

**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA ELEKTROTECHNIKY  
A KOMUNIKAČNÍCH TECHNOLOGIÍ**

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

**ÚSTAV AUTOMATIZACE A MĚŘICÍ TECHNIKY**

DEPARTMENT OF CONTROL AND INSTRUMENTATION

**ANALÝZA DAT ZÍSKANÝCH Z GENOTYPIZAČNÍCH ESEJÍ  
Z REAL-TIME PCR**

ANALYSIS OF DATA OBTAINED FROM GENOTYPING ESSAYS FROM THE REAL-TIME PCR

**BAKALÁŘSKÁ PRÁCE**

BACHELOR'S THESIS

**AUTOR PRÁCE**

AUTHOR

**Martin Bečička**

**VEDOUCÍ PRÁCE**

SUPERVISOR

**Ing. Jiří Sekora**

**BRNO 2018**

# Bakalářská práce

bakalářský studijní obor **Automatizační a měřicí technika**

Ústav automatizace a měřicí techniky

**Student:** Martin Bečička

**ID:** 174191

**Ročník:** 3

**Akademický rok:** 2017/18

**NÁZEV TÉMATU:**

## **Analýza dat získaných z genotypizačních esejí z real-time PCR**

### **POKYNY PRO VYPRACOVÁNÍ:**

- 1) Seznamte se s metodou real-time PCR a jejím využitím pro genotypizování pomocí hydrolyzačních TaqMan sond.
- 2) Seznamte se s real-time PCR data markup language (rdml) a strukturou ukládaných dat v tomto formátu.
- 3) Navrhněte software pro zobrazení dat získaných z real-time PCR cycleru ve formátu RDML umožňující automatickou identifikaci genotypů nejen na základě hodnot end-point fluorescence, ale také na základě uživatelem definovaných parametrů amplifikačních křivek.
- 4) Sestavte software včetně uživatelského rozhraní.
- 5) Otestujte sestavený software na datech z reálného provozu. Výsledky diskutujte.

### **DOPORUČENÁ LITERATURA:**

- [1] Deepak S, Kottapalli K, Rakwal R, et al. Real-Time PCR: Revolutionizing Detection and Expression Analysis of Genes. *Current Genomics*. 2007;8(4):234-251.
- [2] Kemp, Paul F., Cole, Jonathan J., Sherr, Barry F., and Evelyn B. Sherr. *Handbook of Methods in Aquatic Microbial Ecology*. CRC Press, 28. 6. 1993

**Termín zadání:** 5.2.2018

**Termín odevzdání:** 21.5.2018

**Vedoucí práce:** Ing. Jiří Sekora

**Konzultant:** MUDr. Michal Jurajda, Ph.D.

**doc. Ing. Václav Jirsík, CSc.**  
*předseda oborové rady*

### **UPOZORNĚNÍ:**

Autor bakalářské práce nesmí při vytváření bakalářské práce porušit autorská práva třetích osob, zejména nesmí zasahovat nedovoleným způsobem do cizích autorských práv osobnostních a musí si být plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č.40/2009 Sb.

## **Abstrakt**

Bakalářská práce se zabývá vizualizací dat získaných z real-time PCR ve vývojovém prostředí MATLAB. Teoretická část práce obsahuje úvod do tématu PCR a real-time PCR, nástrojů pro vyhodnocení dat získaných z real-time PCR s popisem mezinárodního standardu RDML pro jejich ukládání. Praktická část popisuje funkce vytvořeného grafického rozhraní.

## **Klíčová slova**

Real-time PCR, GUI, end-point analýza, Matlab, RDML

## **Abstract**

Bachelors thesis is concerned with the visualization of real-time PCR data in MATLAB. Theoretical part of the thesis provides introduction to PCR and real-time PCR, describes tools used for the evaluation of data acquired from real-time PCR and the international format RDML used for the storage of such data. The practical part describes the developed graphical interface.

## **Keywords**

Real-time PCR, GUI, end-point analysis, Matlab, RDML

### **Bibliografická citace:**

BEČIČKA, M. Analýza dat získaných z genotypizačních esejí z real-time PCR. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2018. 39 s. Vedoucí bakalářské práce Ing. Jiří Sekora.

## **Prohlášení**

„Prohlašuji, že svou bakalářskou práci na téma „analýza dat získaných z genotypizačních esejí z real-time PCR“ jsem vypracoval samostatně pod vedením vedoucího bakalářské práce a s použitím odborné literatury a dalších informačních zdrojů, které jsou všechny citovány v práci a uvedeny v seznamu literatury na konci práce.

Jako autor uvedené bakalářské práce dále prohlašuji, že v souvislosti s vytvořením této bakalářské práce jsem neporušil autorská práva třetích osob, zejména jsem nezasáhl nedovoleným způsobem do cizích autorských práv osobnostních a jsem si plně vědom následků porušení ustanovení § 11 a následujících autorského zákona č. 121/2000 Sb., včetně možných trestněprávních důsledků vyplývajících z ustanovení části druhé, hlavy VI. díl 4 Trestního zákoníku č. 40/2009 Sb.

V Brně dne: .....

.....

podpis autora

## **Poděkování**

Děkuji vedoucímu bakalářské práce Ing. Jiřímu Sekorovi a konzultantovi MUDr. Michalu Jurajdovi, Ph.D. za odborné vedení, trpělivost a ochotu.

V Brně dne: .....

.....

podpis autora

# Obsah

Úvod .....	10
1 Polymerázová řetězová reakce(PCR) .....	11
1.1 Princip PCR.....	11
1.2 Detekce amplifikovaného produktu PCR .....	12
1.2.1 Gelová elektroforéza.....	12
1.2.2 Hybridizační metody .....	13
1.3 Praktické aplikace PCR.....	13
1.3.1 SNP Genotypování .....	13
1.3.2 Diagnóza onkologických onemocnění.....	14
1.3.3 Využití ve virologii.....	14
1.3.4 Využití v kriminalistice .....	15
1.3.5 Využití v archeologii .....	15
2 Real-time PCR.....	16
2.1 Princip real-time PCR.....	16
2.2 Specifická a nespecifická detekce signálu .....	16
2.3 Automatizace real-time PCR (qPCR) .....	18
2.4 Analýza a interpretace dat .....	19
2.4.1 Amplifikační křivka .....	20
2.4.2 Genotypizace pomocí end-point analýzy .....	21
3 Real-time PCR Data Markup Language .....	23
3.1 XML.....	23
3.1.1 Základní prvky XML .....	23
3.2 Struktura RDML .....	24
4 Realizace programu .....	26
4.1 Analýza požadavků .....	26
4.1.1 Konstrukce amplifikační křivky .....	28
4.2 Grafické uživatelské rozhraní – GUI .....	29
4.2.1 Popis GUI.....	30
4.2.2 Klasifikaci přímek podle fluorescence a čísla cyklu .....	35
5 Diskuze výsledků .....	36
Závěr.....	37
Literatura .....	38

## Seznam symbolů, veličin a zkratk

PCR	Polymerázová řetězová reakce
qPCR	Kvantitativní (real-time) PCR
ssDNA	Jedno vláknová DNA
dsDNA	Dvou vláknová DNA
Taq	<i>Thermus aquaticus</i>
mRNA	Mediátorová ribonukleová kyselina
RDML	Real-time PCR Data Markup Language
SNP	Single-nucleotide polymorphism



## Seznam obrázků

<i>Obr. 1.1: Průběh amplifikačního cyklu PCR [4].</i>	11
<i>Obr. 2.1: Vlevo jsou křivky tání specifického a nespecifických produktů. Vpravo je zobrazena negovaná derivace těchto křivek[21].</i>	17
<i>Obr. 2.2: Systémy specifické a nespecifické detekce[3].</i>	18
<i>Obr. 2.3: End-point analýza pro vyhodnocení genotypizačních dat. Ukázka z programu LightCycler® 96 SW 1.1.</i>	21
<i>Obr. 3.1: Zjednodušené schéma RDML[14].</i>	25
<i>Obr. 4.1: Amplifikační křivky při neideálním průběhu amplifikace. Ukázka z programu LightCycler® 96 SW 1.1.</i>	27
<i>Obr. 4.2: Logistická křivka a její parametry.</i>	28
<i>Obr. 4.3: Příklad rozložení GUI používaného v LightCycler® 96 SW 1.1.</i>	29
<i>Obr. 4.4: Zjednodušený vývojový diagram.</i>	30
<i>Obr. 4.5: Panel pro klasifikaci křivek po provedení analýzy.</i>	32
<i>Obr. 4.6: Panel pro genotypizaci a manuální určení.</i>	34
<i>Obr. 4.7 Klasifikace podle čísla cyklu a fluorescence.</i>	35

# ÚVOD

Polymerázová řetězová reakce (polymerase chain reaction, dále jen PCR) je jednou z nejpoužívanějších metod v molekulární biologii. Jedná se o rychlou metodu in vitro amplifikace DNA, pomocí které při optimálních podmínkách lze vytvořit velké množství kopií DNA i jen z jedné templátové molekuly DNA. Cílem této bakalářské práce je vytvořit grafické rozhraní pro vizualizaci a analýzu dat získaných z kvantitativní PCR.

Úvodní kapitoly práce se zabývají teoretickými znalostmi nutnými pro vytvoření a pochopení funkcí naprogramovaného GUI. První kapitola představuje fungování, způsoby vyhodnocování výsledků a oblasti, ve kterých je PCR využívána.

Druhá kapitola pojednává o real-time PCR, metodě založené na principu PCR. Zde jsou popsány nejvíce používané metody pro vizualizaci a analýzu real-time PCR dat.

Následuje popis standardu RDML, sloužícího pro ukládání dat získaných z real-time PCR.

Samotný popis realizace programu lze najít v předposlední kapitole. Jsou zde uvedené nároky kladené na GUI konzultantem této práce. Následuje detailní popis funkcí výsledného grafického rozhraní.

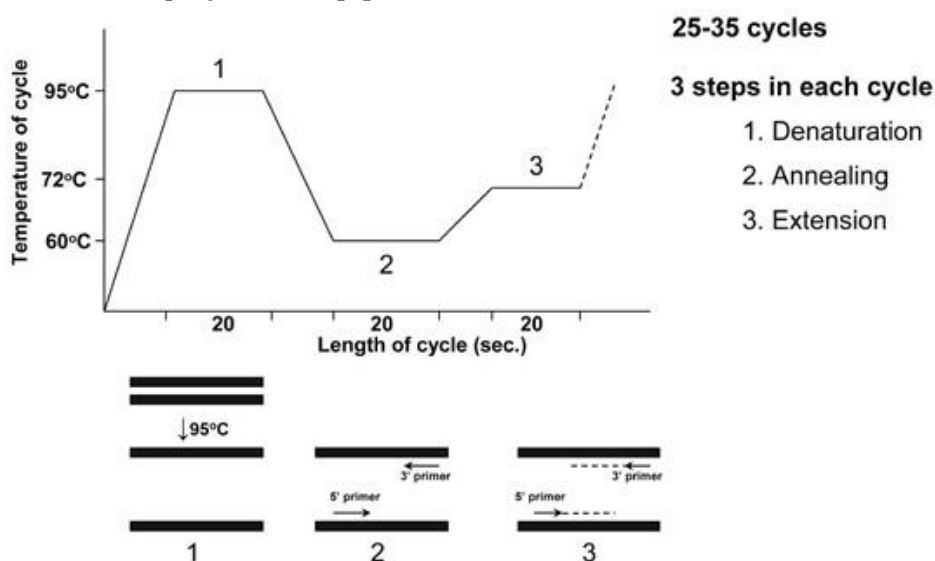
Diskuze dosažených výsledků tvoří poslední kapitolu této práce.

# 1 POLYMERÁZOVÁ ŘETĚZOVÁ REAKCE(PCR)

Polymerázová řetězová reakce je proces amplifikace vybraných úseků DNA. Poprvé byla navržena v roce 1983 Kerry Mullisem, který za to byl v roce 1993 oceněn Nobelovou cenou. Dnes se řadí mezi jednu z nejpoužívanějších technik v molekulární biologii.[1]

## 1.1 Princip PCR

Principem PCR je enzymatická amplifikace nukleových kyselin do ssDNA komplementární k templátové ssDNA. Jeden cyklus PCR sestává z tří kroků, a to z **denaturace dsDNA**, která rozvolní dvouvláknovou DNA a umožní **nasedání primerů**, neboli **annealing**. DNA polymerázy se naváží na nasednuté primery, čímž zahájí samotnou **syntézu komplementární jednovláknové DNA**. Tyto tři kroky, řízené teplotou tvoří amplifikační cyklus PCR, který je graficky znázorněn na Obr. 1.1. Pro syntézu dostatečného množství amplikonu, je amplifikační cyklus PCR opakován zpravidla 25 až 30x. K úspěšnému cyklu PCR je zapotřebí mít v reakční směsi dostatek templátové DNA, dva specifické primery, DNA nukleotidtrifosfáty a termostabilní DNA polymerázu.[3]



Obr. 1.1: Průběh amplifikačního cyklu PCR [4].

Při **denaturaci dvouvláknové DNA** je teplota zvýšena na 94 °C, což způsobí rozpad vodíkových můstků vázajících dvě jednovláknové DNA. Při úvodním cyklu je denaturační teplota držena obvykle 2-5 minut, což zajistí kompletní rozpad nadbytečně dlouhé templátové DNA. V následujících cyklech jsou už přítomny kratší cílové produkty PCR, proto postačí setrvat v denaturační teplotě zpravidla 1 minutu. Při použití Hot Start DNA polymerázy se délka úvodního cyklu zvyšuje až na 15

minut. Tyto polymerázy mají enzymatickou aktivitu reverzibilně blokovanou. K odblokování enzymatické aktivity dojde v prvním cyklu PCR, čímž je zajištěno, že nedojde k amplifikaci primerů nasednutých nespecificky na DNA při nižších teplotách během přípravy reakční směsi.[2]

**Annealing** neboli nasedání primerů je umožněno snížením teploty pod denaturační teplotu použitých primerů, což umožní primerům nasedat ke komplementárním sekvencím ssDNA. Primery jsou oligonukleotidy o délce 16-30 nukleotidů, které jsou komplementární k protilehlým řetězcům DNA určeného k amplifikaci. Je pravidlem, že cílem zájmu je jen určitá sekvence templátové DNA. Teplota v této fázi je závislá na použitých primerech a dosahuje teplot v rozmezí 55 až 72 °C.[2][3]

Optimální teplota pro **elongaci DNA** (prodlužování řetězců DNA) je určena použitou DNA polymerázou, obvykle 72 °C. DNA polymeráza započne syntézu DNA ve směru 5' → 3' od místa, kde je nasednutý primer. Syntéza nového vlákna je založená na komplementaritě bází podle Watson-Crickova párování. Doba trvání elongace je závislá na požadované délce nově vytvořených vláken a na typu použité DNA polymerázy. Pro své termostabilní vlastnosti je v současnosti nejvíce používaná Taq polymeráza bakterie *Thermus aquaticus*. Standardní čas pro elongaci pomocí Taq DNA polymerázy je asi 10 sekund na 1500 párů bází.[2][3]

Při ideálních podmínkách by každý amplifikační cyklus PCR měl zdvojnásobit počet vstupní DNA. V praxi se takto děje na začátku reakce, ke konci již není reakce tak účinná. Snížení účinnosti je způsobeno spotřebováním potřebných reaktantů a nahromaděním koncových produktů. Obvykle je amplifikační cyklus PCR zopakován 25–35 x.[4]

## 1.2 Detekce amplifikovaného produktu PCR

Výstupem správně proběhlé PCR je namnožená sekvence templátové DNA, nebo původní směs, pokud není přítomná v templátovém DNA komplementární sekvence k použitým primerům. Ať je výstup jakýkoliv, je potřeba použít dalších technik na jeho vyhodnocení.

### 1.2.1 Gelová elektroforéza

Gelová elektroforéza je jednou z nejčastějších a nejrychlejších cest, jak analyzovat PCR produkty. Jejím principem je pohyb nabitých částic, v tomto případě negativně nabitě DNA, v elektrickém poli. Každý nukleotid obsahuje záporně nabitou fosfátovou skupinu, proto míra pohybu DNA v nabitém poli závisí jen na délce DNA.[2]

Pro vizualizaci výsledků je nutné obarvit DNA fluorescenčními barvivými. Pro tento účel je v současnosti nejvíce používán ethidium bromid, i přes snahu ho

nahradit jinými barvivy, kvůli jeho karcinogenním účinkům. Po proběhnutí elektroforézy je použita UV lampa a popř. kamera pro archivaci výsledků. Jako molekulové síto slouží agarózové, či polyakrylamidové gely s velikostí pórů závislou na očekávané délce DNA a na požadovaném rozlišení. Pro přibližný odhad délky DNA se do volných jamek gelu může přidat velikostní marker s předem známým počtem párů bází.[2]

### **1.2.2 Hybridizační metody**

Hybridizační metody jsou používány k určení přítomnosti specifických genů. Principem těchto metod je použití specifických oligonukleotidových sond, které za přítomnosti komplementárního úseku ssDNA k němu hybridizují. Denaturace dsDNA na ssDNA probíhá pomocí zvýšení teploty nebo chemicky. Typickými představiteli této metody je např. Southernův blotting, při kterém restriční endonukleázy rozštěpí DNA na kratší úseky, tyto úseky jsou rozděleny podle velikostí pomocí gelové elektroforézy, denaturovány, přeneseny na filtr a následně jsou použité radioaktivní, nebo neradioaktivní hybridizační sondy. Po uvolnění nenavázaných sond pomocí roztoků, je možné vyhodnotit, které sekvence obsahují navázané sondy.[2][9]

## **1.3 Praktické aplikace PCR**

PCR je jedna z nejvyužívanějších metod v molekulární biologii. Proto je možné najít její využití téměř ve všech odvětvích. Nejedná se pouze o medicínskou diagnostiku, ale tato metoda se využívá od archeologie, forenzní vědy a evoluční biologie po populační genetiku a klonování.[8]

### **1.3.1 SNP Genotypování**

Genotypování je proces určení genotypu jedince a jeho následné porovnání se zbytkem populace. Jednonukleotidové polymorfismy (SNPs, single-nucleotide polymorphisms) jsou záměny jedné báze, vyskytující se na určitém místě genomu u více než 1 % jedinců zkoumané populace. SNPs se mohou nalézat v kódujících i nekódujících sekvencích genomu.[17]

V nekódujících sekvencích mohou sloužit např. pro komparativní genetiku, která porovnává genomy různých organismů, či pro evoluční genetiku. SNPs v kódujících oblastech genomu mohou způsobit změnu v expresi genu, a tak vést k různým fenotypům. I když je to nepravděpodobné, je možné že jediný SNP může způsobit dědičné onemocnění. Správné fungování funkcí v organismu zpravidla závisí na více genech, a proto pro se na manifestaci změn ve fenotypu může podílet více SNPs. Pro realizaci testu se používá real-time PCR se dvěma a více fluorescenčně

značenými oligonukleotidovými sondami. Více o oligonukleotidových sondách lze nalézt v kapitole 2.2.[17]

### **1.3.2 Diagnóza onkologických onemocnění**

Onkologická onemocnění jsou druhou nejčastější příčinou smrti, proto je jejich přesná a včasná detekce jedním z největších zdravotních problémů dnešní doby. Nádorová onemocnění jsou způsobena většinou somatickými mutacemi v DNA, které mají za následek ztrátu kontroly nad proliferací a diferenciací buněk. Tyto mutace můžou být většího měřítko jako např. translokace, inverze, delece částí DNA nebo menšího charakteru jako bodové mutace, bodové delece/inzerce. Mutace jsou způsobeny chybou při replikaci DNA, vystavením škodlivému záření, či chemikáliím.[6]

Nejčastější mutací u nádorů, vyskytující se u více než 50 % zhoubných nádorů, je mutace v tumor supresorovém genu p53 kódovaným genem T53. Nedílnou součástí onkologické diagnostiky je také qPCR, často používaná jako validace výsledků z DNA čipů nebo také jako vysoce senzitivní test pro detekci malého počtu mutací. Diagnostika pomocí qPCR probíhá za použití fluorescenčně značených sond komplementárních např. k variantám nemutovaných genů T53. Pozitivní výsledek odpovídá jiným poměrům, či jiným křivkám tání nemutované, referenční DNA a daného vzorku.[6]

Rakovinné buňky také vykazují zvýšenou hodnotu exprese určitých genů, a buňka tedy může obsahovat větší množství určitých proteinů, což může sloužit jako marker onkologického onemocnění. Vyšetření probíhá pomocí real-time PCR mRNA získaného z vyšetřované oblasti. Následně se porovnají průběhy amplifikačních křivek z vyšetřované oblasti oproti referenčním křivkám z mRNA z buněk neobsahující mutaci. Pokud buňky odebrané z vyšetřované oblasti obsahují mutovaný gen HER-2 (mutace hojně se vyskytující u rakoviny prsu), lze zaznamenat dřívější nárůst fluorescence amplifikačních křivek při jejich porovnání s amplifikačními křivkami získaných z referenčních zdravých buněk. To je způsobeno přítomností většího množství cDNA ve vzorku získaném z mutovaných buněk, a tudíž je možné předpokládat větší množství mRNA v těchto buňkách.[6]

### **1.3.3 Využití ve virologii**

Nedávný technologický vývoj umožnil amplifikaci specifické částí virové DNA a jejich následnou analýzu, kde často je i možné přesné stanovení genotypu, subtypu, a mutací těchto virů.[7]

Na vzestupu je použití PCR k diagnóze HIV viru. HIV je RNA virus, který po infiltraci CD4 buněk nebo makrofágů využije reverzní transkriptázy na včlenění vlastní sekvence DNA do hostitelské DNA. Tam jako provirus čeká na transkripci,

která by vytvořila další kopie viru HIV. U viru HIV je testování pomocí amplifikace DNA zvláště užitečné, kvůli faktu že je většinou použité testování antigenů p24 je nemožné použít u dítěte do 18 měsíců, protože je nemožné rozlišit, zda antigen p24 nalezen v krvi doopravdy pochází od dítěte, či od matky. Další nevýhodou klasického testování na přítomnost p24 je jeho nízká hladina u asymptotických pacientů. Pomocí HIV DNA detekce je možné zaznamenat HIV pozitivní o zhruba 1 týden dříve než pomocí detekce p24 antigenu. Kvůli těmto výhodám je při testování HIV čím dál více využívána PCR.[7]

### **1.3.4 Využití v kriminalistice**

V kriminalistice je poměrně častá situace, kdy je k dispozici jen malý vzorek DNA podezřelého na místě činu. Pomocí PCR je možné i z velmi malého vzorku vytvořit miliony kopií a následně je použít na genetickou daktyloskopii. Při genetické daktyloskopii jsou studovány určité části DNA, které vykazují vysokou variabilitu mezi jedinci. V minulosti tato technika byla značně kontroverzní, ale s vývojem technologie a s novými studiemi se dokázala spolehlivost této metody. V současnosti je používána jako spolehlivý důkaz. Tato metoda nemusí být použita jen k důkazu nevin, desítky případů jsou známy, kde nevinní lidé strávili mnoho let za mřížemi. PCR bylo použito v tomto případě na dokázání nevin těchto lidí, i z důkazů starých několik desítek let.[8]

### **1.3.5 Využití v archeologii**

Dalším důležitým využitím je studium archeologické DNA (aDNA). Tyto vzorky mohou být staré až několik milionů let. Omezením studia aDNA je degradace původní DNA, kde v lepším případě degraduje na sekvence čítající několika stovek bází a v horším není možné extrahovat jakoukoli užitečnou část DNA.[8]

## 2 REAL-TIME PCR

Krátce po objevu PCR byly učiněny první pokusy použít PCR ke kvantitativním analýzám. První kvantitativní metoda byla kompetitivní PCR. Tato metoda zahrnovala přidání známého množství jiné DNA (standardu), obsahující homologní sekvenci totožnou s vyšetřovanou DNA. Obě templátové DNA proto při amplifikaci soutěží o reakční komponenty. Z elektroforegramu výsledné směsi DNA bylo možné stanovit poměr stanovované DNA vůči standardu. Jedním z prvních pokusů o real-time PCR bylo odebírání malých vzorků z reakce každých  $n$ -cyklů. Vzorky pak mohly být analyzovány pomocí gelové elektroforézy. Později byla do reakční směsi přidána fluorescenční barva a cycler byl doplněn o zdroj excitačního světla a kameru. Fluorescenční barvivo se naváže na molekuly DNA a tím zvýší svojí fluorescence, která je zaznamenávána kamerou. Každý cyklus PCR je tak možné kvantifikovat na rozdíl od tradiční PCR, kde se dá ke kvantifikaci použít mnohem nepřesnější elektroforéza.[3][20]

### 2.1 Princip real-time PCR

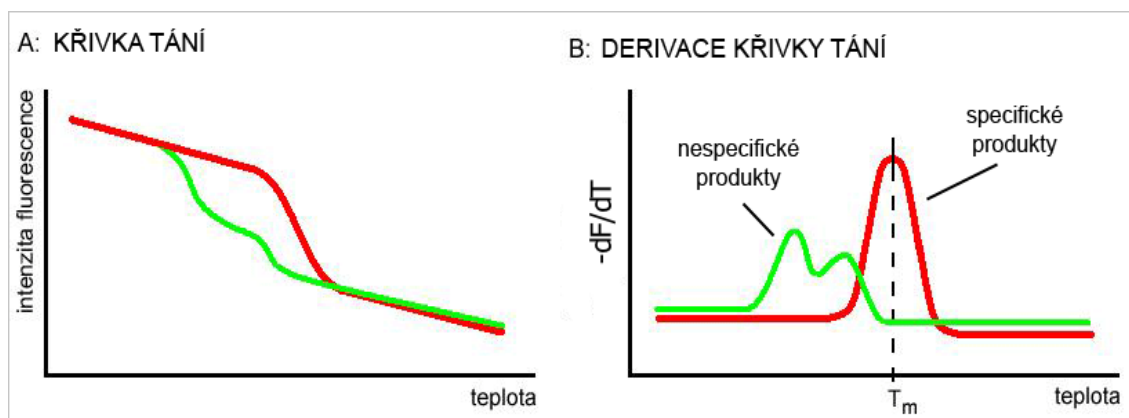
V kvantitativní real-time PCR (qPCR) je množství syntetizovaných produktů zaznamenáváno průběžně. Ke kvantifikaci jsou použita fluorescenční barviva, jejichž fluorescence je přímo úměrná množství PCR produktů. Fluorescence může být takto průběžně kvantifikována v každém cyklu. Techniky detekce jsou rozděleny do dvou kategorií: specifická a nespecifická detekce, podle způsobu vazby fluorescenčního barviva na DNA.[5]

### 2.2 Specifická a nespecifická detekce signálu

Nespecifická detekce využívá schopnosti interkalačních barviv vázat se na jakoukoli dsDNA. Tato metoda tedy nedokáže rozlišit, zda se váže na nespecifické (primer dimery, artefakty), nebo na specifické produkty.

V těchto případech může být specifická reakce vylepšena zvýšením teploty nad denaturační teplotu nespecifických produktů, ale pod denaturační teplotu specifických produktů. To je možné kvůli faktu, že nespecifické produkty se na sebe váží menším počtem vodíkových můstků, proto mají i menší teplotu tání. Přítomnost nespecifických produktů můžeme zjistit pomocí analýzy křivky tání provedené post hoc. Jak je vidět na Obr. 2.1 křivka tání zobrazuje, jak se vyvíjí intenzita fluorescence při zvyšování teploty. Pro názornost jsou křivky tání specifických a nespecifických produktů zobrazeny zvlášť. Při samotné analýze jsou tyto dvě křivky sečteny dohromady. Pro lepší vizuální rozlišení různých křivek tání se používá negovaná derivace křivky tání.[1]





Obr. 2.1: Vlevo jsou křivky tání specifického a nespecifických produktů. Vpravo je zobrazena negovaná derivace těchto křivek[21].

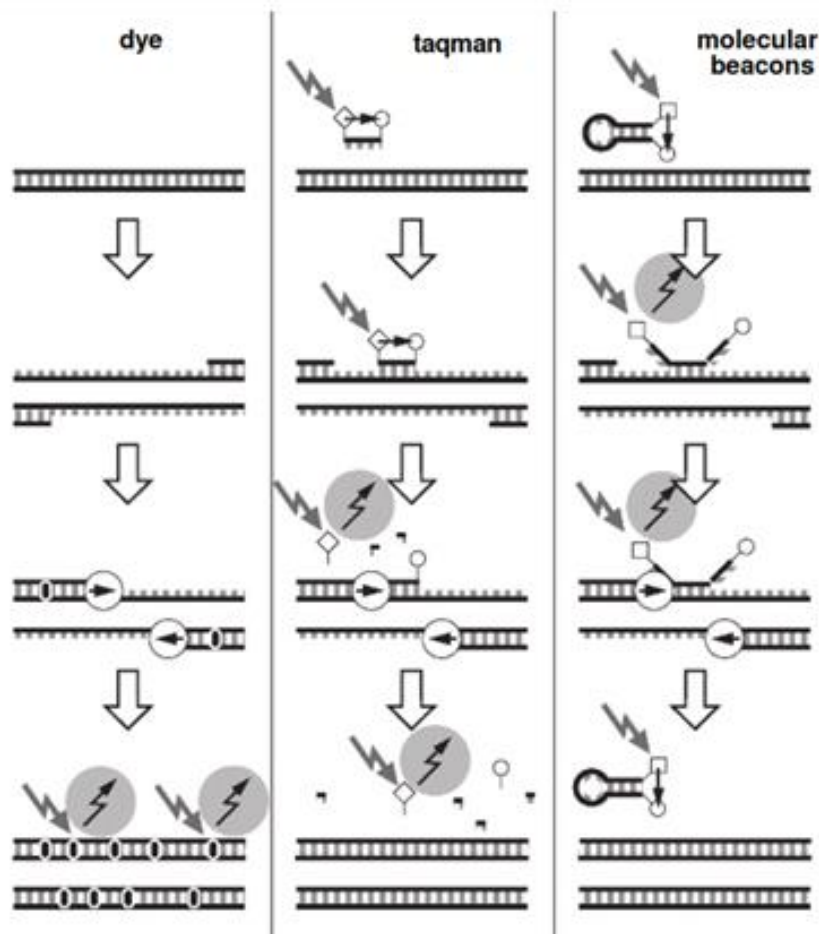
Prvním z interkalačních barviv byl ethidium bromid. Jeho afinita i senzitivita ale nebyla dostatečná pro velmi senzitivní aplikace, a proto dnes nejpoužívanější interkalační barvivo je SYBR Green I, jehož afinita k dsDNA je 100 x větší než afinita ethidium bromidu. Vysoká senzitivita SYBR Green I je zapříčiněna tím, že po navázání na dsDNA se fluorescence zvýší 1000x v porovnání s nenavázaným barvivem.[10]

U specifické detekce dochází ke změně fluorescence jen při amplifikaci cílové sekvence. Specifická detekce využívá oligonukleotidových sond se sekvencí komplementární k templátové DNA. Na sondy je navázán reportér (fluorescenční prvek) a zhášec (quencher). Při excitaci zářením, zhášec utlumí fluorescenci reportéru, tento mechanismus je známý pod pojmem Försterův rezonanční přenos energie (FRET). Poté při procesu elongace DNA dojde k jejich oddálení, což umožní reportéru emitovat fluorescenční záření. Protože oligonukleotidové sondy musejí být komplementární k ssDNA, části sekvence DNA musí být předem známy. Existuje více variant specifické detekce, které se liší použitými oligonukleotidovými sondami.[3]

TaqMan sondy, také známy pod jménem hydrolyzační sondy, jsou oligonukleotidové sondy, které na svém 5' konci mají kovalentně navázaný fluorescenční reportér a zhášec (quencher) na svém 3' konci. TaqMan sondy hybridizované se ssDNA jsou při elongaci rozloženy na reportér, zhášec a jednotlivé nukleotidy prostřednictvím 5' → 3' exonukleázové aktivity Taq DNA polymerázy. Reportér už není v těsné blízkosti quencheru, a proto je zaznamenána vyšší fluorescence vzorku. Fluorescence je měřena každý cyklus a je proporcionálně závislá na množství uvolněných reportérů, a tedy na množství amplifikovaného produktu.[3]

Další systém specifické detekce jsou molecular beacon sondy, specifické oligonukleotidové sondy s navázaným zhášcem a reportérem blízko u sebe. Od TaqMan sond se liší délkou a tvarem sondy. Jak je vidět na Obr. 2.2, molekulární

sondy mají tvar vlásenky, kde jejich střed je komplementární k cílové sekvenci a jejich okraje jsou komplementární sami k sobě. Při nasedání primerů hybridizují ke komplementární ssDNA, což přeruší FRET, poté jsou ve fázi prodlužování řetězců vystrčeny zpátky do roztoku, kde se navrátí do vlásenkového tvaru. Fluorescence je tedy měřena na konci fáze nasedání primerů a je přímo úměrná množství cílové DNA v roztoku.[9]



Obr. 2.2: Systémy specifické a nespecifické detekce[3].

## 2.3 Automatizace real-time PCR (qPCR)

Tradiční PCR byla realizována manuálním přesouváním zkumavek mezi třemi vodními lázněmi ohřátými na teploty používané v amplifikačním cyklu. Tento značně namáhavý proces byl nahrazen v 90. letech vyvinutím prvního termocykleru.[1]

Termocykler je přístroj, který mění a udržuje teplotu v předem naprogramovaných intervalech. V současnosti jsou používány nejvíce termocyklery založené na Peltierově jevu, nebo ohřevu vzduchu. Peltier články používají jako médium pro přenos tepla kovové bloky, ve kterých jsou uloženy vzorky. Nevýhodou

Peltier článků je nerovnoměrná distribuce tepla v ohřívacím bloku držícím vzorky, bohužel při PCR i nepatrná změna teploty může někdy vést k jiným výsledkům. Cyklery založené na ohřevu vzduchu distribuují teplotu rovnoměrněji, a jsou proto používány v aplikacích extrémně senzitivních na teplotu. Moderní cyclery založené na ohřevu vzduchu jsou schopny realizovat 30-40 cyklů v rozmezí 20-30 minut. Peltier články jsou pomalejší kvůli vyšší tepelné kapacitě použitého kovového bloku.[2]

Přístroje pro qPCR jsou rozšířeny o zdroj excitačního světla a detekční systém zaznamenávající fluorescenci, tedy jsou kombinací termocykleru a fluorimetru. Výběr vhodného zdroje excitačního světla je důležitý, protože rozpětí vlnových délek vydávaných tímto zdrojem by mělo spadat do okolí excitačních maxim použitých fluorescenčních barviv. Nejvíce používány jsou halogenové žárovky, které mají široké emisní spektrum od 360 nm do 1000 nm, a xenonová výbojka, která má větší intenzitu světla při výdeji podobného spektra.[1]

Nedílnou součástí qPCR cyklu je také softwarový balíček. Dobrý softwarový balíček by měl obsahovat automatickou analýzu dat, možnost manuálního nastavování proměnných a manuální manipulace s daty, která umožní lepší vyhodnocení výsledků. Důležitým, ale někdy přehlíženým aspektem je možnost exportu dat do Microsoft Excelu, nebo do formátu RDML, který umožňuje otevřít tato data i v softwarovém balíčku od jiného výrobce.[1]

Trendy ve zlepšování cyklerů se pohybují ve třech hlavních směrech. První z nich je zvyšování počtu excitačních filtrů. Filtry umožňují propuštění jen určité vlnové délky, a tedy teoreticky excitaci jen jednoho fluorescenčního barviva. Více filtrů zmírňuje spektrální smíchání při použití více barviv, a umožňuje tak multiplexování esejí.[1]

Druhý trend spočívá ve zvyšování počtu vzorků, který je schopný cykler najednou pojmout. Většinou se udává v počtu jamek, který je blok schopný pojmout. V menších laboratořích se používá většinou 96 jamkový blok, ve větších laboratořích se používá buď 384 jamkový, nebo 1536 jamkový blok.[1]

Třetí spočívá ve zvyšování rychlostí cyklů, sestávající z cyklerů umožňující rychlejší tepelné cykly, vytvořením speciálních master mixů a použitím jiných fluoroforů.[11]

## 2.4 Analýza a interpretace dat

Datovým výstupem z qPCR reakce jsou hodnoty fluorescence jednotlivých jamek a číslo cyklu, popř. časový údaj, kdy byla fluorescence změřena. Na vyhodnocení a získání těchto dat se používá široká paleta metod, které mají každá své výhody a nevýhody. Např. první manipulace s daty probíhá při získávání fluorescenčních hodnot. Pro získání jediné fluorescenční hodnoty, je většinou použito více měření.

Obvykle z měření odstraníme hodnoty s nepřiměřeně velkou odchylkou a následně je zprůměrujeme. Toto nastavení je často zabudované v softwaru a je nepřístupné uživateli.[15]

Pro správné vyhodnocení výsledku je často použita kombinace metod, dávající robustnější výsledky. Nejpoužívanější z těchto metod je amplifikační křivka.[1]

### 2.4.1 Amplifikační křivka

Amplifikační křivka ukazuje závislost fluorescence na počtu proběhlých cyklů. Před konstrukcí křivky, je potřeba odečíst fluorescenční pozadí a zmírnit šum.

Šum může být způsoben náhodnými jevy, samotným měřícím systémem (kvantifikace signálu) a použitím nevhodných reaktantů. Nejčastěji se pro odstranění šumu používá proložení dat křivkou. Pro konstrukci křivky se nejčastěji používá nelineární regrese.[16]

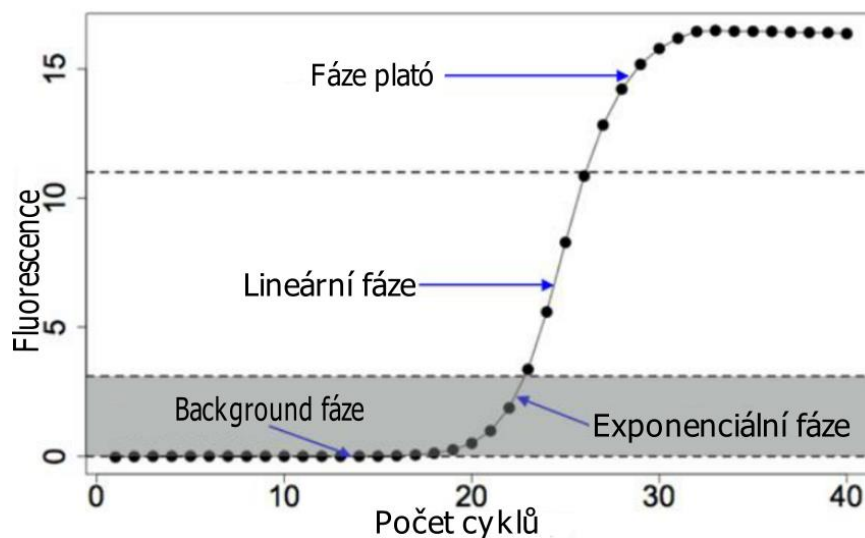
Odečtení fluorescenčního pozadí způsobí, že všechny amplifikační křivky budou začínat od stejného počátečního bodu (nuly) a zároveň je možné odstranit šum způsobený změnou ve fluorescenčním pozadí. Další výhoda plyne z odstranění mezery mezi různými barvivy při realizaci multiplex qPCR. Pro větší komfort uživatele, je většina softwarů vybavena automatickým odečtením pozadí. K odečtení se může použít odečtení minimální naměřené hodnoty, nebo odečtení hodnot získaných z negativní kontroly.[16]

Výsledná amplifikační křivka se podle charakteru růstu rozděluje na čtyři fáze: background, exponenciální, lineární a plateau fázi. V tzv. „**background**“ fázi je nárůst fluorescence nedetekovatelný. Technologie používané v qPCR mají svoje limity, kde jedním z nich je senzitivita. V této fázi je nárůst fluorescence pod senzitivitu přístroje, a proto se podobá šumu. I přes nemožnost detekce probíhá v této fázi exponenciální růst.[17]

**Exponenciální fáze** nastává v momentě, kdy kvůli dostatečnému nahromadění cílového produktu, signál započne exponenciálně růst. V této fázi je křivka velmi podobná ideální amplifikační křivce, která exponenciálně roste po dobu všech cyklů PCR. Tato fáze se proto používá pro stanovení templátové DNA.[1]

Exponenciální růst se začne pomalu zpomalovat a reakce přejde do **lineární fáze**. Amplifikační křivka se začne méně a méně podobat ideální křivce a efektivita reakce se postupně snižuje.[17]

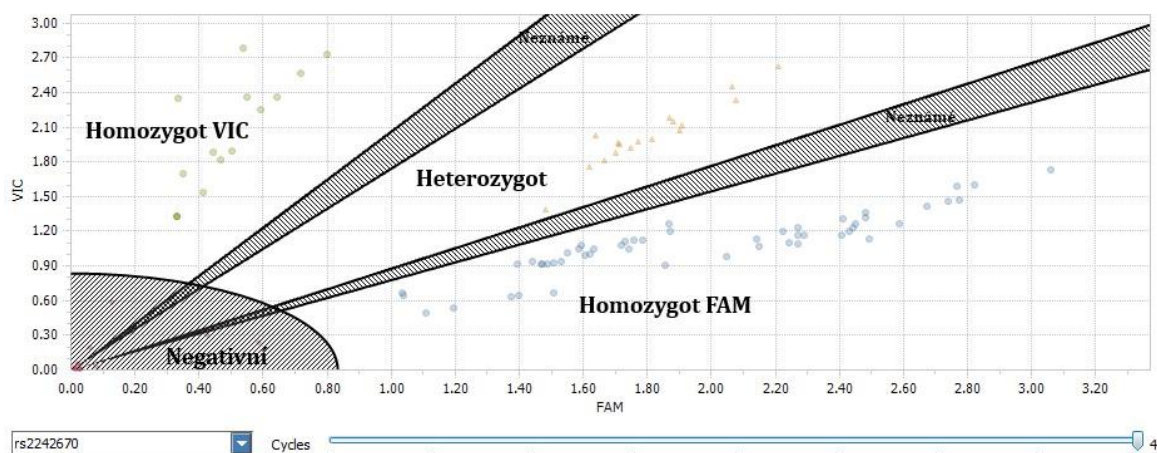
Lineární fáze poté přejde v „**plateau**“. Nárůst fluorescence se zpomaluje, až dojde k úplné „únavě“ reakce a fluorescence zůstává stejná. Zpomalování reakce je způsobeno např. nahromaděním koncových produktů (konkurence koncových produktů s primery o vazbu na templát) a spotřebováváním potřebných reaktantů.[17]



Obr. 2.1: Čtyři fáze PCR vyznačené na amplifikační křivce[17].

## 2.4.2 Genotypizace pomocí end-point analýzy

Při end-point analýze, jak název napovídá, se ke klasifikaci genotypů používají jen poslední změřené hodnoty fluorescence. Genotypizace pomocí qPCR používá zpravidla dvě rozdílné sondy, každou značenou jiným barvivem. Klasifikace je provedena na základě end-point grafu. Jednotlivé body v grafu jsou sestrojeny pomocí hodnot fluorescence při posledním cyklu qPCR. Na ose x jsou konečné hodnoty fluorescence barviva 1, na ose y pak hodnoty fluorescence barviva 2 ze stejné jamky.



Obr. 2.3: End-point analýza pro vyhodnocení genotypizačních dat. Ukázka z programu LightCycler® 96 SW 1.1.

Samotná genotypizace pak sestává ve vyhodnocení poměrů barviv a klasifikaci rozdělí bodů, patřících do jednoho genotypu. Jak je vidět na obrázku Obr. 2.3, v ideálním případě jsou shluky dobře rozlišitelné a je možné je rozdělit pomocí

výsečí na různé oblasti. Rozřazení funguje na manuální bázi. Např. software od fy Roche se snaží udělat počáteční odhad, ale je nezbytné, aby uživatel zhodnotil počáteční odhad a upravil ho podle svých potřeb a znalostí. S výsečí může uživatel manipulovat intuitivně myší, kde je možné změnit velikost výseče, ale už ne její tvar.

Zajímavostí je, že software od fy Roche umožňuje sestavení end-point grafu v kterémkoliv cyklu, ale klasifikace s výsečí je možná jen v posledním cyklu. Pravděpodobný důvod je zabránění nedostatečně znalým uživatelům najít cyklu, ve kterém je výsledný graf lehce klasifikovatelný, ale ve skutečnosti nemá dostatečnou výpovědní hodnotu. Tato metoda tedy nebere v potaz předchozí hodnoty fluorescence, nebo sklony amplifikačních křivek.[17]

## 3 REAL-TIME PCR DATA MARKUP LANGUAGE

Data získaná z qPCR termocykleru jsou ukládána a předávána dále ve formátu, který je unikátní pro každého výrobce. Situaci neusnadňuje to, že často i každý přístroj od stejného výrobce musí mít dedikovaný software. Nejenže tyto softwary často nedokáží splnit komplikované požadavky na analýzu dat, požadovanou vědeckými pracovníky, ale také neumožňují jednoduchý přenos těchto dat do jiného softwaru. To vedlo k vytvoření různých nezávislých programů na analýzu dat. Tyto programy často uměly importovat data jen v určitém formátu, což často vedlo k zadávání dat ručně při malém počtu měření. Situaci dále zkomplikovala velká variace formátů ukládání dat, což také znemožnilo vytvoření knihoven s experimentálními daty, což např. u DNA čipů bylo již řadu let praxí.[12]

Tyto nemalé komplikace volaly po vytvoření všeobecně uznávaného formátu ukládání dat. Otevřená diskuze, trvající dva roky, započala v roce 2005 a aktivním zapojením různých společností (Roche, Applied Biosystems, Biogazelle) vyústila v konsorcium, které brzy přišlo s prvním všeobecně uznávaným standardem se jménem RDML (Real-time PCR Data Markup Language). Kvůli předem zmíněným požadavkům je RDML založený na jazyku XML, specializujícím se na sdílení dat mezi informačními systémy.[12]

### 3.1 XML

eXtensible Markup Language (XML) je v současnosti celosvětově rozšířený standard pro sdílení informací. I přes to že má XML v názvu jazyk, je důležité poznamenat, že XML sám o sobě jazyk není, spíše je to výčet pravidel pro vytváření jazyků. Protože informační technologie významně zjednodušili sdílení velkého množství dat, vznikla potřeba jednotného jazyka ukládání a následného sdílení informací. To se promítlo např. ve společnostech chtějících přesunout své databáze na internet, ale jejich starší nekompatibilní systémy tuto situaci značně zkomplikovaly. XML je navržen tak, aby značně zjednodušil tyto komplikace a mohl být používán v rozličných aplikacích. Jeho pružnost a užitečnost jsou dokázány faktem, že je používán od bankovníctví po správu sítí.[13]

#### 3.1.1 Základní prvky XML

Pro pochopení RDML je nezbytné znát základní pravidla XML, na kterém je formát RDML založen. Základní prvky XML jsou tagy, elementy, atributy. Tyto základní elementy jsou ukázány v praxi v následujícím příkladu:

```

<tag> Tato věta je obsah elementu</tag>
<jmeno_elementu> Obsah elementu </jmeno_elementu>
<jmeno_elementu název_atributu=" obsah atributu"> </jmeno_elementu>

```

Každý celý řádek předchozího kódu se označuje jako element. Tyto elementy jsou vyčleněny počátečními a ukončujícími tagy, kde počáteční tag elementu je vyčleněn znaky <, > a ukončující tag pomocí lomítka uvnitř těchto znaků. Atribut je přiřazení určité hodnoty elementu. Způsob zápisu atributu je možné vidět na posledním řádku. [13]

Zbývá popsat poslední klíčový prvek XML, a to je vnořování elementů. Vnořený element je vlastně podjednotkou nadřazeného elementu. Elementy můžou obsahovat neomezeně mnoho elementů, které mohou mít také vnořený neomezený počet elementů. Názornou ukázkou zajistí příklad kódu vyňatého z RDML souboru[13]:

```

<step>
  <nr>1</nr>
  <temperature>
    <temperature>95</temperature>
    <duration>600</duration>
    <ramp>4.4</ramp>
  </temperature>
</step>

```

Užitečnost XML tedy spočívá v seskupení informací mající stejné jméno/tagy, a tak vlastně napodobuje přirozený proces třídění či ukládání informací.

## 3.2 Struktura RDML

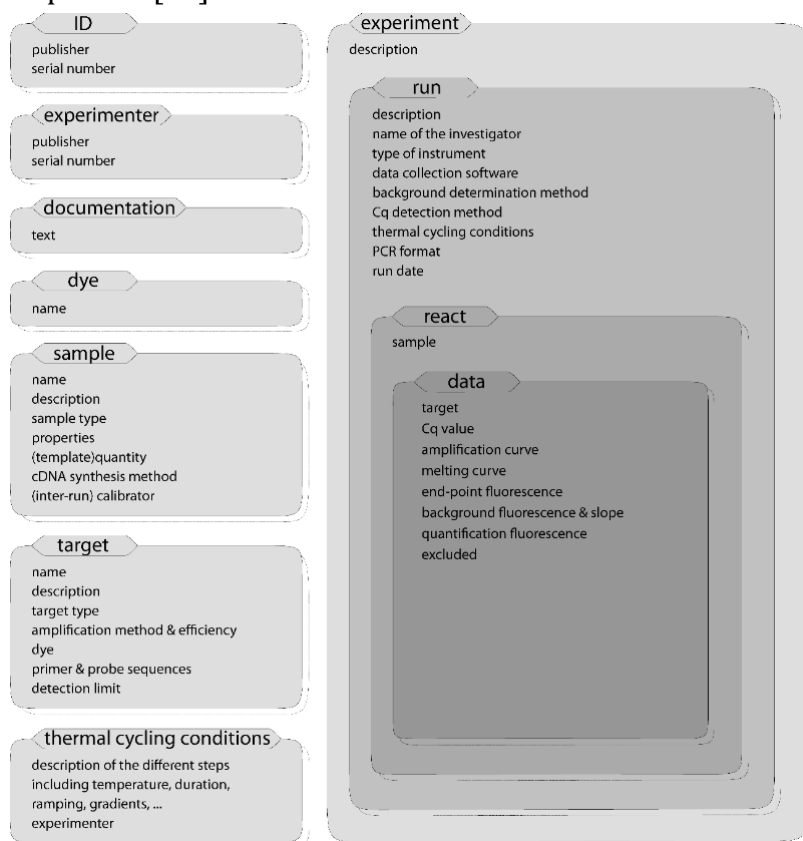
RDML bylo vytvořeno pro uložení dat z více experimentů do jednoho RDML souboru. Jak je vidět ze zjednodušeného schéma na obrázku Obr. 3.1, RDML sestává ze 7 hlavních elementů: ID, experimentator, documentation, dye, sample, target, thermal cycling conditions a experiment. Kromě těchto elementů je připojen i čas vytvoření souboru. Plné nezjednodušené schéma je přiloženo jako příloha.[14]

Element „ID“ slouží k jednoznačné identifikaci dokumentu a autorů. Všechny relevantní informace o použitých vzorcích je možné popsat v elementu „sample“. Element „target“ slouží pro popsání primerů, genů a informací o PCR reakci. Samotné popsání nastavení amplifikačního cyklu PCR probíhá v „thermal cycling



conditions“, zpravidla obsahuje pět „step“ elementů popisujících teplotu, čas a rychlost změny teploty každého kroku. Na velký počet jmenovaných prvků je odkazováno v hlavním bloku popisující experiment, je proto důležité, aby bylo dodrženo správné pořadí schématu. Formát RDML může obsahovat jeden nebo více z jmenovaných elementů, co více všechny elementy jsou dobrovolné a nemusí být vyplněny. Z těchto důvodů je formát RDML velmi flexibilní.[14]

Hlavní část RDML, kde je uložena většina informací o výsledcích proběhlého qPCR, je v elementu experiment. Jak je vidět ze zjednodušeného schématu, experiment může obsahovat více elementů „run“, tedy dat z různých qPCR reakcí, a co více, dokonce i dat z různých termocyklerů. V samotném elementu „run“ se popisuje přístroj i software použitý při kolekci dat, reference na elementy „thermal cycling conditions“, „experimenter“, „documentation“ a popř. další podrobnosti o proběhlé qPCR. Element „react“ pak obsahuje číslo jamky (číslované po řádcích) a podjednotku „data“ do které se pak ukládají naměřené fluorescenční hodnoty, aktuální teplota a číslo cyklu. Tyto hodnoty jsou uloženy v podjednotce „adp“. To už ale není možné vidět na zjednodušeném schématu. Plné schéma je kvůli své velikosti přiloženo jako příloha.[14]



Obr. 3.1: Zjednodušené schéma RDML[14].

## 4 REALIZACE PROGRAMU

Proces realizace programu je možné rozdělit do tří částí

1. Analýza požadavků
2. Konstrukce amplifikační křivky
3. Grafické rozhraní

V následujících částech bude každá část popsána podrobněji.

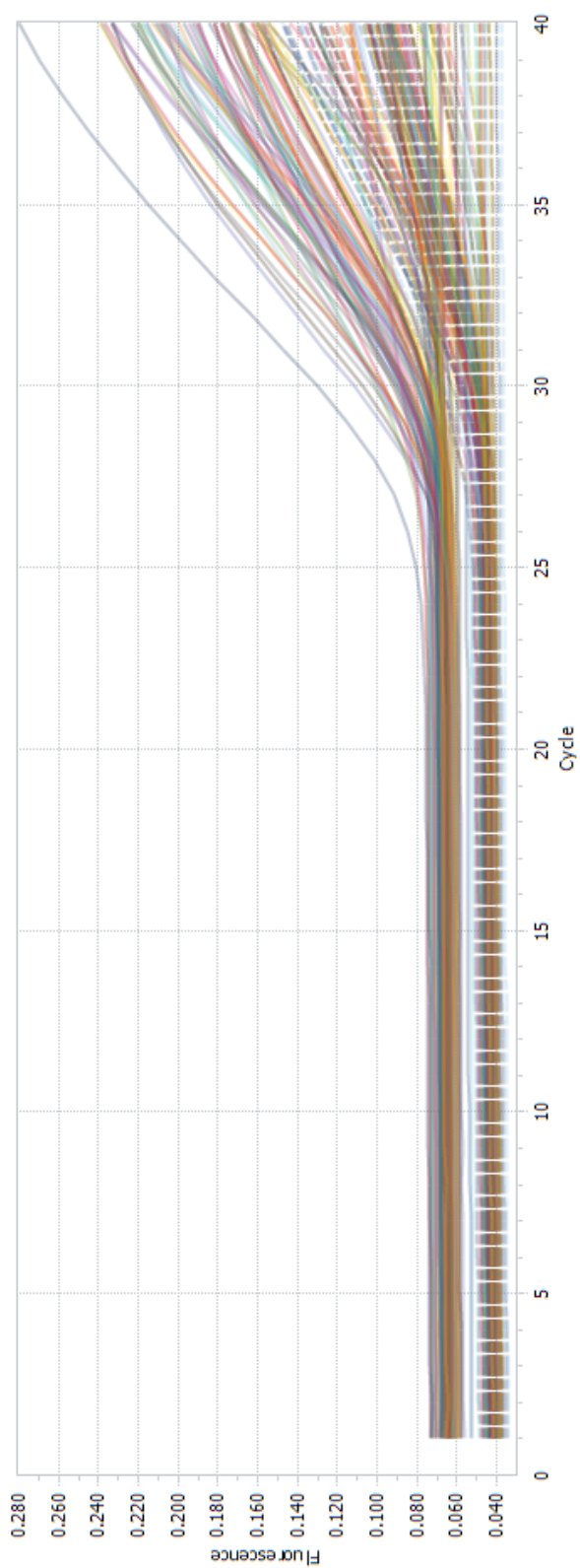
### 4.1 Analýza požadavků

K analýze genotypizačních dat získaných pomocí fluorescenčních sond se používá end-point analýza. Nevýhoda tohoto přístupu se nejvíce projevuje v reakcích, ve kterých dochází k amplifikaci nescifických produktů (primer-dimery, nasedání primerů k částečně homologním sekvencím). V takových případech je často nemožné rozlišit, zda reakce proběhla správně jen fluorescenčních hodnot z posledního cyklu. Ukázka amplifikačních křivek z reakcí, kde dochází k nescifické amplifikaci je možné vidět na Obr. 4.1. Data jsou získaná z reálné PCR proběhlé v přístroji LightCycler® 96 a vyhodnocena jeho dedikovaným softwarem. Jak je vidět některé křivky nemají viditelnou exponenciální fázi, strmost jejich růstu je relativně nízká a lineární. Toto není problém pro křivky, které začnou stoupat relativně brzo. Je možné je sledovat po vícero cyklů a s velkou pravděpodobností vystoupají hodnoty fluorescence nad hranici pochyb o správnosti reakce.

Žádný ze zmíněných problémů ale nemusí být vidět v end-point analýze. Problémem je, že end-point analýza bývá jediná dostupná analýza v dedikovaných softwarech. Uživatel by proto musel procházet velké množství amplifikačních křivek manuálně a jednotlivě rozhodovat o genotypu, což je značně časově náročné.

V dedikovaném softwaru od Roche také není možnost manuálního volení genotypů. Jedinou možností uživatele je vyřadit jamku z analýzy úplně. Takže pokud chce uživatel provést manuální genotypizaci, je nucen dělat si poznámky vedle.

Tyto komplikace byly hlavním impulsem pro vytvoření tohoto GUI.



Obr. 4.1: Amplifikační křivky při neideálním průběhu amplifikace.  
Ukázka z programu LightCycler® 96 SW 1.1.

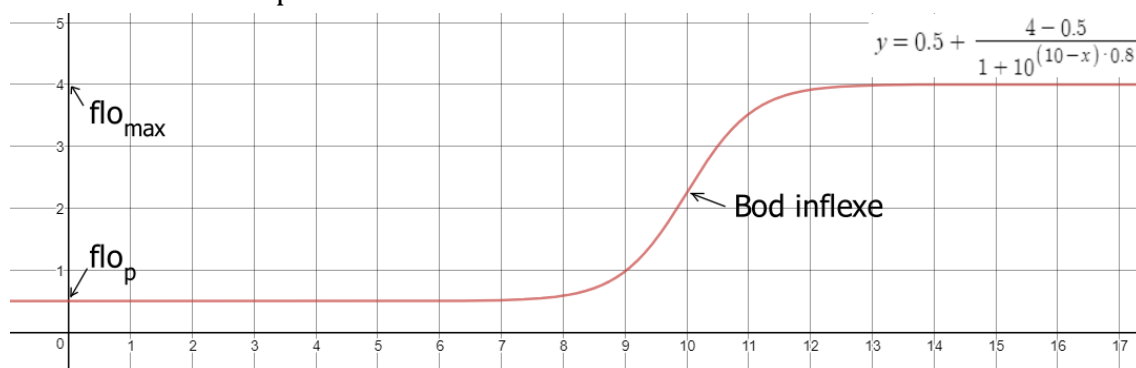
### 4.1.1 Konstrukce amplifikační křivky

Důležitou částí GUI je zpracování naměřených hodnot získaných z real-time PCR. Pro vyhodnocení průběhu reakce je vhodné zkonstruovat amplifikační křivky. Pro aproximaci amplifikační křivky se ukázala jako nejvýhodnější aproximace pomocí logistické funkce s 4 parametry. Logistická funkce má, stejně jako amplifikační křivka představující správně proběhlou amplifikaci, tvar podobný písmenu S. Funkce je definovaná jako

$$flo = flo_p + \frac{flo_{max} - flo_p}{1 + 10^{(x_{50} - x) \cdot s}} \quad (4.1)$$

kde:

- $flo$  = hodnota fluorescence v cyklu  $x$
- $flo_p$  = minimální naměřená fluorescence
- $flo_{max}$  = maximální naměřená fluorescence
- $X_{50}$  = cyklus ve kterém nastane bod inflexe
- $s$  = strmost stoupání



Obr. 4.2: Logistická křivka a její parametry.

Minimální naměřená fluorescence je velmi často rozdílná v každé jamce. Nejčastější příčinou jsou nepřesnosti způsobené pipetováním. To vede k nepřehlednému grafu, kde se amplifikační křivky různě kříží.

Při typickém počtu 96 jamek na jedno qPCR a dvěma barvivami v každé jamce, by výsledný graf měl obsahovat 192 přímk. Takovýto graf by byl nepřehledný. Pro zlepšení přehlednosti grafu lze využít toho, že pro vyhodnocení genotypů nese největší výpovědní hodnotu změna fluorescence v průběhu reakce. Toho je možné využít pro posunutí křivek do společného počátku, v tomto případě nulové hodnoty fluorescence v prvním cyklu real-time PCR. Rovnice tedy dostane tvar:

$$flo = \frac{flo_{max} - flo_p}{1 + 10^{(x_{50} - x) \cdot s}} \quad (4.2)$$

Dalšími výhodami je, že proložením dat křivkou odpadne nutnost odstranění šumu, nebo vybočujících hodnot. Oba problémy se nejvíce vyskytují v background fázi amplifikačních křivek.

## 4.2 Grafické uživatelské rozhraní – GUI

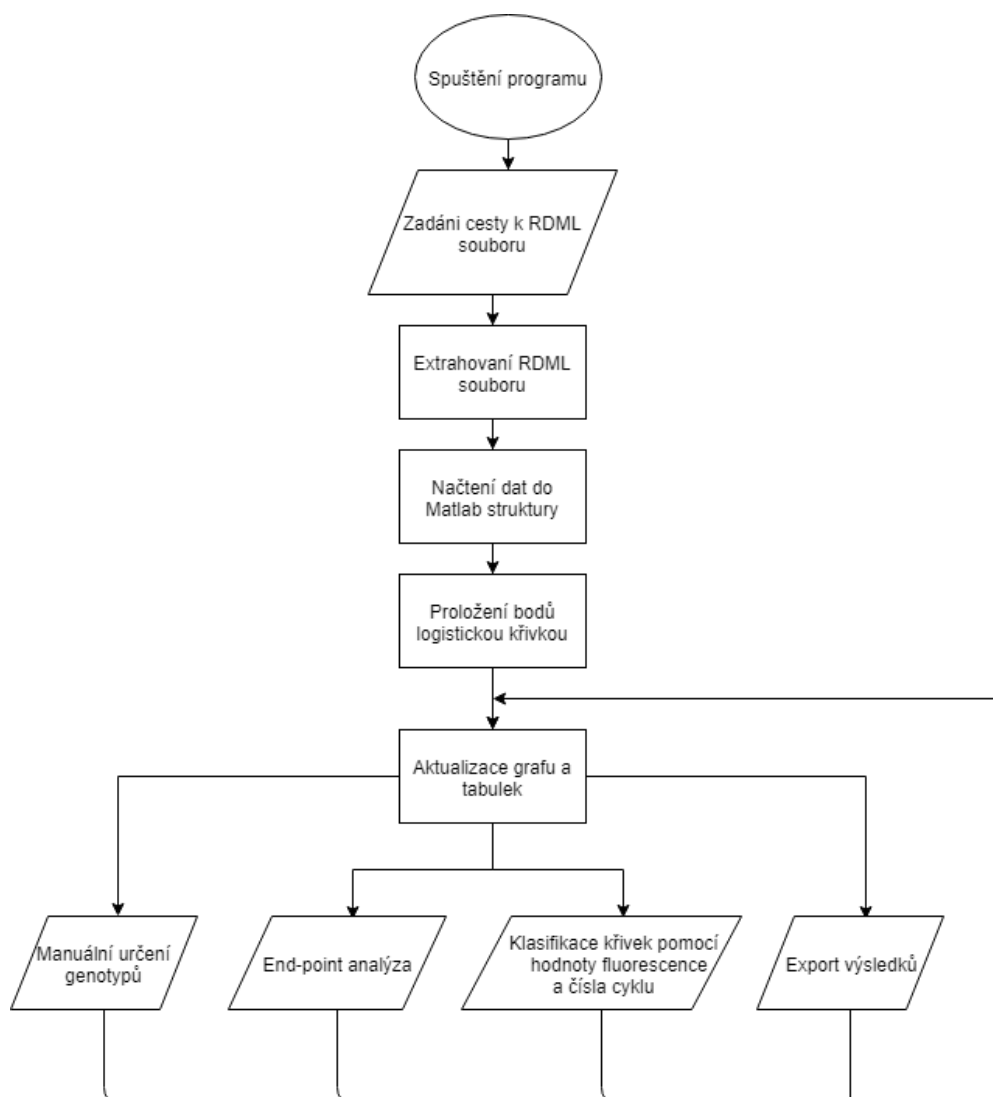
Software je zrealizovaný ve vývojovém prostředí Matlab. Matlab poskytuje široké spektrum funkcí pro analýzu a vizualizaci dat, jak přímo obsažených v instalačním balíčku, tak i jinými uživateli naprogramovaných funkcí stažitelných z oficiálního webu Matlabu. Software byl naprogramován s pomocí GUIDE, vestavěného prostředí pro tvorbu grafických rozhraní. Všechny grafické prvky jsou umístěny relativně, jejich velikost a umístění se tedy bude automaticky měnit s rozlišením obrazovky a velikostí okna. Rozložení GUI je inspirováno programem LightCycler® 96.

GUI umožňuje uživateli volat funkce jednoduše, pomocí stisku myši. Uživatel tedy nemusí znát názvy funkcí a odpadá nutnost je volat pomocí příkazové řádky. Místo toho se při kliku na předem definované interaktivní pole zavolá předem definovaná funkce neboli „callback“ vlastní danému interaktivnímu poli. Všechny funkce callback byly nastaveny jako nepřerušitelné, takže další callback je možné spustit až po dokončení předcházejícího callbacku.



Obr. 4.3: Příklad rozložení GUI používaného v LightCycler® 96 SW 1.1.

Charakter callback funkcí zajišťuje, že všechny proměnné, které jsou v něm vytvořeny, budou na jeho konci smazány. Pro ukládání a předávání dat mezi funkcemi slouží struktura „handles“. Handles obsahuje ukazatele na všechny grafické prvky vytvořené v GUIDE a současně je možné ji použít pro ukládání proměnných.



Obr. 4.4: Zjednodušený vývojový diagram.

GUI se spouští skriptem GUI.m, ve kterém jsou i všechny callbacky a většina hlavních funkcí. GUI obsahuje dva panely, které budou následně popsány.

### 4.2.1 Popis GUI

GUI je rozděleno na dva panely, mezi kterými může uživatel libovolně přepínat. Dva panely byly vytvořeny kvůli lepší správě grafických prvků. Přepínání mezi panely je pak zrealizováno přepínáním viditelnosti panelů uživateli. První panel je určen pro klasifikaci přímků, druhý pro samotnou end-point analýzu a případné manuální určení genotypu. Na Obr. 4.5 lze vidět úvodní panel GUI po načtení RDML souboru a provedení analýzy. Nyní budou krátce popsány oblasti jednotlivých panelů.

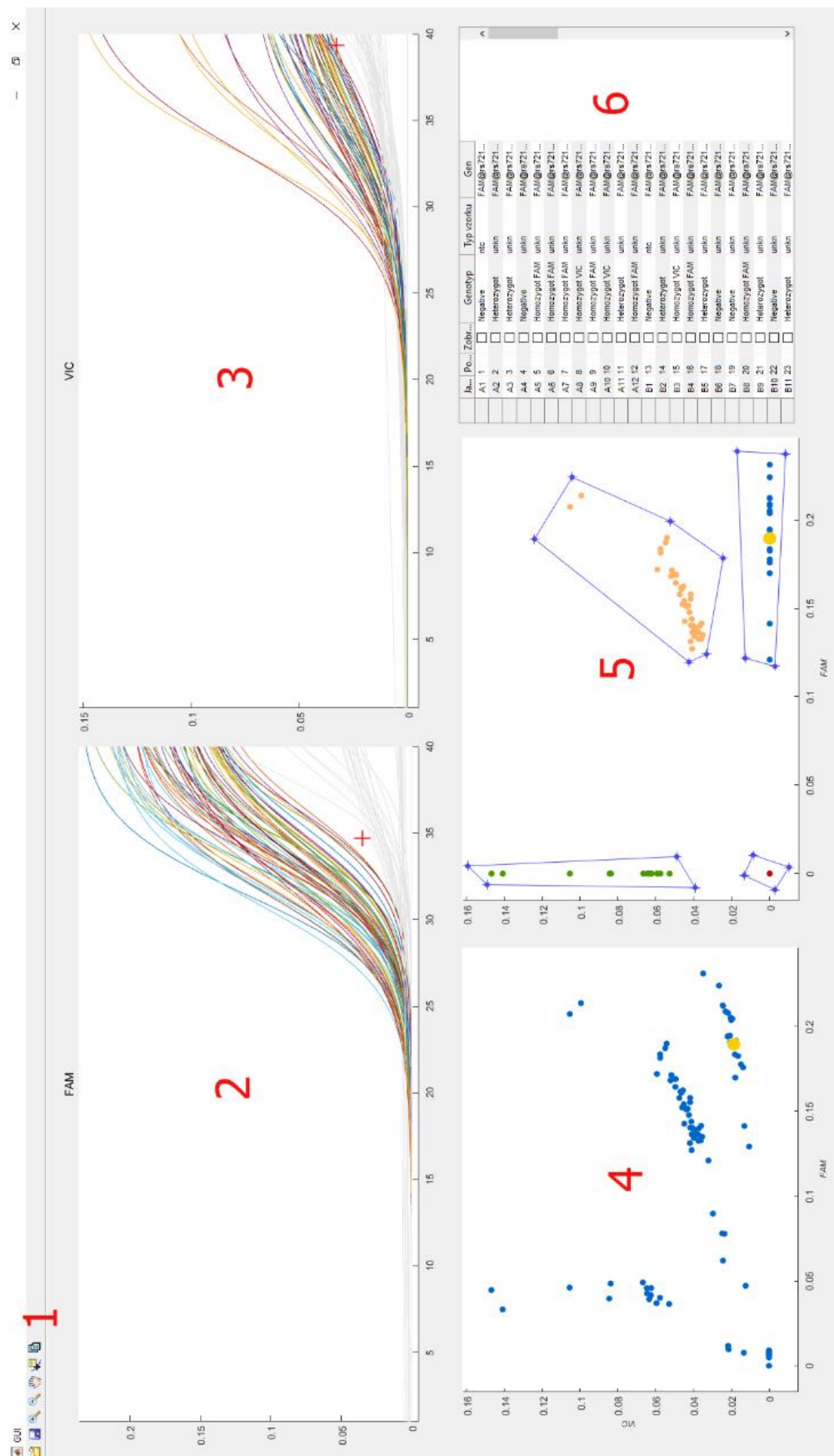
Oblast vlevo od č. 1 obsahuje lištu nástrojů. Prvním nástrojem je ikona pro zadání cesty k RDML souboru. Po zadání cesty se inicializují tabulky a vykreslí grafy. Dále obsahuje ikonu pro export výsledků do textového souboru, jehož struktura je

podobná tabulce v oblasti č. 6. Pro práci s grafy nesmí chybět klasické nástroje typu přiblížení, oddálení, posunutí a data kurzor. Tyto nástroje fungují ve všech objektech typu „Axes“. Data kurzor po kliknutí na křivku, či bod v bodovém grafu zobrazí informace o hodnotách os v daném místě a jamku kterému bod, či křivka náleží. Dále zobrazí bod náležící dané jamce v end-point grafech číslo 4 a 5. To je zvláště užitečné, pokud klasifikace v grafech amplifikačních křivek změnila pozici koncových bodů. Poslední ikona přepne z panelu pro klasifikaci na panel pro genotypizaci a naopak.

Oblast č. 2 a 3 je určena pro třídění křivek pomocí hodnoty fluorescence v určitém cyklu. Třídění je rozděleno do dvou grafů. Každý graf zobrazuje naměřené fluorescenční hodnoty jiného barviva. O které barvivo se jedná je napsáno v titulku grafu. V tomto případě jde o barvivo FAM (fluorescein) vlevo a o barvivo VIC vpravo. Více o procesu klasifikace je v kapitole 4.2.2.

Grafy č. 4 a 5 obsahují grafy koncových bodů. Graf č. 4 vždy zobrazuje původní rozložení koncových bodů, bez ohledu na další klasifikační metody. Oproti tomu graf č. 5 zobrazuje graf po genotypizaci a po provedení klasifikace v grafech č. 2 a 3.

Tabulka v oblasti č. 6 obsahuje informace o jednotlivých jamkách. Konkrétně obsahuje číslo jamky (číslováno po řádcích), označení pozice jamky, výsledný genotyp, pokud je to známo tak i typ vzorku (negativní kontrola – ntc, neznámý – unkn a další) a označení genu jehož alely zkoumáme. Dále tabulka umožňuje zobrazení jedné a více jamek po zaškrtnutí příslušného pole ve sloupci zobrazit.



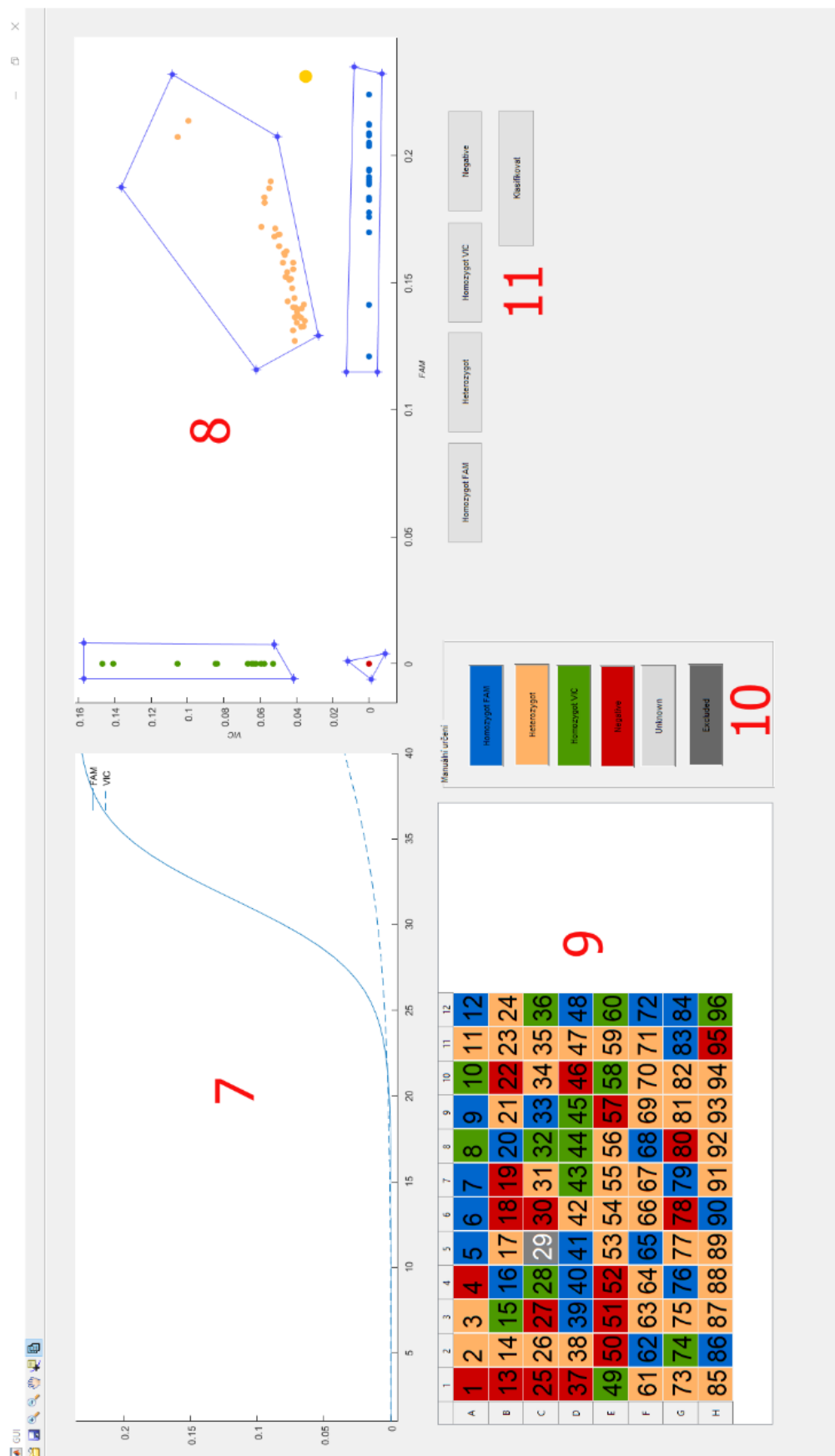
Obr. 4.5: Panel pro klasifikaci křivek po provedení analýzy.



Na obrázku Obr. 4.6 je možné vidět panel, který slouží pro manuální určení genotypu a hromadnou klasifikaci pomocí end-point analýzy.

End-point analýza probíhá v grafu pod číslem 8. Graf zobrazuje stav koncových bodů po klasifikaci pomocí fluorescence a čísla cyklu. Klasifikace do genotypů probíhá pomocí tlačítek v oblasti 11. Tlačítka nesou názvy genotypů, a po jejich stlačení je uživateli umožněno definovat oblast, která bude náležet danému genotypu. Oblasti jsou definované pomocí polygonů. Uživatel definuje body polygonu kliknutím myši v oblasti grafu. Takto vytvořené body jsou mezi sebou automaticky spojeny. Oblasti ležící vně polygonu jsou vyhodnoceny jako neznámý genotyp. Graf v oblasti 5 je kopií tohoto grafu.

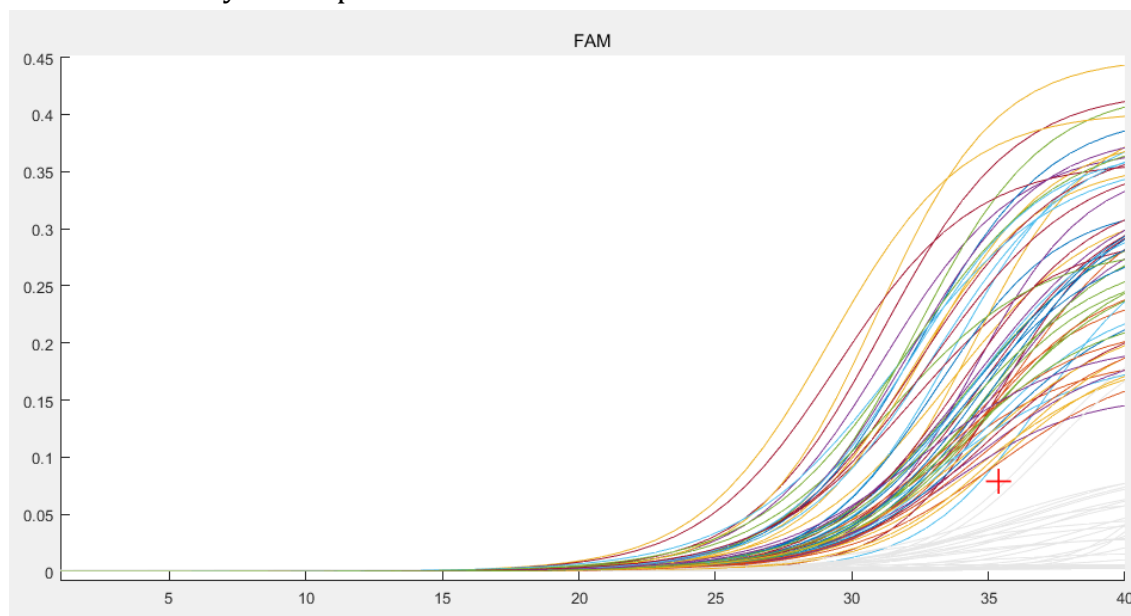
V tabulce v oblasti č. 9 probíhá manuální genotypizace, a také slouží jako přehledná tabulka výsledků. Rozložení čísel v tabulce je stejné jako v qPCR bloku. Legenda barevného kódování je v oblasti č. 10. Manuální určení probíhá klikem na jamku. Po kliku se zobrazí amplifikační křivky náležící aktivní jamce v oblasti číslo 7 navíc se ještě zvýrazní náležitý bod ve všech end-point grafech GUI. Uživatel tedy má před sebou k dispozici všechny dostupné informace. Zvolení genotypu probíhá vpravo pomocí tlačítek v panelu manuálního určení. Manuální určení má vyšší prioritu než end-point klasifikace, takže genotyp jde znovu změnit zase jen manuálně. Tabulka zvládá i vybrání více jamek najednou. Při následném manuálním zvolení genotypu je změněn genotyp poslední označené jamky z výběru.



Obr. 4.6: Panel pro genotypizaci a manuální určení.

### 4.2.2 Klasifikaci přímek podle fluorescence a čísla cyklu

Pro odstranění problémových křivek byl vytvořen červený posuvný kříž, který je možno vidět na Obr. 4.7 v pravém dolním rohu. Kříž není nutno posouvat zadáním hodnot, ale je možné ho rovnou potáhnout myší. Po posunutí kříže uživatelem, jsou za vyřazené křivky považovány ty, kde fluorescence je menší než fluorescence křivky v daném cyklu. Pro odlišení vyřazených křivek je použito světle šedé barvy. Kvůli rozdílným vlastnostem použitých barviv, je nutné analyzovat a vytrdit nevhodné křivky zvlášť pro každou barvu.



Obr. 4.7 Klasifikace podle čísla cyklu a fluorescence.

Vytríděné křivky by nebylo vhodné vyřadit z analýzy, protože s největší pravděpodobností u nich proběhla nespecifická amplifikace. Reakce je tedy možné považovat za negativní (není přítomna sekvence komplementární k TaqMan sondě). Samotné vyřazení křivky je realizováno vynulováním výsledné fluorescence v end-point grafu na 0, což změní výsledné rozpořádání bodů v end-point grafu, a je tak možné lépe rozlišit jednotlivé genotypy.

## 5 DISKUZE VÝSLEDKŮ

Vytvořený software je schopen analyzovat data z jakéhokoliv real-time cycleru uložená ve formátu RDML. Po importu dat zobrazí software amplifikační křivky všech vzorků ve vybraných fluorescenčních kanálech, tedy pro vybrané fluorescenční barvy použité pro označení použitých TaqMan sond. Vedle amplifikačních křivek zobrazí i bodový graf znázorňující fluorescence každého vzorku v obou fluorescenčních kanálech v posledním cyklu PCR. Následně je možno pohyblivým křížem označit hraniční cyklus PCR a hraniční fluorescenci pro eliminaci fluorescenčních signálů, které považujeme za falešně pozitivní vzhledem k pozdnímu a mírnému vzestupu hodnot fluorescence. Jde o ty případy, kdy je z tvaru křivek zřejmé, že jedna ze sond má tendenci reagovat i nespecificky a zvyšovat falešně pozitivně svůj signál. Výsledkem je bodový graf „očištěný“ od těchto dubiozních signálů a snáze interpretovatelný automatickým přiřazením daného genotypu. Bez této možnosti byla obsluha real-time PCR často nucena odečítat výsledky genotypizace z amplifikačních křivek jednotlivě, vzorek po vzorku.

Program nabízí jasné rozhraní umožňující intuitivní ovládání, a to především ve srovnání s jinými obdobnými programy. Co dává prostor pro zlepšení a další rozvoj, je absence možnosti nastavit minimální „slope – strmost“ vzestupu amplifikační křivky jako kritérium pro odlišení skutečně pozitivních od falešně pozitivních vzorků. Bylo by možno snáze eliminovat jako falešně pozitivní signály, které sice v závěru PCR vykazují vysokou hodnotu fluorescence, ale vzestup hodnot fluorescence nevykazuje charakteristický tvar amplifikační PCR křivky.

Dodaný software umožňuje export výsledků genotypizace ve formátu snadno zpracovatelném v běžných tabulkových procesorech jako je třeba MS Excel.

## ZÁVĚR

Cílem práce bylo seznámit se s problematikou real-time PCR a navrhnout software, který by byl schopný klasifikovat data získaná z real-time PCR genotypizačních esejí i jinak než na pomoci fluorescenčních end-point hodnot. Metod pro klasifikaci dat bylo vyzkoušeno více. Nejlepší se ukázala grafická klasifikace pomocí hodnot fluorescence a čísla cyklu. Klasifikace funguje velmi dobře na vyřazení křivek s pozdní neideální amplifikací, které byly identifikovány konzultantem jako hlavní problém. Nakonec byl výsledný software zaslán konzultantovi, jenž s úspěchem ověřil jeho funkci na větším množství dat.

# LITERATURA

- [1] DORAK, M. Tevfik. Real-time PCR. New York: Taylor & Francis Group, c2006. ISBN 978-041-5377-348
- [2] MCPHERSON, M. J. a S. G. MÜLLER. *PCR*. 2nd ed. New York: Taylor & Francis, c2006. Basics (Routledge (Firm)). ISBN 978-0415355476.
- [3] MÜLHARDT, Cornel. *Molecular biology and genomics*. Boston: Elsevier Academic Press, c2007. ISBN 9780120885466.
- [4] HUE-ROYE, K. a S. VEGE. Principles of PCR-based assays. In: Imunohematology. 4. Washington DC: American Red Cross, 2008, s. 170-175. ISSN 0894-203X.
- [5] LIFE TECHNOLOGIES. *Real-time PCR handbook* [online]. [cit. 2017-12-07]. Dostupné z: <https://www.thermofisher.com/fi/en/home/life-science/pcr/real-time-pcr/qpcr-education/real-time-pcr-handbook.html>
- [6] S. BERNARD, Philip a Carl T. WITTEWER. Real-Time PCR Technology for Cancer Diagnostics [online]. 2002 [cit. 2017-12-24]. Dostupné z: <http://clinchem.aaccjnl.org/content/48/8/1178.long>
- [7] FEARON, Margaret. *The laboratory diagnosis of HIV infections* [online]. 2005 [cit. 2017-12-24]. Dostupné z: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2095005/>
- [8] M. POWLEDGE, Tabitha. The polymerase chain reaction [online]. 2004 [cit. 2017-12-24]. Dostupné z: <http://www.physiology.org/doi/full/10.1152/advan.00002.2004>
- [9] Hybridizační metody. Sdfs [online]. [cit. 2017-12-24]. Dostupné z: <http://labguide.cz/metody/hybridizacni-metody/>
- [10] KALTENBOECK, Bernhard a Chengming WANG. Advances in Real-Time PCR: Application to Clinical Laboratory Diagnostics [online]. Elsevier, 2005, 2005, s. 219-259 [cit. 2017-12-25]. Advances in Clinical Chemistry. DOI: 10.1016/S0065-2423(05)40006-2. ISBN 9780120103409. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0065242305400062>
- [11] BUSTIN, Stephen A. How to speed up the polymerase chain reaction. Biomolecular Detection and Quantification [online]. 2017, 12, 10-14 [cit. 2017-12-25]. DOI: 10.1016/j.bdq.2017.05.002. ISSN 22147535. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S2214753517301985>
- [12] LEFEVER, S., J. HELLEMANS, F. PATTYN, D. R. PRZYBYLSKI, C. TAYLOR, R. GEURTS, A. UNTERGASSER a J. VANDESOMPELE. RDML: structured language and reporting guidelines for real-time quantitative PCR data. Nucleic Acids Research [online]. 2009, 37(7), 2065-2069 [cit. 2017-12-26]. DOI: 10.1093/nar/gkp056. ISSN 0305-1048. Dostupné z: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkp056>

- [13] RAY, Erik T. Learning XML. Cambridge, Mass.: O'Reilly, 2001. ISBN 0-59600-046-4.
- [14] Introduction to RDML v1.2 [online]. RDML Consortium, 2008 [cit. 2017-12-26]. Dostupné z: <http://www.rdml.org/files.php>
- [15] LOGAN, Julie, Kirstin EDWARDS a Nick SAUNDERS. Real-time PCR: current technology and applications. Norfolk, UK: Caister Academic Press, c2009. ISBN 978-1-904455-39-4.
- [16] PIAZZA, G. a P. M. RIDKER. Is Venous Thromboembolism a Chronic Inflammatory Disease? Clinical Chemistry [online]. 2015, 61(2), 313-316 [cit. 2018-01-01]. DOI: 10.1373/clinchem.2014.234088. ISSN 0009-9147. Dostupné z: <http://www.clinchem.org/cgi/doi/10.1373/clinchem.2014.234088>
- [17] PAGE, Robert B. a Arnold J. STROMBERG. Linear Methods for Analysis and Quality Control of Relative Expression Ratios from Quantitative Real-Time Polymerase Chain Reaction Experiments. The Scientific World JOURNAL [online]. 2011, 11, 1383-1393 [cit. 2018-01-01]. DOI: 10.1100/tsw.2011.124. ISSN 1537-744x. Dostupné z: <https://www.hindawi.com/journals/tswj/2011/485391/abs/>
- [18] Genotyping - qPCR. Biosearch Technologies [online]. [cit. 2018-01-01]. Dostupné z: <https://www.biosearchtech.com/support/education/snp-genotyping>
- [19] GUESCINI, Michele, Davide SISTI, Marco BL ROCCHI, Laura STOCCHI a Vilberto STOCCHI. A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. BMC Bioinformatics [online]. 2008, 9(1), 326- [cit. 2018-01-05]. DOI: 10.1186/1471-2105-9-326. ISSN 1471-2105. Dostupné z: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-326>
- [20] BERÁNEK, Martin. *Molekulární genetika pro bioanalytiku*. Praha: Karolinum, 2016. ISBN 9788024632247.
- [21] Detekce příčin rezistence protideštičkové léčby u rizikových pacientů [online]. In: . [cit. 2018-05-14]. Dostupné z: <http://atelim.com/detekce-pin-rezistence-protidestikov-lby-u-rizikovch-pacient-d.html?part=3>