 1, which can be used as a positive control for all three RT-PCR assays. The alignment also contains the most closely-related bat virus (Bat SARS-related CoV isolate bat-SL-CoVZC45, GenBank Acc.No. MG772933.1) as well as the most distant member within the SARS-related bat CoV clade, detected in Bulgaria (GenBank Acc. No. NC_014470). Dots represent identical nucleotides compared to sequence Wuhan-Hu 1. Substitutions are specified. More comprehensive alignments in the Appendix.

Real-time reverse-transcription polymerase chain reaction

All assays used the same conditions. Primer and probe sequences, as well as optimized concentrations are shown in Table 1. A 25-μl reaction was set up containing 5 μl of RNA, 12.5 μl of 2 X reaction buffer provided with the Superscript III one step RT-PCR system with Platinum Taq Polymerase (Invitrogen; containing 0.4 mM of each deoxyribonucleotide triphosphates (dNTP) and 3.2 mM magnesium sulfate), 1 μl of reverse transcriptase/Taq mixture from the kit, 0.4 μl of a 50 mM magnesium sulfate solution (Invitrogen – not provided with the kit), and 1 μg of nonacetylated bovine serum albumin (Roche). All oligonucleotides were synthesised and provided by Tib-Molbiol, Berlin. Thermal cycling was performed at 55°C for 10 min for reverse transcription, followed by 95°C for 3 min and then 45 cycles of 95°C for 15 s, 58°C for 30 s.

Table 1. Primers and probes

Optimized concentrations are mol per liter of final reaction mix.

(e.g., 1.5 microliters of a 10 micromolar (uM) primer stock solution per 25 microliter (ul) total reaction volume yields a final concentration of 600 nanomol per liter (nM) as indicated in the table)
-note that standard, non-optimized reaction conditions as indicated by suppliers of one-step RT-PCR kits will generally yield sufficient sensitivity-

Assay/ Use	Oligonucleotide ID	Sequence (5'–3')	Comment
RdRP gene	RdRP_SARSr-F2	GTGARATGGTCATGTGTGGCGG	use 600 nM per reaction
	RdRP_SARSr-R1	CARATGTTAAASACACTATTAGCATA	use 800 nM per reaction
	RdRP_SARSr-P2	FAM-CAGGTGGAACCTCATCAGGAGATGC-BBQ	Specific for 2019-nCoV, will not detect SARS-CoV use 100 nM per reaction and mix with P1
	RdRP_SARSr-P1	FAM-CCAGGTGGWACRTCATCMGGTGATGC-BBQ	Pan Sarbeco-Probe, will detect 2019-nCoV, SARS-CoV and bat-SARS-related CoVs use 100 nM per reaction and mix with P2
E gene	E_Sarbeco_F1	ACAGGTACGTTAATAGTTAATAGCGT	use 400 nM per reaction
	E_Sarbeco_R2	ATATTGCAGCAGTACGCACACA	use 400 nM per reaction
	E_Sarbeco_P1	FAM-ACACTAGCCATCCTTACTGCGCTTCG-BBQ	use 200 nM per reaction

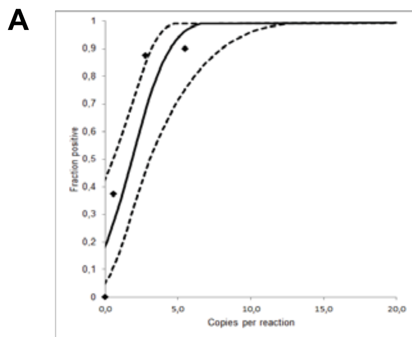
W is A/T; R is G/A; M is A/C ; FAM, 6-carboxyfluorescein; BBQ, blackberry quencher

Technical sensitivity testing

Preliminary assessment of analytical sensitivity for RdRp assay.

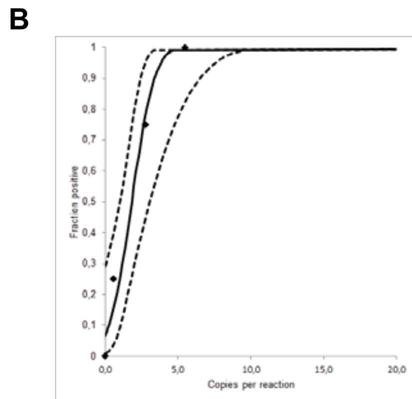
We tested purified cell culture supernatant containing SARS-CoV strain Frankfurt-1 virions grown on Vero cells, and quantified by real-time RT-PCR assay as described in Drosten et al. (2) using a specific *in-vitro* transcribed RNA quantification standard. The results are shown in Figure 3. All assays are highly sensitive.

A



A. First line assay: E gene
Technical limit of detection (LOD) = 5.2 RNA copies/reaction, at 95% hit rate; 95% CI: 3.7-9.6 RNA copies/reaction.

B



B. Confirmatory assay: RdRP gene
Technical LOD = 3.8 RNA copies/reaction, at 95% hit rate; 95% CI: 2.7-7.6 RNA copies/reaction.

Figure 3. A, E-gene assay, B, RdRp gene assay. X-axis shows input RNA copies per reaction. Y-axis shows positive results in all parallel reactions performed, squares are experimental data points resulting from replicate testing of given concentrations (x-axis) in parallels assays (8 replicate reactions per datum point). The inner line is a probit curve (dose-response rule). The outer dotted lines are 95% confidence intervals.

RdRp assay sensitivity with single probe application using the assay variant that only contains the 2019-nCoV specific probe.

SARS 2019-nCoV

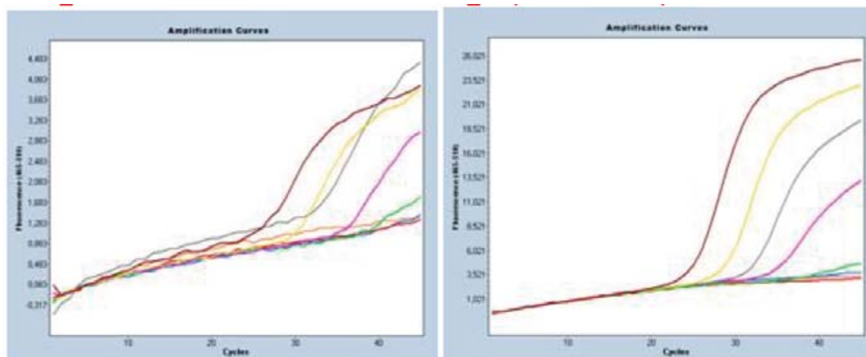


Figure 4. Preliminary experiment comparing single probe assay for SARS-CoV (probe RdRP_SARSr-P1, left panel) with single probe assay for 2019-nCoV (probe RdRP_SARSr-P2, right panel). Note that the fluorescent signal in these assays is suboptimal due to the use of PCR-generated targets.

Berlin, Jan 17th, 2020

Breadth of detection

To show that the assays will detect other bat-associated SARS-related viruses, we tested bat-derived fecal samples available from Drexler et al., (3) und Muth et al., (4) using the novel assays.

KC633203, Betacoronavirus BtCoV/Rhi_eur/BB98-98/BGR/2008
KC633204, Betacoronavirus BtCoV/Rhi_eur/BB98-92/BGR/2008
KC633201, Betacoronavirus BtCoV/Rhi_bla/BB98-22/BGR/2008
GU190221 Betacoronavirus Bat coronavirus BR98-19/BGR/2008
GU190222 Betacoronavirus Bat coronavirus BM98-01/BGR/2008
GU190223, Betacoronavirus Bat coronavirus BM98-13/BGR/2008

All samples were successfully tested positive by the E gene assay. Detection of these relatively distant members of the SARS-related CoV clade suggests that all Asian viruses are likely to be detected.

Specificity testing

1. Chemical stability
To exclude non-specific reactivity of oligonucleotides among each other, both assays were tested 40 times in parallel with water and no other nucleic acid except the provided oligonucleotides. In none of these reactions was any positive signal detected.

2. Cross-reactivity with other coronaviruses
Cell culture supernatants containing human coronaviruses (HCoV)-229E, -NL63, -OC43, and -HKU1 as well as MERS-CoV were tested in all three assays (Table 2). For the non-cultivable HCoV-HKU1, supernatant from human airway culture was used. Virus RNA concentration in all samples was determined by specific real-time RT-PCRs and in-vitro transcribed RNA standards designed for absolute viral load quantification.

Table 2. Cell-culture supernatants tested by all assays

Cell culture supernatants	Tested concentration	Result
Alphacoronaviruses		
Human coronavirus NL63	4x10^9 RNA copies/ml	No reactivity with any of three assays
Human coronavirus 229E	3x10^9 RNA copies/ml	No reactivity with any of three assays
Betacoronaviruses		
Betacoronavirus 1 (strain HCoV-OC43)	1x10^10 RNA copies/ml	No reactivity with any of three assays
Human coronavirus HKU1 (HCOV-HKU1)	1x10^5 RNA copies /ml	No reactivity with any of three assays
Middle East respiratory syndrome-related coronavirus (strain EMC/2012)	1x10^8 RNA copies/ml	No reactivity with any of three assays

3. Tests of human clinical samples previously tested to contain respiratory viruses

Both assays were applied on human clinical samples from our own diagnostic services, previously tested positive for the viruses listed in Table 3. All tests returned negative results.

Table 3. Tests of known respiratory viruses and bacteria in clinical samples

Clinical samples with known viruses	Number of samples tested
HCoV-HKU1	2
HCoV-OC43	5
HCoV-NL63	5
HCoV-229E	5
MERS-CoV	5
Influenza A (H1N1/09)	6
Influenza A (H3N2)	5
Influenza A(H5N1)	1
Influenza B	3
Rhinovirus/Enterovirus	3
Respiratory syncytial virus (A/B)	6
Parainfluenza 1 virus	3
Parainfluenza 2 virus	3
Parainfluenza 3 virus	3
Parainfluenza A or -B virus	5
Human metapneumovirus	3
Adenovirus	3
Human Bocavirus	3
Legionella spp.	3
Mycoplasma spp.	3
Total clinical samples	75

Berlin, Jan 17th, 2020

References

1. Corman VM, Eckerle I, Bleicker T, Zaki A, Landt O, Eschbach-Bludau M, et al. Detection of a novel human coronavirus by real-time reverse-transcription polymerase chain reaction. Euro Surveill. 2012;17(39).
2. Drosten C, Gunther S, Preiser W, van der Werf S, Brodt HR, Becker S, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med. 2003;348(20):1967-76.
3. Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, Goettsche M, et al. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. J Virol. 2010;84(21):11336-49.
4. Muth D, Corman VM, Roth H, Binger T, Dijkman R, Gottula LT, et al. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. Sci Rep. 2018;8(1):15177.

Berlin, Jan 17th, 2020

Annex:

Annex figure. Non-redundant alignments of SARS-related CoVs focused on oligonucleotide binding sites of all assays (top to bottom: RdRp, E, N). Viruses not present in these alignments have been removed because their binding sites are 100% identical to one of the members of the alignment. ("--") means sequence gaps not covered by oligonucleotides. Note that these alignments contain only one sequence of 2019-nCoV while Figure 2 above contains all presently released sequences. We will fuse this into one figure.

Berlin, Jan 17th, 2020

Workflow Protocol

1. First line screening assay

E assay:

MasterMix:	**Per reaction**	
H_2O (RNAse free)	2.6 µl	
2x Reaction mix*	12.5 µl	
$MgSO_4$(50mM)	0.4 µl	
BSA (1 mg/ml)**	1 µl	
Primer E_Sarbeco_F1 (10 µM stock solution)	1 µl	ACAGGTACGTTAATAGTTAATAGCGT
Primer E_Sarbeco_R2 (10 µM stock solution)	1 µl	ATATTGCAGCAGTACGCACACA
Probe E_Sarbeco_P1 (10 µM stock solution)	0.5 µl	FAM-ACACTAGCCATCCTTACTGCGCTTCG-BBQ
SSIII/Taq EnzymeMix*	1 µl	
Total reaction mix	20 µl	
Template RNA, add	5 µl	
Total volume	25 µl	

* Thermo Fischer/Invitrogen: SuperScriptIII OneStep RT-PCR System with Platinum® Taq DNA Polymerase
** MgSO4 (50 mM) [Sigma], This component is not provided with the OneStep RT-PCR kit
*** non-acetylated [Roche].

Cycler:

55°C	10’	
94°C	3’	
94°C	15’’	45x
58°C	30’’	

If assay No 1 is positive, continue to assay No 2.

Berlin, Jan 17th, 2020

2. Confirmatory assay

RdRp assay:

MasterMix:	**Per reaction**	
H_2O (RNAse free)	0.6 µl	
2x Reaction mix*	12.5 µl	
$MgSO_4$(50mM)	0.4 µl	
BSA (1 mg/ml)**	1 µl	
Primer RdRP_SARSr-F2 (10 µM stock solution)	1.5 µl	GTGARATGGTCATGTGTGGCGG
Primer RdRP_SARSr-R1 (10 µM stock solution)	2 µl	CARATGTTAAASACACTATTAGCATA
Probe RdRP_SARSr-P1 (10 µM stock solution)	0.5 µl	FAM-CCAGGTGGWACRTCATCMGGTGATGC-BBQ
Probe RdRP_SARSr-P2 (10 µM stock solution)	0.5 µl	FAM-CAGGTGGAACCTCATCAGGAGATGC-BBQ
SSIII/Taq EnzymeMix*	1 µl	
Total reaction mix	20 µl	
Template RNA, add	5 µl	
Total volume	25 µl	

* Thermo Fischer/Invitrogen: SuperScriptIII OneStep RT-PCR System with Platinum® Taq DNA Polymerase
** MgSO4 (50 mM) [Sigma], This component is not provided with the OneStep RT-PCR kit
*** non-acetylated [Roche].

Cycler:

55°C 10’
94°C 3’
94°C 15’’
58°C 30’’ | 45x

If assay No 2 is positive, continue to assay No 3.

3. Discrimatory assay

RdRp assay:

MasterMix:	**Per reaction**	
H_2O (RNAse free)	1.1 µl	
2x Reaction mix*	12.5 µl	
$MgSO_4$(50mM)	0.4 µl	
BSA (1 mg/ml)**	1 µl	
Primer RdRP_SARSr-F2 (10 µM stock solution)	1.5 µl	GTGARATGGTCATGTGTGGCGG
Primer RdRP_SARSr-R1 (10 µM stock solution)	2 µl	CARATGTTAAASACACTATTAGCATA
Probe RdRP_SARSr-P2 (10 µM stock solution)	0.5 µl	FAM-CAGGTGGAACCTCATCAGGAGATGC-BBQ
SSIII/Taq EnzymeMix*	1 µl	
Total reaction mix	20 µl	
Template RNA, add	5 µl	
Total volume	25 µl	

* Thermo Fischer/Invitrogen: SuperScriptIII OneStep RT-PCR System with Platinum® Taq DNA Polymerase
** MgSO4 (50 mM) [Sigma], This component is not provided with the OneStep RT-PCR kit
*** non-acetylated [Roche].

Cycler:

55°C	10’	
94°C	3’	
94°C	15’’	
58°C	30’’	45x

Assay No 3 is specific for 2019-nCoV

Note: Other generic real-time RT-PCR reagents can be used for all assays. In this case, use oligonucleotides at concentrations indicated. If using Light Cycler instrument with glass capillaries, use Light Cycler-specific reagents or add BSA as indicated in the detailed documentation above.

DNA sequencing

In this assignment you will develop a technique to determine DNA sequences. Please follow this assignment step by step.

If you are not sure of how to progress, please feel free to formulate the problem you are facing. For example: "I would like X to happen as this is the requirement to perform Y. I am not sure how to make X happen but if this would be possible I can continue with …). It is often more important to recognize a problem than knowing a solution!

A. DNA-polymerase synthesis the new DNA strands during replication. Should you not be familiar with this, briefly recapitulate (see Fig. 1). Describe the mechanism of DNA polymerase. Now you would like to use DNA polymerase in a test tube to copy DNA. Which components do you have to add to the test tube? Assume that you wish to copy only a short piece of DNA of a few hundred nucleotides (bases).

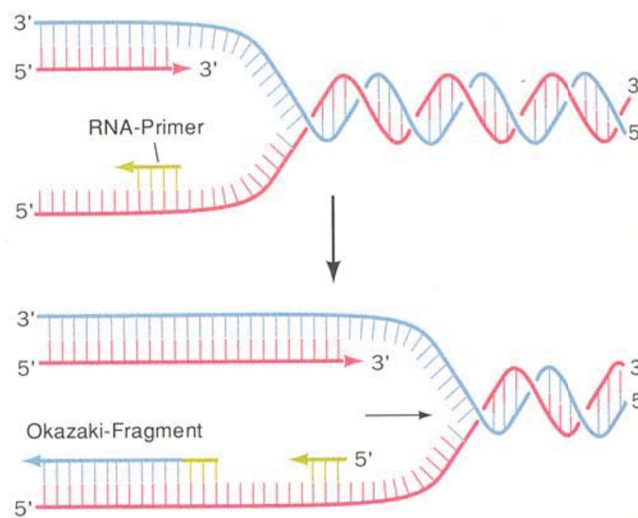


Fig. 1. Replication of DNA by DNA-polymerase. Figure from Voet/Voet, Biochemie

B. Analyse the nucleotides in Fig. 2A and the single DNA strand in Fig. 2B. What would happen if you add ddATP but no dATP to your test tube?
What would happen, if you add a mixture of dATP and ddATP to your test tube (reaction A). What effect would different ratios of dATP and ddATP. If you like, treat this problem by statistical means. Remember: Your test tube does not contain only one DNA fragments but many identical DNA fragments.

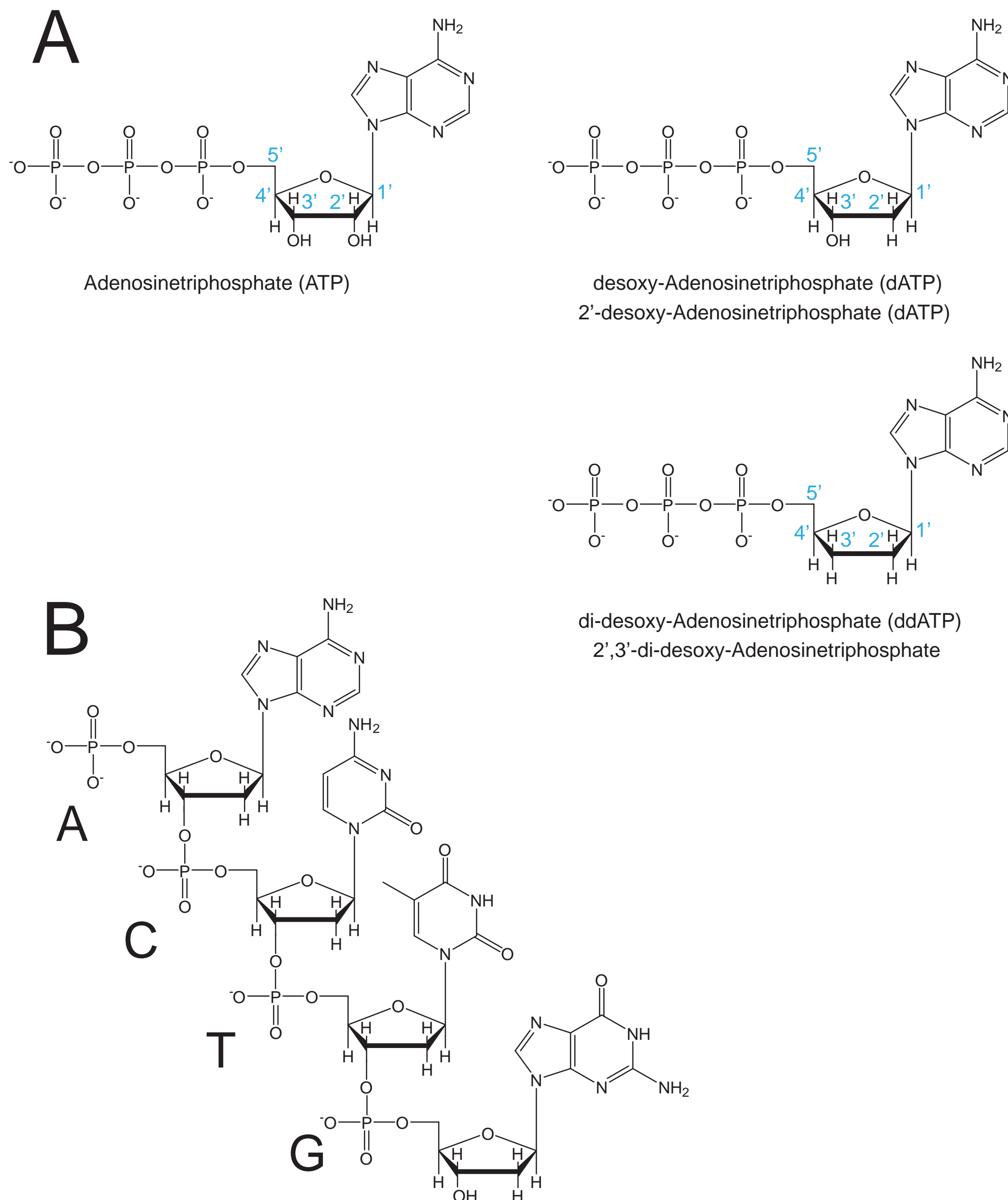


Fig. 2. Nucleotides. A. Show are ATP and two related nucleotides. Which nucleotide is used to build DNA? Analog variants exist also of GTP, CTP and TTP. **B.** A short single stranded DNA.

C. Fig. 3 introduces the concept of gel electrophorese. We will discuss details of this technique in another lecture, but these are not relevant for now. This method allows the separation of molecules based on size.
What would happen if the outcome from reaction A is subject to electrophorese (assume that you are able to see the DNA)?

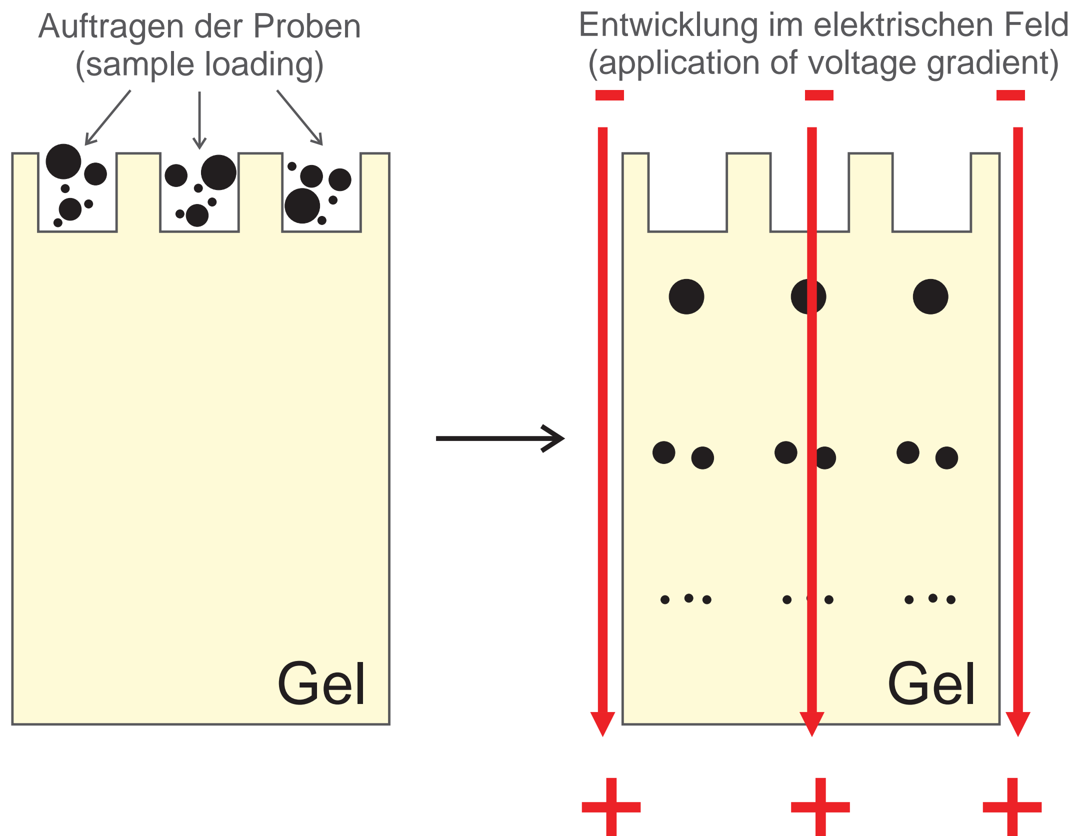


Fig. 3. Gel electrophorese. This techniques allows the separation of molecules based on the size in a voltage gradient.

D. Now we modify reaction A by using radioactive ddATP. The reaction will be subjected to electrophorese. Afterwards the gel will be exposed to an x-ray film. The x-ray film will visualise radioactive spots on the gel. What does such a spot mean? Which conclusions can you derive?

E. Based on these experiments, try to develop a method to determine the sequence of DNA.
Start with describing the concept you are proposing.
Now try to apply this concept to an example. You wish to determine the sequence of the following DNA:
TGCCATCGCTAAATGCTAGCCTAATTTGCTAGGTCATTAG
In agreement with general convention the sequence is given from the 5' to the 3' end.
Which chemical would you need? Prepare a drawing of the expected outcome – how would the x-ray film look like? Explain how you can derive the sequence of the DNA.

F. After you have developed a sequence technique you now wish to increase the throughput of the technique. Also you would like to prevent the use of radioactivity. It is possible to couple nucleotide such as ddATP with dyes. Furthermore, it is possible to prepare gels in thin capillaries (Fig. 4). Could you make use of this techniques?

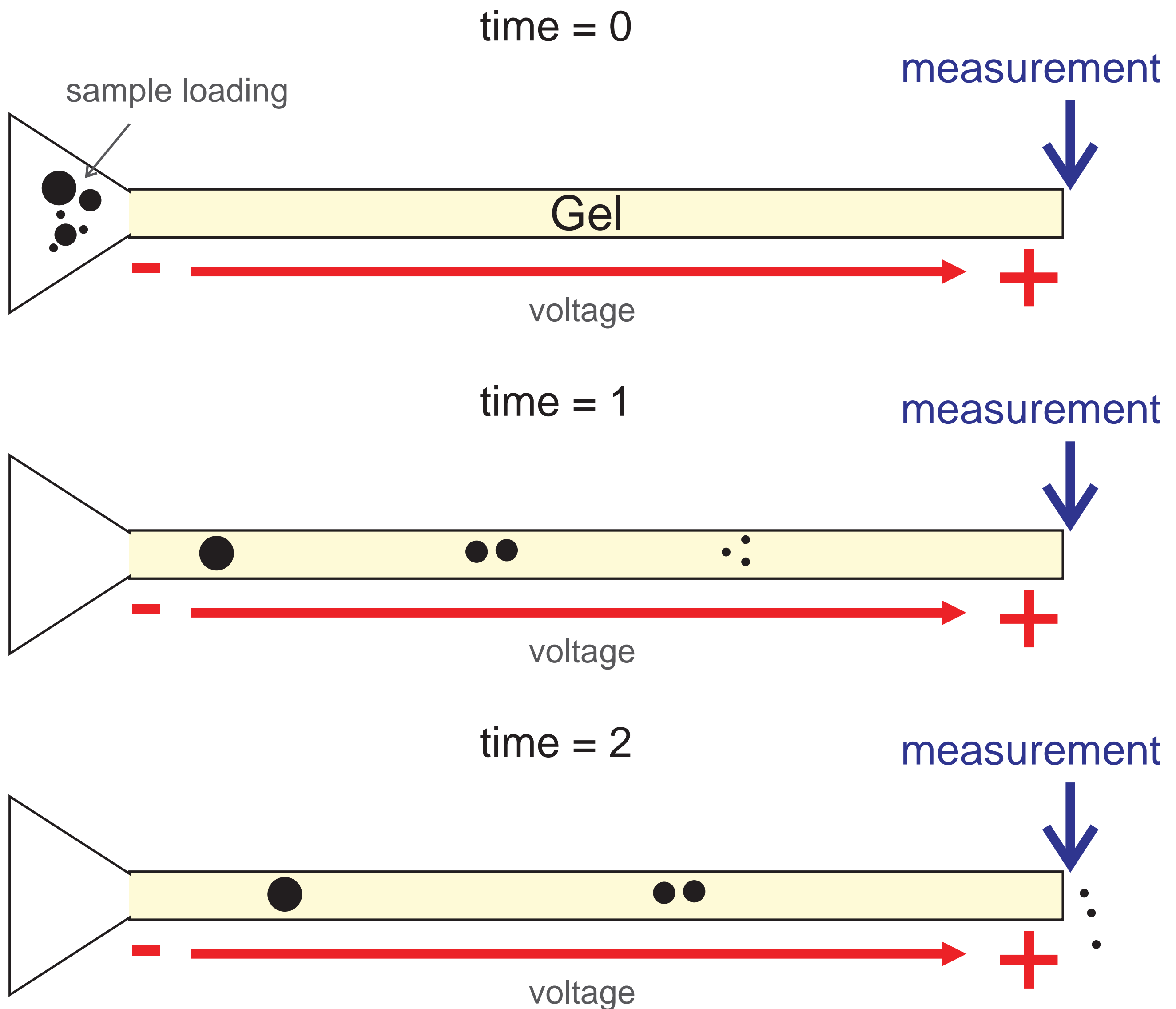


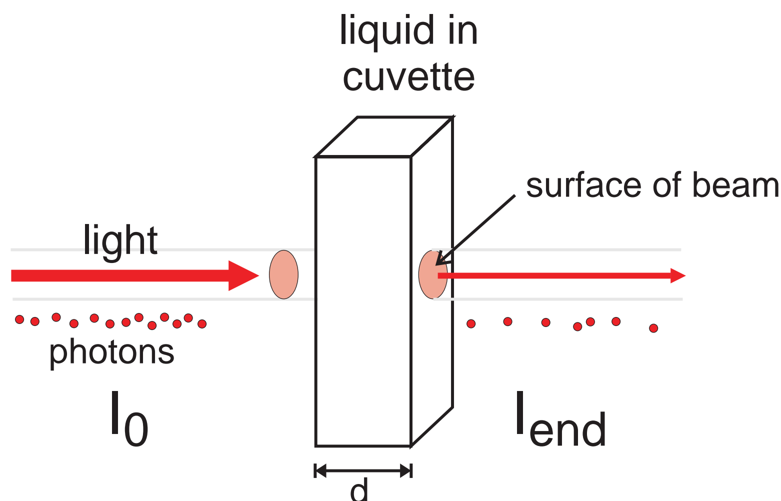
Fig. 4. Capillary Electrophoreses. The gel is prepared in thin capillaries. The sample is loaded to one end of the capillary and driven through the capillary by the application of the voltage gradient. Detection occurs at the other end of the capillary.

Determiation of concentration of proteins and nucleotide acids

In assignment 1 and 2 we frequently assumed that we would be able to "see" DNA or proteins. It is, however, the task of an analytical chemist to "make molecules visible", such that they can be measured.

Many molecules absorbed light (you should be familiar with this principles, which are repeated here only briefly). The ability to absorbed light depends on the structure of the molecule as well as on the wavelength λ of the light. Thus the molecule can absorbed only photons of specific energy. This is what you experience in daily live and the reason why the world appears colourful to us. A surface appears as blue to us, if it absorbs the green and the red part of the light but reflects the blue. Likewise, the colour of a liquid is determined based on the parts of the light that are absorbed when passing through the liquid. If the ability of a liquid to absolved light of different wavelengths is systematically tested, a spectrum of absorption is obtained.

The experimental set-up is as follows: A liquid is placed into a cuvette and light of a given wavelength is passed through. Before entering the cuvette the intensity of the light be I_0. The intensity of the light leaving the cuvette be I_{end}. I_{end} is smaller than I_o if the liquid or devolved molecules have absorbed photons.

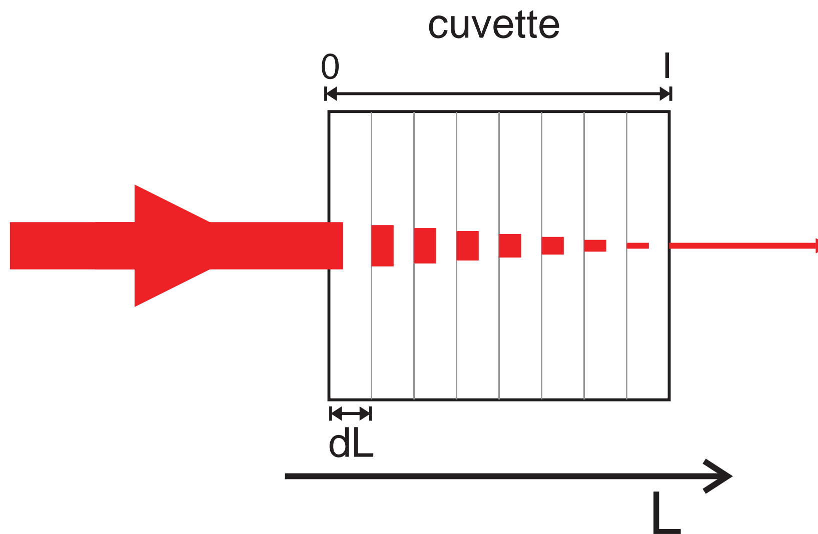


We aim to determine the decay of the intensity while the light passed through the cuvette assuming that a substance of concentration c is dissolved in the liquid and that the liquid itself is not absorbing. We realise: As higher the concentration as higher the probability that a photon hits a molecule and get absorbed. As longer the distances the light travels through the cuvette, as more photons get absorbed. Therefore:

$$\Delta I = I_{start} \cdot \beta \cdot c \cdot d \qquad (1)$$

ΔI is the loss of intensity if light of intensity I_{start} enters a solution of concentration c and travels distance d. β is a constant of proportionality.

This suggests that the intensity decays linearly with the distance. This is, however, true only if the distance of travel is infinitesimal short. We therefore divide the distance of travel in short intervals (stages) with the length of dL:



We now understand why (1) is only valid for infinitesimal short stage: After the light has passed the first stage dL, parts of the photons are absorbed and thus the insanity is slightly reduced. Thus, the intensity at start for the following stage dL is not I_{start} any longer. Thus we have to use the local intensity as I_{start} for the following stage dL always.

(1) should therefore be written as:

$$dI = -I \cdot \beta \cdot c \cdot dL \tag{2}$$

dI is the local change of intensity in the next stage
I is the local intensity that enters the next stage
dL is the infinitesimal short length of a stage
The algebraic sign was chosen to be negative as the intensity decays.

(2) is a diffential equation that can be solved by integration after separation of the variables.

$$\frac{1}{I}\, dI = -\beta \cdot c \cdot dL \tag{3}$$

$$\int \frac{1}{I}\, dI = \int -\beta \cdot c \cdot dL \tag{4}$$

$$\ln(I) + const1 = -\beta \cdot c \cdot L + const2 \tag{5}$$

*const*1 and *const*2 are the constants of integration. We define:
const = *const*2 – *const*1 and obtain:

$$\ln(I) = -\beta \cdot c \cdot L + const \tag{6}$$

Const needs to be determined form the boundary conditions.

Boundary condition 1: I_0 is the intensity that enters the cuvette (Fig. 1).

Boundary condition 2: We arbitrarily set the position at which the light enters the cuvette to be the start of our axis. Therefore $L = 0$ at the outside of the cuvette.

With $I = I_0$, $L = 0$, and (6) we obtain:

$$\ln(I_0) = const \qquad (7)$$

(7) and (6) can now be combined and solved for *I*:

$$\ln(I) = -\ \beta \cdot c\ \cdot L + \ln(I_0) \qquad (8)$$

$$\ln\left(\frac{I}{I_0}\right) = -\ \beta \cdot c\ \cdot L \qquad (9)$$

$$\boldsymbol{I = I_0\, e^{-\beta \cdot c \cdot L}} \qquad \textbf{(10)}$$

(10) allows to determine the intensity of the light beam at any point L in the cuvette.

Of interest is the intensity I_{end} that leaves the cuvette (Fig. 1). As the cuvette has the length d (Figs. 1,2) we have to set L=d in (10):

$$I_{end} = I_0\, e^{-\beta \cdot c \cdot d} \qquad (11)$$

This allows determining the fraction of photons that was not absorbed:

$$\frac{I_{end}}{I_0} = \frac{I_0 e^{-\beta \cdot c \cdot d}}{I_0} = e^{-\beta \cdot c \cdot d} \qquad (12)$$

We now define the **extinction E**:

$$E \equiv -\lg\left(\frac{I_{end}}{I_0}\right) = -\lg\left(e^{-\beta \cdot c \cdot d}\right) = \beta \cdot c\ \cdot\ d\ \cdot \lg(e) = \beta \cdot c\ \cdot\ d\ \cdot 0.434 = \varepsilon\ \cdot c\ \cdot\ d \qquad (13)$$

(13) is also referred to as the law of Lambert and Beer:

$$\boldsymbol{E \equiv -\lg\left(\frac{I_{end}}{I_0}\right) = \varepsilon\ \cdot c\ \cdot\ d} \qquad \textbf{(14)}$$

By combining some constants we have in addition introduced the coefficient of extinction ε. ε is specific for the substance of investigation and depends on the wavelength. The ε of DNA at 260 nm is typically given as:

ε_{260}(DNA)

ε can be given in relation to the amount of the substance (c in mol/l) or to the mass of the substance (c in g/l). Unit of ε is therefore either:

l/mol/cm or l/g/cm

cm is used for practical reasons as most standard cuvettes have a length of 1 cm.
If ε is known, the concentration of the substance can be determined by measuring the amount of the absorb light. Most standard spectrometers directly report E.

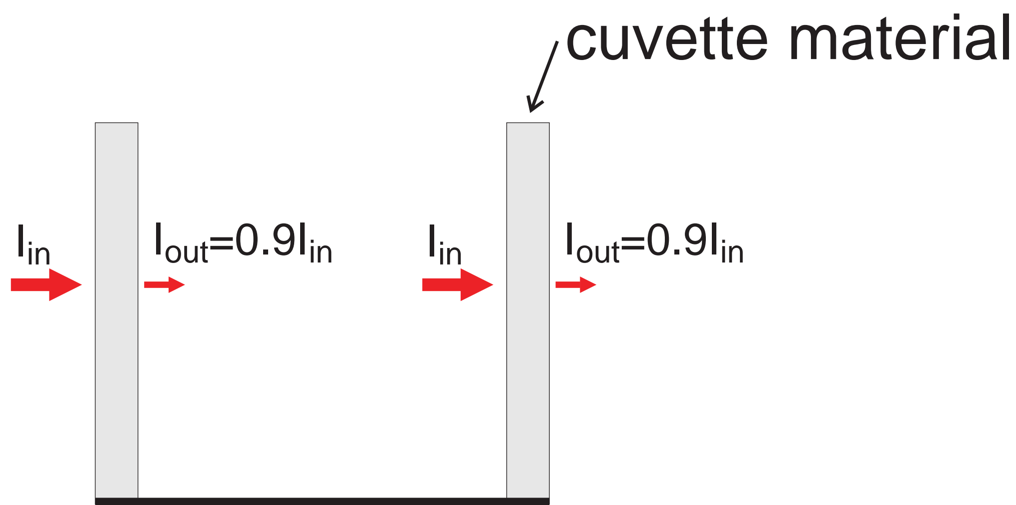
A. Design an apparatus capable of measuring the extinction of a solution. Which elements are required? Which data should be collected?

B. ε_{259}(ATP) = 15400 l/mol/cm
The molecular mass of ATP is 507.18 g/mol.
What is ε if the mass concentration should be used?

C. What is the concentration of an ATP solution (measured in a 1 cm cuvette) if E = 0.559?

D. Determine the extinction E of an ATP solution of 64,9 µM (1 cm cuvette).
Assume I_0 = 1. Determine I_{end}.
Use (13) to determine the intensity at different positions in the cuvette. Plot the obtained data.
Perform the same calculations for a concentration of 0.649 µM, 6,49 µM, and 649 µM
Which values for E can be measured in practice with sufficient confidence? Discuss your answer.

E. We assume that it is possible to determine the intensity of the light that enters the cuvette and of the light the leaves the cuvette. This assumption is reasonable. However, the light also has to pass the walls of the cuvette. It is likely that the material the cuvette is made from will absorb a fraction of the light. It is therefore common laboratory praxis that the extinction of the empty cuvette (E_{empty}) is subtracted from filled cuvette (E_{filled}), such that the extinction of the sample is considered to be E_{sample} = E_{filled} – E_{empty}. Based on Fig. 3 and (14) prove that this approach is mathematically correct.



F. A similar problem is absorption of light by solvent. ε of the solvent should be small in comparison to ε of the substance to be analysed. Nevertheless, absorption by the solvent may contribute to the measured E. It is therefore common to subtract the extinction of the cuvette filled with solvents from the cuvette with the sample. Demonstrate based on (2) that this approach is valid. From (2) you should obtain an expression in analogy to (14) without in depth calculations.

G. The isolation of DNA is a standard technique in biochemical laboratories. DNA concentrations are typically determined be measuring the extinction at 260 nm. The following coefficients of extinction are used:

double stranded DNA ε_{260} (dsDNA) = 20 l/g/cm
single stranded DNA ε_{260} (ssDNA) = 27 l/g/cm
mRNA ε_{260} (mRNA) = 25 l/g/cm

Does the differences in the coefficients in extinction for double and single stranded DNA meet the expectations? Explain and discuss your expectations.

The coefficient of extinction for ATP (task B) was given in mol/l/cm. Why is the coefficient of extinction for DNA related to the mass and not to the amount?

Intrigued by the difference for ds and ssDNA you find the following coefficients of extinction in the literature:

ε_{260} (ATP) = 15400 l/mol/cm
ε_{260} (GTP) = 11500 l/mol/cm
ε_{260} (TTP) = 8700 l/mol/cm
ε_{260} (CTP) = 7400 l/mol/cm

Assume human DNA with a GC content of 40%. Try to calculate the coefficient of extinction for DNA based on the coefficients for the single nucleotides. Discuss your results.

The GC content varies between different microorganisms from 20% to 80%. Try to estimate how big the error is, if you used the coefficient as provided above (which is for human DNA).

H. For protein typically this coefficient of extinction is used:

ε_{280} (protein) = 0,7 l/g/cm

As you see, E is measure at 280 nm. Consider also these coefficients:

ε_{260} (protein) = 0,4 l/g/cm
ε_{260} (dsDNA) = 20 l/g/cm
ε_{280} (dsDNA) = 11 l/g/cm

Why is DNA measured at 260 nm and protein at 280 nm? Consider also a formal treatment of the problem. This formal treatment is applicable to any measurement. The measured signal Γ (in our case the extinction) depends on the quantity to be measured η (in our case the concentration). The decency of Γ on η be mathematically known (in our case (14)). The sensitivity S of the measurement method is defined as:

$$S = \frac{d\Gamma}{d\eta} \qquad \textbf{(15)}$$

In words: The sensitivity of a method is the change in signal depending on the change in the quantity to be measured. This is the first derivative of the signal. What is the sensitivity based on (14)? What determines the sensitivity?

I. In F you have demonstrated that extinctions are additive. That means the extinction of a solution containing two different substances is the sum of the extinction of the extinction of the individual substances at the same concentration. Consider a mix solution of DNA and protein for which these values were measured:

E_{260} = 0.760
E_{280} = 0.850

Determine the concentration of DNA and protein.

J. For pure DNA solution E_{260}/E_{280} = 1.8 is observed.
For pure protein solution E_{260}/E_{280} = 0.57 is observed.
The ratio E_{260}/E_{280} is commonly used as a measure for the purity of isolated DNA. Try to estimate how much protein a DNA sample must contain to observe a clear change in ratio.

K. In G you tried to predict the coefficient of extinction of DNA base on the single nucleotides. The same could be done for protein. This would require knowledge of the coefficients of extinction of the amino acids. It has be found, that only tryptophan and tyrosine and slightly cysteine absorb light:
ε_{280} (Trp) = 5690 l/mol/cm
ε_{280} (Tyr) = 1280 l/mol/cm
ε_{280} (Cys) = 125 l/mol/cm
Which consequence does this have if you are interested in individual proteins?

Assignment 3
qPCR – quantitative PCR

As many of you proposed, it is indeed possible to detect the Corona Virus with qPCR. Some of you are already in part familiar with this technique. Here we will discuss qPCR step by step and learn about different flavours of qPCR.

A. We start with a “normal“ PCR.

Fig. 1 shows a DNA-fragment and two primers for use in in PCR reaction. (A primers is often referred to as oligo (from oligonucleotide)). In addition the reaction contains dNTPs (dATP, dGTP, dCTP and dTTP) (Fig. 2) and DNA-polymerase. If you are not familiar with the function of DNA-polymerase, check your textbooks of biochemistry).

The reaction is executed under these conditions (duration and temperature might be chosen slightly different by the experimenter):

1. Heating to 95 °C for 30 seconds – melting
   The DNA-fragment is melted, meaning that both DNA strands are separated from each other such that single stranded DNA is obtained. Compare the PCR reaction with the process of DNA replication in cells. What is substituted by heating?
2. Cooling to e.g. 50 °C for 30 seconds – annealing
   The primers bind to the single DNA strands. A rule of thumb allows to determine the melting temperature of the primer (the temperature at which the primer is just able to bind):
   (count of A and T) * 2 + (count of G and C) * 4 – 10 = annealing temperature in °C
3. Heating to about 70 °C for x seconds
   The temperature is chosen such that the polymerase can operate optimally. This is 68 °C for Taq, and 72 °C for Pfu.
   The duration depends on the length of the DNA fragment to be amplified. Taq requires one minute for 1000 bp, Pfu two minutes.

Step 1 to step 3 result in on copy of the DNA.

Step 1 to step 3 can be repeated several times, such that also copies of copies are generated. In an „normal“ PCR typically 30 cycles are used, in a qPCR up to 45 cycles.

Where do the primers bind to the DNA fragment?

Draw the product of the first cycle.

Continue up to cycle 5.

DNA-polymerase is a protein. What characteristics should this protein have if you look a the temperature profile of the PCR cyclus.

In vivo, DNA-polymerase requires RNA primers to start the synthesis reaction. How can a biochemist make use of this at his advantage?

Fig. 1 shows primers made form DNA and not from RNA. It seems that this is not a problem for DNA-polymerase (why?). Why do biochemist prefer DNA over RNA primers?

It is possible to combine step 2 and step 3? Analyse the protocol provided by The Charité (assignment 1) and discuss.

DNA fragment:

5′ TCTGGATAATGTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGA 3′
3′ AGACCTATTACAAACGCGGCTGTAGTATTGCCAAGACCGTTTATAAGACTTTACTCGACAACTGTTAATTAGTAGCCGAGCATATTACACACCTTAACACTCGCCT 5′

Primer 1:
5′ ATGTTTGCGCCGACATCATAAC 3′

Primer 2:
5′ AATTCCACACATTATACGAGCC 3′

Fig. 1. A double stranded DNA fragment is shown. Normally only the upper strand would be shown to provide sequence information. In addition, the sequences of two primers are shown, which should be used for the PCR reaction. The length of the primer is real and could be used in the laboratory. The DNA fragment is for simplicity uncommonly short. Typically 70 to 200 bp are amplified in a qPCR. Amplification of regions up to 5000 bp (and often also longer) is routine in normal PCR reaction used in the context of molecular biology.

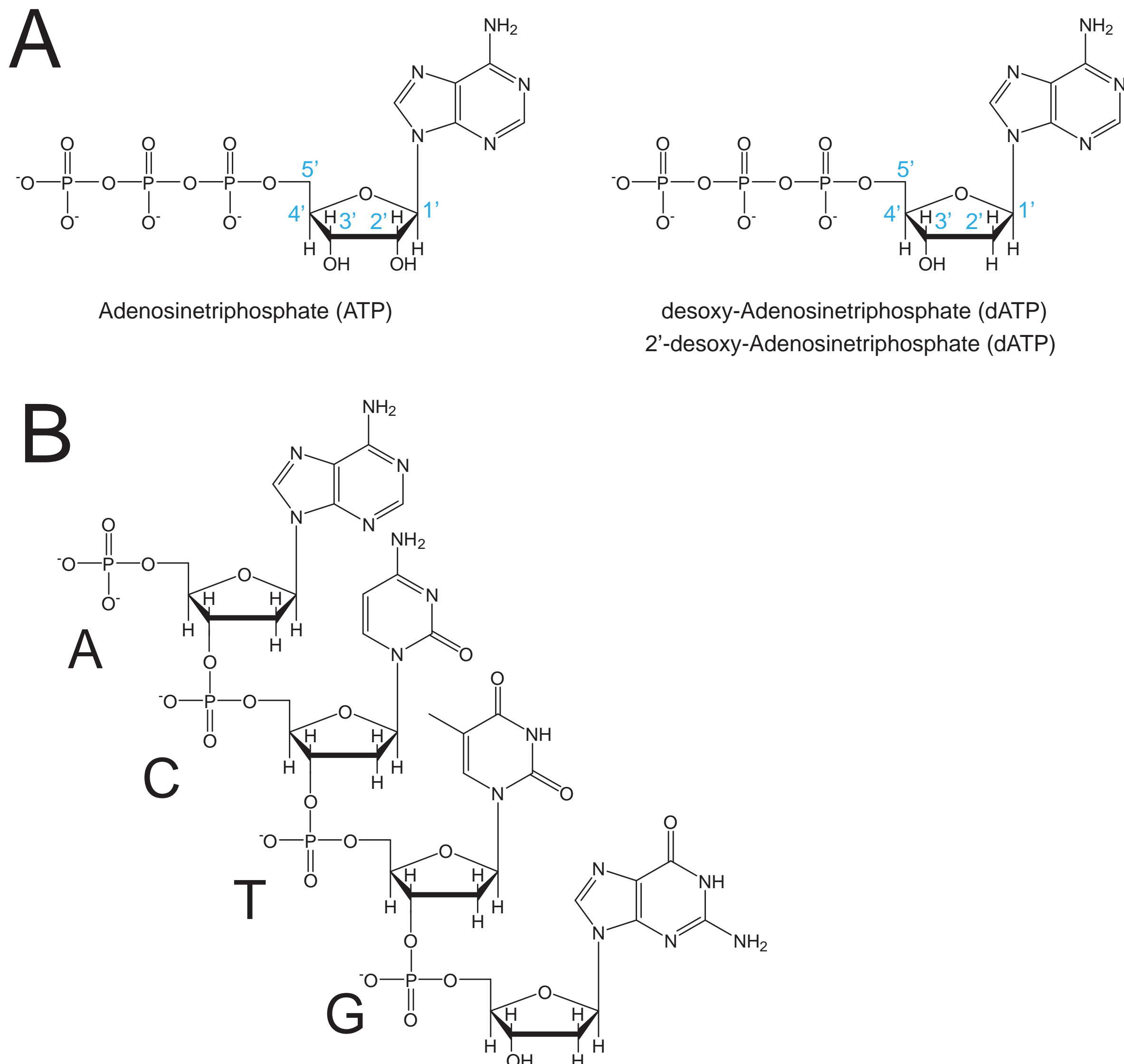


Fig. 2. Nucleotides. A. shown are ATP and dATP. B. a short single strand composed of all nucleotides.

B. Calculate the number of DNA molecules after 5, 10, 15, 20, …, 45 cycles if the reaction was started with a single DNA molecule.
Calculate the number of DNA molecules after 45 cycles if the reaction started with 10, 1,000, or 100,000 DNA molecules.
Assume that the reaction is performed in a volume of 25 µl. Determine the obtained DNA concentration in in mol/l and g/l. Assume that the amplified DNA fragment has a size of 100 bp. Discuss your results.

C. Calculate the number of DNA molecules at the start of the PCRE if 6,002,049,024 molecules were obtained after 20 cycles (the accuracy of this number is not realistic).

Task X for D, F and G.

So far we discussed a „normal" PCR. In B and C we learned that it is possible to deduce the amount of DNA present at the beginning of the reaction from the amount of the PCR product. We will discus three possibilities to quantify the PCR product. Try to compare these methods. List pro and cons.

- How easy and how precise in the quantification? Consider potential sources of errors.
- Any quantitative analysis requires a signal Γ that is related to the measure (here the amount of PCR product). Try to define a mathematical equation for the dependency of Γ.
- qPCR aims to detect a specific DNA sequence on the background of a waste excess of background sequences. How is specificity reached?
- Distinguish two objectives:
  a. You want to conclude if a specific sequence (e.g. a virus) is present or not. How do you have to analyse the data?
  b. You want to gain information on the concentration of a DNA sequence of interest.

D. Method 1: The PCR product is analysed on an agarose gel, which resolves DNA fragments according to size. Some of you have suggested this approach. In this way, a "normal" PCR becomes a qPCR. How does the gel look like? Consider this situation: sample 1, absence of DNA of interest; sample 2, a very low concentration of the DNA of interest: sample 3, concentration 4 times the concentration of sample 2; sample 4, concentration 1000 times the concentration of sample 2.

E. Intermezzo: This task should help you to understand to origin of the specificity of a qPCR. A PCR reaction is performed as in A. This time three DNA fragments and 4 primers are present in the reaction (Fig. 4). Where do the primers bind? The product of the reaction is analysed by agarose gel. Draw and explain the expected outcome.

The following to methods are referred to as real time PCR. Aim is to monitor the formation of the PCR product life (in real time).

F. Method 2: A fluorophore such as SYBR-green is added to the reaction described in A. SYBR-green binds in particular to double stranded DNA. If a fluorophore is excited with light of wavelength λ_1 it emits light at a longer wavelength λ_2. The PCR machine is designed such that the fluorescence single can be measured in each cycle. The intensity depends on the environment of the fluorophore and is larger if the fluorophore is bound to DNA.
At which point of the cycle should the fluorescence be measured?
Draw the fluorescence depending of the number of cycles.

DNA fragment1:

5′ AGCGTGCAGAGATTTCAATGCTTGAAGGAGCGGTTTTGGATATTAGATACGGTGTTTCGAGAATTGCATATAGTAAAGACTTTGAAACTCTCAAAGTTGATTTTCT 3′
3′ TCGCACGTCTCTAAAGTTACGAACTTCCTCGCCAAAACCTATAATCTATGCCACAAAGCTCTTAACGTATATCATTTCTGAAACTTTGAGAGTTTCAACTAAAAGA 5′

DNA fragment2:

5′ TGTGCCTGGATGCGTTCCCAAAATTAGTTTGTTTTAAAAAACGTATTGAAGCTATCCCACAAATTGATAAGTACTTGAAATCCAGCAAGTATATAGCATGGCCTTT 3′
3′ ACACGGACCTACGCAAGGGTTTTAATCAAACAAAATTTTTTGCATAACTTCGATAGGGTGTTTAACTATTCATGAACTTTAGGTCGTTCATATATCGTACCGGAAA 5′

DNA fragment3:

5′ AAATGCTGAAAATGTTCGAAGATCGTTTATGTCATAAAACATATTTAAATGGTGATCATGTAACCCATCCTGACTTCATGTTGTATGACGCTCTTGATGTTGTTTT 3′
3′ TTTACGACTTTTACAAGCTTCTAGCAAATACAGTATTTTGTATAAATTTACCACTAGTACATTGGGTAGGACTGAAGTACAACATACTGCGAGAACTACAACAAAA 5′

Primer 1:
5′ atggtgatcatgtaacccatc 3′

Primer 2:
5′ GAGATTTCAATGCTTGAAGGAG 3′

Primer 3:
5′ AATCAACTTTGAGAGTTTCAAAG 3′

Primer 4:
5′ acaacatcaagagcgtcatac 3′

Fig. 3

G. Method 3: Next to the two primers used in A an additional oligo is added to the reaction. This oligo is modified at the 5' and 3' end. At the 5'end a FAM group and at the 3'end a BBQ group is attached (Fig. 4A):

FAM-TATTCTGAAATGAGCTGTTGA-BBQ.

Where does the oligo bind to the DNA. What would be the consequence for the polymerase reaction? Assume you are using Taq Polymerase. This polymerase has additionally 5'-3' exo nuclease activity. What are the consequences of this?

FAM is fluorescein with linker. What is the function of the linker?

Fluorescein is like SYBR-green a fluorophore.

Excurse:

A fluorophore absorbs a photon of a specific wavelength. The energy of the photon in then stored in the fluorophore. The fluorophore is able to emit a portion of the stored energy again as light. It thus emits a photon with lower energy (longer wavelength) than the absorbed photon. The interval between abortion and emission is in the order of 10 ns. The remaining energy stays with the fluorophore and can be transfer to water molecules by collision. If a molecule that is able to absorb light of the same energy as the fluorophore emits comes close to the fluorophore prior to emission, the energy can be transfer without emission of light to that molecule. This process is referred to as fluorescence resonance energy transfer (FRET). Its physical description requires quantum mechanical considerations. As a consequence, the fluorophore lost its energy without emission of light. It is said that the fluorophore war quenched. BBQ is such a molecule and referred to as quencher. BBQ then transfers the energy by collision to water. For efficient quenching orientation and distance between fluorophore and quencher needs to be optimal. The distance should not be longer than 5 nm. As FAM and BBQ are coupled to one oligo both are in rather close proximity. A smart design of the oligo can increase the relative proximity further (Fig 4B).

Explain why the fluorescence in increasing during upon formation of PCR product.

H. The Charité uses method 3. Analyse the oligos presented in table 1 of the document provided by The Charité.

Makes The Charité use of the trick shown in Fig. 4B? Why?

What is the meaning of the letters W, R and M in the oligos? What are the consequence? What is the intention behind this design?

Analyse figure 4 in the document provided by The Charité. The figure shows the time course of a qPCR. Try to explain the shown curves. Are the curves as you expected them to be?

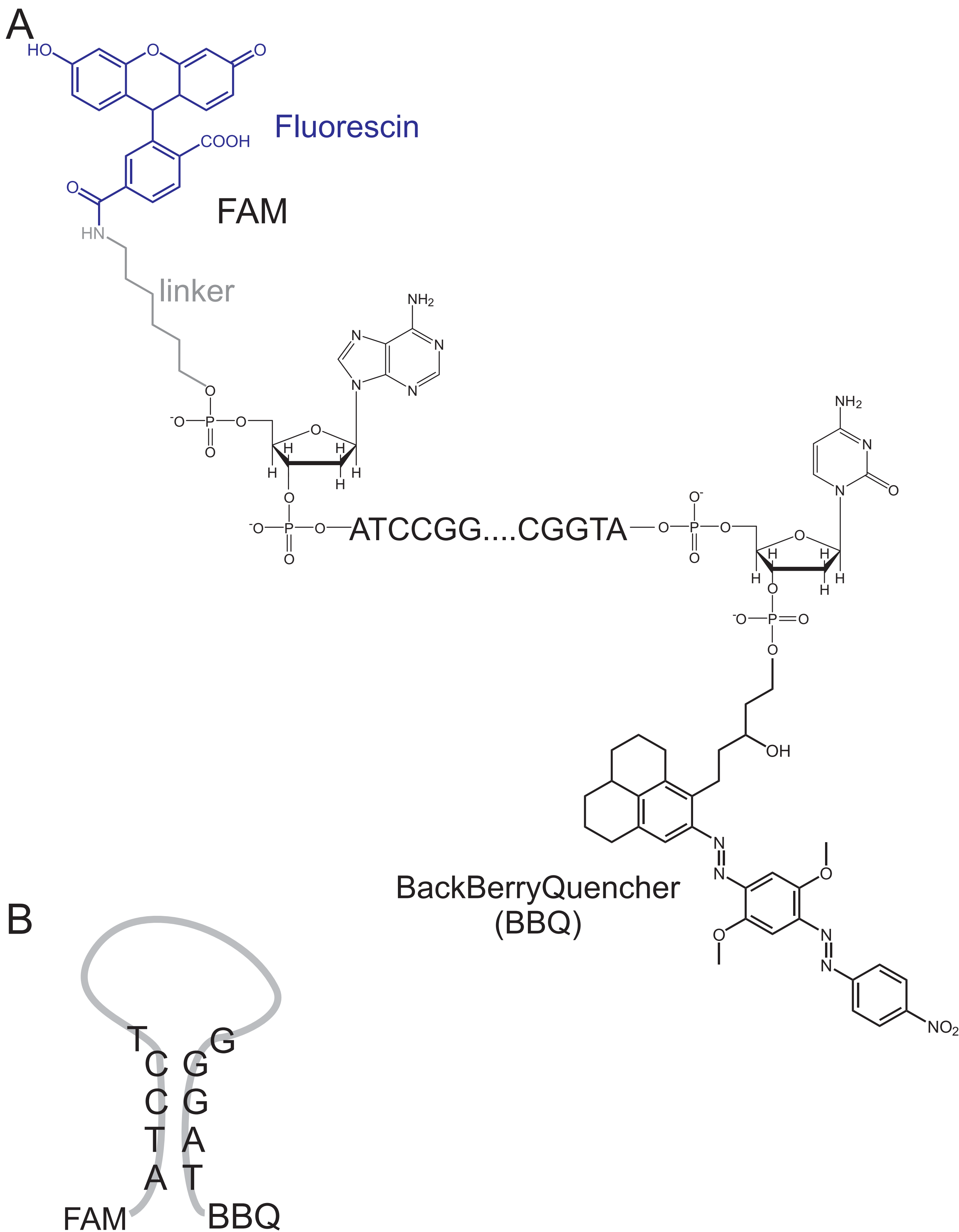


Fig. 4