

Univerzita Karlova
Přírodovědecká fakulta

Studijní program: Chemie
Studijní obor: Medicinální chemie



Tereza Neuwirthová

Bez buněčná exprese proteinových knihoven z různých repertoárů aminokyselin
Cell-free expression of protein libraries made of different amino acid repertoires

Bakalářská práce

Vedoucí závěrečné práce: Mgr. Klára Hlouchová, Ph.D.

Praha, 2017

Prohlášení

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, dne 26. 5. 2017

Podpis

Abstrakt

Tato bakalářská práce je úvodem do projektu zabývajícího se vlivem evoluce genetického kódu, tedy i v čase se měnícího repertoáru aminokyselin v proteinech, na vývoj struktury a funkce bílkovin. Prvotní aminokyselinová abeceda byla zřejmě pouze poloviční velikosti abecedy současné a zahrnovala podle dostupných analýz okolo 10 prebioticky přítomných aminokyselin, pocházejících z exogenních či endogenních zdrojů.

Obsahem této práce je mRNA transkripce a bezbuněčná exprese pěti randomizovaných proteinových knihoven (obsahujících zhruba 10^{11} variant). Uvedené knihovny se liší svým aminokyselinovým složením a odráží tak jednotlivé předpokládané fáze evoluce genetického kódu. Hlavním cílem tedy je převést knihovny z DNA formy do formy proteinů a poskytnout tak zdroj pro purifikaci knihoven a jejich následnou biofyzikální charakterizaci (výskyt sekundární a terciární struktury) a analýzy výskytu funkce v takto navržených knihovnách.

Celý výzkum by tak teoreticky mohl přispět k zodpovězení základních otázek nejen z oboru proteinové evoluce, ale i syntetické biologie či proteinového inženýrství.

Klíčová slova: Bezbuněčná exprese, proteiny, evoluce genetického kódu

Abstract

This bachelor thesis is an introduction to a project dealing with the influence of genetic code evolution (and the consequent amino acid repertoire in proteins) on the evolution of protein structure and function. The initial amino acid alphabet was apparently only half the size of the present alphabet and according to available metaanalyses it included only about 10 prebiotically present amino acids derived from potential exogenous or endogenous sources.

This work includes mRNA transcription and cell-free expression of five randomized protein libraries (containing approximately 10^{11} variants). These libraries differ in their amino acid composition and reflect individual predicted stages of genetic code evolution. The main goal is to convert libraries from DNA to protein form to provide a source for library purification and their subsequent biophysical (occurrence of secondary and tertiary structure) and functional analyses.

The whole research could theoretically contribute to a few basic questions not only in the fields of protein evolution but also in areas of synthetic biology or protein engineering.

Keywords: Cell-free expression, proteins, genetic code evolution

This thesis is written in Czech.

Obsah

Abstrakt	1
Abstract	2
Použité zkratky	5
1. Úvod	8
2. Cíle práce.....	9
3. Přehled problematiky	10
3.1. Evoluce aminokyselin a genetického kódu	10
3.1.1. Genetický kód a jeho vývoj.....	11
3.1.1.1. Stereochemický původ	12
3.1.1.2. Teorie adaptace	12
3.1.1.3. Koevoluční teorie	12
3.1.2. Původ a výskyt aminokyselin.....	13
3.1.2.1. Exogenní a endogenní zdroje aminokyselin.....	13
3.1.2.2. Začleňování aminokyselin do genetického kódu	16
3.1.3. Sekvenční repertoár bílkovin	21
3.2. Proteinové knihovny	23
3.3. Syntéza proteinů <i>in vitro</i>	27
3.3.1. Srovnání CFPS s ostatními metodami syntézy proteinů	28
4. Výsledky a diskuze.....	31
4. 1. Návrh knihoven	31
4.2. Příprava DNA knihoven.....	34
4.2.1. Připojení stabilizačních sekvencí	35
4.3. Transkripce knihoven	36
4.4. Translace knihoven.....	37
4.5. Diskuze	39
5. Experimentální práce.....	41
5.1. Přístroje a chemikálie	41
5.1.1. Pufry a ostatní roztoky	41
5.1.2. Přístroje	42
5.1.3. Chemikálie	42
5.2. Obecné postupy	44
5.2.1. Práce s DNA a RNA.....	44
5.2.1.1. Polymerázová řetězcová reakce (PCR).....	44

5.2.1.2.	Horizontální agarózová elektroforéza.....	46
5.2.1.3.	Akrylamidová denaturující elektroforéza	46
5.2.2.	Práce s proteiny	47
5.2.2.1.	SDS elektroforéza	47
5.2.2.2.	Elektropřenos proteinů na nitrocelulózovou membránu s chemiluminiscenční detekcí protilátky	48
5.2.3.	Práce s knihovnami	48
5.2.3.1.	Příprava knihoven	48
5.2.3.2.	Transkripce knihoven.....	49
5.2.3.3.	Precipitace RNA	51
5.2.3.4.	Translace knihoven	51
6.	Závěr.....	52
7.	Poděkování.....	53
8.	Literatura	54
	Dodatek	56
	Sekvence oligonukleotidů použitých pro syntézu knihoven	56
	Sekvence primerů použitých pro přidání stabilizačních sekvencí ke knihovnám.....	57

Použité zkratky

10E	knihovna složená z aminokyselin A, D, E, G, I, L, P, S, T, V
10L	knihovna složená z aminokyselin R, N, C, Q, H, K, M, F, W, Y
10R	knihovna složená z aminokyselin A, E, G, L, P, S, C, H, K, F
14SL	knihovna složená z aminokyselin A, D, E, G, I, L, P, S, T, V, R, H, K, F
20F	knihovna složená ze všech 20 proteinogenních aminokyselin
A	adenin
Ala (A)	alanin
AMK	aminokyselina
APS	peroxodisíran draselný
Arg (R)	arginin
Asn (N)	asparagin
Asp (D)	kyselina asparagová
ATP	adenosintrifosfát
B	degenerovaný nukleotid (G, T nebo C)
bp	pár bází (z angl. base pair)
C	cytosin
CET	koevoluční teorie (z angl. coevolution theory)
CFPS	bezbuňková syntéza proteinu (z angl. cell-free protein synthesis)
Cys (C)	cystein
D	degenerovaný nukleotid (G, A nebo T)
DNA	deoxyribonukleová kyselina
dNTPs	deoxyribonukleotidy
dsDNA	dvouvláknová DNA (z angl. double strand)
DTT	dithiotreitol
FLAG	specifická sekvence pro afinitní purifikaci (aminokyselinová sekvence DYKDDDDK)
G	guanin
GCP II	enzym glutamát karboxypeptidáza II
Gln (Q)	glutamin
Glu (E)	kyselina glutamová
Gly (G)	glycin
H	degenerovaný nukleotid (A, C nebo T)

His (H)	histidin
HPLC	vysoko účinná kapalinová chromatografie
IDPs	prachové částice (z angl. interplanetary dust particles)
Ile (I)	isoleucin
IVC	<i>in vitro</i> kompartmentalizace (z angl. <i>in vitro</i> compartmentalization)
K	degenerovaný nukleotid (G nebo T)
Leu (L)	leucin
LUCA	poslední společný předek (z angl. last universal common ancestor)
Lys (K)	lysin
M	degenerovaný nukleotid (A nebo C)
MALDI-TOF	ionizace desorpcí laserem v přítomnosti matrice s analyzátozem doby letu
Met (M)	methionin
mRNA	mediátorová ribonukleová kyselina
N	degenerovaný nukleotid (A, C, G nebo T)
NBPs	nepřirodní randomizované proteinové sekvence (z angl. never born proteins)
NEB	New England Biolabs
NTPs	ribonukleotidy
ORF	otevřený čtecí rámec (z angl. open reading frame)
PCR	polymerázová řetězová reakce (z angl. polymerase chain reaction)
Phe (F)	fenylalanin
Pro (P)	prolin
PURE	syntéza proteinů využívající rekombinantních prvků (z angl. The Protein synthesis Using Recombinant Elements)
Pyl	pyrrolysin
R	degenerovaný nukleotid (G nebo A)
RM-PCR	náhodná multi-rekombinantní PCR (z angl. random multi-recombinant PCR)
RNA	ribonukleová kyselina
S	degenerovaný nukleotid (G nebo C)
SDS	dodecylsírán sodný
Sec	selenocystein
Ser (S)	serin
ssRNA	jednovláknová RNA (z angl. single-strand)
stop	terminační kodon
T	thymin

Thr (T)	threonin
TEMED	Tetramethylethylenediamin
t-RNA	transferová ribonukleová kyselina
Trp (W)	tryptofan
Tyr (Y)	tyrosin
U	uracil
UV	ultrafialové záření
V	degenerovaný nukleotid (G, C nebo A)
Val (V)	valin
W	degenerovaný nukleotid (A nebo T)
X	kterákoliv ze čtyř bází dle Cricka
Y	degenerovaný nukleotid (T nebo C, IUPAC) nebo kterákoliv ze čtyř bází dle Cricka

1. Úvod

Na počátku procesu vzniku života na Zemi bylo pravděpodobně prebioticky dostupné velké množství aminokyselin, pocházejících z exogenních či endogenních zdrojů, z nichž se jen asi pouhých deset, označovaných jako rané, začlenilo do tehdejšího genetického kódu¹⁻³. Postupem času genetický kód expandoval a začlenily se zbývající aminokyseliny, označované jako pozdní^{4,5}. Není dosud zřejmé, co tyto nově začleněné aminokyseliny přináší proteinové struktuře. Je možné získat strukturované a funkční proteiny pouze z uvedených deseti původních aminokyselin? V minulosti byly provedeny pokusy zabývající se charakterizací struktury randomizovaných proteinů skládajících se z redukované aminokyselinové abecedy. Pro konstrukci funkčních proteinů stačí pravděpodobně pouhá polovina přírodou využívané aminokyselinové abecedy⁶. Malé proteiny SH3 domény se sekundární strukturou β -skládaného listu byly syntetizovány z velké části jen z pěti aminokyselin⁷. Aminokyseliny však v těchto případech byly vybrány na základě svých fyzikálně-chemických vlastností, a nikoliv na základě jejich evolučního pořadí. Zda by bylo možné vytvořit proteiny skládající se pouze z raných aminokyselin a jaká by byla struktura takových proteinů, stále zůstává otázkou.

Cílem této práce je vytvořit knihovny náhodných proteinových sekvencí z různých repertoárů aminokyselin tak, aby vhodně odrážely evoluční stupně vývoje genetického kódu či jejich kontroly. Těchto výše zmíněných knihoven je navrženo celkem pět, z nichž jedna obsahuje všech 20 proteinogenních aminokyselin (20F), druhá pouze 10 raných aminokyselin (10E) a třetí 10 raných aminokyselin obohacených o 4 pravděpodobně začleněné do genetického kódu nejdříve ze skupiny 10 pozdních aminokyselin (14SL). Další dvě knihovny jsou používány jako kontrolní a obsahují pouze 10 pozdních aminokyselin (10L) a dále 10 aminokyselin vybraných za účelem obsáhnout co nejširší spektrum fyzikálně-chemických vlastností (10R).

Hlavním cílem této práce je tedy poskytnout materiál vhodný pro další charakterizace struktur a funkcí proteinů z uvedených knihoven. Knihovny náhodných sekvencí představují statisticky významný soubor a mohou sloužit ke srovnání potenciálu tvorby struktury a funkce mezi jednotlivými podskupinami aminokyselin.

2. Cíle práce

- Seznámení se s literaturou týkající se evoluce genetického kódu a bezbuněčné (*in vitro*) exprese.
- Příprava pěti DNA knihoven kódujících náhodné sekvence proteinů složené z různých repertoárů aminokyselin.
- Transkripce a bezbuněčná exprese knihoven do formy proteinů.

3. Přehled problematiky

3.1. Evoluce aminokyselin a genetického kódu

Centrální dogma molekulární biologie, tak jak bylo představeno v roce 1958 F. Crickem, je základním kamenem a zároveň spojovacím článkem téměř veškerého života na Zemi, i přes jeho obrovskou rozmanitost⁸. Říká nám, že genetická informace uložená v DNA je nejprve prepisována do RNA a následně je přeložena do proteinů, které jsou základní stavební jednotkou organismů, nebo v podobě enzymů katalyzují reakce nezbytné pro metabolismus. Většina přirozeně se vyskytujících genomů přímo kóduje v DNA 20 aminokyselin (viz Tab. 1).

Tab. 1 Tabulka genetického kódu s přiřazením aminokyselin k příslušným kodonům. Převzato a upraveno².

		2. pozice											
		U		C		A		G					
1. pozice	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	3. pozice	U		
		UUC		UCC			UAC				UGC	C	
		UUA	Leu	UCA			UAA	stop			UGA	stop	A
		UUG		UCG			UAG	stop			UGG	Trp	G
	C	CUU	Leu	CCU	Pro	CAU	His	CGA	Arg		U		
		CUC		CCC				CAC				CGC	C
		CUA				CCA		CAA			Gln	CGA	A
		CUG		CCG			CAG					CGG	G
	A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser		U		
		AUC		ACC				AAC				AGC	C
		AUA	Met	ACA			AAA	Lys			AGA	A	
		AUG		ACG			AAG					AGG	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U				
	GUC		GCC				GAC			GGC	C		
	GUA				GCA		GAA		Glu	GGA	A		
	GUG		GCG			GAG				GGG	G		

Kromě těchto 20 standardních aminokyselin bylo v genomech některých organismů objeveno kódování pro 21. a 22. aminokyselinu. Jedná se o selenocystein (Sec), vyskytující se u archebakterií, eubakterií i eukaryot, která mezi standardních 20 není řazena kvůli abnormalitám při její translaci, a pyrrolysin (Pyl), vyskytující se převážně u methanogenních archebakterií⁹.

Konkrétněji jsou aminokyseliny kódovány tříznakovou soustavou, tedy genetickým kódem. Dle D. P. Snustada a M. J. Simmonse ho můžeme definovat jako nepřekrývající se kód, který definuje pozici jednotlivých aminokyselin, iniciaci a terminaci translace pomocí RNA-

kodonů¹⁰. Přiřazení aminokyselin ke konkrétním kodonům vidíme v Tabulce 1 (Tab. 1, strana 10). Označení „stop“ znamená terminační kodon.

3.1.1. Genetický kód a jeho vývoj

Genetický kód je, jak již bylo řečeno, složen z tripletů nukleotidů. Každý nukleotid je tvořený fosforylovanou cukernou složkou (ribóza nebo deoxyribóza) a jednou z pěti bází, jmenovitě adenin, cytosin, guanin a thymín (DNA) nebo uracil (RNA).

Výhodou existence třípísmenného kódu je, už podle předpovědi amerického fyzika G. Gamowa, celkový počet možností, které nám třípísmenný kód poskytuje^{2,11}. Kdyby genetický kód byl pouze dvoupísmenný, poskytoval by jen 16 (4^2) reprezentací pro 20 aminokyselin, na druhou stranu třípísmenný kód nám zajistí celkem 64 (4^3) reprezentací. Většina aminokyselin (s výjimkou methioninu a tryptofanu) je kódována několika nukleotidovými triplety a proto nazýváme genetický kód degenerovaným tedy redundantním². Díky této vlastnosti je minimalizován počet chyb při translaci v případě bodových mutací, a zároveň genetický kód stále poskytuje prostor pro evoluci genetické informace¹².

Genetický kód můžeme také charakterizovat jako univerzální, což znamená, že až na výjimky je stejný pro buňky téměř všech prokaryotních i eukaryotních organismů. Toto pravidlo porušuje například genetický kód mitochondrie, ve kterém je tryptofan kódován univerzálním terminačním kodonem UGA. Další drobné odchylky jsou známé u některých druhů mikroorganismů^{13–15}.

Dalším rysem genetického kódu je logika přiřazení párů kodon-aminokyselina. Přiřazení aminokyselin ke kodonům není náhodné a zde uvádím pro příklad několik pravidel, kterými se řídí:

1. XYZ a XZY vždy kódují stejnou aminokyselinu.
2. XYA a XYG často kódují stejnou aminokyselinu. Výjimkou z tohoto pravidla tvoří methionin a tryptofan, protože jsou kódovány pouze jedním kodonem.
3. Všechny kodony reprezentující jednu z aminokyselin začínají stejnou bází. Pouze leucin, arginin a serin jsou výjimky.
4. Pokud seskupíme kodony do sad podle podobnosti, můžeme si všimnout určité systematiky v přidělování aminokyselin ke kodonům z těchto sad dle fyzikálně-chemické podobnosti mezi aminokyselinami. Jako příklad lze uvést, že kodony obsahující uracil na druhém místě jsou přiděleny hydrofobním aminokyselinám.

(Písmeny X a Y rozumíme kteroukoliv ze čtyř možných bází¹⁶.)

Soubor těchto „pravidel“ lze chápat jako průvodní jev toho, že genetický kód nevznikl najednou, ale postupně v čase pravděpodobně expandoval a zmíněná „pravidla“ jsou tedy jeho evoluční stopou.

Principy přiřazení kodonů jednotlivým aminokyselinám poprvé popsal již F. Crick v roce 1968¹⁶. Postupně byly rozvinuty možné teorie, jak je popsáno níže.

3.1.1.1. Stereochemický původ

Tato teorie postuluje, že každá aminokyselina byla ke svým tripletům přiřazena na základě fyzikálně-chemické afinity, jinak řečeno triplety byly s aminokyselinou spojeny z určitých stereochemických důvodů. Teze je platná hlavně pro prvotní formu kódu.

3.1.1.2. Teorie adaptace

Genetický kód se téměř s určitostí v průběhu času vyvíjel směrem k minimalizaci vlivu translačních chyb. Teorie se opírá o fakt, že téměř ve všech organismech genetický kód určuje aminokyselinové sekvence mnoha evolučně vysoce vyvinutých proteinů, pro které by jakákoliv jeho změna už byla nevýhodná¹⁷.

Na základě toho tedy tato teorie tvrdí, že genetický kód je univerzální, protože jakákoliv změna v něm byla evolucí velmi striktně odmítnuta a od posledního společného předka všech organismů („last universal common ancestor“ = LUCA) se již nemění.

Teorie nevysvětluje princip přiřazení kodonů k aminokyselinám na úplném počátku.

3.1.1.3. Koevoluční teorie

Protože ani jedna z teorií nám neposkytla systematické řešení problému, byla J. T. Wongem navržena Koevoluční teorie (CET = „Coevolution theory“). Ta poprvé rozděluje aminokyseliny na aminokyseliny fáze 1 (rané) a fáze 2 (pozdí). Tato teorie je postavena na čtyřech základních principech¹⁸:

1. Prebiotické prostředí nebylo zásobeno všemi 20 proteinogenními aminokyselinami. Ty, které nebyly v prostředí dostupné, musely být postupně doplněny biosyntézou.
2. „Pretran“ syntéza - přeměna prekurzorové aminokyseliny na žádanou aminokyselinu během vazby na tRNA poskytla mechanismus pro včlenění aminokyselin fáze 2 do genetického kódu.

3. Tento biosyntetický vztah mezi prekurzorovou aminokyselinou a aminokyselinou žádanou je faktorem ovlivňujícím logiku přiřazení aminokyselin ke kodonům v genetickém kódu.
4. Genetický kód je proměnlivý a dovoluje mutace, proto došlo k jeho expanzi a připojení aminokyselin fáze 2.

Závěrem lze říci, že tyto tři teorie se vzájemně spíše doplňují a pravděpodobně platí všechny současně, i přestože často bývají chápány jako tři navzájem se vylučující alternativy¹⁹.

3.1.2. Původ a výskyt aminokyselin

„Přehrát evoluční pásku“ je z dnešní perspektivy neproveditelné. Tedy identifikovat přesně prebioticky dostupné aminokyseliny a zjistit, které z nich byly začleněny do genetického kódu a v jakém prostředí či pořadí, je dnes nemožné. Musíme také vzít v úvahu alternativu, že genetický kód byl v rané fázi mnohem proměnlivější – některé aminokyseliny mohly být v průběhu evoluce nahrazeny jinými atp. Nicméně, v několika posledních desítkách let vědecká komunita nashromáždila velké množství teoretických a experimentálních studií na toto téma a jejich metaanalýzami můžeme dospět alespoň k určité aproximaci evoluce genetického kódu a výskytu aminokyselin.

3.1.2.1. Exogenní a endogenní zdroje aminokyselin

Potenciální zdroje prebiotických aminokyselin rozdělujeme na exogenní a endogenní.

Exogenními zdroji rozumíme partikule vesmírného původu, hlavně meteority (uhlíkaté chondrity) a prachové částice (IDPs z angl. interplanetary dust particles), které na povrch Země v jejím raném období pravděpodobně dopadaly častěji než dnes. Konkrétně na meteoritech bylo nalezeno velké množství aminokyselin, z nichž pouze malá část se vyskytuje mezi již uvedenými 20 proteinogenními. Na obrázku 1 (viz Obr. 1, strana 14) jsou uvedeny aminokyseliny detekované analýzou meteoritů¹.

Tak velké množství aminokyselin, z nichž většina není biologicky aktivních, lze považovat za jeden z důkazů jejich eventuálního mimozemského původu – případná kontaminace meteoritu zemských povrchem je téměř vyloučena. Dalším důležitým poznatkem analýz meteoritů je přítomnost obou chirálních forem proteinogenních aminokyselin v poměru blížícím se racemické směsi (1:1). Co se týče polohy aminoskupiny v molekulách aminokyselin – kromě biologických alfa-aminokyselin byly pozorovány i beta-, gama- nebo delta-aminokyseliny. Aminokyseliny s menším počtem uhlíků byly obecně přítomny ve vyšších koncentracích než ty víceuhlíkaté¹.

C2

Glycin

C3

Alanin

 β -alanin

Serin

Sarkosin

C4

Threonin

allo-threonin α -aminomáselná kyselina β -aminomáselná kyselina α -aminoisomáselná kyselina β -aminoisomáselná kyselina*N*-ethylglycin*N,N*-dimethylglycin*N*-methylalanin*N*-methyl- β -alanin

Asparagová kyselina

2,3-diaminomáselná kyselina

2,4-diaminomáselná kyselina

3,3'-diaminoisomáselná kyselina

C5

Valin

Isovalin

Norvalin

3-aminovalerová kyselina

3-amino-2-methylomáselná kyselina

allo-3-amino-2-methylomáselná kyselina

3-amino-2,2-dimethylpropionová kyselina

3-amino-2-ethylpropionová kyselina

4-aminovalerová kyselina

4-amino-2-methylomáselná kyselina

4-amino-3-methylomáselná kyselina

5-aminovalerová kyselina

Glutamová kyselina

2-methylasparagová kyselina

3-methylasparagová kyselina

allo-3-methylasparagová kyselina*N*-methylasparagová kyselina

4,4'-diaminoisovalerová kyselina

C6

Leucin

Isoleucin

allo-isoleucin

Norleucin

Pseudoleucin

Cykloleucin

2-methylnorvalin

Piperidin-2-karboxylová kyselina

2-amino-2-ethylomáselná kyselina

2-amino-2,3-dimethylomáselná kyselina

3-amino-2-ethylomáselná kyselina

3-amino-2,3-dimethylomáselná kyselina

3-methylaminvalerová kyselina

4-aminokapronová kyselina

4-amino-3,3-dimethylomáselná kyselina

4-amino-2-methylvalerová kyselina

4-amino-3-methylvalerová kyselina

4-amino-4-methylvalerová kyselina

6-aminokapronová kyselina

 α -aminoadipová kyselina β -aminoadipová kyselina

2-methylglutamová kyselina

ostatní izomery byly detekovány ale nebyly přesně identifikovány

C7

2-amino-2,3,3-trimethylomáselná kyselina

2-amino-2-ethyl-3-methylomáselná kyselina

2-amino-2-ethylvalerová kyselina

2-amino-3-ethylvalerová kyselina

2-amino-2,3-dimethylvalerová kyselina

2-amino-2,4-dimethylvalerová kyselina

2-amino-3,3-dimethylvalerová kyselina

2-amino-3,4-dimethylvalerová kyselina

2-amino-4,4-dimethylvalerová kyselina

allo-2-amino-2,3-dimethylvalerová kys.*allo*-2-amino-3,4-dimethylvalerová kys.

2-amino-2-methylkapronová kyselina

2-amino-3-methylkapronová kyselina

allo-2-amino-3-methylkapronová kyselina

2-amino-4-methylkapronová kyselina

allo-2-amino-4-methylkapronová kyselina

2-amino-5-methylkapronová kyselina

2-aminoenanthová kyselina

 α -aminopimelová kyselina

1-aminocyklohexanová kyselina

ostatní izomery byly detekovány ale nebyly přesně identifikovány

C8

izomery byly detekovány ale nebyly přesně identifikovány

C9

Fenylalanin

Tyrosin

ostatní izomery byly detekovány ale nebyly přesně identifikovány

Obr. 1 Seznam aminokyselin detekovaných na meteoritech. Převzato^{1,20}.

Zdroji endogenními rozumíme výskyt nebo vznik aminokyselin přímo na Zemi z tehdy dostupných sloučenin za příhodných podmínek. První teorie zabývající se vznikem aminokyselin z jednoduchých anorganických látek formulovali A. I. Oparin a J. B. S. Haldane²¹. Ti postulovali, že první cukry a aminokyseliny mohly vznikat vlivem vhodného energetického zdroje (například UV záření nebo elektrického výboje) ve vodném prostředí z jednoduchých anorganických látek jako je methan, amoniak a vodík. Podmínkou by byla redukující prebiotická atmosféra, ve které by nedocházelo k oxidačnímu rozkladu organických meziproductů.

V roce 1953 byla tato teorie podložena experimentem S. L. Millera a H. Ureyho. Do izolované aparatury byly přes elektrické výboje přiváděny plynné složky prebiotické atmosféry (dle Oparinovy teorie - methan, amoniak, vodní pára a vodík). Po několika dnech byly v aparatuře přidružené varné nádobě simulující praoceán pomocí papírové chromatografie detekovány aminokyseliny glycin, α -alanin, β -alanin, kyselina asparagová a α -amino-*n*-máselná kyselina³. Následně byly pokusy provedeny s mnoha modifikacemi. Například byla urychlena cirkulace reakčních plynů od výbojky díky zavádění vodní páry vstříkovaním. Tato modifikace zajistila méně redukující podmínky, než když byla vodní pára přivedena jednorázově při zahájení reakce. Další modifikací byla výměna jiskrového výboje za výboj tichý²².

V roce 2008 byla znovu provedena analýza Millerových vzorků z roku 1955. Vzorky použité pro tuto analýzu pocházely právě ze tří výše popsaných aparatur. Díky současným moderním technologiím, HPLC s fluorescenčním detektorem a MALDI-TOF hmotnostní spektrometrii, bylo tentokrát ve vzorcích identifikováno více sloučenin. Ve vzorcích pocházejících z aparatury s méně redukujícími podmínkami bylo identifikováno 22 aminokyselin a 7 aminů. Některé z těchto sloučenin byly detekovány tedy až novou analýzou uvedených starých vzorků. Ve vzorcích z aparatur s odlišnými podmínkami bylo nalezeno podobné množství aminokyselin, ale s menší diverzitou než u aparatury se slabším redukčním prostředím. Dále z výsledků analýzy lze vyčíst, že hydroxylované sloučeniny byly více syntetizovány v méně redukujících podmínkách. Aminokyseliny s chirálními centry při těchto experimentech vznikaly v poměru racemických směsí, zde je patrná podobnost s analýzami meteoritů²³.

Syntéza aminokyselin pravděpodobně mohla probíhat i hluboko pod mořskou hladinou v prostředí tzv. hydrotermálních zřidel. Z podmořských „komínů“ neboli „černých kuřáků“ (z angl. „black smokers“) vyvěrá přehřátá voda bohatá na minerální látky (zejména na síru). Při styku této vody se studeným okolním prostředím dochází ke krystalizaci minerálních látek a jejich následné sedimentaci na mořském dně. V okolí těchto „černých

kuřáků“ panují extrémní podmínky (vysoký tlak, teplota a nedostatek světla), a přesto zde byly objeveny rozmanité formy života, především bohatá společenství termofilních bakterií, které čerpají energii právě ze sirných minerálů. Kvůli nepřítomnosti kyslíku v tomto prostředí a současné přítomnosti některých minerálů s energeticky bohatými chemickými vazbami (například pyrit), mohlo toto prostředí zapříčinit vznik prebiotických aminokyselin².

V níže uvedené tabulce (viz Tab. 2, strana 17) jsou uvedeny aminokyseliny detekované na příkladu exogenního a endogenního zdroje.

3.1.2.2. Začleňování aminokyselin do genetického kódu

Z velkého množství prebioticky dostupných aminokyselin, ať už z endogenních či exogenních zdrojů (viz Obr. 1, strana 14), se do genetického kódu zabudovalo pravděpodobně jen deset z nich (viz Tab. 3, strana 18). Tyto aminokyseliny označujeme jako rané. Zbýlých deset proteinogenních aminokyselin ve výsledcích analýz zabývajících se prebiotickou dostupností aminokyselin nenacházíme (v Tab. 3 označeny červeně).

Tab. 2 Aminokyseliny zjištěné analýzou směsi z Millerova experimentu a chemickým rozborem Murchisonova meteoritu. Převzato a upraveno²⁴.

Aminokyseliny	Murchisonův meteorit	Millerův experiment
Glycin	++++	++++
Alanin	++++	++++
α -aminomáselná kyselina	+++	++++
α -aminoisomáselná kyselina	++++	++
Valin	+++	++
Norvalin	+++	+++
Isovalin	++	++
Prolin	+++	+
Piperidin-2-karboxylová kyselina	+	+
Asparagová kyselina	+++	+++
Glutamová kyselina	+++	++
β -alanin	++	++
β -aminomáselná kyselina	+	+
β -aminoisomáselná kyselina	+	+
γ -aminomáselná kyselina	+	++
Sarkosin	++	+++
<i>N</i> -ethylglycin	++	+++
<i>N</i> -methylalanin	++	++

Molární poměr vzhledem ke glycinu (= 100): + = 0,05-0,5; ++ = 0,5 - 5; +++ = 5-50; ++++ = více než 50.

Tab. 3 Tabulka 20 proteinogenních aminokyselin a jejich prebiotický výskyt. Převzato a upraveno ²⁴.

Aminokyseliny	Meteorit ²⁵	Millerův experiment ²⁶
Glycin	+	+
Alanin	+	+
Serin	+	+
Asparagová kyselina	+	+
Glutamová kyselina	+	+
Threonin	+	+
Prolin	+	+
Valin	+	+
Isoleucin	+	+
Leucin	+	+
Lysin	-	-
Cystein	-	-
Asparagin	-	-
Glutamin	-	-
Histidin	-	-
Methionin	-	-
Arginin	-	-
Fenylalanin	-	-
Tyrosin	-	-
Tryptofan	-	-

Symboly: + = detekováno; - = nedetekováno

Byly provedeny desítky studií zabývající se chronologií začlenění aminokyselin do genetického kódu, ze kterých byly nakonec vypracovány dvě metaanalýzy. První z nich v roce 2000 E. N. Trifonovem, který srovnal seřazení aminokyselin dle 40 kritérií týkajících se jejich začlenění do genetického kódu a došel k závěru, že první aminokyseliny v genetickém kódu byly ty prebioticky dostupné. Mezi uvedených 40 kritérií patří mimo jiné teorie zabývající se studiem původu genetického kódu, termostabilita prvotních nukleových kyselin, zastoupení aminokyselin v pokusech simulujících prebiotické podmínky, analýzy meteoritů, komplexnost vlastností aminokyselin, genomové analýzy nebo i stabilita kodonové-antikodonové interakce pro různé kodony. Trifonov se domnívá, že v rozvoji tripletového kódu sehrála termostabilita kodonové-antikodonové interakce hlavní roli. Nové kodony se podle této studie objevovaly v komplementárních párech a ty, které se objevily později, byly odvozené z kodonů evolučně starších. V souladu s těmito čtyřmi pravidly byly aminokyseliny seřazeny do časové posloupnosti:

Gly/Ala, Val/Asp, Pro, Ser, Glu/Leu, Thr, Arg, Asn, Lys, Gln, Ile, Cys, His, Phe, Met, Tyr, Trp⁵.

V roce 2009 byla podobná analýza provedena Paulem G. Higgsem a Ralphem E. Pudritzem, kteří ale uvažovali kritérií méně⁴. Nicméně obě tyto metaanalýzy se téměř shodují v rozdělení aminokyselin, jak můžeme vidět v tabulce 4 (Tab. 4, strana 20), na skupinu evolučně raných a evolučně pozdních, podobně jako předestřel Wong ve své koevoluční teorii¹⁸. I pro účely této práce mezi rané aminokyseliny řadíme Gly, Ala, Val, Asp, Pro, Ser, Glu, Leu, Ile a Thr.

Tab. 4 Tabulka aminokyselin dle jejich relativního stáří podle metaanalýz Higgse & Pudritze (2009)⁴ a Trifonova (2004)²⁷. Rané aminokyseliny jsou označeny modře, pozdní aminokyseliny zeleně. V závorce je vždy uvedeno pořadí aminokyselin.

Aminokyseliny	Higgs & Pudritz (2009)	Trifonov (2004)
Glycin	0 (1)	0 (1)
Alanin	0,13 (2)	0,04 (2)
Asparagová kyselina	0,24 (3)	0,19 (3)
Glutamová kyselina	0,44 (4)	0,35 (7)
Valin	0,56 (5)	0,22 (4)
Serin	0,57 (6)	0,32 (6)
Isoleucin	0,61 (7)	0,61 (12)
Leucin	0,63 (8)	0,49 (9)
Prolin	0,68 (9)	0,29 (5)
Threonin	0,81 (10)	0,45 (8)
Lysin	0,88 (11)	0,75 (14)
Fenylalanin	0,92 (12)	0,84 (16)
Arginin	0,93 (13)	0,58 (10)
Histidin	0,93 (13)	0,73 (13)
Asparagin	1 (14)	0,60 (11)
Glutamin	1 (14)	0,61 (12)
Cystein	1 (14)	0,79 (15)
Tyrosin	1 (14)	0,90 (17)
Methionin	1 (14)	0,92 (18)
Tryptofan	1 (14)	1 (19)

Hodnoty z výše uvedené tabulky byly vypočítány normalizací dat uvedených v citovaných člancích pomocí vzorce: $H_{Tab.} = \frac{H_{AMK} - H_{min}}{H_{max} - H_{min}}$, ve kterém $H_{Tab.}$ je hodnota relativního stáří z výše uvedené tabulky, H_{AMK} je hodnota stáří určité aminokyseliny z daného článku, a H_{max} a H_{min} jsou maximální a minimální hodnoty stáří aminokyselin uvedené v daném článku.

Dvacet proteinogenních aminokyselin pravděpodobně nebylo do genetického kódu včleněno pouhou náhodou. Ve Weberově a Millerově studii z 80. let 20. století byly zvažovány tyto faktory, charakteristické pro všechny proteinogenní aminokyseliny²⁸:

- (1) Proč právě aminokyseliny jsou biochemickými katalyzátory, a ne například hydroxykyseliny nebo jiné organické sloučeniny?
- (2) Proč jsou v proteinech vázány téměř výhradně jejich α -formy?
- (3) Proč mají všechny nejméně jeden vodík v α -poloze?
- (4) Proč žádná z nich, kromě prolinu, není N-alkyl aminokyselina?
- (5) Proč jsou všechny v L-formě (R-enantiomerní na α -uhlíku kromě cysteinu, který je S-enantiomerní, a glycinu, který není chirální)?
- (6) Dle jakých kritérií byly vybrány postranní řetězce?
- (7) Proč je jich jen dvacet, přestože máme k dispozici 64 kodonů?

Studie A. L. Webera a S. L. Millera dokonce tvrdí, že za předpokladu vzniku života na jiné planetě by nejspíše 15 aminokyselin z těchto 20 bylo stejných jako na naší planetě. Toto tvrzení ještě konkretizují, a to tak, že by bylo velmi překvapivé, pokud by aminokyselin bylo stejných méně než 10, ale téměř s určitostí jich nebude stejných všech dvacet²⁸. Tato tvrzení však byla S. Freelandem v roce 2010 zpochybněna⁹.

3.1.3. Sekvenční repertoár bílkovin

Kdybychom uvažovali polypeptidový řetězec, který by obsahoval 100 aminokyselin z výběru všech 20 proteinogenních, měli bychom v principu matematickým výpočtem 20^{100} možných kombinací, jak by mohl takový polypeptidový řetězec vypadat. Na Zemi skutečně existuje několik tisíc miliard proteinů, takže řádově vzato bývá uvažováno číslo 10^{15} . Z poměru těchto dvou čísel vyplývá, že musela být vybrána pouze velmi malá část z obrovského množství možností²⁹.

Předpoklad, že na existujících proteinech je něco speciálního, podporuje myšlenkový směr determinismu, který tvrdí, že proteiny mohly vzniknout pouze takové, jaké jsou dnes známé. Termodynamicky totiž žádná jiná možnost není možná. Zastáncem této teorie je například Charles de Duve³⁰.

Druhou možností je vznik proteinů řízený především nahodilostí, tedy proteiny ani nukleové kyseliny, které existují, nejsou z termodynamického hlediska ničím speciální. Pokud by tedy život vznikl znovu za stejných či velmi podobných podmínek, mohly by se produkovat proteiny úplně jiné²⁹.

Pod názvem Never Born Proteins (NBPs), neboli v doslovném překladu „nikdy nenarozené proteiny“, rozumíme proteiny ze souboru všech možných kombinací (jak bylo uvedeno výše), které nebyly evolučně vybrány a přirozeně se na Zemi nevyskytují.

Dnes víme, že pro biologickou funkčnost proteinu je velmi důležitá jeho konformace, sekundární a terciální struktura. Tvoří never born proteins také některé ze známých stabilních konformací a mohly by eventuálně být součástí živé hmoty? Na tuto otázku se ve svém výzkumu pokusil odpovědět C. Chiarabelli³¹.

Spolu se svými spolupracovníky vytvořil knihovnu náhodných DNA sekvencí o délce 150 bází, templátů pro proteiny o délce 50 aminokyselin. Každý DNA templát obsahoval okolo svého středu předem definovanou sekvenci kódující tri-peptid, prolin-arginin-glycin, který je substrátem serinové proteázy trombinu. Jeho cílem bylo tyto náhodné proteiny připravit a poté štěpit uvedeným enzymem. Proteiny tvořící stabilní konformaci totiž ono místo, štěpitelné trombinem, s větší pravděpodobností skryjí uvnitř vytvořené struktury, a proto nebudou štěpeny. K selekci těchto proteinů byla použita metoda „phage display“.

Překvapivým výsledkem tohoto pokusu byla skutečnost, jak velké procento struktur zřejmě tvoří stabilní konformery (přibližně 20%³¹).

Další výzkum v této oblasti by mohl pomoci objasnit vznik prvotních bílkovin a roli evolučních procesů při jejich výběru a optimalizaci.

3.2. Proteinové knihovny

V uplynulém desetiletí byly vyvinuty metody tvorby lépe definovaných a rozsáhlejších proteinových knihoven. To umožnilo rozvoj celé řady nových přístupů k funkčnímu screeningu a selekčním technikám potřebným k izolaci a charakterizaci proteinů v odvětví proteinového inženýrství, kde knihoven využíváme³². Další využití proteinových knihoven nacházíme například v odvětví zabývajícím se *in vitro* proteinovou evolucí³³.

Velikosti proteinových knihoven se liší s použitím různých *in vitro* selekčních metod. Největších knihoven (okolo 10^{13} variant) využívají metody mRNA display či ribozomální display a naopak nejmenších (okolo 10^8 variant) využívají metody IVC (z angl. *in vitro* compartmentalization), „phage display“ a „microbead display“³³.

Proteinové knihovny jsou často kódovány chemicky syntetizovanou DNA, ve které je část jejího otevřeného čtecího rámce (z angl. open reading frame, ORF) randomizovaná³⁴. Otevřeným čtecím rámcem rozumíme sekvenci nukleové kyseliny, která kóduje samotný protein. Se syntézou dlouhých ORF částí deoxyribonukleové kyseliny, především těch randomizovaných, jsou spojeny mnohé komplikace. Inzerce nebo delece jednoho či více párů bází, které se během syntézy objevují, bodové mutace a nezamýšlené terminační kodony v regionu náhodných sekvencí mohou zapříčinit předčasnou terminaci translace při expresi dané DNA³⁴.

Jednou z možností snížení výskytu terminačních kodonů v knihovně náhodných sekvencí je použití degenerovaných kodonů. To znamená, že namísto tripletů NNN (N = A, C, T, G) lze použít kodony NNK (K = G, T) a NNS (S = C, G) nebo úplnou kodonovou abecedu sníženou o kodony NDT (D = A, G, T). Použití těchto kodonů snižuje pravděpodobnost zahrnutí nezamýšlených terminátorů translace v randomizovaném úseku. Konkrétně kodony NNN zahrnují tři stop kodony, zatímco kodony NNK a NNS pouze jeden³³. Pokud je cílem vybrat na určitá místa pouze určitý repertoár aminokyselin, lze použít další možné degenerované kodony dle tabulky 5 (Tab. 5, strana 24).

Existují dva typy knihoven s randomizovanými úseky. Knihovny využívající randomizace pouze v určitém specifickém regionu celé své sekvence, např. v aktivním místě enzymu, a knihovny využívající randomizace po celé své délce kromě sekvencně definovaných specifických úseků kompatibilních s používanou analyzační technikou³². Pro naše účely využíváme druhé uvedené možnosti, proto se ve své práci zaměřuji převážně na takto vypadající knihovny.

Tab. 5 Tabulka nomenklatury nukleotidů. Zdroj IUPAC.

symbol	nukleotid	symbol	nukleotid
A	adenosin	M	A nebo C
C	cytidin	S	G nebo C
G	guanosin	W	A nebo T
T	thymidin	B	G, T nebo C
U	uridin	D	G, A nebo T
R	G nebo A	H	A, C nebo T
Y	T nebo C	V	G, C nebo A
K	G nebo T	N	A, C, G nebo T

Metoda využívající výše popsaného využití degenerovaných kodonů byla navržena v roce 1990 J. K. Scottem a G. P. Smithem. Randomizovaná sekvence hexapeptidů byla sestavena pomocí degenerovaných kodonů NNK. Fragmenty dlouhé 33 bp (z angl. base pair, párů bází) mohly obsahovat 32^6 (přibližně 10^9) různých nukleotidových sekvencí³⁵. Nevýhodou takto syntetizované knihovny náhodných sekvencí je pevná délka randomizovaného regionu, který je pro mnohé účely příliš krátký³⁶.

Další možnou metodou je výslednou knihovnu poskládat z fragmentů již předem selektovaných tak, aby obsahovaly co nejméně terminačních kodonů a bodových mutací³³. Tuto strategii syntézy knihovny zvolili G. Cho a kolektiv v roce 2000. Aby se vyhnuli předčasným terminačním kodonům, bodovým mutacím a problémům s délkou randomizovaného úseku, prováděli *in vitro* selekci. Tyto selekce prováděli pomocí metody mRNA display³⁴. Takto syntetizovaná knihovna sice poskytuje flexibilitu ve své délce, ale standardní restriční enzymy zanechávají jistou přebytečnou sekvenci (z angl. „scar sequence“) mezi sekvencemi, které spojují³⁶.

Metoda s názvem náhodná multi-rekombinantní PCR (angl. random multi-recombinant PCR, RM-PCR), také použitelná pro syntézu proteinových knihoven, je založena na principu PCR a umožňuje přeskupování většího množství fragmentů DNA, i bez homologních sekvencí, během klasické PCR reakce³⁷. Metoda však vyžaduje nejprve přípravu dlouhých, překrývajících se dvouvláknových DNA molekul nebo oligonukleotidů, které limitují flexibilitu návrhu

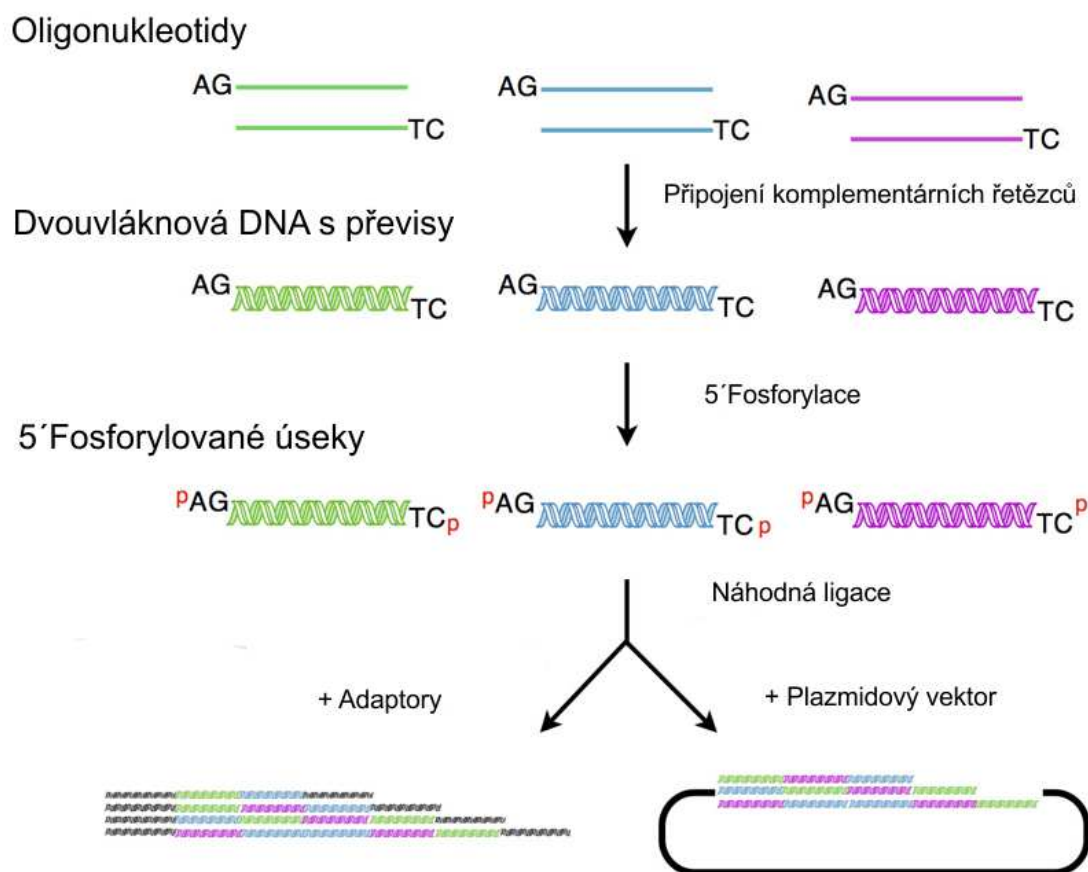
náhodných sekvencí. Kromě toho také s každým přidaným PCR cyklem se zvyšuje pravděpodobnost vzniku bodových mutací³⁶.

Metoda využívající DNA čipů (angl. microarrays) pro syntézu oligonukleotidů nabízí ekonomicky výhodnou syntézu milionů DNA sekvencí dlouhých okolo 200 bp, avšak metoda vyžaduje použití speciální instrumentace a také kvalita a délka takto získaných oligonukleotidů není pro účely získávání sekvencí DNA knihoven složených z NBP dostatečná^{36,38}.

Za účelem zjednodušit celý proces syntézy a zároveň minimalizovat nedostatky byla v roce 2015 vyvinuta K. Fujishimou jedna z nejnovějších metod syntézy proteinových knihoven. Flexibilita této metody, jak v návrhu sekvencí, tak i v jejich délce, dovoluje následné *in vivo* i *in vitro* analýzy vytvořené knihovny.

Postup při syntéze proteinové knihovny výše uvedenou metodou:

1. Nejprve byly syntetizované degenerované oligonukleotidy o délce 16-30 bp s převisy (1-2 báze).
2. Poté byly komplementární oligonukleotidy s převisem na 5' konci spojeny do dvouvláknových molekul DNA.
3. Ve třetím kroku proběhla fosforylace dsDNA úseků na 5' konci.
4. Následovala ligace oligonukleotidových úseků, která probíhá v místech s fosforylovanými převisy.
5. Nakonec byly přidány adaptory obsahující i restrikční místa pro použití *in vitro*, nebo plazmidové vektory pro použití *in vivo*³⁶. Celý uvedený princip je znázorněn na obrázku 2 (Obr. 2, strana 26).



Obr. 2 Metoda konstrukce knihovny založená na principu slučování DNA úseků s převisy (v tomto případě AG-TC). Převzato a upraveno³⁶.

Uvedené převisy jsou záměrně dvounukleotidové, protože převisy tvořené pouze jednotlivými nukleotidy selhaly při tvorbě delších ligačních produktů a dva nukleotidy ještě neurčují konkrétní aminokyselinu – proto není narušena randomizace.

Tento přístup k syntéze udržuje, ve srovnání s metodou syntézy pomocí degenerovaných NNK kodonů, kontrolu nad zastoupením jednotlivých aminokyselin v knihovně, což umožňuje poskytnout uživateli návrh specifických DNA bloků přesně pro účely dalšího výzkumu. Jednou z klíčových výhod této metody je produkce proteinové knihovny s výskytem pouze některých aminokyselin ze všech 20 proteinogenních. Z tohoto důvodu lze metodu využít ke studiu raných proteinů, které pravděpodobně neobsahovaly všechny aminokyseliny³⁶.

Aby bylo možné použít knihovny pro specifické *in vitro* evoluční analýzy, sekvence na obou koncích DNA musí být kompatibilní s používanou technikou. Na 5' konci musí být obsažen promotor a tzv. zesilovací sekvence (z angl. enhancer sequences) nezbytné pro usnadnění transkripce a translace DNA knihovny do podoby proteinů. Bližší specifikace těchto

zesilovacích sekvencí závisí na typu systému použitého pro *in vitro* transkripci a translaci. Mezi další sekvenční prvky, které je možno zahrnout, patří specifické sekvence (z angl. tags) pro afinitní purifikaci, například FLAG-kotva nebo histidinová kotva (His-tag), terminační kodon či stabilizační vlásenky (z angl. hairpins)³³.

Pro syntézu knihoven, se kterými nakládáme v této práci, byla použita adaptace výše uvedeného postupu K. Fujishimy.

3.3. Syntéza proteinů *in vitro*

Syntéza proteinů *in vitro*, neboli bezbuněčná syntéza proteinů (z angl. cell-free protein synthesis, CFPS) je velmi účinnou technikou pro produkci proteinů jednoduchou a efektivní cestou a tím pádem i technikou se širokým spektrem možných využití - například při syntéze farmaceutických proteinů, při produkci proteinových knihoven pro výzkum evoluce proteinů a strukturní genomiky nebo v proteinovém inženýrství^{39,40}.

Podobně jako technika PCR, která využívá buněčné „nástroje“ pro replikaci DNA, bezbuněčná syntéza proteinů využívá buněčný systém určený k transkripci a translaci. Mnoho z těchto bezbuněčných systémů pro syntézu proteinů je postaveno na základě extraktů z různých typů buněk. Mezi nejvíce využívané patří například *Escherichia coli*, králičí retikulocyty, pšeničné klíčky a buňky hmyzu. Nejrozšířenějším z uvedených CFPS systémů je systém využívající prokaryotické buňky *Escherichia coli*, ostatní uvedené jsou ty nejvíce využívané ze systémů eukaryotických a jejich výhodou oproti prokaryotickému je, že mohou produkovat některé komplexní proteiny a post-translační modifikace, které bakterie neposkytují. Všechny tyto systémy jsou komerčně dostupné³⁹.

Uvedený buněčný extrakt zahrnuje všechny komponenty potřebné pro transkripci, translaci a správné sbalení proteinu (například ribozomy, aminoacyl-tRNA syntetázy, iniciační a elongační faktory, ribozom uvolňující faktory, enzymy obnovující nukleotidy, chaperony, foldázy atd.). Aby docházelo k úspěšné produkci výsledného proteinu, CFPS systém obsahuje i katalytický modul nutný pro tvorbu energie spotřebovanou při syntéze polypeptidu, která je dodávána většinou ve formě molekul ATP. Poslední důležitou složkou systému, kromě templátu DNA nebo mRNA kódujícího žádaný protein, jsou substráty pro probíhající reakce jako aminokyseliny, nukleotidy, energetické substráty, kofaktory a anorganické soli³⁹.

3.3.1. Srovnání CFPS s ostatními metodami syntézy proteinů

Ribozomální syntéza proteinů (*in vivo* i *in vitro*) poskytuje vysokou přesnost inkorporace aminokyselin i relativně vysokou rychlost elongace, což umožňuje produkci dlouhých proteinových řetězců oběma těmito technikami. Naopak technika syntézy proteinů na pevné fázi umožňuje syntézu pouze kratších řetězců, navíc vyžaduje chráněné aminokyselinové deriváty a je technikou relativně pomalou⁴⁰. Srovnání těchto tří přístupů je uvedeno v Tab. 6.

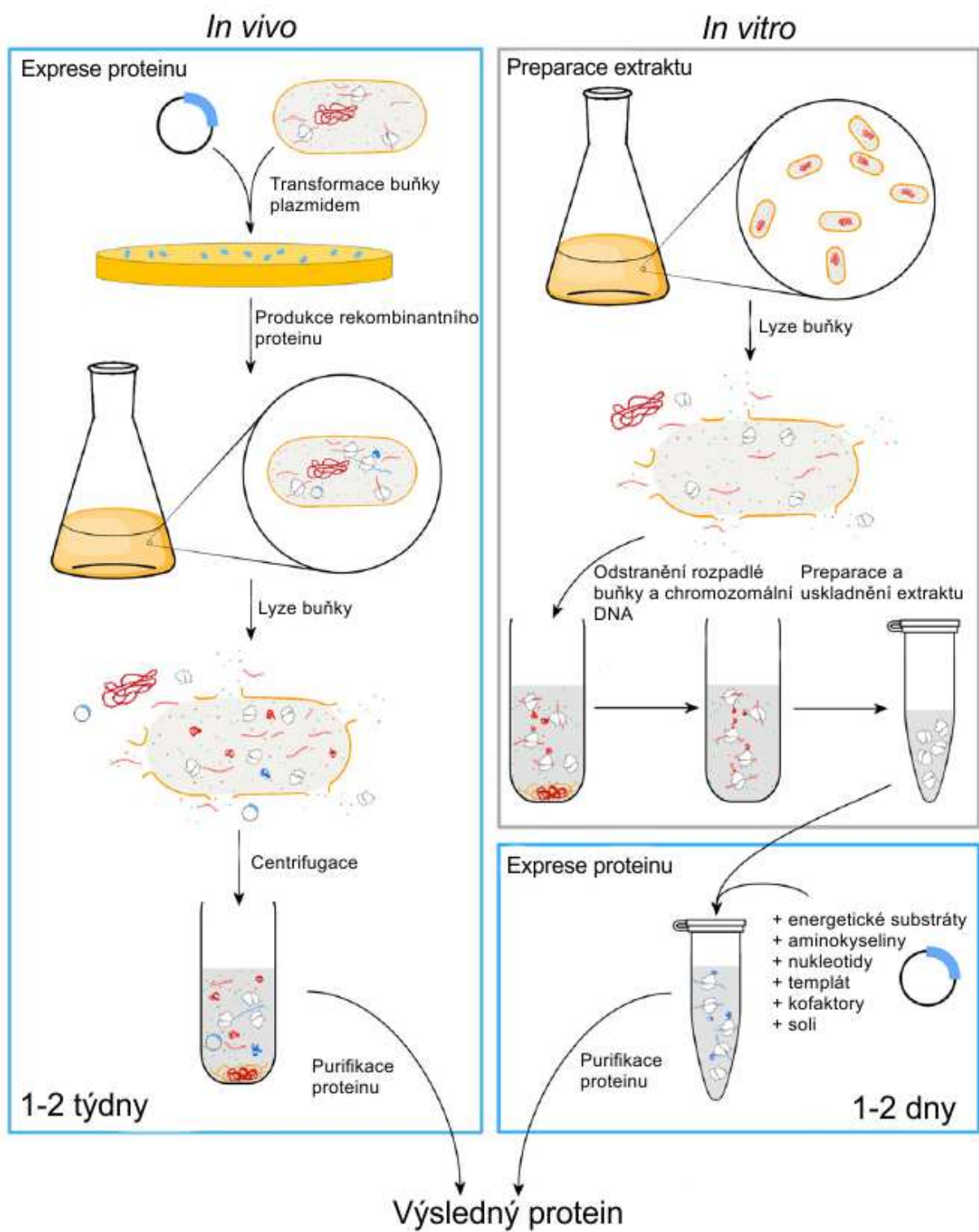
Tab. 6 Srovnání technik proteinové syntézy. Převzato a upraveno⁴⁰.

Technika	Substrát	Rychlost elongace [AMK/s]	Délka polypeptidového řetězce [AMK]
CFPS	20 α -aminokyselin	1 až 15	3 až 2 700
In vivo	Glukóza, anorganické soli	5 až 22	25 až 35 000
Neribozomální syntéza proteinů na pevné fázi	20 chráněných α -aminokyselin	0,001	2 až 40

AMK = počet aminokyselin inkorporovaných do polypeptidu

Pokud srovnáme *in vitro* a *in vivo* ribozomální syntézu mezi sebou, obě mají nesporně své výhody i nevýhody. Technika *in vitro* ve srovnání s „buněčnými továrnami“ umožňuje inkorporaci i jiných aminokyselin než pouze dvaceti geneticky kódovaných, dále mohou být produkovány i proteiny pro buňky toxické nebo buňkou vyhodnocené jako atypické. Také značení rekombinantních proteinů stabilními izotopy nebo radionuklidy není při *in vitro* expresi proteinů znehodnoceno post-translačními modifikacemi molekuly. *In vitro* metoda umožňuje produkci celých knihoven proteinů, zatímco *in vivo* exprese (závislá na transformaci genetického materiálu do buňky) by významně omezila variabilitu produkováných sekvencí. Další nespornou výhodou je doba trvání celého procesu exprese. Mezi jednu z největších nevýhod CFPS je produkce žádaných sloučenin pouze v malém měřítku⁴⁰.

Celý proces *in vitro* i *in vivo* syntézy proteinů je znázorněn na obrázku 3 (Obr. 3, strana 29).



Obr. 3 Celý proces *in vitro* a *in vivo* syntézy proteinů. Převzato a upraveno³⁹.

3.3.2. Systém PURE

Systém označovaný jako PURE (z angl. The **P**rotein synthesis **U**sing **R**ecombinant **E**lements) je oproti jiným systémům odlišný v tom, že proteinové složky systému jsou značené histidinovou kotvou a po expresi je lze odstranit z reakční směsi afinitní purifikací. Rekombinantní komponenty, dohromady 36 enzymů zapojených do procesu transkripce a translace, jsou přidány k ribozomům a transferovým ribonukleovým kyselinám, izolovaným z buněk speciálně pro tyto účely modifikovaných kmenů bakterie *Escherichia coli* a dalším nezbytným sloučeninám jako jsou nukleotidy, aminokyseliny, rekombinantní RNA polymeráza a modul generující molekuly ATP. Celý tento systém dohromady tvoří funkční translační aparát, ke kterému pro úspěšnou expresi proteinu stačí přidat jen templát DNA kódující žádaný protein^{40,41}.

Takto vytvořený systém není znečištěn proteázami a ribonukleázami, což jsou enzymy pro *in vitro* translaci a transkripci nežádoucí. Další výhodou PURE systému je, že nedochází ke kontaminaci reakčního systému membránami, které se z lyzátu těžko odstraňují⁴¹. Složení PURE směsi lze modifikovat - je možné vynechat například ribozom uvolňující faktory, PURE se proto často využívá ve spojení s metodou ribozomální display⁴⁰. Tato metoda však není vhodná pro přípravu vzorků ve velkém měřítku a také ji není možné využít pro syntézu proteinů, které vyžadují post-translační modifikace, jako například metylaci či fosforylaci⁴¹.

Systémy PURE jsou dnes komerčně dostupné (například PURESYSTEM, Cosmo Bio, Tokyo, Japonsko nebo PURExpress, New England Biolabs, Beverly MA, USA), takže jsou využitelné pro různé aplikace laboratorního výzkumu⁴⁰.

4. Výsledky a diskuze

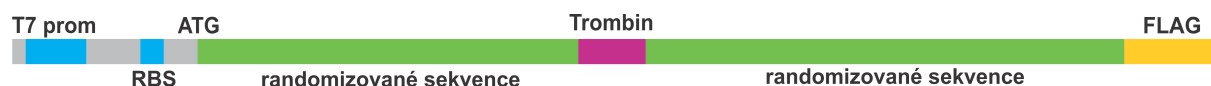
4. 1. Návrh knihoven

Tato práce se zabývá přípravou pěti knihoven s různým aminokyselinovým složením. Příslušné označení knihoven, kritérium výběru aminokyselin a konkrétní aminokyselinové složení knihoven je uvedeno v Tab. 7. Knihovny 10E, 14SL a 20F představují postupné evoluční etapy genetického kódu, alespoň podle dostupných studií (viz. Tab. 4, strana 20). Kontrolní knihovna 10R představuje výběr 10 aminokyselin na základě zachování co nejširšího spektra fyzikálně-chemických vlastností⁶ a knihovna 10L zahrnuje pouze pozdní aminokyseliny.

Tab. 7 Charakterizace studovaných knihoven

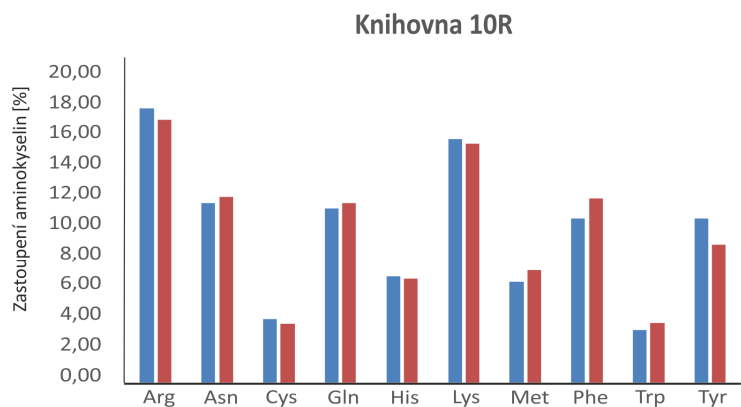
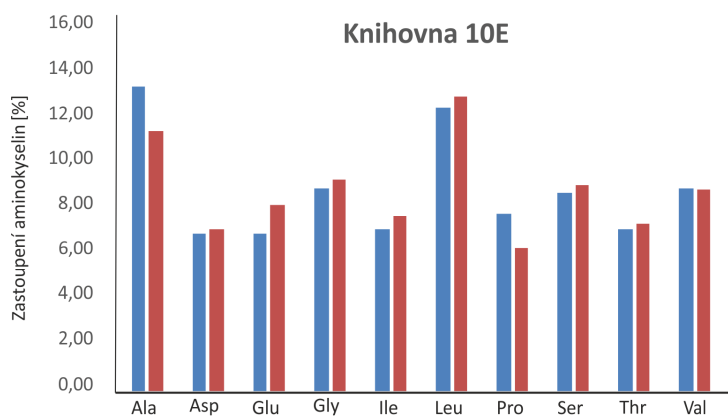
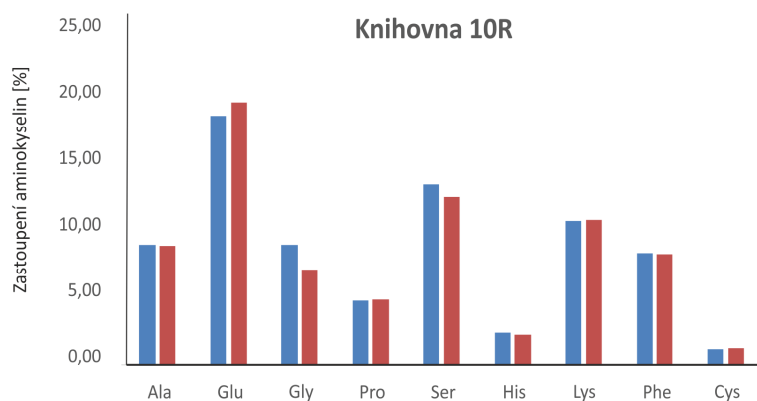
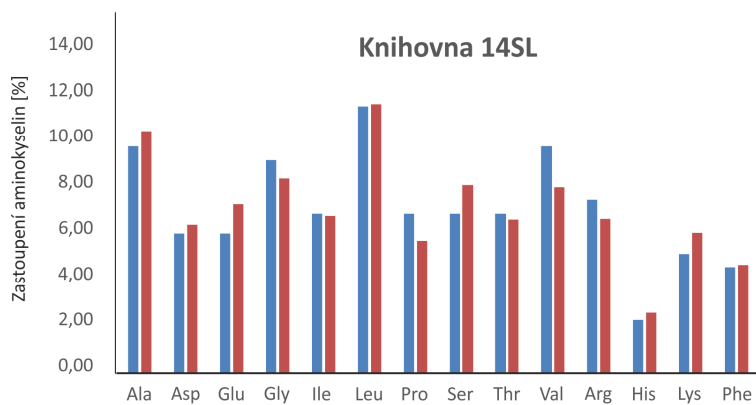
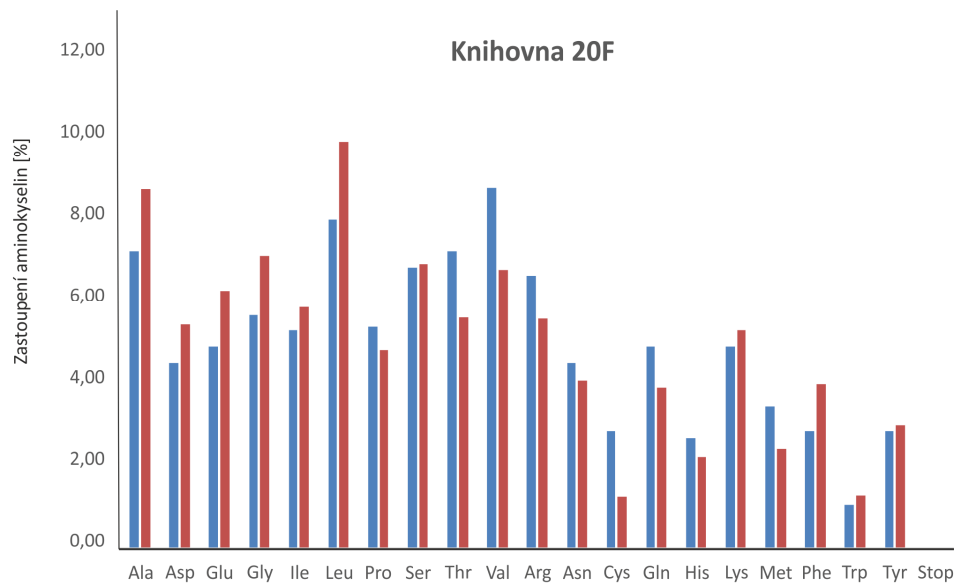
Označení knihovny	Kritérium výběru	Konkrétní aminokyseliny
20F	všechny proteinogenní AMK (z angl. Full)	A, G, R, N, D, C, Q, E, H, I, L, K, M, F, P, S, T, W, Y, V
10E	rané AMK (z angl. Early)	A, D, E, G, I, L, P, S, T, V
10L	pozdní AMK (z angl. Late)	R, N, C, Q, H, K, M, F, W, Y
14SL	rané AMK a čtyři evolučně následující (z angl. Semi-Late)	A, D, E, G, I, L, P, S, T, V, R, H, K, F
10R	co nejpestřejší výběr AMK z hlediska fyzikálně-chemických vlastností (z angl. Rational)	A, E, G, L, P, S, C, H, K, F

Jednotlivé sekvence tvořící DNA knihovny obsahují na svém začátku T7 promotor (T7 prom) pro iniciaci transkripce a místo pro nasednutí ribozomu (RBS), samotné randomizované sekvence začínají iniciačním kodonem ATG a v polovině obsahují jediný definovaný úsek, štěpné místo pro enzym trombin (Trombin). Tento úsek by měl později sloužit ke srovnání přibližné frekvence sbalování jednotlivých knihoven. Na svém C' konci tyto sekvence obsahují 1×FLAG kotvu (aminokyselinová sekvence DYKDDDDK) Celková délka sekvencí je 358 bází (Obr. 4, strana 32).



Obr. 4 Jednotlivé sekvence v DNA knihovnách

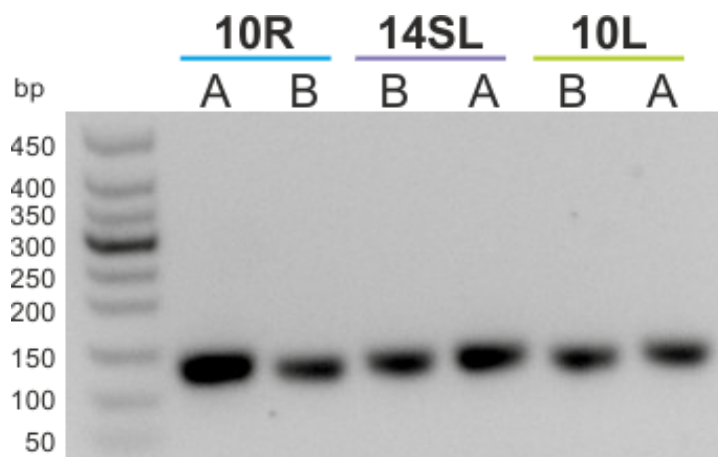
V randomizovaných úsecích, které kódují náhodné proteiny, bylo použito několik kombinací degenerovaných kodonů. Popis těchto kodonů je uveden v sekci 3.2. (viz Tab. 5, strana 24) Každý z těchto kodonů kóduje pouze určitý počet aminokyselin z 20 proteinogenních. Změnou zastoupení vybraných degenerovaných kodonů v randomizovaném úseku bylo dosaženo poměrově téměř shodného aminokyselinového složení, jaké pozorujeme v přírodě (viz Obr. 5, strana 33) a zároveň se zamezilo výskytu předčasných terminačních kodonů. Algoritmus pro výběr těchto sekvencí byl vyvinut kolegy v naší laboratoři ve spolupráci s Dr. K. Fujishimou (laboratoř Prof. Lynn Rothschild, NASA Ames Institut, USA). Tímto postupem bylo dosaženo variability v knihovnách přibližně 10^{11} .



Obr. 5: Procentuální zastoupení jednotlivých aminokyselin v experimentálních knihovnách 20F, 14SL, 10R, 10E a 10L (modrá) a jejich vzájemné poměry v biologických proteinech (červená, viz databáze Uniprot⁴²).

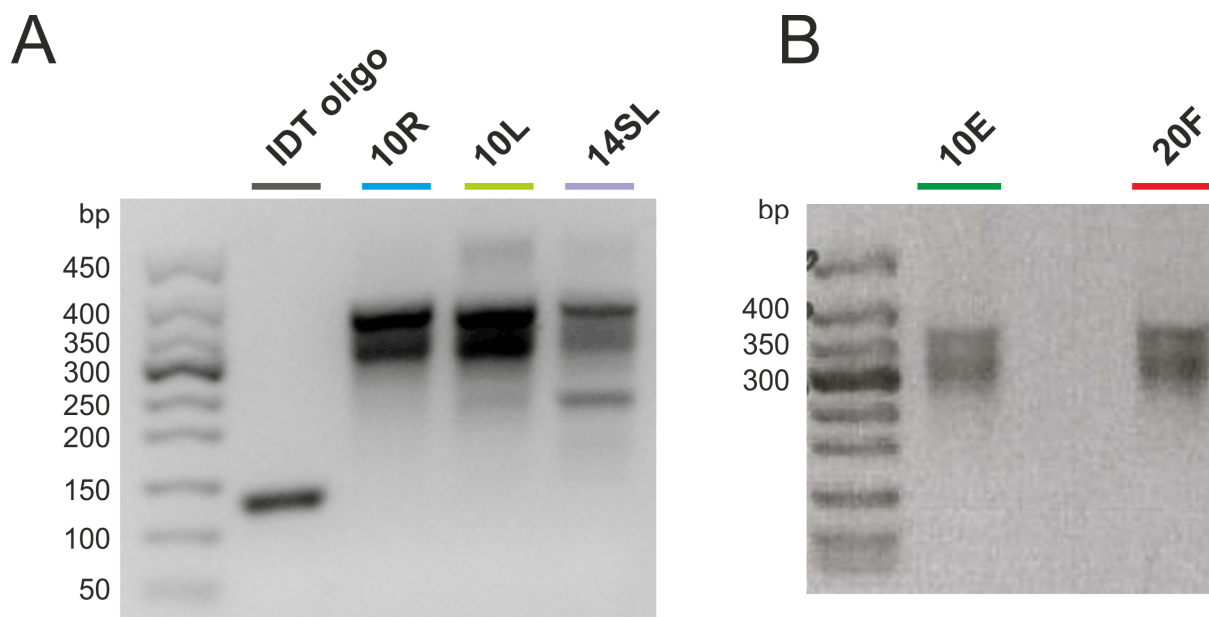
4.2. Příprava DNA knihoven

DNA knihovny 10R, 14SL a 10L byly připraveny vždy ze dvou komerčně syntetizovaných degenerovaných oligonukleotidů v naší laboratoři dle postupu popsáno v sekci 5.2.3.1. Na obrázku 6 jsou znázorněny zakoupené oligonukleotidy s překryvem v trombinovém štěpném místě, 191 (oligonukleotid A) a 189 (oligonukleotid B) bází dlouhé, od firmy IDT (USA).



Obr. 6: 2,5% agarózový gel znázorňující délku a kvalitu zakoupených oligonukleotidů. Zleva: 3 μ l markeru GelPilot 50 bp Plus Ladder; 10R, 14SL, 10L označení knihoven viz Tab.7, strana 31; A, B – 6 pmol navzájem komplementárních oligonukleotidů navržených pro každou z knihoven

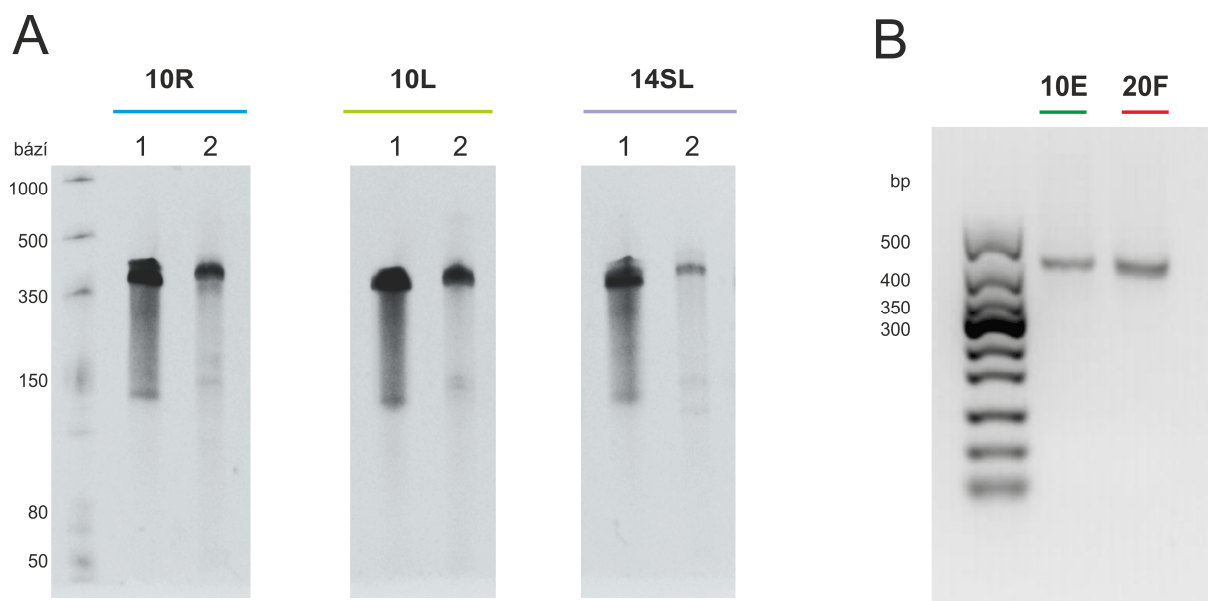
Na Obr. 7A, strana 35, jsou výsledné DNA knihovny. DNA knihovny s označením 20F a 10E byly syntetizovány stejným postupem ve spolupracující laboratoři Prof. Lynn Rothschild (NASA Ames Institut, USA). Tyto knihovny jsou zobrazeny na Obr. 7B, strana 35. Očekávaná délka výsledných knihoven je 358 bází. Sekvence všech oligonukleotidů použitých pro syntézu jsou uvedeny v dodatku práce.



Obr. 7: 2,5 % agarózový gel znázorňující **(A)** DNA knihovny **10R, 10L a 14SL** (označení knihoven viz Tab. 7, strana 31); zleva: 3 μ l markeru GelPilot 50 bp Plus Ladder; IDT oligo – 6 pmol jednoho ze zakoupených oligonukleotidů pro srovnání velikosti; 10R, 10L, 14SL – 6 pmol výsledných DNA knihoven; **(B)** DNA knihovny **10E a 20F**; zleva: 3 μ l markeru GelPilot 50 bp Plus Ladder; 10E, 20F – 6 pmol DNA knihoven

4.2.1. Připojení stabilizačních sekvencí

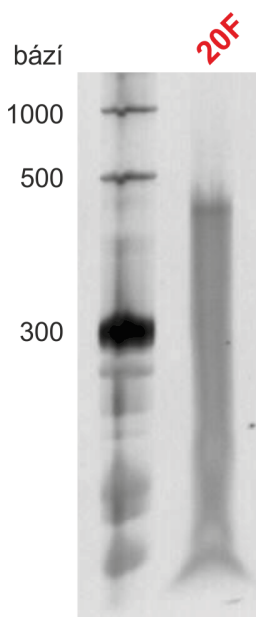
Vzniklé DNA knihovny začínají T7 promotorem na 5' konci a končí STOP kodonem. Pro stabilizaci knihoven po transkripci a jejich efektivní translaci byly navrženy primery zavádějící přídavné sekvence (zčásti poly-A) na oba konce sekvencí DNA knihoven. V případě nedokonalé komerční syntézy oligonukleotidů se tak zároveň zajistí přítomnost neporušené sekvence (viz. Obr. 4, strana 32) T7 promotoru a STOP kodonu na obou koncích knihovny. PCR bylo provedeno dle postupu popsaného v sekci 5.2.1.1. Výsledná délka sekvence se tak navýšila o 21 bp a celkem je templát před transkripcí tvořen 379 bp (viz. Obr. 8, strana 36).



Obr. 8: (A) 5% akrylamidový denaturující gel znázorňující knihovny 10R, 10L a 14SL (označení knihoven viz Tab. 7, strana 31) po syntéze (1) a po přidání stabilizačních sekvencí (2), 2,5 μ l low range ssRNA markeru (B) 2,5 % agarózový gel znázorňující knihovny 10E a 20F po přidání stabilizačních sekvencí, 3 μ l markeru GelPilot 50 bp Plus Ladder

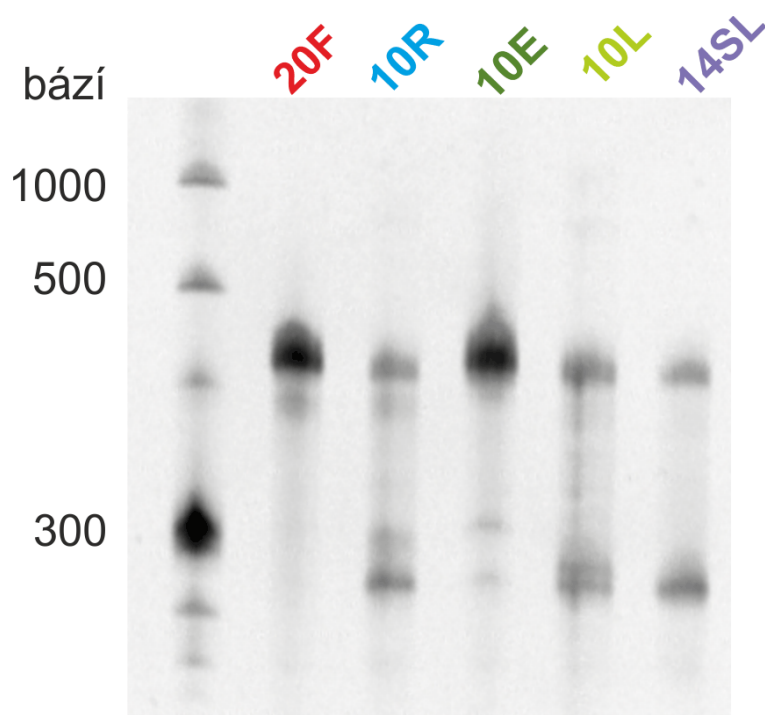
4.3. Transkripce knihoven

Nejprve byla reakční směs pro transkripci připravována za použití T7 RNA polymerázy od firmy New England Biolabs (NEB, USA) (viz sekce 5.2.3.2). Tento postup byl použit pouze pro jednu z knihoven (20F), přičemž vznikající RNA byla směsí různých délek (Obr. 9).



Obr. 9: 5% akrylamidový denaturující gel znázorňující produkt transkripce 20F knihovny T7 RNA polymerázou (NEB, USA). Zprava – 3 μ l Low Range ssRNA markeru; 20F – 500 ng transkribované knihovny

Druhý postup transkripce je popsán v sekci 5.2.3.2. Transkripce byla v tomto případě prováděna pomocí transkripčního kitu Ampliscribe T7 Flash Transcription (výrobce Epicentre, USA). Tímto postupem se podařilo připravit mRNA ze všech pěti knihoven. Vzorek mRNA přečištěné pomocí precipitace (viz sekce 5.2.3.3.) je znázorněn na Obr. 10. Transkripcí byla získána koncentrace mRNA pohybující se okolo 3 $\mu\text{g}/\mu\text{l}$. Tato koncentrace byla stanovena spektrofotometricky na přístroji NanoDrop (Thermo Scientific, USA). Celkový výtěžek reakce po precipitaci mRNA se tedy pohyboval okolo 60 μg (objem reakce 20 μl). Požadovaná délka výsledných transkriptů je 379 bází.

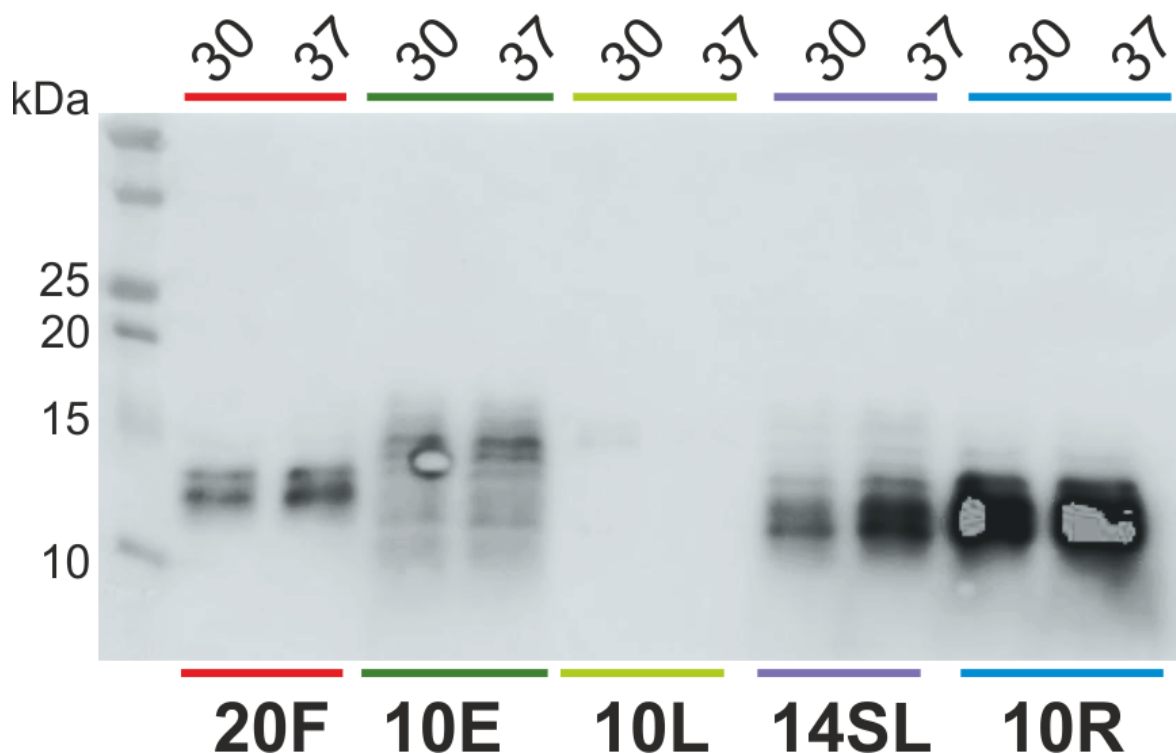


Obr. 10: 5% akrylamidový denaturující gel s RNA ze všech pěti knihoven. Zprava: 2,5 μl low range ssRNA markeru; 500 ng RNA z příslušných knihoven, označení viz Tab.7, strana 31.

4.4. Translace knihoven

Dle postupu popsaného v sekci 5.2.3.4. s využitím komerčního kitu PURExpress (NEB, USA) byla provedena optimalizace translace knihoven při teplotách 30 $^{\circ}\text{C}$ a 37 $^{\circ}\text{C}$. Pro tuto optimalizaci byla vždy provedena 1/5 uvedené reakce, tedy celkový objem reakce byl 5 μl . Analýza produkce proteinů byla provedena pomocí metody Western blot (viz sekce 5.2.2.2). Na obrázku 11 (Obr. 11, strana 38), lze vidět, že knihovny 20F, 14SL a 10R se produkují lépe při teplotě 37 $^{\circ}\text{C}$. Kvantifikace produkce byla provedena pomocí vzorku proteinu GCP II (enzym glutamát karboxypeptidáza II, z angl. glutamate carboxypeptidase II) s 1 \times FLAG kotvou o známé koncentraci (stanovené na základě

aminokyselinové analýzy), který nám laskavě poskytli pracovníci laboratoře doc. Jana Konvalinky z ÚOCHB, AV ČR, v. v. i. Koncentrace knihoven 20F a 10E byly odhadnuty při obou teplotách na 2 mg/l. Knihovna 10L se produkovala pouze při teplotě 30 °C, ale velmi slabě (výsledná koncentrace byla odhadnuta na < 0,2 mg/l). U knihovny 14SL lze vidět nejmarkantnější rozdíl mezi produkcemi při různých teplotách. Při teplotě 37 °C byla koncentrace odhadnuta na 20 mg/l. Nejlépe se produkovala knihovna 10R, při obou teplotách o koncentraci zhruba 200 mg/l.



Obr. 11: Western blot s translačním produktem všech knihoven. Zleva: 2,5 µl markeru Precision Plus Protein Dual Color Standard, produkty translace při dvou teplotách (30 °C, 37 °C) všech pěti knihoven (označení viz Tab. 7, strana 31). Na gel byla nanesena vždy celá reakce (5 µl).

4.5. Diskuze

Cílem práce bylo připravit 5 proteinových knihoven o různém aminokyselinovém složení pro jejich následnou biofyzikální charakterizaci. Práce tedy zahrnovala syntézu DNA knihoven, transkripci do mRNA a bezbuněčnou translaci pro získání proteinové knihovny.

Důležitým parametrem pro syntézu knihoven byl vzájemný poměr vybraných aminokyselin a také fixní délka výsledných bílkovin. Knihovny byly syntetizovány vždy ze dvou komerčně syntetizovaných oligonukleotidů, které obsahovaly v oblasti kódující randomizované úseky degenerované kodony. Tyto sekvence byly navrženy za pomoci algoritmu vyvinutého spolupracovníky v naší laboratoři. Tak bylo například zabráněno nežádoucímu výskytu STOP kodonů uvnitř sekvencí. Kromě sekvencí pro randomizované úseky oligonukleotidy obsahovaly rovněž sekvence potřebné pro následnou translaci (T7 promotor, ribozomální vazebné místo), purifikaci knihoven (C-terminální FLAG kotvu) a charakterizaci (rozpoznávací sekvenci pro trombin).

Spojením komerčně syntetizovaných oligonukleotidů jsme získali všech pět DNA knihoven. Jejich výsledná délka se podle analýzy na agarózové elektroforéze pohybovala v rozmezí 300-400 bp, zatímco očekávaná délka DNA knihovny je 358 bp. Toto rozmezí může být dáno různým složení oligonukleotidů variabilní knihovny v porovnání s markery molekulové hmotnosti. V případě knihovny 14SL byl ale patrný i produkt o délce okolo 250 bp, což může svědčit o snížené kvalitě 14SL DNA knihovny.

Naším původním záměrem bylo převést knihovny ve formě DNA molekul do formy proteinů pomocí kitu PURExpress (NEB, USA), který byl navržen pro vkládání DNA templátu do reakce. V našem případě tento postup přípravy selhal, což mohlo být způsobeno malou účinností prvního kroku reakce, tedy transkripce. Navíc tento kit neobsahuje v základní reakci inhibitory RNáz a při jejich absenci může docházet k odbourávání mRNA knihoven a snižování tak celkové účinnosti reakce. Proto byla DNA forma knihoven nejprve transkribována do formy mRNA a až tento templát byl v optimálním množství vkládán do reakce, a to v přítomnosti inhibitorů RNáz. Zároveň byla před tímto krokem sekvence DNA knihoven obohacena na obou koncích stabilizačními sekvencemi pomocí PCR pro zvýšení stability vzniklé RNA a pro zajištění správnosti funkční sekvence na obou koncích, které mohly být neúplné po prvotní syntéze oligonukleotidů.

Transkripce byla prováděna nejprve pomocí T7 RNA polymerázy (NEB, USA), avšak tento postup byl neúspěšný – výsledná RNA byla směsí různých délek, mohla tedy být degradovaná

anebo byl zčásti neúspěšný již krok transkripce (viz Obr. 9, strana 36). Později byl zakoupen kit pro transkripci Ampliscribe T7 Flash Transcription (výrobce Epicentre, USA) a tímto postupem byla produkce mRNA úspěšnější. Důvodem mohla být přítomnost již zmíněného inhibitoru proti RNázám, které mohly kontaminovat vzorek DNA knihoven. Výrobce uvedené soupravy udává, že z vloženého templátu o hmotnosti 0,5 μg je dvouhodinovou inkubací při 37 $^{\circ}\text{C}$ možné vyrobit až 8,2 $\mu\text{g}/\mu\text{l}$ RNA. Naše výtěžky se pohybovaly okolo 3 $\mu\text{g}/\mu\text{l}$ RNA. Výsledné množství transkriptu je velmi variabilní a závisí na kvalitě vkládaného templátu a i na jeho velikosti, protože kratší templáty, jako je tomu v našem případě, se dle výrobce produkuje s menší efektivitou. Nižší výtěžky v případě knihoven mohou být způsobeny také vysokou variabilitou prepisovaných sekvencí. U knihoven 14SL, 10R a 10L byly patrné i kratší transkripty, než je požadovaná délka (379 nukleotidů). Je možné, že u některých sekvencí těchto knihoven docházelo k předčasné terminaci transkripce. Z nich vzniklé proteiny by ovšem neobsahovaly FLAG kotvu a nebyly tedy detekovatelné pomocí Western blotu. Po následné purifikaci (pomocí FLAG kotvy) by tedy ani neznečistily výsledný preparát.

Po vložení mRNA templátu do reakce kitu PURExpress (NEB, USA) se produkovaly všechny knihovny. Na Obr. 11 (strana 38) můžeme vidět širší spektrum produktů s různou molekulovou hmotností. Nejvíce je to patrné u knihovny 10E, což pravděpodobně odráží opět různé aminokyselinové složení peptidů v knihovně. Toto pozorování lze považovat za důkaz, že alespoň částečná variabilita knihoven zůstala zachována.

Jako průměrný výtěžek translační reakce s kitem PURExpress (NEB, USA) výrobce uvádí výslednou koncentraci proteinu mezi 10-200 mg/ml . Při provedení optimalizační reakce při dvou teplotách se nám odhadem u jedné z knihoven (10R) podařilo dosáhnout i horní hranice uvedené meze. Knihovna 14SL dosahuje výtěžku odhadem 20 mg/l , knihovny 20F a 10E 2 mg/l a 10L má jen hraničně detekovatelný výtěžek (odhadem 0,2 mg/l). Rozdíly mezi produkcí jednotlivých knihoven mohou být dány už uváděnou variabilitou knihoven na úrovni DNA templátu používaného pro transkripci, dále pak do reakce vstupujícím mRNA templátem, kde jsou patrné rozdíly v kvalitě transkriptu (viz Obr. 10, strana 37). Dalším možným důvodem různého množství produkovaných proteinů je také jejich stabilita. Například knihovna 10L bude mít v důsledku svého složení zřejmě výraznou tendenci k agregaci a je možné, že část produktu se tedy vysráží již během translační reakce (vzorek byl před analýzou na SDS PAGE vždy centrifugován). Zároveň lepší výtěžky u knihoven s omezeným množstvím aminokyselin při zachování repertoáru fyzikálně chemických vlastností (10R a 14SL) v porovnání s knihovnou 20F jsou v souladu s dřívější studií, ve které byla pozorována vyšší rozpustnost

proteinů podobného složení. Její autoři navrhovali nižší komplexitu aminokyselinového složení proteinů jako lepší zdroj sekvencí pro potřeby proteinového inženýrství a biotechnologie⁴³.

Výše zmíněné rozdíly v rozpustnosti/agregaci studovaných knihoven zřejmě také ovlivní proces jejich následné purifikace. V rámci předběžných pokusů jsme provedli purifikaci na základě molekulové hmotnosti (pro odstranění ribozomálního komplexu byla využita ultracentrifugace), dále reverzní purifikaci s využitím histidinové kotvy (kterou jsou označeny komponenty kitu PURExpress) a přímou purifikaci s využitím FLAG-kotvy, kterou jsou označeny proteiny z knihoven. Tyto pokusy nebyly předmětem předkládané práce. Každá z purifikovaných knihoven (20F, 14SL a 10E) se chovala rozdílně a pro úspěšnou purifikaci bude nutná optimalizace protokolu pro každou z knihoven zvlášť. Jako nejlepší protokol se zatím jeví přímá purifikace s využitím FLAG-kotvy v kombinaci s detergenty. Než budou knihovny po purifikaci biofyzikálně charakterizovány (např. měření cirkulárního dichroismu a štěpení trombinem), bude třeba otestovat jejich variabilitu. Je pravděpodobné, že variabilita knihoven se při PCR a při přepisu z DNA do mRNA snížila. Toto ověření bude provedeno sekvenací transkribované mRNA všech používaných knihoven.

5. Experimentální práce

5.1. Přístroje a chemikálie

5.1.1. Pufry a ostatní roztoky

- TAE pufr: 57,1 ml 99% CH₃COOH, 100 ml 0,5 M EDTA, 242 g Tris-HCl, doplněno vodou na 1l, pH 8
- 44% roztok akrylamidu: 42,8 g akrylamidu, 1,2 g bisakrylamidu, doplněno vodou do 100 ml
- Pufr A: 250 ml HPLC vody, 116,25 g akrylamidu, 7,5 g bisakrylamidu
- Zásobní roztok pro akrylamidovou denaturující elektroforézu: 10 g močoviny, 2 ml 10× koncentrovaného TBE pufru, 1 ml Pufru A, 1 ml 44% akrylamidu, doplněno vodou na 20 ml
- TBE pufr: 89 mM Tris-HCl (pH 7,6), 89 mM kyselina boritá, 2 mM EDTA
- 6×vzorkový pufr: 350 mM Tris-HCl, pH 6,8, 30% (v/v) glycerol, 350 mM SDS, 4% (v/v) 2 mM 2-merkптоethanol, 180 μM bromfenolová modř
- Elektrodový pufr: 25 mM Tris-HCl, 250 mM glycin, 0,1% SDS, pH 6

- Přenosový pufr: 10% (v/v) methanol, 25 mM Tris-HCl, 192 mM glycin, 0,1% SDS, pH 8,2
- 10×PBS pufr: 1,37 M NaCl, 27 mM KCl, 100 mM Na₂HPO₄, 18 mM KH₂PO₄, pH neupravováno
- Promývací pufr: 1×PBS pufr
- Roztok kaseinu: Casein Blocker (ThermoScientific) rozředěn v poměru 1:20 v promývacím pufru

5.1.2. Přístroje

- Amersham Imager 600, GE Healthcare (UK)
- aparatura na horizontální elektroforézu Owl, Thermo Scientific (USA)
- aparatura na vertikální elektroforézu Bio-Rad (USA)
- blokový termostat MB – 102, Bioer Technology (Čína)
- centrifugy
 - Eppendorf 5415 R (Německo)
 - Biofuge Pico, Heraeus Instrument (Německo)
 - Allegra® X-15R, Beckman Coulter (USA)
 - Sorvall Evolution RC, Thermo Scientific (USA)
- inkubátor IPP 400, Memmert (Německo)
- laboratorní váhy HL-400, A&D Engineering, Inc. (USA)
- pH metr pH 50+ Benchtop, XS Instruments (Itálie)
- skenovací přístroj Typhoon 9410 Variable Mode Imager, GE Healthcare (USA)
- spektrofotometr NanoDrop 1000, Thermo Scientific (USA)
- UV transiluminátor QUANTUM ST4-1100, Vilber Lourmat (Německo)
- zdroj napětí PowerPac™ Basic power supply, Bio-Rad (USA)

5.1.3. Chemikálie

Bio-Rad, USA

- marker All Blue Precision Plus Protein Standard
- marker Precision Plus Protein Dual Color Standard

Biotium, Hayward, USA

- fluorescenční barvivo GelRed

Epicetre, USA

- AmpliScribe™ T7-Flash™ Transcription Kit

Fluka, Německo

- 2-merkapt ethanol
- Akrylamid
- Dodecylsír an sodný
- TEMED

New England Biolabs (NEB), Ipswitch, USA

- DNA Polymerase I, Large (Klenow) Fragment
- Phusion DNA polymeráza
- Low Range ssRNA marker
- PURExpress
- 2×RNA Loading Dye
- směsný roztok deoxynukleotidů (dNTPs)
- T7 RNA polymeráza

Penta, Praha, Česká republika

- Agaróza
- Boritá kyselina
- Dihydrogenfosforečnan draselný
- Ethanol
- Ethylendiamintetraoctová kyselina
- Chlorid draselný
- Methanol
- Močovina

Qiagen, Německo

- GelPilot 100 bp Plus Ladder
- 5×GelPilot Loading Dye

Serva, Heidelberg, Německo

- Bromfenolová modř

Sigma-Aldrich, Německo

- Anti-FLAG protilátka
- Bisakrylamid
- Glycerol
- Hydrogen fosforečnan sodný dihydrát
- Chlorid sodný
- Octan amonný
- Peroxodisíran draselný
- Tris-HCl

Takara

- PCR buffer

ThermoFisher Scientific, USA

- Casein Blocker
- Nanášecí pufr pro agarózovou elektroforézu: 6×DNA Loading Dye
- SuperSignal West Femto Maximum Sensitivity Substrate Chemiluminiscent Solution

USB, Cleveland, USA

- Glycin

Zymo Research, USA

- DNA (PCR) Clean-up & Concentrator kit

5.2. Obecné postupy

5.2.1. Práce s DNA a RNA

5.2.1.1. Polymerázová řetězcová reakce (PCR)

PCR reakce byla použita k připojení primeru 1 a primeru 2 k připravené DNA knihovně a pro amplifikaci DNA templátu.

Postup PCR:

Do PCR mikrokumavky byla pipetována reakční směs o celkovém objemu 50 μl . Složení reakční směsi je uvedeno v Tab. 8.

Tab. 8 Složení reakční směsi pro PCR reakci

Složka	Objem
DNA knihovna	0,5 μl (= 50 ng)
dNTPs (10 mM)	10 μl
DNA polymeráza Phusion	0,5 μl
DNA primery (primer 1 a primer 2) (10 μM)	po 2,5 μl
Pufř k DNA polymeráze Phusion, 5 \times koncentrovaný	10 μl
Sterilní voda	24 μl

Tab. 9 Průběh fází PCR reakce

Fáze	Teplota [$^{\circ}\text{C}$]	Čas [s]	Počet cyklů
1	98	30	1
2	98	10	30
3	55	15	
4	72	15	
5	72	300	1
6	4	30	1

Po PCR reakci byla DNA přečištěna pomocí kitu DNA (PCR) Clean-up & Concentrator (Zymo Research, USA).

Popsaného PCR postupu bylo využito při připojení přímých a zpětných stabilizačních sekvencí k DNA knihovně. Sekvence primerů jsou uvedeny v dodatku práce.

5.2.1.2. Horizontální agarózová elektroforéza

2,5% agarózový gel byl připraven rozpuštěním 1 g agarózy ve 40 ml 1×koncentrovaného TAE (viz sekce 5.1.1.) pufru. Aby se agaróza rozpustila, byla směs zahřáta v mikrovlnné troubě. Poté byla směs ochlazená na teplotu asi 50 °C a byly přidány 4 µl fluorescenčního barviva GelRed (BIOTIUM, USA, 10000× ve vodě). Směs byla promíchána a nalita na elektroforetickou desku. Do tekuté směsi byl vložen elektroforetický hřeben, kvůli vytvoření příslušných jamek k nanesení vzorků. Gel ztuhl při pokojové teplotě a poté byl hřeben vyjmut, gel byl vložen do elektroforetické vany a přelit 1× TAE pufrům po rysku (viz složení pufrů v sekci 5.1.1.). Do první jamky bylo nanášeno 5 µl markeru GelPilot 100bp Plus Ladder (Qiagen, Německo). Do ostatních jamek byly nanášeny vzorky s nanášecím pufrům 1×GelPilot Loading Dye (Qiagen, Německo). Elektroforéza probíhala při napětí 120 V a byla ukončena, když vzorky putující čelem doputovaly do tří čtvrtin délky gelu. Po ukončení elektroforézy byl gel umístěn na UV transiluminátor a pod UV světlem při 312 nm byly pozorovány DNA fragmenty.

5.2.1.3. Akrylamidová denaturující elektroforéza

Pro provedení akrylamidové elektroforézy byla nejprve sestavena aparatura a poté byl připraven gel složený z 6 ml zásobního roztoku pro akrylamidovou denaturující elektroforézu (viz složení pufrů v sekci 5.1.1., 10 µl TEMED a 60 µl 10% APS. Tento gel byl nalit až po okraj mezi dvě očištěná skla umístěná v držáku, a nakonec do něj byl okamžitě ponořen elektroforetický hřeben. Po ztuhnutí gelu byl hřeben vyjmut, skla byla umístěna do elektroforetické aparatury a převrstvena 1× TBE pufrům (viz sekce 5.1.1.). Do krajní jamky byl nanášen Low Range ssRNA marker (NEB, USA). K jednotlivým vzorkům byl přidán RNA loading Dye (NEB, USA) a také byly nanášeny do jamek v gelu. Elektroforéza probíhala při napětí 260 V a byla ukončena přibližně po 20 minutách, jakmile modrá barva doputovala ke spodnímu okraji separačního gelu. Poté byl gel přenesen do 50 ml 1× koncentrovaného TBE pufru a bylo k němu přidáno 5 µl fluorescenčního barviva GelRed (BIOTIUM, USA, 10000× ve vodě). Po patnáctiminutové inkubaci byl umístěn na UV transiluminátor a pod UV zářením při 312 nm byly pozorovány RNA fragmenty.

5.2.2. Práce s proteiny

5.2.2.1. SDS elektroforéza

Stejně jako u akrylamidové denaturující elektroforézy byla nejprve sestavena aparatura a poté byly připraveny gely, jejichž složení je uvedeno v Tab. 10.

Tab. 10 Složení jednotlivých gelů pro SDS elektroforézu

Separační gel (18%)		Zaostřovací gel (6,6%)	
Složka	Objem	Složka	Objem
1,5 M Tris-HCl, pH 8,8	2,5 ml	1 M Tris-HCl, pH 6,8	1,25 ml
44% AA	4,0 ml	44% AA	750 μ l
10% SDS	100 μ l	10% SDS	50 μ l
10% APS	100 μ l	10% APS	50 μ l
TEMED	10 μ l	TEMED	10 μ l
doplněno destilovanou vodou do 10 ml		doplněno destilovanou vodou do 5 ml	

Složky 10% APS a TEMED byly vždy přidány až těsně před nalitím gelu mezi dvě skla. Separační gel byl pipetován zhruba do tří čtvrtin výšky skel a poté byl převrstven ethanolem. Po 30 minutách stání gelu za laboratorní teploty byl ethanol odlit a byl napipetován zaostřovací gel na již ztuhlý separační gel. Okamžitě byl vsunut hřeben pro vytvarování jamek potřebných k nanášení vzorků. Po dalších 30 minutách polymerace gelu byl hřeben vyjmut a skla byla umístěna do elektroforetické aparatury a přelita elektrodovým pufr (složení viz sekce 5.1.1.). K jednotlivým vzorkům byl přidán vzorkový pufr (složení viz sekce 5.1.1.). Takto připravené vzorky byly zahřáty na 98 °C po dobu 5 minut. Vzorky a marker byly nanášeny na zaostřovací gel. Elektroforéza byla připojena ke zdroji napětí 200 V po dobu přibližně 1 hodiny. Elektroforéza byla ukončena, jakmile bromfenolová modř doputovala ke spodnímu okraji separačního gelu. Gel byl barven negativním barvením, nejprve 10 minut v 0,2 M imidazolu, poté 0,3 M síranem zinečnatým do objevení elektroforeogramu. Barvení bylo zastaveno promytím v destilované vodě. Pokud byl gel přenášen na membránu, byl odbarven 2% kyselinou citronovou třikrát po dobu tří minut.

5.2.2.2. Elektropřenos proteinů na nitrocelulózovou membránu s chemiluminiscenční detekcí protilátky

Při elektropřenosu metodou Western blot byly proteiny nejprve separovány metodou SDS elektroforézy popsanou výše (viz 5.2.2.1.). Poté byla sestavena aparatura pro elektropřenos a gel byl přichycen k nitrocelulózové membráně (7,5 x 5 cm) tak, aby mezi nimi nebyl žádný vzduch, a upevněn v aparatuře. Aparatura byla až po rysku naplněna přenosovým pufrem a připojena ke zdroji napětí 100 V po dobu jedné hodiny. Následně byla nespécifická místa na membráně blokována v roztoku kaseinu (složení viz sekce 5.1.1.) po dobu jedné hodiny při laboratorní teplotě za stálého míchání. Membrána byla inkubována v roztoku anti-FLAG protilátky (5000× ředěna, koncentrace zásobního roztoku 1 mg/ml, Sigma-Aldrich, Německo) v 5 ml roztoku kaseinu (viz sekce 5.1.1.). V připraveném roztoku byla membrána inkubována přes noc při 4 °C za stálého míchání. Po ukončení inkubace byla membrána třikrát po dobu pěti minut promyta v promývacím pufru (viz sekce 5.1.1.). Membrána byla vyvolána 600 µl vyvolávacího roztoku SuperSignal West Femto Maximum sensitivity substrate chemiluminiscent solution (ThermoFisher, USA). Membrána byla inkubována s vyvolávacím roztokem po dobu pěti minut. Nakonec byl blot vizualizován v přístroji Amersham Imager 600 (GE Healthcare, UK).

5.2.3. Práce s knihovnami

5.2.3.1. Příprava knihoven

Navržené randomizované oligonukleotidy (výrobce IDT, USA, každý 0,05 mg) byly naředěny na koncentraci 20 µM. Reakce nejprve probíhala bez účasti polymerázy během teplotního gradientu od 92 °C do 10 °C. Teplota se snižovala vždy po 5 minutách skokem o 4 °C. Posledním krokem bylo chlazení po dobu 5 minut při 4 °C. Složení reakční směsi je uvedeno v Tab. 11, strana 49.

Tab. 11 Reakční směs

Složka	Objem
oligonukleotidy (výsledná koncentrace 2 μ M)	2 \times 5 μ l
dNTPs (10 mM)	4 μ l
PCR pufr (Takara, 10 \times koncentrovaný)	5 μ l
sterilní voda (prostá RNáz a DNáz)	32 μ l

Následně byly ke směsi přidány 2 μ l polymerázy DNA Polymerase I, Large (Klenow) Fragment (NEB, USA, 5000 units/ml) a reakce probíhala dle Tab. 12.

Tab. 12 Teplotní a časové fáze reakce po přidání polymerázy

Fáze	Teplota [$^{\circ}$ C]	Čas [min]
1	25	5
2	37	60
3	50	15
4	4	5

Celkový objem reakce i s přidanou polymerázou byl 50 μ l. Nakonec byla výsledná DNA přečištěna kitem DNA (PCR) Clean-up & Concentrator (Zymo Research, USA).

5.2.3.2. Transkripce knihoven

Nejprve bylo složení reakční směsi pro transkripci knihoven připravováno podle Tab. 13, strana 50. Celkový objem reakční směsi byl 20 μ l. Polymeráza byla do směsi přidána až nakonec. Reakční směs byla inkubována 1 h při 37 $^{\circ}$ C.

Tab. 13 Složení reakční směsi pro transkripci

Složka	Objem
10× koncentrovaný pufr k T7 RNA polymeráze (NEB, USA)	2 μ l
NTPs (10mM)	4 μ l
T7 RNA polymeráza (NEB, USA, 50000 units/ml)	2 μ l
DNA templát	0,8 μ l (=400 ng)
sterilní voda (prostá RNáz a DNáz)	11,2 μ l

Později byla transkripce prováděna pomocí Ampliscribe T7 Flash Transcription kitu (výrobce Epicentre, USA), složení reakční směsi je uvedeno v Tab. 14. Reakce byla míchána při laboratorní teplotě v uvedeném pořadí jako je uvedeno v Tab. 14. Celkový objem jedné reakce byl 20 μ l. Reakční směs byla v tomto případě inkubována po dobu 2 h při 37°C.

Tab. 14 Složení reakční směsi při transkripci pomocí kitu

Složka	Objem
sterilní voda (prostá RNáz a DNáz)	3,3 μ l
DNA templát (120 ng/ μ l)	3 μ l
10× koncentrovaný pufr (Ampliscribe T7 Flash Transcription kit)	2 μ l
NTPs (100 mM)	7,2 μ l
DTT (100 mM)	2 μ l
RNAzový inhibitor (Ampliscribe T7 Flash Transcription kit)	0,5 μ l
T7 Flash roztok enzymů	2 μ l

5.2.3.3. Precipitace RNA

Ke standardní reakci pomocí Ampliscribe T7 Flash Transcription kitu (Epicentre, USA) (20 μ l) bylo přidáno 20 μ l 5M octanu amonného. Poté byla směs inkubována na ledu po dobu 15 minut. Po ukončení inkubace byla směs centrifugována (10000 \times g) 15 minut při 4 °C. Vzniklá peleta byla promyta 70% ethanolem, který se po centrifugaci (10000 \times g, 15 minut při 4°C) a odebrání supernatantu nechal odpařit, a RNA byla rozpuštěna ve sterilní vodě. Orientační koncentrace RNA po precipitaci byla měřena spektrofotometricky.

5.2.3.4. Translace knihoven

V této práci byl využit k translaci knihoven kit PURExpress (NEB, USA). Celkový objem jedné reakce byl 25 μ l. Složení reakční směsi je uvedeno v Tab. 15. Vždy byl přidáván roztok B do roztoku A a nakonec RNA templát. Reakce byla míchána na ledu. Reakční směs byla inkubována při 30 °C, nebo 37 °C po dobu 4 hodin. Poté byl vzorek centrifugován (10000 \times g) po dobu 15 minut.

Tab. 15 Složení reakční směsi pro translaci

Složka	Objem
Roztok A	10 μ l
Roztok B	7,5 μ l
RNA templát (1,2 μ g/ μ l, výsledné množství v reakci 3 μ g)	2,5 μ l
sterilní voda (prostá RNáz a DNáz)	Doplněno na celkový objem 25 μ l

6. Závěr

Byla zpracována rešerše o evoluci genetického kódu a o bezbuněčné expresi proteinů.

Podářilo se úspěšně připravit mRNA od všech pětĩ navržených knihoven. Dále se podařilo převést všechny knihovny do formy proteinů, avšak knihovnu 10L pouze s velmi nízkým výtěžkem. Čtyři z knihoven tedy mohou být použity k purifikaci a následným charakterizacím.

7. Poděkování

Nejprve bych ráda poděkovala své školitelce Mgr. Kláře Hloučové, Ph.D za pomoc, ochotu, trpělivost, cenné rady a podnětné prostředí v její laboratoři.

Dále bych ráda poděkovala Mgr. Vjačeslavu Treťjačenkovi a Mgr. Heleně Jindrové za jejich neocenitelnou pomoc a rady během mé laboratorní práce i při vypracovávání této práce.

Děkuji také doc. RNDr. Janu Konvalinkovi, CSc. za jeho vřelý přístup ke studentům a příležitost, které se mi u něj dostalo.

V neposlední řadě děkuji také své rodině, především Mgr. Jiřímu Štoplovi, za velkou podporu během celé doby mého studia.

8. Literatura

- (1) Burton, A. S.; Stern, J. C.; Elsilá, J. E.; Glavin, D. P.; Dworkin, J. P. *Chem. Soc. Rev.* **2012**, *41*, 5459–5472.
- (2) Kejnovský, E. *Tajemství genů: Od vzniku života po genom člověka*; Academia, 2015, pp 25-44. ISBN 978-80-200-2478-7.
- (3) Miller, S. L. *Science* **1953**, *117* (3046), 528–529.
- (4) Higgs, P. G.; Pudritz, R. E. *Astrobiology* **2009**, *9* (5), 483–490.
- (5) Trifonov, E. N. *Gene* **2000**, *261* (1), 139–151.
- (6) Murphy, L. R.; Wallqvist, A.; Levy, R. M. *Protein Eng.* **2000**, *13* (3), 149–152.
- (7) Riddle, D. S.; Santiago, J. V.; Bray-Hall, S. T.; Doshi, N.; Grantcharova, V. P.; Yi, Q.; Baker, D. *Nat. Struct. Biol.* **1997**, *4*, 805–809.
- (8) Crick, F. H. C. *Symp Soc Exp Biol* **1958**, *12*, 138–163.
- (9) Freeland, S. *In Amino Acids, Peptides and Proteins in Organic Chemistry*; 2009; Vol. 1, pp 43–75. ISBN 978-3-527-32096-7.
- (10) Snustad, D. P.; Simmons, M. J. *Genetika*; Masarykova univerzita: Brno, 2009, pp 333. ISBN 978-80-210-4852-2.
- (11) Gamow, G. *Nature* **1954**, *173*, 318.
- (12) Zhu, W.; Freeland, S. *J. Theor. Biol.* **2006**, *239*, 63–70.
- (13) Santos, M. A. S.; Tuite, M. F. *Nucleic Acids Res.* **1995**, *23* (9), 1481–1486.
- (14) Crick, F. H. C.; Orgel, L. E. *Icarus* **1973**, *19* (3), 341–346.
- (15) Barrell, B. G.; Bankier, A. T.; Drouin, J. *Nature* **1979**, *282* (5735), 189–194.
- (16) Crick, F. H. C. *J. Mol. Biol.* **1968**, *38*, 367–379.
- (17) Koonin, E. V.; Novozhilov, A. S. *IUBMB Life* **2009**, *61* (2), 99–111.
- (18) Wong, J. T. *Orig. Life Evol. Biosph.* **2007**, *37*, 403–408.
- (19) Yarus, M.; Caporaso, J. G.; Knight, R. *Annu. Rev. Biochem.* **2005**, *74*, 179–189.
- (20) Ptáčková, B. Komplementační peptid LacZ-alfa jako nástroj pro studium molekulární evoluce, 2015. Bakalářská práce.
- (21) Lazcano, A. *Cold Spring Harb. Lab. Press* **2010**, 1–17.
- (22) Miller, S. L. *J. Am. Chem. Soc.* **1955**, *77* (7), 2351–2361.
- (23) Johnson, A. P.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. *Science* **2008**, *322*, 1.
- (24) Cleaves, H. J. *J. Theor. Biol.* **2010**, *263* (4), 490–498.

- (25) Pizzarello, S.; Huang, Y.; Fuller, M. *Geochim. Cosmochim. Acta* **2004**, 68 (23), 4963–4969.
- (26) Miller, S. L. *Organic Geochemistry*; Springer US, 1993; pp 625–637. ISBN 978-1-4613-6252-4.
- (27) Trifonov, E. N. *J. Biomol. Struct. Dyn.* **2004**, 22 (1), 1–11.
- (28) Weber, A. L.; Miller, S. L. *J. Mol. Evol.* **1981**, 17 (5), 273–284.
- (29) Luisi, P. L. *THE EMERGENCE OF LIFE: From Chemical Origins to Synthetic Biology*; Cambridge University Press, 2006; pp 68-70. ISBN 13 978-0-521-82117-9.
- (30) de Duve, C. *Life Evolving: Molecules, Mind, and Meaning*; Oxford University Press, 2002; pp 98. ISBN 978-0195156058.
- (31) Chiarabelli, C.; De Lucrezia, D.; Franchi, M.; Gallori, E.; Luisi, P. L. *Chem. Biodivers.* **2006**, 3 (8), 869–877.
- (32) Denault, M.; Pelletier, J. N. *Methods Mol. Biol.* **2006**, 352, 127–154.
- (33) Golynskiy, M. V; Iii, J. C. H.; Morelli, A.; Morrone, D.; Seelig, B. *Methods Mol. Biol.* **2013**, 978, 73–92.
- (34) Cho, G.; Keefe, A. D.; Liu, R.; Wilson, D. S.; Szostak, J. W. *J. Mol. Biol.* **2000**, 297, 309–319.
- (35) Scott, J. K.; Smith, G. P.; Scorr, J. K.; Smith, G. P. *Science* **2017**, 249 (4967), 386–390.
- (36) Fujishima, K.; Venter, C.; Wang, K.; Ferreira, R.; Rothschild, L. J. *Scientific Reports* **2015**, 5, 12–16.
- (37) Tsuji, T.; Onimaru, M.; Yanagawa, H. *Nucleic Acids Res.* **2001**, 29 (20), 97e–97.
- (38) Baker, M. *Nat. Methods* **2011**, 8, 457–460.
- (39) Carlson, E. D.; Gan, R.; Hodgman, C. E.; Jewett, M. C. *Biotechnol. Adv.* **2012**, 30 (5), 1185–1194.
- (40) Whittaker, J. W. *Biotechnol. Lett.* **2013**, 35 (2), 143–152.
- (41) Kuruma, Y.; Ueda, T. *Nat. Protoc.* **2015**, 10 (9), 1328–1344.
- (42) Consortium, U. *Nucleic Acids Res.* **2017**, 45 (D1), D158–D169.
- (43) Tanaka, J.; Yanagawa, H.; Doi, N. *PLoS One* **2011**, 6 (3).

Dodatek

Sekvence oligonukleotidů použitých pro syntézu knihoven

20F

>oligonukleotid A

CTGTAATACGACTCACTATAGGGACACCAATAGAGAAAGAGGAGAAATACTAGATGVWKVWK
NBGSHRVWKNNTDNCNBGDNCNBGVVAVVASHRVWKVWVADNCVWVADNCDNCVVDN
CDNCNBGNNTDNCNHRVVANNTVWKDNCVVASHRVVANBGDNCGCGTTGGTTCCACGTGGCA
GC

>oligonukleotid B

TTTTTTTTTTCATTACTTATCGTCGTCATCTTTATAGTCANNTBBMWBCVNGNHANNTBBCVN
GNHTBBANNMWBGNHCVNYDSANNTBBTBBCVNMWBANNCVNTBBMWBMWBANNTBBMWBTB
BYDSGNHANNGNHANNTBBGNHMWBGNHMWBMWBYDSCVNTBBGNHYDSMWBGNHGCTGCCA
CGTGGAACCAACGC

10E

>oligonukleotid A

CTGTAATACGACTCACTATAGGGACACCAATAGAGAAAGAGGAGAAATACTAGATGMYYYYR
GMNRBYGMNYYRGNYYRGMNYYRBYRBYYYRGMNRBYGMNGMNMYYGNRBYYYRYRGN
NRBYMYYYRBYGNMYYYYRBYRBYGMNGMNGMNMYYGMNGCGTTGGTTCCACGTGGCA
GCG

>oligonukleotid B

TTACTTATCGTCGTCATCTTTATAGTCYRRRVYNKCNKCNKCNCRKRKYRRKNNCYRRNK
CRVYYRRRKRKRKCRVYYRRNKCRKRKRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYRKYR
VYNKCYRRNKCRKNNCYRRRVYYRRNRCYRRYRRRKYRNNCGCTGCCACGTGGAACCAA
CGC

10R

>oligonukleotid A

CTGTAATACGACTCACTATAGGGACACCAATAGAGAAAGAGGAGAAATACTAGATGYTVRAR
GRRCWCKCMGRRGMAYTVGMARARYTVRARTYVGMAGMATYVSCATYVSCAYTVTYVSCAGR
RSCARARTKCGRRTKCRARGRRSCAYTVRARRARYTVSCASCARARGCGTTAGTCCCAGCGTG
GGAGC

>oligonukleotid B

CTATTACTTATCATCATCATCTTTATAATCYCYTYBRAYTYYYCGWGDAAGWGYCTGSYT
YGWGBRABRATGSBRAYTYGWGYCKGMYCBARKGMBRABARTGSTKCBARGWGMABARY
TYTYGMAYYCBRAYTYGWGTGSKGMTGSTGSYTYTGSYTYTGSYTYTGSYTYTGSYTYTGSY
CGC

10L

>oligonukleotid A

CTGTAATACGACTCACTATAGGGACACCAATAGAGAAAGAGGAGAAATACTAGATGTWCYGG
TWCAABTWCTWCTWCMAGMAGYGGAABYGGTWCAGAKGMAGMAGTWCMAGMAGYGGCRWCR
WADGAKGMATMATTWCMATTWCADGYGGTKYMAGTKYAABTWCADGGCGTTAGTCCCGCGTG
GGAGC

>oligonukleotid B

CTATTACTTATCATCATCATCTTTATAATCCTKATKGWACMTVTTGWACMTCMTGWACTKCM
TCMTCMTCTKCHTATKATKCCRGWAWYGWYGCTKRMACMTCTKCTKVTTCTKCCRCMTCTKW
YGWYGATKCCRATKGWACHTVTTCTKRMAGWAVTTVTTCTKGWAGCTCCCACGCGGGACTAA
CGC

14SL

>oligonukleotid A

CTGTAATACGACTCACTATAGGGACACCAATAGAGAAAGAGGAGAAATACTAGATGWYYMKC
SRTSRTYYDRWAGNCSRTRVRRWAGNCWYYGHSSRTGNCWYMKCWYYYDGHSRVRGHSN
CGNCYYDGNCSRTGNCYDGHSMKCMYTYYSRTRVRRWARVRSRTGCGTTAGTCCCGCGTG
GGAGC

>oligonukleotid B

CTATTACTTATCATCATCATCTTTATAATCHRRARKTWYRRWDAHBYBTWYSDCSDCTWYTW
YARKGNCGNCYBYAYSDAHTWYARKSDCHRRYBYAYSBYSDCRRWHRRGNCGMKTWYSDCG
NCHRRYBYGNCGMKAYSAYSBYSDCHRRGNCAYSTWYBYARKGCTCCCACGCGGGACTAA
CGC

Sekvence primerů použitých pro přidání stabilizačních sekvencí ke knihovnám

> primer 1

AAAAAGTGGCCTGTAATACGACTCACTATAGG

>primer 2

TTTTTTTTTCACTTATCGTCGTCATCTTTATAGTC