

MASARYKOVA UNIVERZITA
Přírodovědecká fakulta

BAKALÁŘSKÁ PRÁCE

Brno 2017

Lenka Libichová



**MASARYKOVA
UNIVERZITA**
PŘÍRODOVĚDECKÁ FAKULTA
Ústav experimentální biologie
Oddělení genetiky a molekulární
biologie



Metody konstrukce chytrých knihoven mutantních enzymů v proteinovém inženýrství

Bakalářská práce

Lenka Libichová

Bibliografický záznam

Autor: Lenka Libichová
Přírodovědecká fakulta, Masarykova univerzita
Ústav experimentální biologie

Název práce: Metody konstrukce chytrých knihoven mutantních enzymů
v proteinovém inženýrství

Studijní program: Experimentální biologie

Studijní obor: Molekulární biologie a genetika

Vedoucí práce: Mgr. Radka Chaloupková, Ph.D.

Akademický rok: 2017

Počet stran: 55

Klíčová slova: Chytré knihovny; cílená mutageneze; enzymy; semi-rationální design; proteinové inženýrství; halogenalkandehalogenasy

Bibliographic Entry

Author: Lenka Libichová
Faculty of Science, Masaryk University
Department of Experimental Biology

Title of Thesis: Methods for construction of smart mutant libraries in protein engineering

Degree Programme: Experimental Biology

Field of Study: Molecular Biology and Genetics

Supervisor: Mgr. Radka Chaloupková, Ph.D.

Academic Year: 2017

Number of Pages: 55

Keywords: Smart libraries; focused mutagenesis; enzymes; semi-rational design; protein engineering; haloalkane dehalogenases

Abstrakt

Tato práce se soustředí na modifikaci a vylepšování vlastností enzymů pomocí metod proteinového inženýrství. Nejprve jsou stručně popsány tři hlavní přístupy proteinového inženýrství: řízená evoluce, racionální design a semi-racionální design. Vývoj v oblasti molekulární biologie a bioinformatiky vede k upřednostňování semi-racionálního designu a metod cílené řízené evoluce za účelem konstrukce tzv. chytrých knihoven, které vykazují vyšší kvalitu než rozsáhlé knihovny generované při náhodné mutagenezi. Navýšení poměru vylepšených a testovaných variant zvyšuje pravděpodobnost získání biotechnologicky zajímavého enzymu. Hlavním cílem práce je vytvořit přehled metod konstrukce chytrých knihoven a vzájemně je porovnat. Přičemž většina z nich je založena na saturační mutagenezi a využití tzv. redukované aminokyselinové abecedy, která zvyšuje kvalitu a snižuje velikost konstruované knihovny. Halogenalkandehalogenasy slouží v této práci jako modelové enzymy. Je navržena redukovaná aminokyselinová abeceda pomocí počítačového nástroje CASTER určená pro souběžnou saturační mutagenezi tří pozic v sekvenci enzymu DhaA. Na základě této redukované aminokyselinové abecedy dochází k radikálnímu snížení velikosti výsledné teoretické knihovny.

Abstract

This thesis focuses on modification and improvement of enzymatic properties by methods of protein engineering. Firstly, three main approaches of protein engineering are briefly described: directed evolution, rational design and semi-rational design. Developments in molecular biology and bioinformatics lead to the prioritization of the semi-rational design and methods of targeted and directed evolution. The semi-rational design allows design of so-called smart libraries with better quality rather than large libraries generated by random mutagenesis. Increase in the proportion of better and tested variants leads to higher probability of obtaining biotechnologically interesting variants of enzymes. The main goal of this thesis is to provide an overview and comparison of current methods for construction of smart libraries. These methods are often based on the saturation mutagenesis and the use of reduced amino acid alphabet that enhances quality and reduces the size of constructed libraries. Haloalkane dehalogenases serve as model enzymes. A reduced amino acid alphabet is designed for simultaneous saturation mutagenesis in three positions of amino acid sequence of protein DhaA by CASTER computer tool. This reduced amino acid alphabet radically decreases the size of final theoretical library.



MASARYKOVA UNIVERZITA

Přírodovědecká fakulta

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

Akademický rok: 2016/2017

Ústav: Ústav experimentální biologie

Studentka: Lenka Libichová

Program: Experimentální biologie

Obor: Molekulární biologie a genetik

Ředitel Ústavu experimentální biologie PŘF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje bakalářskou práci s názvem:

Název práce: Metody konstrukce chytrých knihoven mutantních enzymů v proteinovém inženýrství

Název práce anglicky: Methods for construction of smart mutant libraries in protein engineering

Oficiální zadání:

Enzymy se uplatňují v narůstajícím počtu biotechnologických procesů díky tomu, že jejich vlastnosti (stabilitu, aktivitu, specifitu a selektivitu) lze vylepšovat metodami proteinového inženýrství. Vývoj v oblasti molekulární biologie, počítačového modelování a bioinformatiky vede k upřednostňování semi-rationálního designu a metod cílené řízené evoluce před náhodnou mutagenézí. Semi-rationální design umožňuje navrhnout tzv. chytré knihovny, což jsou malé knihovny obsahující mutace v místech s potenciálem ovlivňovat danou vlastnost enzymu a kódující sadu racionálně vybraných substitucí. Chytré knihovny vykazují vyšší kvalitu než rozsáhlé knihovny generované při náhodné mutagenézi, neboť navýšení poměru vylepšených a testovaných variant vede ke zvýšení pravděpodobnosti získání biotechnologicky zajímavého enzymu. V rámci této bakalářské práce bude vypracován přehled současných metod přípravy chytrých knihoven. Metody budou vzájemně porovnány a vybraná metoda bude použita k experimentálnímu designu modelového enzymu.

Jazyk závěrečné práce: čeština

Vedoucí práce: Mgr. Radka Chaloupková, Ph.D.

Konzultant: prof. Mgr. Jiří Damborský, Dr.
Mgr. Táňa Koudeláková, Ph.D.

Datum zadání práce: 14. 10. 2016

V Brně dne: 8. 11. 2016

Souhlasím se zadáním (podpis, datum):

Lenka Libichová
studentka

Mgr. Radka Chaloupková, Ph.D.
vedoucí práce

prof. RNDr. Jan Šmarda, CSc.
ředitel Ústavu experimentální
biologie

Poděkování

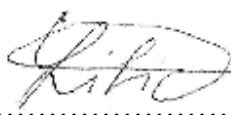
V první řadě bych chtěla poděkovat doktorce Táně Koudelákové, která přispěla s množstvím přínosných rad a připomínek a vždy si ochotně našla čas se mi věnovat. Velice děkuji také doktorce Radce Chaloupkové za skvělé vedení této bakalářské práce a rovněž za užitečné podmínky k jejímu zpracování.

V závěru děkuji i panu profesoru Jiřímu Damborskému a výzkumné skupině Loschmidtových laboratoří za rozvoj mých znalostí ve vědecké praxi a především v oblasti molekulární biologie. Nabyté znalosti značně přispěly k vypracování mé bakalářské práce.

Prohlášení

Prohlašuji, že jsem svoji bakalářskou práci vypracovala samostatně s využitím informačních zdrojů, které jsou v práci citovány.

Brno 4. května 2017



.....

Lenka Libichová

Obsah

ÚVOD.....	10
1 ENZYMY.....	11
1.1 Funkce a vlastnosti enzymů.....	11
1.2 Struktura enzymů.....	12
1.2.1 Primární struktura.....	12
1.2.2 Sekundární struktura.....	15
1.2.3 Terciární struktura.....	16
1.2.4 Kvartérní struktura.....	16
2 PROTEINOVÉ INŽENÝRSTVÍ.....	18
2.1 Racionální design.....	19
2.2 Řízená evoluce.....	20
2.3 Semi-racionální design.....	21
3 CHYTRÉ KNIHOVNY.....	22
3.1 Bioinformatické nástroje pro konstrukci chytrých knihoven.....	22
3.2 Redukce aminokyselinové abecedy.....	24
3.3 Význam PCR pro konstrukci chytrých knihoven.....	26
3.4 Metody tvorby chytrých knihoven.....	27
3.4.1 Saturační mutageneze.....	28
3.4.1.1 Saturační mutageneze v jednom místě.....	28
3.4.1.2 Iterativní saturační mutageneze.....	31
3.4.1.3 Souběžná saturační mutageneze.....	32
3.4.2 ISOR.....	33
3.4.3 Parsimoniální mutageneze.....	35
3.4.4 OSCARR.....	36
3.4.5 MAX randomizace.....	38
4 MODELOVÉ ENZYMY A NÁVRH EXPERIMENTÁLNÍHO DESIGNU.....	41
4.1 Modelové enzymy.....	41
4.2 Vlastnosti halogenalkandehalogenas.....	41

4.3 Halogenalkandehalogenasa DhaA31.....	42
4.4 Návrh redukce aminokyselinové abecedy.....	44
ZÁVĚR.....	47
SEZNAM POUŽITÉ LITERATURY.....	48

Úvod

Enzymy jsou katalytické proteiny, které jsou podstatnou součástí všech živých organizmů. Hrají klíčovou roli v existenci a funkci biologických struktur a dějů. Vlastnosti enzymů, jako vysoká specifita a selektivita, umožňují jejich široké využití v biotechnologickém průmyslu, medicíně či při ochraně životního prostředí. Ostatní vlastnosti, jako funkce a stabilita za fyziologických podmínek, však mohou jejich využití limitovat.

Funkci a stabilitu enzymů za reakčních podmínek definovaných zamýšlenou aplikací lze vylepšit pomocí proteinového inženýrství. Náhodné či cílené změny na úrovni DNA se promítají v modifikované struktuře enzymů a v žádoucím ovlivnění jejich enantioselektivity či substrátové specifity, aktivity se substrátem zájmu nebo stability během katalytické reakce. V současné době se v rámci proteinového inženýrství rozvíjí hlavně přístup zvaný semi-rationální design využívající dostupných informací o struktuře a funkci enzymů k přípravě tzv. chytrých knihoven. Cílení mutací do míst, které mají potenciál ovlivnit klíčové vlastnosti enzymů, vede k významnému snížení počtu jejich testovaných variant, a šetří tak čas i náklady na získání biotechnologicky zajímavých biokatalyzátorů.

Cílem této práce je vytvořit stručný přehled metod proteinového inženýrství využitelných při konstrukci chytrých knihoven mutantních enzymů, a připravit tak teoretický základ pro design praktického experimentu. Bude navržena chytrá knihovna halogenalkandehalogenasy DhaA31 pro vylepšení její aktivity vůči substrátu 1,2,3-trichlorpropanu.

1 Enzymy

Enzymy jsou proteiny s katalytickou aktivitou. Řadí se mezi biopolymery, které se skládají z aminokyselin vzájemně propojených peptidovou vazbou. Enzymy určují povahu a rychlost chemických reakcí, a řídí tak většinu biochemických procesů v živých organismech. Pro efektivní modifikaci a vylepšení enzymů na úrovni DNA je nezbytné znát nejen jejich funkci, ale také strukturu a strukturně-funkční vztahy.

1.1 Funkce a vlastnosti enzymů

Enzymy jsou reakčně specifické – katalyzují určitý typ chemické reakce. Důsledkem evoluce vznikly také enzymy specifické pro určitý substrát či skupinu podobných substrátů. V ostatních případech se enzymy ukazují spíše jako promiskuitní. Za substrátovou specifitu odpovídá hlavně tvar a velikost substrátu, vzhled a umístění aktivního místa enzymu a jejich vzájemná komplementarita. Svou roli hraje také náboj či přístupové cesty do aktivního místa a jeho hydrofobní nebo hydrofilní charakter. Specifita enzymu pro určitý substrát je důležitá pro jeho funkci a využití.

Enantioselektivita enzymů našla své uplatnění hlavně v chemickém a farmaceutickém průmyslu. Souvisí s chirálním charakterem některých substrátů, kdy se jedna chemická sloučenina může vyskytovat ve dvou různých enantiomerních formách (R/S), z nichž pouze jedna vykazuje požadovaný biologický účinek. Rozlišit dvě různé enantiomerní formy je důležité například při výrobě léčiv nebo činidel.

Pro praktickou aplikaci enzymů je velmi podstatné zlepšení jejich stability. Většina enzymů vykazuje optimální aktivitu ve vodném roztoku při neutrálním pH a za mírných teplotních podmínek. Výjimkou jsou enzymy pocházející z tzv. mezofilních organismů, které si zachovávají svoji funkci i při vyšších teplotách, což je výhodné pro uplatnění v biotechnologickém průmyslu. Patří mezi ně například halogenalkandehalogenasy, o kterých se zmíním později v této práci.

Aktivita enzymů spočívá v ovlivnění rychlosti chemických reakcí snižováním jejich aktivační energie. Souvisí s koncentrací substrátu, teplotou, pH prostředí či přítomností solventů, aktivujících nebo inhibujících látek. Aktivitu enzymů je možné ovlivnit na úrovni jejich struktury modifikací přístupových cest nebo aktivního místa.

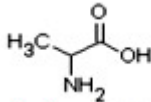
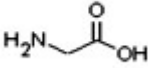
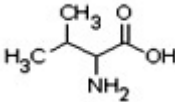
1.1 Struktura enzymů

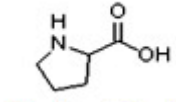
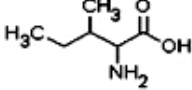
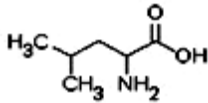
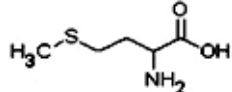
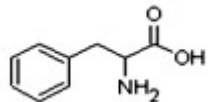
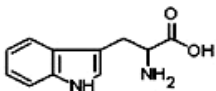
Funkce a vlastnosti enzymů jsou podmíněny jejich strukturou. Strukturu proteinů lze klasifikovat do čtyř úrovní: primární, sekundární, terciární a kvartérní (Obr. 1).

1.1.1 Primární struktura proteinů

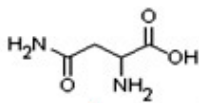
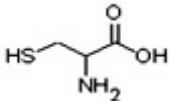
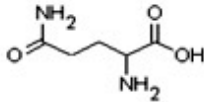
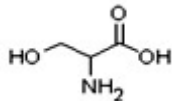
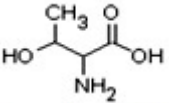
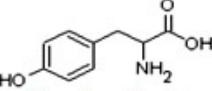
Primární struktura proteinů je dána pořadím a počtem aminokyselin v polypeptidovém řetězci. Jedná se o lineární řetězec aminokyselin, který se zapisuje od volné aminoskupiny (N-konce) po volnou karboxylovou skupinu (C-konec). Aminokyseliny jsou mezi sebou propojeny kovalentní peptidovou vazbou, a tvoří tak proteinovou páteř. Ty aminokyseliny, ze kterých se proteiny skládají, nazýváme proteinogenní. Jedná se především o L- α -aminokyseliny. Dělíme je na základě jejich polaritě a náboje na nepolární (Tab. 1), polární (Tab. 2), kyselé (Tab. 3) a bazické (Tab. 4).

Tabulka 1 - Nepolární aminokyseliny (hydrofobní)

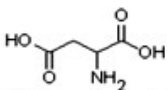
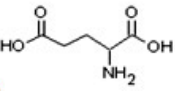
Název	Zkratka	Vzorec	Vlastnosti	Význam
Alanin	Ala, A		<ul style="list-style-type: none">- nejčastější- prostorově nevýrazná- málo reaktivní- vzácně nese specifickou funkci	<ul style="list-style-type: none">- hydrofobní interakce- tvorba a stabilizace α-helixů- otevření tunelu
Glycin	Gly, G		<ul style="list-style-type: none">- nejjednodušší- prostorově nevýrazná- opticky neaktivní- evolučně stabilní- nemá zcela nepolární charakter	<ul style="list-style-type: none">- nižší výskyt glycinu vede k vyšší stabilitě proteinu
Valin	Val, V		<ul style="list-style-type: none">- málo reaktivní- poměrně rigidní	<ul style="list-style-type: none">- hydrofobní interakce- nepodporuje konformaci α-helix

Prolin	Pro, P		<ul style="list-style-type: none"> - rigidní - imino skupina umožňuje prolinu konformaci cis (u ostatních aminokyselin je trans) 	<ul style="list-style-type: none"> - vysoce stabilizující (vyšší stabilita = nižší aktivita) - nepodporuje konformaci α-helix a β-list - výskyt v β-ohybech
Isoleucin	Ile, I		<ul style="list-style-type: none"> - málo reaktivní - větvení řetězce na β-uhlíku 	<ul style="list-style-type: none"> - hydrofobní interakce - omezuje konformační volnost - výskyt uvnitř proteinu
Leucin	Leu, L		<ul style="list-style-type: none"> - málo reaktivní - větvení řetězce na γ-uhlíku 	<ul style="list-style-type: none"> - hydrofobní interakce - větší konformační volnost - preferuje konformaci α-helix před β-listem - hydrofobní interakce - výskyt uvnitř proteinu - tvorba strukturálních motivů
Methionin	Met, M		<ul style="list-style-type: none"> - donor methylové skupiny 	<ul style="list-style-type: none"> - iniciační („start“) kodon (AUG) - vyskytuje se vzácně - koenzym mnoha methylačních enzymů
Fenylalanin	Phe, F		<ul style="list-style-type: none"> - prostorově výrazná - aromatická - aromatický kruh funguje jako chromofor 	<ul style="list-style-type: none"> - hydrofobní interakce - uzavírá tunely a zabraňuje průniku vody (vyšší aktivita)
Tryptofan	Trp, W		<ul style="list-style-type: none"> - prostorově výrazná - aromatická 	<ul style="list-style-type: none"> - uzavírá tunely - může vytvářet molekulární brány

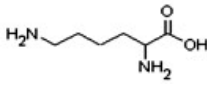
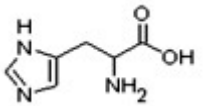
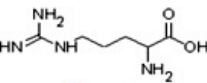
Tabulka 2 - Polární aminokyseliny (bez náboje)

Název	Zkratka	Vzorec	Vlastnosti	Význam
Asparagin	Asn, N		- amin kyseliny asparagové	- donor/akceptor protonu - vyskytuje se na povrchu proteinu - počátek a konec α -helixů, ve smyčkách β -listů - protein-proteinové interakce - interakce v aktivním místě
Cystein	Cys, C		- thiolová -SH skupina	- tvorba disulfidických můstků - stabilizace terciární struktury
Glutamin	Gln, Q		- amid kyseliny glutamové	- vyšší konformační entropie ve srovnání s asparaginem
Serin	Ser, S		- hydroxylová -OH skupina	- vyskytuje se v aktivním místě (serinové proteasy)
Threonin	Thr, T		- hydroxylová -OH skupina - 2 chirální uhlíky	- podléhá často postranlační modifikaci
Tyrosin	Tyr, Y		- prostorově výrazná - aromatická	- uzavírá tunely

Tabulka 3 - Kyselé aminokyseliny (v neutrálním prostředí negativně nabitě)

Název	Zkratka	Vzorec	Vlastnosti	Význam
Kyselina asparagová	Asp, D		- hydrofilní - více rigidní než kyselina glutamová	- iontové interakce - vyskytuje se ve vazebných a aktivních místech
Kyselina Glutamová	Glu, E		- hydrofilní	- iontové interakce - vyskytuje se ve vazebných a aktivních místech

Tabulka 4 - Bazické aminokyseliny (v neutrálním prostředí pozitivně nabitě)

Název	Zkratka	Vzorec	Vlastnosti	Význam
Lysin	Lys, K		- prostorově výrazná	- iontové interakce - nezbytný stavební prvek proteinů
Histidin	His, H		- prostorově výrazná - aromatická - běžná	- iontové interakce - histidinová kotva (detekce a afinitní purifikace) - přenáší náboj (snadná protonace a deprotonace) - vyskytuje se ve vazebných a aktivních místech
Arginin	Arg, R		- prostorově výrazná - bazická guanidinová skupina	- iontové interakce

1.2.2 Sekundární struktura proteinů

Sekundární struktura proteinů představuje sbalení polypeptidového řetězce na základě vodíkových interakcí mezi vodíkovým atomem amidové skupiny a atomem kyslíku karbonylové skupiny hlavního proteinového řetězce (Pauling a kol., 1951). Rozlišujeme tři základní typy stavebních motivů: α -helixy, β -skládané listy, smyčky a ohyby.

Prvním z motivů, α -helix, je uspořádán do šroubovice. Na jednu otočku šroubovice připadá zhruba 3,6 aminokyselin, které jsou uspořádány tak, že jejich postranní řetězce směřují ze šroubovice ven. Tento strukturní motiv tvoří především aminokyseliny methionin, alanin, leucin, glutamová kyselina a lysin. Naopak prolin, glycin či asparagová kyselina mají velmi nízké předpoklady pro uspořádání do α -helixu. Velmi rigidní prolin láme řetězec a vytváří smyčky. Naopak glycin je schopný rozrušit struktury helixu na základě své konformační flexibility. V uspořádání α -helixu se tudíž tyto aminokyseliny vyskytují velice zřídka (Davies, 1964, Guzzo, 1965). V β -skládaném listu propojují vodíkové můstky dva paralelní řetězce, a tak dochází ke stabilizaci struktury. Řetězce bývají často stočené do tvaru písmene U. Na rozdíl od α -helixu je tento motiv méně kompaktní.

Ohyby („bend“) a smyčky („loops“) spojují α -helixy nebo β -skládané listy, přičemž mění směr polypeptidového řetězce. Navíc se struktura proteinu tak stává kompaktnější. Tyto strukturní motivy obvykle obsahují hydrofilní zbytky a nacházejí se na povrchu proteinů.

Hlavním rozdílem mezi nimi je především jejich délka. Smyčky jsou delší, ohyby kratší. Ohyby sestávají ze čtyř nebo pěti aminokyselin a mají vnitřní vodíkové vazby.

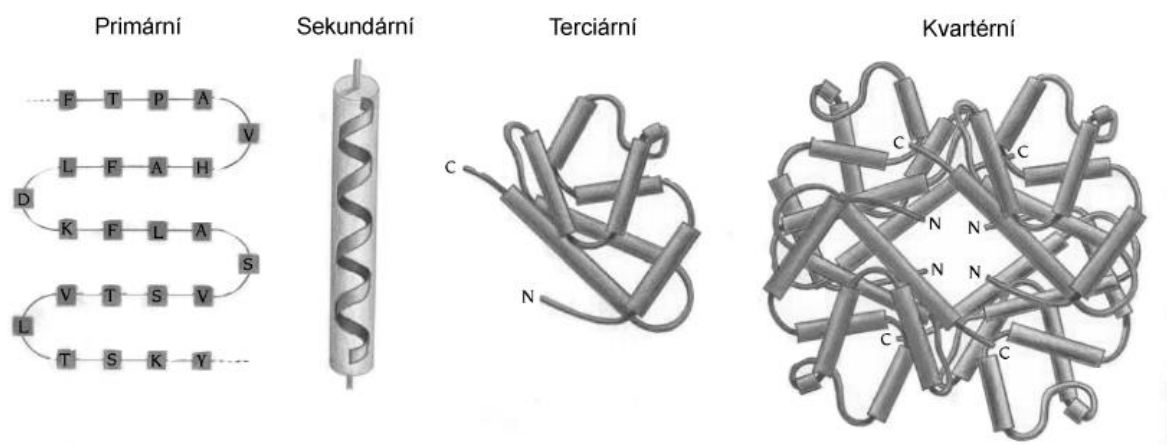
1.2.3 Terciární struktura proteinů

Jedná se o trojrozměrnou strukturu polypeptidového řetězce, která vzniká na základě působení hydrofobních interakcí, především van der Waalsových sil, v rámci postranních řetězců aminokyselin. Při skládání řetězců do terciální struktury hraje roli také konformační stabilita, teplota, pH, míra solvatace, vazba kofaktoru a ligandu, primární a sekundární struktura, chaperony či postranní modifikace. Například hydrofobní interakce jsou významné spíše u větších proteinů (Pace a kol., 2011). Terciární struktura je navíc stabilizována pomocí kovalentních vazeb mezi postranními řetězci, jedná se o tzv. disulfidické můstky (vazba S-S), které představují z důvodu své kovalentní povahy nejsilnější typ interakce. Svou významnou roli zde hrají i iontové síly. Tento typ struktury tedy není udržován pouze vodíkovými můstky, jako je tomu u sekundárních struktur, jelikož stabilita terciární struktury je velice významná pro funkci proteinu.

Mezi terciární struktury řadíme proteinové domény a sklady („foldy“). Doména je strukturně i funkčně nezávislá oblast proteinu. „Fold“ je obecnou architekturou proteinu, kterou dělíme na základě biochemických vlastností na globulární, fibrilární a membránové, podle strukturních vlastností na α , β , α/β a $\alpha+\beta$. Patří sem také supersekundární struktury. Jedná se o opakující se kombinace sekundárních motivů, které představují přechod mezi sekundární a terciární strukturou proteinu.

1.2.4 Kvartérní struktura proteinů

Kvartérní struktura je spojení několika proteinových řetězců, tedy uspořádání podjednotek v proteinových komplexech. Podjednotky jsou mezi sebou propojeny prostřednictvím kovalentních vazeb, především pomocí disulfidických můstků. Svou roli zde hrají také slabé nekovalentní interakce, vodíkové můstky nebo hydrofobní efekt. Uspořádání do kvartérní struktury zaujímají pouze oligomerní proteiny.



Obrázek 1 - Strukturní úrovně proteinu (Branden a Tooze, 1999)

2 Proteinové inženýrství

Proteinové inženýrství je poměrně mladá vědní disciplína, která zkoumá souvislosti mezi strukturou a funkcí proteinů a usiluje o vylepšení jejich vlastností, případně o konstrukci zcela nových variant. Vznik proteinového inženýrství byl logickým vývojem v oblasti rekombinantních DNA technologií. Přestože se o principech modifikace DNA prostřednictvím syntetických oligonukleotidů hovořilo již dříve (Hutchinson a Edgell, 1971), proteinové inženýrství přišlo na svět až v druhé polovině 70. let minulého století.

V souvislosti s tím, jestli informace o struktuře a vlastnostech proteinů máme či ne, vznikly tři základní přístupy proteinového inženýrství: racionální design, řízená evoluce a semi-racionální design (Obr. 2). Pro výběr vylepšených variant enzymů ze souboru/knihoven získaných mutantů je uplatňována selekce nebo screening.

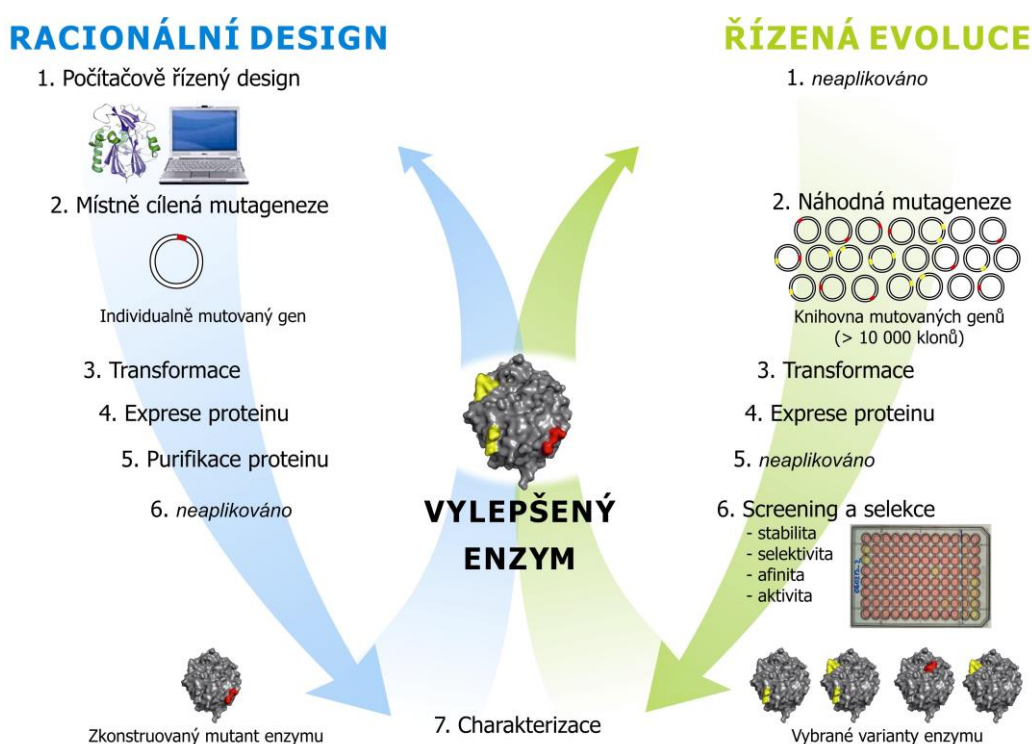
Screening se v proteinovém inženýrství využívá pro výběr nejlepších variant proteinů. Náročnost jeho provedení je dána velikostí knihovny. Čím méně informací o daném proteinu máme, tím je výsledná knihovna větší. Pro prohledání obrovských knihoven jsou nezbytné vysokoúčinné metody („high throughput screening methods“). Účinnost screeningu je dána možností rychle vyhodnotit například optické vlastnosti (barvu, fluorescenci, luminiscenci nebo zákal). Screening založený na fluorescenci se využívá často, jelikož je rychlý a snadný. Podobně jako fotometrický screening se provádí v mikrotitračních destičkách (Bornscheuer, 2005). Většina biomolekul ale není spojena přímo s pozorovatelnými fenotypy, proto je potřeba využít fluorescenční nebo kolorimetrické sondy. V současné době jsou velmi populární také mikrofluidní systémy založené na čípech – „FACS on a chip“ (Agresti a kol., 2010). Pro screening je možné využít také nukleární magnetickou rezonanci (NMR), HPLC, plynovou chromatografií či hmotnostní spektroskopií. U těchto metod je přímo monitorována spotřeba substrátu a vznik produktu. V případě, že je výsledná knihovna mutantních variant opravdu enormní velikosti, v řádu statisíců až milionů, není možné ji prohledat celou, jelikož není k dispozici natolik účinná screeningová metoda.

Selekci je možné využít ke snížení velikosti knihoven, které se dále testují pomocí screeningu. Zajišťuje kontrolu fenotypu a výběr funkčních variant mutantních proteinů i v knihovnách obrovské velikosti. Selekcční metody spočívají v propojení aktivity proteinu s fyzikální separací nebo přežitím organismu, který produkuje aktivní varianty proteinů.

Například selekce založená na vazebné afinitě spočívá ve fyzikální separaci proteinů. Varianty proteinů s požadovanou vazebnou afinitou jsou zachyceny pomocí imobilizovaného

nosiče, zatímco nenavázané proteiny se odplaví. Pro zachycení proteinů se využívá například povrch buněk, ribozomů nebo bakteriofágů.

Selekce založená na přežití organismu je historicky nejstarší metodou výběru aktivních variant enzymu. Spočívá v přežití (pozitivní selekce) či nepřežití (negativní selekce) organismů obsahujících aktivní varianty genů či enzymů důležitých pro život a replikaci. Pokud organismus takové geny či enzymy nemá, není schopen přežít v selekčních podmínkách a dále se replikovat. Rezistence vůči antibiotikům (Orencia a kol., 2001), tj. schopnost neutralizovat nebo exportovat antibiotikum ven z buňky, je nejčastější příklad využití selekce.



Obrázek 2 - Srovnání racionálního designu a řízené evoluce (©Zbyněk Prokop a Jiří Damborský, Loschmidtovi laboratoře, Masarykova univerzita, Brno, Česká republika)

2.1 Racionální design

Racionální design neboli přístup „top-down“ spočívá v místně-specifických změnách daného proteinu. Velkou nevýhodou a zároveň limitujícím krokem této metody je potřeba detailní znalosti struktury a funkce proteinu, kterou lze získat pomocí strukturní a biochemické analýzy. Taková analýza je velmi náročný a především zdlouhavý proces, který může trvat měsíce nebo také několik let. Na základě této znalosti proteinu lze vybrat

tzv. „hot-spoty“, místa ve struktuře proteinu vhodná pro mutagenézi. Výběr „hot-spotů“ se provádí pomocí počítačových metod a nástrojů. Výhodou tohoto přístupu je poměrně jednoduchá charakterizace vzniklých variant proteinů. Mezi molekulárně biologické metody využívané při racionálním designu pro přípravu mutantů se řadí místně-řízená mutagenéze, několikanásobná místně-řízená mutagenéze a genová syntéza.

2.2 Řízená evoluce

Řízená evoluce (*in vitro* evoluce, evoluce ve zkumavce) neboli přístup „bottom-up“ napodobuje vývojové procesy, které probíhají v živé přírodě. Jedná se více méně o náhodný proces, avšak na rozdíl od přirozené evoluce probíhá „ve zkumavce“ a ve velmi zrychlené podobě.

Jedná se o zcela náhodnou modifikaci proteinu, která zahrnuje opakované cykly mutagenéze, exprese a selekce/screeningu (Parra a kol., 2013). Kromě „templátové“ neboli vzorové DNA není k jejímu provedení potřeba téměř žádná znalost struktury a funkce daného proteinu. Mutagenézí dochází ke vzniku tzv. knihoven, souborů velkého množství náhodných variant proteinů. Nové varianty mohou mít vlastnosti lepší, stejné, nebo dokonce horší než původní varianta proteinu. Je proto velmi důležité najít mezi statisíci variant ty, u kterých došlo k vylepšení požadované vlastnosti. K tomu slouží testování vzniklých variant pomocí selekce nebo screeningu. Proveditelnost selekce či screeningu představují hlavní limitující faktor řízené evoluce.

Metody řízené evoluce se dělí na nerekombinantní a rekombinantní. Mezi nerekombinantní metody patří saturační mutagenéze a k chybám náchylná („error-prone“) PCR, která je jednou z nejpoužívanějších metod řízené evoluce (Cadwell a Joyce, 1992). Další známé metody jsou inzerční/deleční mutagenéze, kazetová mutagenéze, oligonukleotidově-řízená mutagenéze, chemická a fyzikální mutagenéze, segmentální mutagenéze nebo saturační mutagenéze celého genu („gene site-saturation mutagenesis“).

Rekombinantní metody se dále dělí na homologii závislé a nezávislé. Výběr metody je dán procentem homologie DNA. V případě, že je homologie DNA větší než 70 %, použijeme metody na homologii závislé, mezi které patří často využívaná metoda míchání DNA („DNA-shuffling“) (Stemmer, 1994) nebo metoda StEP – „Staggered Extension Process“ (Arnold a Georgiou, 2003). K dalším metodám se řadí například RETT – „Recombined Extension on Truncated Templates“ (Lee a kol., 2003), MURA – „Mutagenic and Unidirectional Reassembly“ (Song a kol., 2002), MUPREC – „Multiplex-PCR-Based

Recombination“ (Eggert a kol., 2005) nebo DOGS – „Degenerate Oligonucleotide Gene Shuffling“ (Gibbs a kol., 2001). Pokud je homologie DNA nižší než 70 %, využívají se metody nezávislé na homologii. Mezi nejznámější z nich patří RACHITT – „Gene Family Shuffling by Random Chimeragenesis on Transient Templates“ (Coco a kol., 2001), ITCHY – „Incremental Truncation for the Creation of Hybrid Enzymes“ (Ostermeier a kol., 1999) a SCRATCHY (Lutz a kol., 2001). Ty méně známé jsou například SHIPREC – „Sequence Homology-Independent Protein Recombination“ (Sieber a kol., 2001), SCOPE – „Structure-based Combinatorial Protein Engineering“ (O’Maille a kol., 2002) či cirkulární permutace (Heinemann a Hahn, 1995).

2.3 Semi-rationální design

Semi-rationální design neboli cílená řízená evoluce, či také datově-řízené proteinové inženýrství, je kombinací řízené evoluce a racionálního designu (Obr. 2). Má za cíl zvýšení kvality a snížení kvantity výsledného počtu mutantních variant proteinů. Využívá výhod obou přístupů, a tvoří tak velmi účinnou metodu pro získávání nových forem proteinů s výhodnými vlastnostmi (Tab. 5). Zaměřuje se na specifická rezidua nebo určitou oblast určenou na základě základní znalosti struktury a funkce proteinu (Hidalgo a kol., 2008; Chaparro-Riggers, 2007; Chica a kol., 2005). V této vybrané oblasti dochází k randomizaci jen určitých a pečlivě vybraných fragmentů DNA (Hidalgo a kol., 2008) a k následné tvorbě tzv. „chytrých knihoven“, které jsou menší ale bohatší na žádoucí varianty a nevyžadují tolik účinnou selekční/screeningovou metodu, jako je tomu u řízené evoluce. Zároveň dochází ke zvýšení pravděpodobnosti vzniku výhodně modifikovaných vlastností. V podstatě se jedná o řízenou evoluci úseků proteinů vybraných na základě znalosti struktury a funkce proteinu.

Tabulka 5 - Srovnání přístupů proteinového inženýrství

	Racionální design	Semi-rationální design	Řízená evoluce
Výhody	- žádná selekce či screening	- základní znalost struktury a funkce - selekce/screening definovaného množství variant	- znalost struktury a funkce není třeba
Nevýhody	- detailní znalost struktury a funkce		- selekce/screening velkého množství variant

3 Chytré knihovny

Podstatou konstrukce chytrých knihoven je výběr nejvhodnějších pozic a oblastí pro mutagenезi ve struktuře proteinu, výběr tzv. „hot-spotů“. Pro výběr těchto míst se využívá řada sofistikovaných počítačových metod a programů, například CAVER (Chovancová a kol., 2012), HotSpot Wizard (Pavelka a kol., 2009) nebo FoldX (Guerois a kol., 2002). Druhým důležitým aspektem při vytváření chytrých knihoven je redukce aminokyselinové abecedy, která má za následek snížení velikosti výsledné knihovny mutantních variant, aniž by byla snížena pravděpodobnost úspěšnosti mutagenезe. Tvorba chytrých knihoven variant proteinů tak s sebou přináší nesčetné množství výhod. Jednou z nich je zvýšení pravděpodobnosti, že se nám podaří získat lepší variantu proteinu. Konstrukce chytrých knihoven také šetří práci a náklady. Navíc umožňuje provádět experimenty, které by dříve nebyly vůbec možné.

3.1 Bioinformatické nástroje pro konstrukci chytrých knihoven

Potřeba konstrukce malých a cílených knihoven mutantů v definovaných pozicích je stále více aktuální (Nimrod a kol., 2005). Dříve nebylo možné takové knihovny připravovat, neboť nebyly dostupné sofistikované modelovací a bioinformatické nástroje, které by přehledně pracovaly s dostupnými informacemi o struktuře a funkci proteinů. I v případě, že pro daný protein není vyřešena krystalová struktura a nejsou tedy žádné informace o jeho vlastnostech a funkci k dispozici, přichází na řadu počítačové nástroje – porovnávání příbuzných sekvencí a výpočet strukturního modelu. Je možné uplatnit také tzv. *de novo* design (Jiang a kol., 2008).

Počítačových nástrojů pro určení a vyhodnocení „hot-spotů“ vznikla již spousta (Tab. 6) a vývoj v této oblasti není zdaleka u konce. Univerzální nástroj pro určení „hot-spotů“ za účelem vylepšení katalytických vlastností a termostability enzymu a jejich následné evaluaci je HotSpot Wizard (Pavelka a kol., 2009). Výběr „hot-spotů“ umožňuje cílit mutagenезi pouze na místa vhodná pro úspěšnou změnu a vylepšení požadované vlastnosti proteinu. Kombinace bioinformatiky s proteinovým inženýrstvím umožnila vznik mnoha cílených metod mutagenезe (Gustafsson a kol., 2003).

Tabulka 6 - Přehled vybraných bioinformatických nástrojů pro určení a evaluaci „hot-spotů“ nebo teoretických knihoven

Název	Reference	Funkce
3D-SURFER	La a kol., 2009	
CASTp	Dundas a kol., 2006	Identifikace „hot-spotů“ pro vylepšení katalytických vlastností
metaPocket	Zhang a kol., 2011	
SITEHOUND	Hernandez a kol., 2009	
Pocket-Finder	Laurie a Jackson, 2005	
MOLE	Berka a kol., 2012	Identifikace „hot-spotů“ pro vylepšení katalytických vlastností a termostability
CAVER	Chovancová a kol., 2012	
3DM	Kuipers a kol., 2010	Evaluace „hot-spotů“
ConSurf	Ashkenazyet a kol., 2010	
Evolutionary Trace	Valdar, 2002	
SIFT	Ng a Henikoff, 2003	
GLUE	Firth a Patrick., 2005	Určení diverzity teoretické knihovny
CASTER	Reetz a Carballeira, 2007	
AA-Calculator	Firth a Patrick, 2008	
B-FITTER	Reetz a Carballeira, 2007	Identifikace a vyhodnocení „hot-spotů“ pro vylepšení termostability
FoldX	Guerois a kol., 2002	
PoPMuSiC	Dehouck a kol., 2011	
HotSpot Wizard	Pavelka a kol., 2009	Identifikace a evaluace „hot-spotů“ pro vylepšení katalytických vlastností a termostability

3.2 Redukce aminokyselinové abecedy

Redukované aminokyselinové abecedy jsou podstatnou součástí konstrukce chytrých knihoven. Eliminují inkorporaci nadbytečných kodonů a zabraňují vzniku nefunkčních variant proteinů (Reetz, 2011).

Mutageneze za použití kodonu, který kóduje pouze jednu aminokyselinu, není ve většině případů zcela ideální. Úspěšnost mutageneze lze zvýšit použitím degenerovaných kodonů obsahujících v cílovém kodonu směs bází (Sullivan a kol., 2013). To, že se jednotlivé báze nezačleňují při syntéze se stejnou pravděpodobností, lze vyřešit pomocí speciálních dinukleotidů a trinukleotidů fosforamiditů, které umožňují inkorporovat více různých/vybraných aminokyselin na určitou pozici (Neylon, 2004).

Aminokyselinové abecedy zapisujeme pomocí kodonových setů značených podle standardních kódů pro sekvenci nukleových kyselin na základě pravidel IUB/IUPA (Tab. 7) (například NNB, NNS či NNK kodony, kde N = A/C/G/T, B = C/G/T, S = C/G a K = G/T). Kompletní směs všech bází představuje degenerovaný kodon NNN. Ten však nevede k produkci chytrých knihoven, jelikož kóduje všech 64 kodonů (Tab. 8), a umožňuje tak získat teoreticky všechny možné mutantní varianty enzymů, tedy i nefunkční nebo synonymní. Při použití degenerovaného kodonu NNN jsou navíc aminokyseliny zastoupeny nerovnoměrně. Zatímco methionin je kódován pouze jednou, arginin hned šestkrát (Hudges a kol., 2003). Mnohem častěji se využívá redukovaná aminokyselinová abeceda NNK, která kóduje ve srovnání s NNN pouze 32 kodonů.

Použití degenerovaných kodonů NNK kódujících všech 20 aminokyselin není však také zcela praktické, jelikož výsledný počet transformantů k otestování je stále vysoký (Reetz a Carballeira, 2007; Acevedo-Rocha a kol., 2014). Z tohoto důvodu se začaly používat redukované aminokyselinové abecedy (Walter a kol., 2005). Například degenerovaný kodon NDT kóduje pouze 12 aminokyselin (Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, Gly).

Je potřeba mít na paměti, že při návrhu redukované abecedy je vhodné zachovat aminokyseliny, které se objevují u divokého typu na daných pozicích (Sun a kol., 2016). Volba redukované aminokyselinové abecedy se řídí strukturálními, výpočetními a/nebo sekvenčními daty (Sun a kol., 2016). Při volbě redukované aminokyselinové abecedy je možné postupovat dvěma různými způsoby. Buď použijeme stejnou aminokyselinovou abecedu pro všechny mutované pozice, nebo navrheme pro každé místo jinou. Druhý způsob

je využíván méně, jelikož náročnost designu a cena oligonukleotidů je větší. Jeho využití bylo poprvé popsáno ve studii rezistence vůči antibiotikům (Hayes a kol., 2002).

V souvislosti s návrhem redukované aminokyselinové abecedy se často provádí skenovací mutagenese pomocí jedné aminokyseliny - cysteinu, alaninu, serinu nebo argininu pro průzkum specifických interakcí s postranním řetězcem. Třeba tzv. alaninové skenování, při kterém je každé reziduum v dané oblasti proteinu nahrazeno alaninem. Následně jsou měřeny důsledky této substituce např. na vazebnou afinitu, a tak pozorována role jednotlivých postranních řetězců reziduí. Skenovací mutagenese představuje silnou strategii také pro odkrytí funkce proteinů, vazby ligandu či substrátu, proteinové aktivity a enantioselektivity.

Tabulka 7 - Standardní kódy pro sekvence nukleových kyselin

Kód	Nukleová kyselina
A	Adenosin
C	Cytidin
G	Guanidin
T	Thymidin
U	Uridin
R	puRin (G/A)
Y	pYrimidin (T/C)
K	nukleosid s Keto skupinou (G/T)
M	nukleosid s aMino skupinou (A/C)
S	silná vazba (G/C), „Strong“
W	slabá vazba (A/T), „Weak“
B	bez A (G/T/C)
D	bez C (G/A/T)
H	bez G (A/C/T)
V	bez T (G/C/A)
N	jakákoli (A/G/C/T)

Tabulka 8 - Standardní kódy v DNA pro aminokyseliny

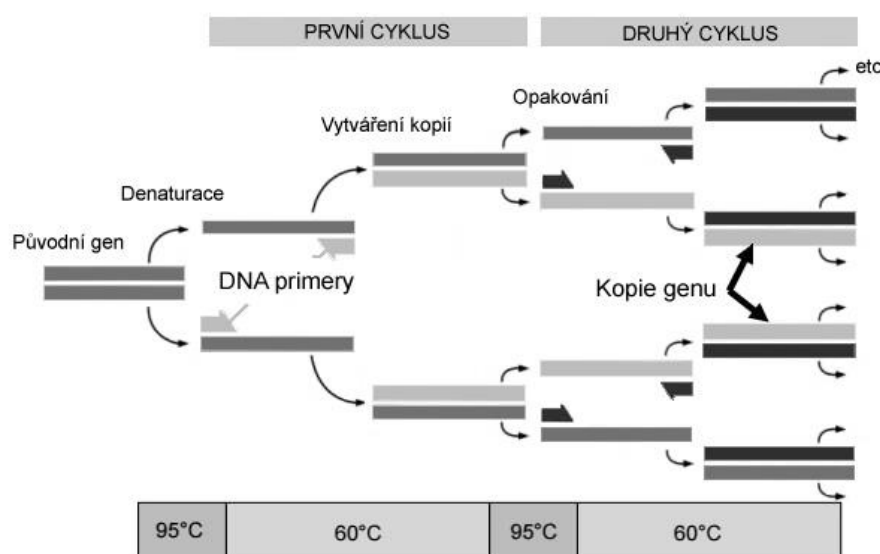
	T	C	A	G	
T	TTT fenylalanin	TCT serin	TAT tyrosin	TGT cystein	T
	TTC fenylalanin	TCC serin	TAC tyrosin	TGC cystein	C
	TTA leucin	TCA serin	TAA stop kodon	TGA stop kodon	A
	TTG leucin	TCG serin	TAG stop kodon	TGG tryptofan	G
C	CTT leucin	CCT prolin	CAT histidin	CGT arginin	T
	CTC leucin	CCC prolin	CAC histidin	CGC arginin	C
	CTA leucin	CCA prolin	CAA glutamin	CGA arginin	A
	CTG leucin	CCG prolin	CAG glutamin	CGG arginin	G
A	ATT isoleucin	ACT threonin	AAT asparagin	AGT serin	T
	ATC isoleucin	ACC threonin	AAC asparagin	AGC serin	C
	ATA isoleucin	ACA threonin	AAA lysin	AGA arginin	A
	ATG methionin	ACG threonin	AAG lysin	AGG arginin	G
G	GTT valin	GCT alanin	GAT kys. asparagová	GGT glycin	T
	GTC valin	GCC alanin	GAC kys. asparagová	GGC glycin	C
	GTA valin	GCA alanin	GAA kys. glutamová	GGA glycin	A
	GUG valin	GCG alanin	GAG kys. glutamová	GGG glycin	G

3.3 Význam PCR pro konstrukci chytrých knihoven

Molekulárně biologické metody tvorby chytrých knihoven jsou často založené na cílené mutagenезi využívající polymerázovou řetězovou reakci – PCR (Karry Mullis, 1983). PCR slouží k namnožení určitého úseku DNA. Spočívá v opakované řízené denuraci dvouřetězcové DNA a následné renaturaci samostatných řetězců se specifickými oligonukleotidy (Obr. 3). Oligonukleotidy slouží jako primery. Jako primer se označuje řetězec nukleových kyselin, který představuje počáteční místo replikace DNA nebo RNA a umožňuje zahájení syntézy nového řetězce pomocí enzymu DNA polymerázy.

Oligonukleotidy je možné nazývat také jako oligodeoxyribonukleotidové primery nebo oligoprimery.

PCR začíná denaturací DNA na jednotlivá vlákna, ke které dochází při 95°C. Následné ochlazení na teplotu 50 – 60°C způsobí renaturaci jednořetězcové DNA a nasedání primerů (tzv. fáze hybridizace). Výrazný nadbytek oligonukleotidů způsobí, že DNA renaturuje rychleji právě s oligonukleotidy místo s dlouhými vlákny jednořetězcové DNA. V této fázi má správná teplota velmi důležitou roli a často je nutné ji optimalizovat. Po nasednutí primerů dochází k syntéze nových řetězců při teplotě 65 – 75°C. Celý proces je znázorněn na obrázku 3. Reakce probíhá v termocykleru, který pružně mění teplotu v potřebných intervalech.



Obrázek 3 - Průběh polymerázové řetězové reakce (Saiki a kol., 1988)

3.4 Metody konstrukce chytrých knihoven

Cílem metod pro konstrukci chytrých knihoven je zvýšit jejich kvalitu a minimalizovat velikost. Většina těchto metod spočívá v randomizaci přesně daných míst a oblastí, která je řízena navrženými oligonukleotidy za využití redukované aminokyselinové abecedy. Oligonukleotidy i celé geny je možné komerčně syntetizovat. S délkou syntetizované sekvence však narůstá i cena. Randomizovat mnoho míst najednou může být komplikované také z důvodu nedostatečné účinnosti screeningových metod. V takovém případě je potřeba zvolit metodu, která je schopná konstrukce malých a chytrých knihoven. Je také možné oblast zájmu rozdělit za vzniku několika samostatných knihoven. To je však nevýhodné, neboť některé substituce by spolu mohly za současné mutageneze vzájemně příznivě kooperovat. Následující metody usilují o tvorbu chytrých knihoven a snížení nároků na screening (Tab. 9).

3.4.1 Saturační mutageneze

Saturační mutageneze označovaná také jako přímá randomizace je hojně používanou metodou pro konstrukci chytrých knihoven. V dnešní době tato metoda existuje již v několika méně či více od sebe odlišných variantách, mezi které patří například saturační mutageneze jedné pozice, souběžná či iterativní saturační mutageneze.

Podstatou saturační mutageneze je využití syntetických oligonukleotidů jako primerů, které obsahují pozměněné báze v určitých pozicích. Oligonukleotidy nesou buď přesně danou záměnu aminokyseliny, která má za následek požadovanou změnu kodonu, nebo uplatňují speciální redukované kodonové sety. Ty naopak umožní vnést do cílového místa různé aminokyseliny (Nov, 2012). Mutace jsou vnášeny například do míst kódujících vazebné místo enzymu pro ligand, katalytické místo nebo přístupové tunely v závislosti na vlastnostech, které chceme pozměnit. Tato metoda se uplatňuje především pro zlepšení substrátové specifity nebo enzymové aktivity (Geddie a Matsumura, 2004).

Saturační mutageneze byla poprvé představena doktorem Wellsem a jeho spolupracovníky (Wells a kol., 1985) jako tzv. kazetová mutageneze. Jednou z jejích modifikací je oligonukleotidově řízená saturační mutageneze, při které se syntetická kazeta inkorporuje do genu pomocí PCR a megaprimerů (Landt a kol., 1990). Další modifikací je tzv. inkorporace syntetických oligonukleotidů pomocí skládání genů (ISOR = Incorporating Synthetic Oligonucleotides via Gene Reassembly) (Herman a Tawfik, 2007) nebo jednoduchá metodologie pro kazetovou randomizaci a rekombinaci v jedné zkumavce (OSCARR = „One-pot Simple methodology for Cassette Randomization and Recombination“) (Hidalgo a kol., 2008).

3.4.1.1. Saturační mutageneze v jednom místě (S-SM = Site-saturation mutagenesis)

Saturační mutageneze v jednom místě je účinnou a jednoduchou metodou. Spočívá v cílení na určité místo ve struktuře proteinu, které má potenciál pozitivně ovlivnit jeho vlastnosti a funkci. Podstatou je využití degenerovaných kodonových setů NNN, NNK nebo NNS, které kódují všech 20 standardních aminokyselin. V takovém případě však dochází ke vzniku tzv. „redundantních“ neboli nadbytečných kodonů a k nerovnoměrné distribuci aminokyselin v rámci knihovny mutantů.

Z tohoto důvodu byla vyvinuta tzv. „small-intelligent“ strategie (Tang a kol., 2012), kde navržená směs primerů obsahuje přesně jeden kodon na jednu aminokyselinu, a umožňuje

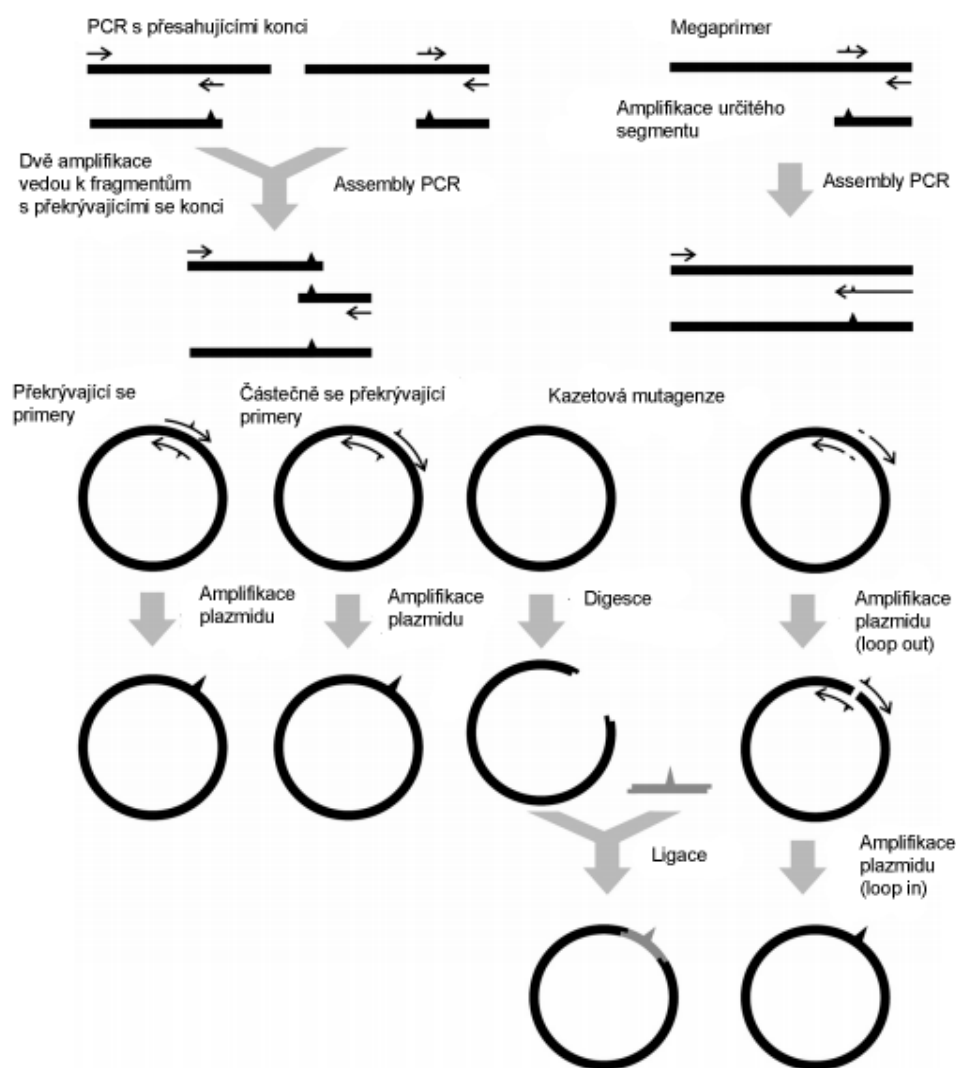
tak tvorbu méně početných a zároveň chytrých knihoven bez stop kodonů, vzácných kodonů nebo nerovnoměrné distribuce aminokyselin. Tyto nepříznivé faktory totiž vedou k produkci neaktivních a duplicitních variant, a snižují tak pravděpodobnost úspěšné mutagenese. Prvním krokem je návrh primerů, pro který byl v rámci „small-intelligent“ strategie speciálně vyvinut počítačový nástroj DC-Analyzer. Podmínkou je, aby degenerované kodonové sety kódovaly všechny navržené aminokyseliny se stejnou pravděpodobností výskytu (Tang a kol., 2012).

Saturační mutagenese v jednom místě byla například použita k prozkoumání substrátové specifity eukaryotické signální peptidasy (Folz a kol., 1988) za využití mutantní varianty lidského preproapolipoproteinu A-II. Varianty konstrukce knihovny mutantních enzymů pomocí saturační mutagenese v jednom místě jsou znázorněny na obrázku 4.

Vývoj této metody vedl ke vzniku saturační mutagenese po celé délce genu (GSSM = „gene site saturation mutagenesis“) (Kretz a kol., 2004). GSSM zavádí bodové mutace do každé pozice v cílovém genu za použití degenerovaných primerů obsahujících NNK nebo NNS kodony. Nakonec se zkombinují mutanti s pozitivními výsledky do finální několiknásobné mutantní varianty. Pomocí této metody je možné prozkoumat všech dvacet aminokyselinových substitucí v daných pozicích sekvence proteinu (Kretz a kol., 2004). Využívá dvou randomizovaných oligonukleotidových primerů, pro každou pozici zvlášť. První primer obsahuje NNK kodon v pozici, která má být mutována. Druhý primer obsahuje komplementární kodon MNN (K = G/T, M = A/C). NNK kodon umožňuje vznik 32 různých sekvencí, které kódují všech 20 aminokyselin a jeden stop kodon. Primery se následně využijí při amplifikaci (pomnožení) za použití kruhového dvouvláknového vektoru DNA. Nevýhodou této metody je potřeba screeningu obrovské knihovny výsledných variant.

Saturační mutagenesi v jednom místě je možné provést několika způsoby, všechny však využívají cirkulární plazmid jako templát pro PCR s primery nesoucími žádoucí mutace. Nejstarší a časově nejnáročnější způsob konstrukce je pomocí inverzní PCR (Ochman a kol., 1988). Inverzní PCR využívá dvou primerů směřujících v sekvenci od sebe, přičemž jeden z nich nese mutaci. Dochází k amplifikaci celého řetězce plazmidu a zavedení mutace. Následuje štěpení methylovaného templátového řetězce pomocí *DpnI*. Vzniklý lineární dvouvláknový plazmid je cirkularizován ligací T4 DNA ligasou. Ligace často vyžaduje optimalizaci podmínek z důvodu nízké účinnosti ligace tupých konců. Poté je vzniklý plazmid transformován do kompetentních buněk *Escherichia coli*. Další možností je využití komerčních sad/kitů, např. QuikChange od společnosti Stratagene (Wang a Malcolm, 2002). PCR s reverzně komplementárními primery nesoucími požadované mutace probíhá

ve speciálně upraveném pufru. Stejně jako v předchozím případě následuje štěpení *DpnI*, jednořetězcové zlomy jsou spojeny enzymy hostitelské bakterie po transformaci. Jedná se zřejmě o nejdražší metodu konstrukce. Je však optimalizovaná, takže snadná a rychlá. Nejnovější alternativa je konstrukce knihovny pomocí akademické aplikace OneClick (Warburton a kol., 2015). Je možné využít komerčně dostupné polymerasy a dva nepřekrývající se, částečně se překrývající, či reverzně komplementární primery. PCR reakce probíhá nejprve odděleně pro každý primer zvlášť. Po deseti cyklech se obě reakce smíchají. Štěpení probíhá opět pomocí *DpnI*, spojení jednořetězcových zlomů se děje prostřednictvím enzymů hostitelské bakterie po transformaci. Tato metoda však může vyžadovat optimalizaci. Cenově se ale jedná o přijatelnou variantu, přičemž časová náročnost této metody je jen o něco větší než u metody QuickChange.



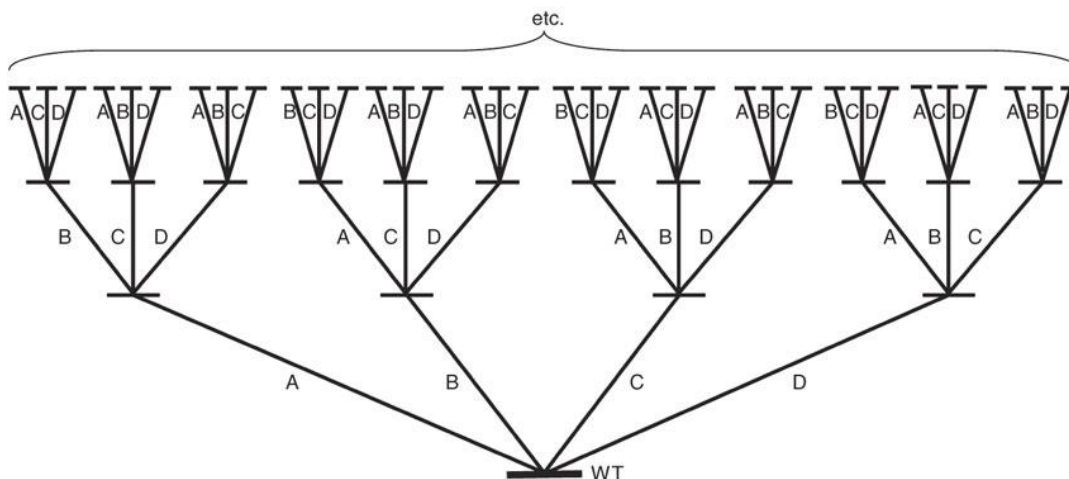
Obrázek 4 - Varianty konstrukce knihovny mutantních enzymů pomocí saturační mutagenese v jednom místě (Siloto, 2012)

3.4.1.2 Iterativní saturační mutagenese (ISM)

Iterativní saturační mutagenese spočívá v postupných (iterativních) cyklech mutagenese (Obr. 5). Je soustředěna do racionálně vybraných míst nebo oblastí (Reetz a Carballeira, 2007). Výběr těchto pozic se řídí tím, jakou vlastnost je potřeba vylepšit. V případě stereoselektivity, regioselektivity, substrátové specifity nebo enzymové aktivity se jedná převážně o pozice v blízkosti aktivního místa, protože tyto mutace s největší pravděpodobností ovlivní výše zmíněné vlastnosti (Morley a Kazlauskas, 2005). Pro iterativní saturační mutagenesi se využívá iterativní saturační test aktivního místa – iterativní CASTing (Combinatorial Active-site Saturation Test) (Parra a kol., 2013) nebo tzv. test B-FIT („B-Factor Iterative Test“) pro určení chemické a tepelné stability proteinu při určitých substitucích.

Iterativní saturační mutagenese začíná návrhem vhodných oligonukleotidů. Poté přichází na řadu randomizace vybraných pozic pomocí saturační mutagenese. Každá pozice vede ke vzniku jedné knihovny. Knihovny jsou následně otestovány pomocí screeningu a nejlepší varianty jsou v dalším kole mutagenese použity jako templát neboli vzor (Reetz a Carballeira, 2007, Parra a kol., 2013). V novém kole je však saturační mutagenese provedena v pozicích, které nebyly saturovány v kroku předešlém. Je možné uplatnit všechny možné kombinace iterativní saturace vybraných pozic, avšak podmínkou to není.

Pomocí metody ISM je možné zlepšovat termostabilitu, aktivitu i substrátovou specifitu enzymu. Příkladem zlepšení termostability pomocí ISM je mutagenese lipasy A z bakterie *Bacillus subtilis* – Lip A (Reetz a kol., 2006a). Deset vybraných aminokyselin určených na základě programu B-FITTER (Reetz a Carballeira, 2007) bylo vybráno pro randomizaci pomocí standardní saturační mutagenese metodou QuikChange. Výsledkem bylo celkem osm knihoven. Nejlepší varianta byla poté použita pro další krok ISM. Celkem došlo k otestování pouhých 8000 variant se získáním dvou hypertermostabilních mutantů X a XI (Reetz a Carballeira, 2007). Enantioselektivita a substrátová specifita byla zlepšena pomocí iterativního CASTingu například u epoxid hydrolasy pocházející z bakterie *Aspergillus niger* (Reetz a kol., 2006b).



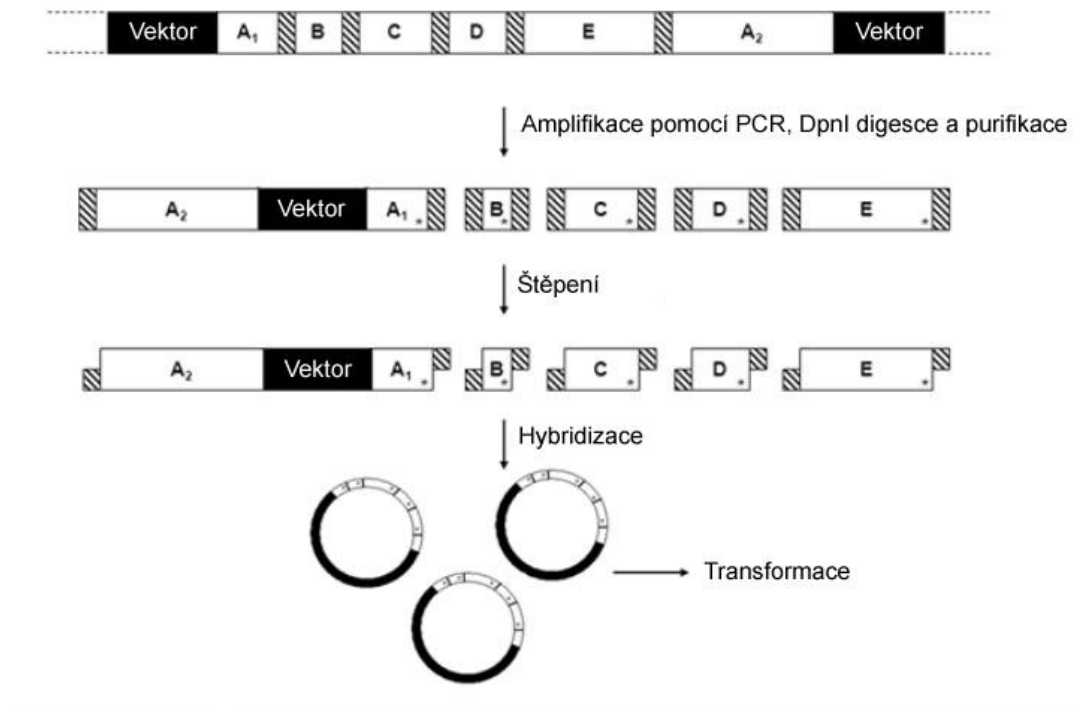
Obrázek 5 - Navazující kroky iterativní saturační mutagenese (Reetz a Carballeira, 2007)

3.4.1.3 Souběžná saturační mutagenese (Simultaneous saturation mutagenesis)

Souběžná saturační mutagenese spočívá na rozdíl od iterativní saturační mutagenese v saturaci různých pozic v jeden čas, tedy zároveň. To umožní docílit možného kooperačního efektu substitucí. Což je výhodné, jelikož zvýšíme pravděpodobnost úspěšnosti mutagenese. Je však důležité v rámci této metody uplatnit redukovanou aminokyselinovou abecedu, jinak získáme příliš mnoho mutantních variant.

Prvním krokem této metody je výběr kodonů a míst pro hybridizaci. Následuje návrh oligonukleotidů, přičemž je potřeba ověřit, že vybraná hybridizační místa jsou unikátní v rámci sekvence vybraného vektoru. Dalším krokem je PCR za použití navržených oligonukleotidů. Klíčovým krokem je vysoce účinné chemické štěpení fosforothiolovaných nukleotidů jódu ve směsi s ethanolem za vzniku dvanáct nukleotidů dlouhých přesahů na 5' konci v dvouvláknové DNA. Závěrečnou fází je hybridizace vektoru a inzerť pro vytvoření plazmidů o plné délce a transformace sestavené OmniChange knihovny do *E. coli* (Dennig a kol., 2011).

Metoda OmniChange (Obr. 6, Dennig a kol., 2011) byla vyvinuta pro mnohonásobnou místně-řízenou souběžnou saturační mutagenesi. Soustředí se především na vylepšení aktivity a selektivity enzymů a kooperačních vztahů jednotlivých substitucí. Úspěšně bylo saturováno například pět nezávislých aminokyselin ve fytase pocházející z bakterie *Yersinia mollaretii* za použití NNK degenerovaných kodonových setů (Dennig a kol., 2011).



Obrázek 6 - Souběžná saturační mutagenese pěti nezávislých pozic pomocí metody OmniChange (Dennig a kol., 2011)

3.4.2 ISOR („Incorporating Synthetic Oligos via Gene Reassembly“)

Metoda ISOR, v českém překladu inkorporace syntetických oligonukleotidů pomocí skládání genů, se řadí mezi metody založené na kombinatorice („combinatorial methods“). Spolupráce racionálního designu a počítačových nástrojů v rámci této metody značně minimalizuje výsledné množství substitucí v sekvenci. Spočívá v částečné diverzifikaci velkého množství residuí. Je odvozena od genového míchání („shufflingu“) a umožňuje provádět změny residuí pomocí substituce, inserce nebo delece. Jedná se o poměrně levný způsob tvorby chytrých knihoven, který navíc umožňuje mutovat 30 až 45 pozic v sekvenci proteinu, což je opravdu velké množství.

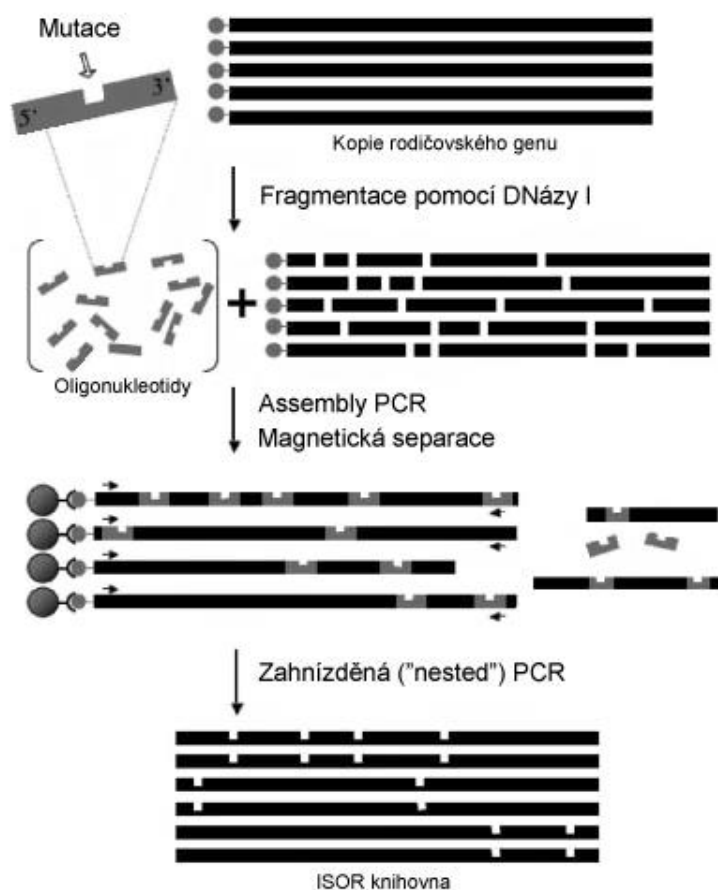
Nejprve je potřeba připravit DNA pomocí restrikčního enzymu DpnI a následné PCR. Následuje digesce a purifikace výsledných fragmentů. Reakce je zahájena smícháním oligonukleotidů, které obsahují randomizované kodony, s fragmenty rodičovského genu. Poté následuje inkorporace kodonů mezi fragmenty genu a další polymerázová řetězová reakce (Obr. 7).

Metoda ISOR byla využita například pro konstrukci knihoven mutantů cytosin-C5 methyltransferasy (Herman a Tawfik, 2007), kdy bylo randomizováno 45 individuálních

pozic. Tato metoda našla své uplatnění také při tvorbě knihoven séra paraoxonasy PON1 pomocí inzercí a delecí v rámci aktivního místa. U některých variant enzymu došlo k vylepšení substrátové specifity.

Metoda ISOR je vysoce účinná na rozdíl od klasické saturační mutagenese, při které získáme obrovskou knihovnu, kde téměř všechny varianty jsou inaktivní. Síla této metody spočívá ve vysoké koncentraci oligonukleotidů a vysoké frekvenci modifikací v každé pozici (Herman a Tawfik, 2007). Směs bývá většinou ekvimolární, což znamená, že koncentrace všech mutací je vyvážená (Rockah-Shmuel a kol., 2014).

Alternativní varianta této metody spočívá v separaci PCR pro každou mutaci. PCR následně vede ke vzniku velkého množství různě dlouhých fragmentů, z nichž každý nese různou mutaci. Následně jsou fragmenty smíchány pomocí tzv. „shufflingu“ se začleněním diverzifikovaných oligonukleotidů (Rockah-Shmuel a kol., 2014).



Obrázek 7 - Systematické znázornění průběhu metody ISOR (Herman a Tawfik, 2007)

3.4.3 Parsimoniální mutagenese (PM = „parsimonious mutagenesis“)

Parsimoniální mutagenese znamená v přímém překladu „skoupá mutagenese“. Má za cíl minimalizovat synonymní kodony a omezit množství aminokyselin, které nezachovávají znaky a vlastnosti původních reziduí (Schier a kol., 1996). Jedná se o počítačově řízenou metodu, která probíhá prostřednictvím tzv. „skenování“ (Krauss a kol., 2010). To je řízeno oligonukleotidem. Oligonukleotid postupně prohledává oblast genu za účelem objevit lepší varianty proteinů. Návrh těchto oligonukleotidů je řízen počítačovou metodou a oblast zájmu je určena na základě komplementarity bází.

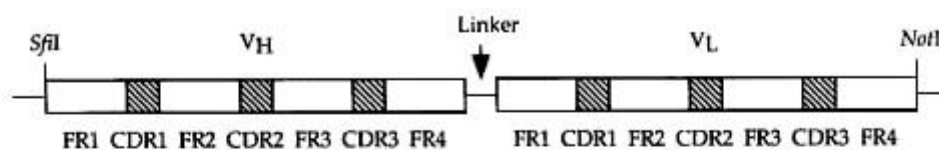
Metoda byla poprvé použita při mutagenези variabilního konce protilátky (Balint a Larrick, 1993). V rámci tohoto projektu byl vyvinut jednoduchý počítačový program PM-CAD pro návrh oligonukleotidů. Program vypočítá optimální složení oligonukleotidové směsi pro každou pozici, která má být randomizována. Příklad konstrukce knihovny pomocí parsimoniální mutagenese je znázorněn na obrázku 8.

Při návrhu oligonukleotidů se přebytečným kodonům vyhneme použitím redukované aminokyselinové abecedy. Frekvence sekvencí je v předem zvolených poměrech. K nukleotidům „rodičovské“ sekvence jsou přidány také tzv. „spiked“ oligonukleotidy. „Spiked“ oligonukleotidy obsahují v určité pozici neekvimolární poměr jednotlivých bází. Tyto oligonukleotidy se používají v případě, když je požadován určitý poměr bází v dané pozici. Frekvence sekvencí s předem zvolenými vlastnostmi nebo sety „nerodičovských“ reziduí je maximální.

Po návrhu oligonukleotidů přichází na řadu skenování mutovaného enzymu. Oligonukleotidy při tom napodobují mutagenези a využívají tendenci genetického kódu upřednostňovat chemicky nebo stéricky konzervované aminokyselinové substituce. To umožní prozkoumat, například v případě mutagenese protilátek, povrch antigenu s jeho minimálním narušením (Balint a Larrick, 1993). Po skenování oligonukleotidem následuje dvoukroková PCR mutagenese a amplifikace mutovaných kruhových plazmidů, které mohou být přímo transformovány do expresních kmenů bakterie *E. coli*.

Parsimoniální mutagenези se však nedostalo širokého uplatnění jako metodě spadající mezi místně-řízenou mutagenези. Použitelnost této metody byla totiž dříve limitována z důvodu vysoké ceny oligonukleotidů.

A Konstrukt genu

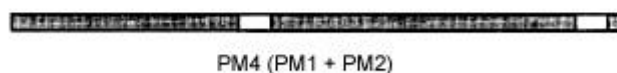


B PCR

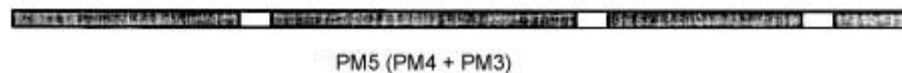
První kolo PCR: amplifikace genových segmentů PM1, PM2 a PM3



Druhé kolo PCR: PM1 a PM2 slouží jako templát pro PCR a vznik PM4



Třetí kolo PCR: spojení PM4 s PM3 za vzniku PM5, následná amplifikace



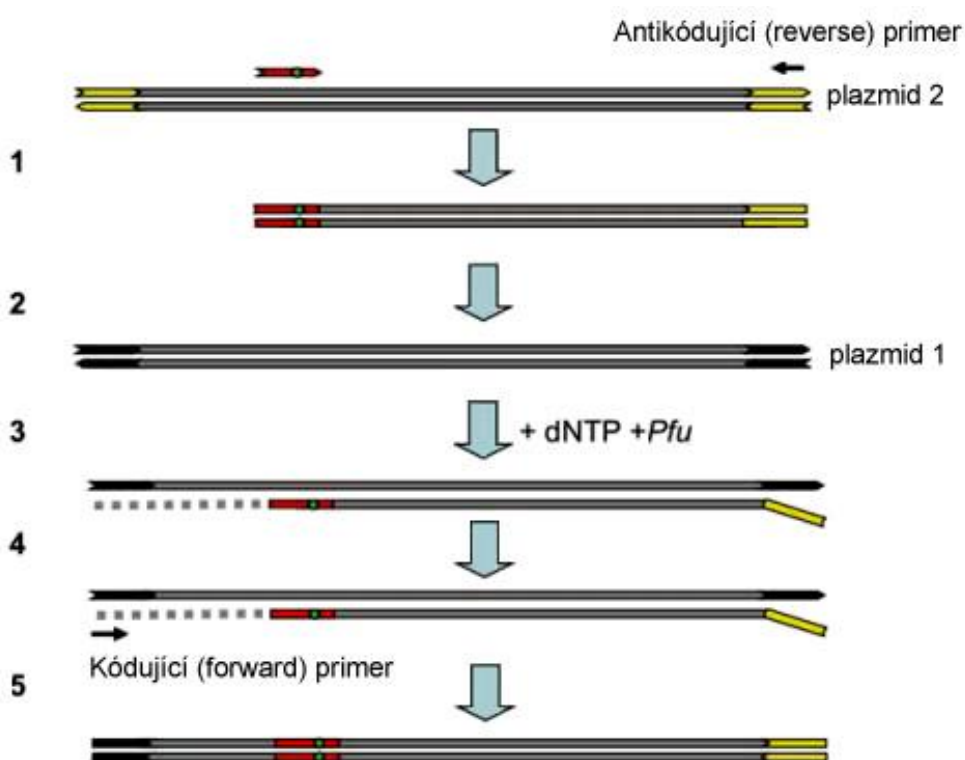
Obrázek 8 - Příklad konstrukce knihovny pomocí parsimoniální mutagenese (Schier a kol., 1996)

3.4.4 OSCARR („One-pot Simple methodology for Cassette Randomization and Recombination“)

Metoda OSCARR, v českém překladu jednoduchá metodologie pro kazetovou randomizaci a rekombinaci v jedné zkumavce, je založená na pečlivém návrhu mutagenních kazet a optimalizaci megaprimerů pro PCR. Je speciálně určena ke konstrukci knihoven s více než deseti randomizovanými kodony. Vyznačuje se 97% úspěšností ve vylepšení mutovaných produktů (Hidalgo a kol., 2008). Pro tuto metodu je potřeba nejprve vytvořit dva konstrukty, do kterých je klonován gen za použití restrikčních enzymů. Degenerované oligonukleotidy jsou navrženy tak, aby mutovaly všechna cílová residua pomocí inkorporace NNN či NNK degenerovaných kodonů. Výpočet vhodného množství každého nukleotidu je možný pomocí

speciálního algoritmu. Mutované kodony by měly být vždy obklopeny aspoň deseti bázemi nemodifikované sekvence na každé straně, aby došlo k jejich bezproblémové inkorporaci. V prvním kroku generuje PCR z mutagenických oligonukleotidů megaprimery. V druhém kroku jsou tyto megaprimery asymetricky prodlouženy do plné délky produktu. Posledním krokem je amplifikace produktu (Obr. 9, Hidalgo a kol., 2008).

Vysoká účinnost této metody byla potvrzena konstrukcí cílených knihoven mutantů esterázy I (PFEI) pocházející z bakterie *Pseudomonas fluorescens* (Hidalgo a kol., 2008). Výsledné mutantní varianty byly prozkoumány pomocí screeningu s cílem objevit varianty s pozměněnou substrátovou specifitou. Byly identifikovány dvě mutantní varianty s desetkrát vyšší katalytickou účinností vůči *p*-nitrofenyl dodekanoátu.



Obrázek 9 - Systematické znázornění metody OSCARR: 1) Vznik megaprimerů z plazmidu číslo 2 pomocí PCR za pomoci vnitřních mutagenických oligonukleotidů a antikódujícího (reverse) primeru. 2) a 3) Megaprimery asymetricky prodlouženy do plné délky produktu použitím plazmidu číslo 1 jako „templátu“. 4) a 5) Prodloužené megaprimery amplifikovány díky externím primerům (Hidalgo a kol., 2008).

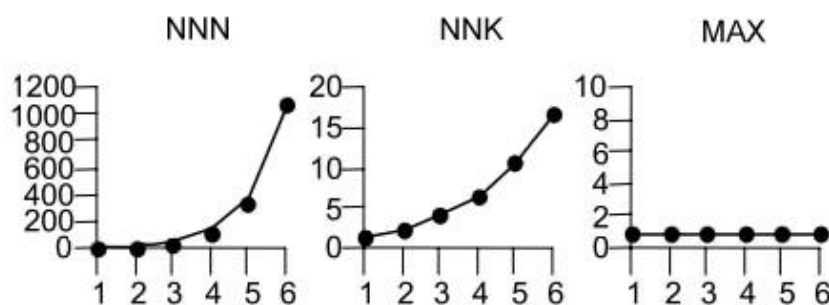
3.4.5 MAX randomizace (MAX = „maximum efficiency“)

MAX randomizace (Obr. 11, Hughes a kol., 2003) neboli randomizace s maximální účinností je založena na použití dvaceti takzvaných MAX kodonů. Každý z nich představuje vybraný kodon pro expresi jedné z dvaceti proteinogenních aminokyselin v *E. coli*. Mezi MAX kodony se řadí: GCG (A), TGC (C), GAT (D), GAA (E), TTT (F), GGC (G), CAT (H), ATT (I), AAA (K), CTG (L), ATG (M), AAC (N), CCG (P), CAG (Q), CGC (R), AGC (S), ACC (T), GTG (V), TGG (W) and TAT (Y). MAX randomizace vede ke konstrukci knihoven s maximální diverzitou a minimální velikostí (Hughes a kol., 2003). Má za cíl eliminovat synonymní kodony, a zabránit tak efektu nárůstu množství kodonů pro randomizaci (Obr. 10). Díky této metodě je teoreticky možné randomizovat více než 12 kodonů v rámci jedné kazety.

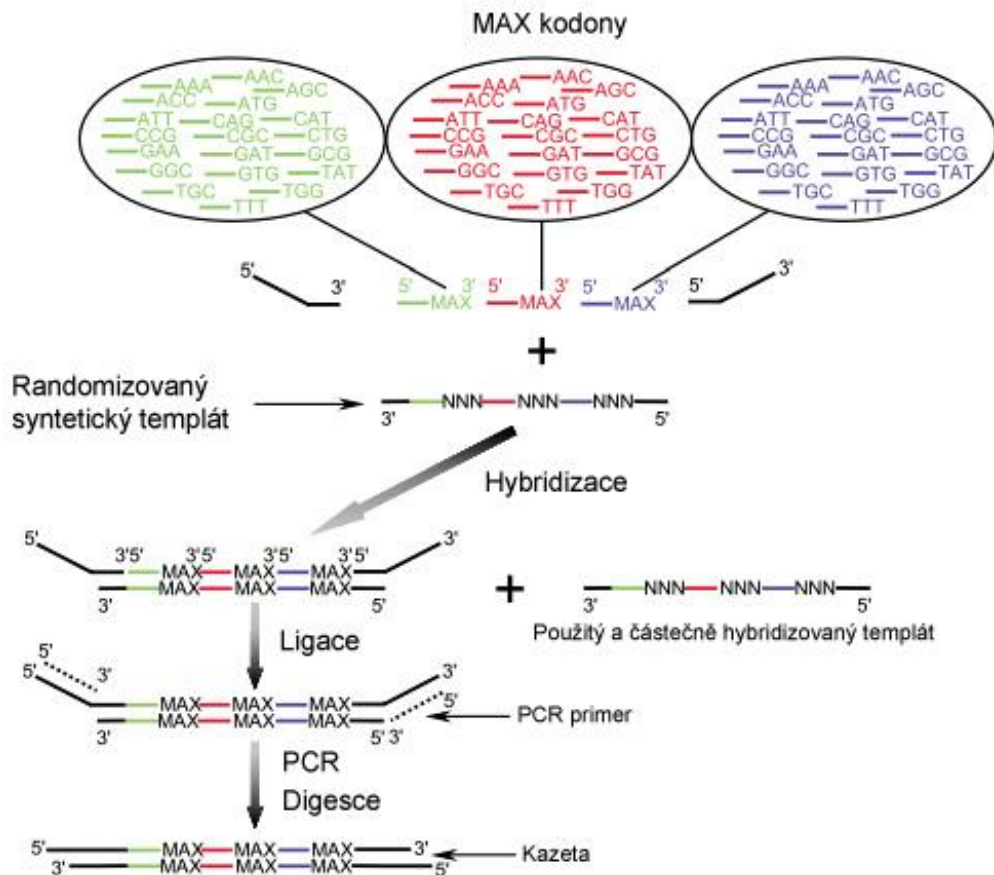
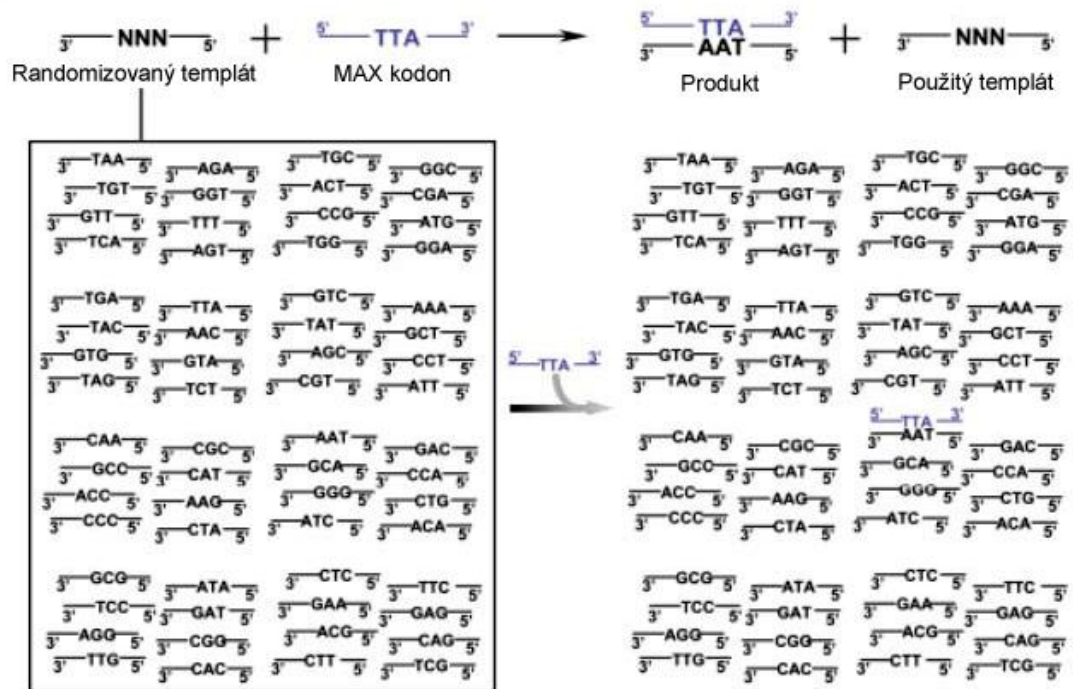
MAX randomizace spočívá v selekční hybridizaci, kdy vybrané MAX kodony hledají své komplementární sekvence v templátovém řetězci DNA, který slouží jako „dokovací“ stanice. Pokud dojde k nesprávnému zařazení některého MAX kodonu, není to problém, jelikož pouze řetězec ligovaný do vektoru je amplifikován následnou polymerázovou řetězovou reakcí. V případě nesprávného zařazení nedojde k ligaci, a tak ani k vytvoření chybného produktu.

Prvním krokem je syntéza oligonukleotidů pro každý randomizovaný kodon. Tyto primery jsou následně použity k hybridizaci „templátového“ řetězce pomocí NNN nebo NNK kodonů. Vzniklé oligonukleotidy jsou ligovány do vektoru a asymetrická PCR zajistí jejich pomnožení.

Pomocí MAX randomizace mohou být randomizovány mnohonásobné kodony, ale maximálně dva smí být přilehlé z důvodu zachování adresovací funkce konzervativní části vybraných oligonukleotidů. Ačkoli tato metoda využívá jednoduché primery, jedná se o poměrně drahou metodu ve srovnání s klasickou NNN/NNK randomizací.



Obrázek 10 - Efekt nárůstu počtu randomizovaných kodonů. Na ose x je znázorněn počet randomizovaných kodonů, na ose y podíl genů (proteinů) (Hudges a kol., 2003).



Obrázek 11 - Schématické znázornění průběhu MAX randomizace (Hudges a kol., 2003)

Tabulka 9 - Srovnání jednotlivých metod pro konstrukci chytrých knihoven

Metoda	Maximální počet kodonů	Výhody	Nevýhody	Reference
S-SM	1 – 2	- cílení na určitou pozici	- mutace maximálně dvou sousedních kodonů - mutace maximálně šesti bází	Chronopoulou a Labrou, 2011
ISM	3 – 12	- postupné cykly umožňují všechny možné varianty - menší velikost knihoven - mutace mnoha pozic	- potlačení kooperačního efektu mutací - velké množství mutantních variant → větší počet menších knihoven	Reetz a Carballeira, 2007
SSM	1 – 2	- mutace v jeden čas - vzájemný kooperační efekt mutací	- velké knihovny - mutace maximálně šesti bází	Dennig a kol., 2011
ISOR	30 – 45	- vysoká frekvence modifikací v každé pozici - ekvimolární směs oligonukleotidů - cenově dostupná metoda	- velké knihovny → potřeba použít redukované abecedy	Herman a Tawfik, 2007
PM	~18	- přesně stanovené poměry bází - minimální narušení povrchu antigenu	- dříve vysoká cena oligonukleotidů - příliš neprozkoumáno	Krauss a kol., 2010
OSCARR	>10	- 97% úspěšnost - dvoukroková PCR a využití megaprimerů	- velké knihovny - potřeba dvou plazmidů	Hidalgo a kol., 2008
MAX	>12	- žádné synonymní kodony - žádný exponenciální efekt nárůstu počtu kodonů - jednoduché primery - knihovny s maximální diverzitou a minimální velikostí	- velký počet primerů - vysoká cena	Hudges a kol., 2003

4 Modelové enzymy a návrh experimentálního designu

4.1 Modelové enzymy

Halogenalkandehalogenasy (EC 3.8.1.5) jsou převážně mikrobiální enzymy, které katalyzují hydrolytickou přeměnu primárních a sekundárních chloralkanů, bromalkanů a jódalkanů. Podstatou celého mechanismu je štěpení vazby mezi uhlíkem a halogenem za vzniku alkylenzymového intermediátu, který je následně hydrolyzován molekulovou vodou. Výsledkem reakce je odpovídající alkohol, halogenový ion a proton (Janssen, 2004). Halogenalkandehalogenasy patří do skupiny α/β hydrolas. Jejich struktura je tvořena jak z α -helixů, tak z β -listů. Do této skupiny se řadí i lipasy, epoxidhydrolasy, thioesterasy, acetylcholinesterasy, diene-laktonhydrolasy, serinkarboxypeptidasy, haloperoxidasy a další.

Mikroorganismy s geny pro halogenalkandehalogenasy někdy využívají halogenované sloučeniny jako jediný zdroj uhlíku a energie. Geny pro halogenalkandehalogenasy tak byly objeveny u bakterií, které osídlují prostředí kontaminované halogenovanými uhlovodíky. Halogenalkandehalogenasy těchto bakterií hrají významnou roli v biodegradačních a bioremediačních procesech. Dále byly halogenalkandehalogenasy popsány u symbiotických bakterií rostlin či mořských kmenů. Putativní neboli domnělé halogenalkandehalogenasy byly nalezeny dokonce i u některých patogenních organismů jako *Mycobacterium tuberculosis* (Jesenská a kol., 2002).

Halogenalkandehalogenasy se skládají z hlavní a víčkové domény. Aktivní místo leží mezi nimi a je spojeno s povrchem pomocí přístupových tunelů. Katalytická pentáda aktivního místa se skládá z katalytické triády a dvou halogenid-stabilizujících residuí. Dehalogenasová reakce se odehrává právě v aktivním místě, přičemž substrát se váže na enzym a dochází ke vzniku komplexu substrát-enzym. Nukleofilní báze katalytické pentády atakuje atom uhlíku, který je součástí substrátu a vzniká kovalentní alkylenzymový intermediát za současného štěpení vazby mezi uhlíkem a halogenem. Uvolní se molekula halogenu a alkylenzymový intermediát je hydrolyzován molekulou vody, která je aktivovaná bází, která je součástí katalytické pentády (Pavlova a kol., 2009).

4.2 Vlastnosti halogenalkandehalogenas

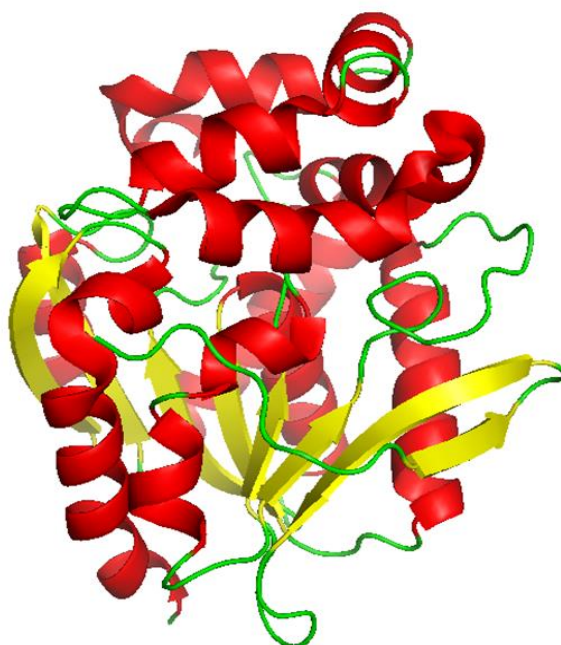
Mezi vlastnosti zájmu u halogenalkandehalogenas patří jejich aktivita, substrátová specifita, termostabilita a enantiosektivita. Aktivita halogenalkandehalogenas je spojena

s aktivním místem a přístupovými tunely. Některé halogenalkandehalogenasy dávají přednost bromoalkanům jiné chloroalkanům. Další upřednostňují spíše menší substráty. Stabilita enzymů je důležitá pro jejich využití v biotechnologických procesech a průmyslu, jelikož stabilita biokatalyzátoru během reakce a možnost recyklace jsou nesmírně důležité. Stabilita zahrnuje termostabilitu a odolnost vůči organickým solventům. Enantiosektivita halogenalkandehalogenas umožňuje produkci opticky čistých sloučenin důležitých pro farmaceutický a chemický průmysl. Velmi zajímavé je využití enzymu DhaA31 při biodegradaci toxického substrátu 1,2,3-trichlorpropanu.

4.3 Halogenalkandehalogenasa DhaA31

Divoký typ enzymu („wild type“) DhaA byl izolován z grampozitivní půdní bakterie *Rhodococcus rhodochrous* NCIMB 13064 (Kulakova a kol., 1997). Aktivní místo tohoto enzymu se skládá z triády tvořené nukleofilem Asp106, bází His272 a katalytickou kyselinou Glu130. Halid-stabilizující residua Asn41 a Trp107 doplňují katalytickou triádu na pentádu (Lahoda a kol., 2011), která je esenciální pro funkci enzymu.

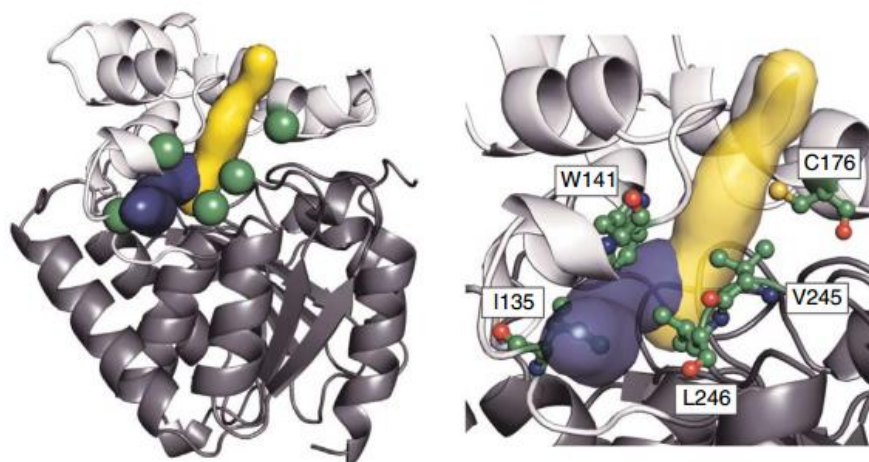
S cílem vylepšit vlastnosti enzymu DhaA bylo vytvořeno několik variant tohoto enzymu. Nakonec byla úspěšně získána varianta DhaA31 (Obr. 12).



Obrázek 12 - Struktura halogenalkandehalogenasy DhaA31 (PDB ID: 3RK4)

Vývoj enzymové varianty DhaA31 byl spojen s výzkumem možné biodegradace 1,2,3-trichloropropanu (TCP). TCP je toxická a pravděpodobně karcinogenní látka, která vzniká jako vedlejší produkt při výrobě jiných halogenovaných sloučenin, například epichlorhydrinu, propylenoxidu či butylenoxidu při výrobě pesticidů, chlorovaných rozpouštědel nebo plastů, a znečišťuje tak podzemní vodu. Dříve se 1,2,3-trichloropropan používal jako odlakovač, odmašťovač nebo rozpouštědlo.

Divoký typ enzymu DhaA nebyl dostatečně aktivní při degradaci TCP, proto bylo třeba se soustředit na vylepšení jeho vlastností a získání nových a aktivnějších variant. Bosma a kol. (2002) aplikoval na divoký typ DhaA jedno kolo DNA míchání („shufflingu“) a error-prone PCR. Byly objeveny dvě nové mutantní formy tohoto enzymu. První mutant (M1) obsahoval mutaci C176Y a vykazoval čtyřikrát vyšší katalytickou aktivitu vůči TCP. Druhý (M2) obsahoval navíc mutaci Y273F. Díky tomu byl až osmkrát aktivnější. Oba modifikované zbytky se nacházely v blízkosti aktivního místa, a činily ho tak méně přístupnější pro vodu, která působila nepříznivě při přeměně TCP, jelikož byla kompetitivní vůči nukleofilu katalytické pentády (Asp106). Mutant M2 byl následně obohacen o třetí substituci, která vedla ke vzniku další mutantní varianty M3 (W141F, C176Y, Y273F). Významnou roli při biodegradaci TCP hrají přístupové tunely, které byly analyzovány prostřednictvím molekulového dokování. Následně došlo k analýze přístupových tunelů také pomocí RAMD simulace – „Random Acceleration Molecular Dynamics“ (Klvana a kol., 2009). Na základě této analýzy byly identifikovány potenciální „hot-spoty“ pro cílenou mutagenézi (Obr. 13). Zkonstruovány byly dvě knihovny, ve kterých se saturovaly identifikované pozice. Otestováno pomocí screeningu bylo celkem 2 568 a 2 705 klonů z každé knihovny (Pavlova a kol., 2009).



Obrázek 13 - „Hot-spoty“ v okolí přístupových tunelů získané na základě RAMD simulace (Pavlova a kol., 2009)

Na základě screeningu bylo identifikováno 51 variant s vylepšenou aktivitou vůči TCP, z toho 25 mělo unikátní sekvenci. Při následné charakterizaci bylo odhaleno, že nejúspěšnější varianta DhaA31 vykazuje až 26krát vyšší katalytickou účinnost vůči TCP. DhaA31 obsahovala celkem pět mutací oproti divokému typu: I135F, C176Y, V245F, L246I a Y273F (Obr. 14, Pavlova a kol., 2009).

Varianta	Reziduum					
	135	141	176	245	246	273
WT	I	W	C	V	L	Y
M2	I	W	Y	V	L	F
M3	I	F	Y	V	L	F
17	F	F	Y	M	I	F
19	Y	F	Y	M	I	F
21	L	F	Y	F	I	F
27	V	W	Y	F	I	F
31	F	W	Y	F	I	F
33	C	W	Y	Y	L	F

Obrázek 14 - Mutantní varianty enzymu DhaA vykazující vylepšenou aktivitu vůči TCP (Pavlova a kol., 2009)

4.4 Návrh redukce aminokyselinové abecedy

V návaznosti na předchozí studii jsem vytvořila teoretický návrh redukce aminokyselinové abecedy za použití bioinformatického nástroje CASTER (Reetz a Carballeira, 2007) s cílem snížit velikost výsledné knihovny.

Varianta DhaA31 byla získána kombinací místně cílené a souběžné saturační mutagenese tří pozic. Pro konstrukci varianty DhaA31 pomocí saturační mutagenese byly použity dva různé templáty nesoucí mutace C176Y+Y273F (M2) a W141F+C176Y+Y273F (M3). Další tři identifikované pozice, I135F, V245F a L246I, byly souběžně saturovány. S těmito třemi substitucemi jsem se rozhodla pracovat také v rámci teoretického návrhu redukované aminokyselinové abecedy taktéž pro souběžnou saturační mutagenesi.

Pro tento teoretický návrh jsem zvolila počítačový nástroj CASTER (Obr. 15). Je volně přístupný na webové stránce <https://www.kofo.mpg.de/en/research/biocatalysis>

a poměrně jednoduchý. Stejně intuitivní a přístupný je i nástroj HotSpot Wizard (Pavelka a kol., 2009), který je možné nalézt na adrese <https://loschmidt.chemi.muni.cz/hotspotwizard>.

The screenshot displays the CASTER software interface with several key sections:

- I. SELECT DEGENERATE CODON:** A table for selecting degenerate codons based on DNA degeneracies. It lists bases (T, C, A, G, Y, R, S, W, K, M, B, D, H, V, N) and their corresponding amino acids or properties.
- II. SET % COVERAGE and % WT background:** Input fields for coverage (95%) and WT background (0%).
- III. Amino acids [AA]:** A list of amino acids with their corresponding codon counts (e.g., Ala [A] 2, Arg [R] 3, Stop 1, total 20).
- IV. Genetic Code - Codon Distribution:** A detailed table showing the distribution of codons across the second position of the codon (T, C, A, G) for various amino acids.
- V. Amino acid Distribution - Chemical Classification:** A table classifying amino acids by side chain (Acidic, Basic, Non-polar aliphatic, Aromatic, Polar, Special Features) and their percentages.
- VI. DNA BASES & BASE PAIRS:** A section for selecting degenerate codons, currently set to 'A. Select Degenerate Codon'.

Obrázek 15 - Uživatelné prostředí počítačového nástroje CASTER

Na základě předchozí saturační mutagenese DhaA31 bylo možné předpovědět, jaké aminokyseliny daná pozice preferuje. Pro pozici 135 jsem proto uvažovala především nepolární aminokyseliny. Stericky nevýrazný alanin a glycin, rigidní prolin, vzácně se vyskytující methionin a tryptofan byl vynechán. V úvahu tedy nakonec připadal pouze valin, leucin, izoleucin a fenylalanin. Jako další možné aminokyseliny pro pozici 135 byl identifikován tyrosin a cystein na základě předchozího souběžné saturační mutagenese. Pozici 245 bylo opět vhodné mutovat za nepolární aminokyseliny stejně jako u první pozice (valin, leucin, izoleucin a fenylalanin). Zařadila jsem také methionin a tyrosin. Valin, leucin a izoleucin jsem uvažovala pro mutagenesi také z toho důvodu, že se tyto aminokyseliny často vyskytovaly v sekvenci divokého typu DhaA na těchto pozicích. Pro pozici 246 byly vybrány dvě aminokyseliny, izoleucin a leucin.

Prostřednictvím nástroje CASTER a myšlenkové úvahy jsem navrhla následující redukovanou aminokyselinovou abecedu: pozice 135 (NDT), 245 (DWK) a 246 (MTT). Výsledná knihovna mutantů by měla obsahovat pouze 288 variant. Pro 95% pokrytí knihovny

by bylo potřeba otestovat jen 861 klonů (Obr. 16). Redukovaná aminokyselinová abeceda by však kódovala jeden stop kodon, který nebylo možné z výběru vyřadit. V případě, že bych při souběžné saturační mutagenезi nepoužila redukovanou aminokyselinovou abecedu, bylo by potřeba otestovat 785 312 variant při uplatnění NNN kodonů a pokrytí knihovny 95 %, s kodonem NNK by se počet snížil na 98 163. V obou případech by redukované aminokyselinové abecedy kódovaly všechny tři stop kodony, což je značně nevýhodné.

Soubor Úpravy Zobrazit Vložit Formát Nástroje Data Okno Nápověda											
A27 =											
A	B	C	D	E	F	G	H	I	J	K	
I. SIMULTANEOUS RANDOMIZATION OF DIFFERENT POSITIONS USING DIFFERENT DEGENERATE CODONS						Amino acids [AA] encoded in each position					
2	Position 1	N	D	T		1	2	3	4	5	
3	Position 2	D	W	K		0	0	0	0	0	
4	Position 3	M	T	T		1	0	0	0	0	
5	Position 4					1	1	0	0	0	
6	Position 5					1	1	0	0	0	
7						0	0	0	0	0	
8	II. SET % COVERAGE		95			Cys [C]	1	0	0	0	0
9	and % WT background		0			Gln [Q]	0	0	0	0	0
10	Positions Randomized	Codons	Colonies			Glu [E]	0	1	0	0	0
11	1					Gly [G]	1	0	0	0	0
12	1 + 2					His [H]	1	0	0	0	0
13	1 + 2 + 3	288	861			Ile [I]	1	1	1	0	0
14	1 + 2 + 3 + 4					Leu [L]	1	1	1	0	0
15	1 + 2 + 3 + 4 + 5					Lys [K]	0	1	0	0	0
16						Met [M]	0	1	0	0	0
17						Phe [F]	1	1	0	0	0
18	IMPORTANT: When different positions are randomized simultaneously it is important to select in each position a degeneracy that contains the correspondent wild-type aminoacid, otherwise the evaluation of the possible effects of the individual libraries would not be properly analyzed, as one of the positions is forced to be mutated.					Pro [P]	0	0	0	0	0
19						Ser [S]	1	0	0	0	0
20						Thr [T]	0	0	0	0	0
21						Trp [W]	0	0	0	0	0
22						Tyr [Y]	1	1	0	0	0
23						Val [V]	1	2	0	0	0
24	The creation of libraries randomizing more than one codon using "mutation forced" degeneracies is as well interesting as the diversity generated differs strongly to the wild-type amino acid sequence.					Stop	0	1	0	0	0
25						Codons	12	12	2	0	0
26						AA	12	10	2	0	0

Obrázek 16 - Výpočet teoretické velikosti knihovny a jednotlivých kodonů za pomoci počítačového nástroje CASTER

Závěr

Funkce a vlastnosti enzymů, jako substrátová specifita, enantioselektivita, stabilita a aktivita, významně souvisí s jejich strukturou. Vlastnosti enzymů je možné modifikovat na úrovni DNA pomocí přístupů a metod proteinového inženýrství. Pro tvorbu cílených knihoven proteinových variant je často a úspěšně uplatňován semi-racionální design. Pracuje pouze se základními znalostmi o struktuře daného proteinu ve srovnání s racionálním designem, zároveň však neklade tak vysoké nároky na screeningové metody, jako je tomu u řízené evoluce. Modifikace vybraných pozic ve struktuře proteinu značně zvyšuje účinnost mutagenese, a přispívá tak k pravděpodobnosti vylepšení průvodního proteinů. Proto se začaly v rámci proteinového inženýrství vyvíjet nové metody zaměřené na konstrukci chytrých knihoven.

Vývoj těchto metod byl a stále je důležitou součástí pokroku v získávání aktivnějších a stabilnějších variant enzymů. Většina z nich je založena na saturační mutagenesi a následné polymerázové řetězové reakci za využití degenerovaných oligonukleotidů. Pro snížení velikosti knihoven se často uplatňují redukované aminokyselinové abecedy, které je možné navrhnout pomocí nejrůznějších počítačových nástrojů (například CASTER nebo HotSpot Wizard). Počítačové nástroje umožňují také výběr specifických míst nebo oblastí ve struktuře proteinu vyznačujících se velkou pravděpodobností pro získání vylepšené varianty enzymu. Úzká spolupráce molekulární biologie a bioinformatiky je v této oblasti zájmu více než důležitá.

Jako modelové enzymy pro tuto práci byly vybrány halogenalkandehalogenasy, které se vyznačují schopností hydrolyzovat primární a sekundární halogenované alkany. U halogenalkandehalogenasy DhaA bylo pozorováno odbourávání atomů chlóru z toxického substrátu TCP, a tedy možné uplatnění tohoto enzymu v bioremediacích. Nízká účinnost a aktivita enzymu divokého typu byla postupně vylepšována. Kombinací racionálního designu, místně cílené a souběžné saturační mutagenese tří pozic byla získána varianta DhaA31 s až 26krát vyšší katalytickou účinností vůči TCP ve srovnání s divokým typem enzymu. Následná teoretická úvaha spočívala v uplatnění souběžné saturační mutagenese dříve vytipovaných pozic DhaA a využití redukované aminokyselinové abecedy, čímž by se výrazně snížily nároky na screening vzniklé knihovny.

Seznam použité literatury

- 1) **Acevedo-Rocha C. G., Kille S. a Reetz M. T.** 2014. Iterative Saturation Mutagenesis: A Powerful Approach to Engineer Proteins by Systematically Simulating Darwinian Evolution. *Methods and Protocols*, 1179: 103-128.
- 2) **Agresti J. J., Antipov E., Abate A. R., Ahn K., Rowat A. C., Baret J.-C., Marquez M., Klibanov A. M., Griffiths A. D. a Weitz D. A.** 2010. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences*, 107(9): 4004-4009.
- 3) **Arnold F. H. a Georgiou G.** 2003. Staggered Extension Process (StEP) In Vitro Recombination. *Directed Evolution Library Creation: Methods and Protocols*, 231: 105-110.
- 4) **Ashkenazy H., Erez E., Martz E., Pupko T. a Ben-Tal N.** 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Research*, 38(2): W529-W533.
- 5) **Balint R. F. a Larrick J. W.** 1993. Antibody engineering by parsimonious mutagenesis. *Gene*, 131(1): 109-118.
- 6) **Berka K., Hanák O., Sehnal D., Banás P., Navrátilová V., Jaiswal D., Ionescu C. M., Svobodová Vareková R., Koca J. a Otyepka M.** 2012. MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Research*, 40(W1): W222-W227.
- 7) **Bornscheuer U. T.** 2005. Trends and Challenges in Enzyme technology. *Advances in Biochemical Engineering/Biotechnology*, 100: 181-203.
- 8) **Bosma T., Damborský J., Stucki G. a Janssen D. B.** 2002. Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene. *Applied and Environmental Microbiology*, 68: 3582-3587.
- 9) **Branden C. I. a Tooze J.** 1999. *Introduction to Protein Structure*. Garland: New York.
- 10) **Cadwell R. C. a Joyce G. F.** 1992. Randomization of genes by PCR mutagenesis. *PCR Methods and Applications*, 2(1): 28-33.
- 11) **Coco W. M., Levinson W. E., Crist M. J., Hektor H. J., Darzins A., Pienkos P. T., Squires C. H. a Monticello D. J.** 2001. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nature Biotechnology*, 19: 354-359.

- 12) **Davies D. R.** 1964. A correlation between amino acid composition and protein structure. *Journal of Molecular Biology*, 9: 605-609.
- 13) **Dehouck Y., Kwasigroch J. M., Gilis D. a Rooman M.** 2011. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC Bioinformatics*, 12(1): 151.
- 14) **Dennig, A., Shivange, A. V., Marienhagen, J. a Schwaneberg, U.** 2011. OmniChange: the sequence independent method for simultaneous site-saturation of five codons. *PLoS One*, 6: e26222.
- 15) **Dundas J., Ouyang Z., Tseng J., Binkowski A., Turpaz Y. a Liang J.** (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research* 34: W116-W118.
- 16) **Eggert T., Funke S. A., Rao N. M., Acharya P., Krumm H., Reetz M. T. a Jaeger K. E.** 2005. Multiplex-PCR-based recombination as a novel high-fidelity method for directed evolution. *ChemBioChem*, 6: 1062-1067.
- 17) **Firth A. E. a Patrick W. M.** 2005. Statistics of protein library construction. *Bioinformatics*, 21(15): 3314-3315.
- 18) **Firth A. E. a Patrick W. M.** 2008. GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries. *Nucleic Acids Research*, 36: W281-W285.
- 19) **Folz R. J., Nothwehr S. F. a Gordon J. I.** 1988. Substrate specificity of eukaryotic signal peptidase. Site-saturation mutagenesis at position -1 regulates cleavage between multiple sites in human pre (delta pro) apolipoprotein A-II. *The Journal of Biological Chemistry*, 263(4): 2070-2078.
- 20) **Geddie M. L. a Matsumura I.** 2004. Rapid evolution of beta-glucuronidase specificity by saturation mutagenesis of an active site loop. *The Journal of Biological Chemistry*, 279(25): 26462-26468.
- 21) **Gibbs M. D., Nevalainen K. M. H. a Bergquist P. L.** 2001. Degenerate oligonucleotide gene shuffling (DOGS): a method for enhancing the frequency of recombination with family shuffling. *Gene*, 271: 13-20.
- 22) **Guerois R., Nielsen J. E. a Serrano L.** 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of Molecular Biology*, 320(2): 369-387.

- 23) **Gustafsson C., Govindarajan S. a Minshull J.** 2003. Putting engineering back into protein engineering: bioinformatic approaches to catalyst design. *Current Opinion in Biotechnology*, 14(4): 366-370.
- 24) **Guzzo A. V.** 1965. The influence of amino acid sequence on protein structure. *Biophysics Journal*, 5(6): 809-822.
- 25) **Hayes R. J., Bentzien J., Ary M. L., Hwang M. Y., Jacinto J. M., Vielmetter J., Kundu A. B. a Dahiyat B. I.** 2002. Combining computational and experimental screening for rapid optimization of protein properties. *Proceedings of the National Academy of Sciences*, 99: 15926-15931.
- 26) **Heinemann U. a Hahn M.** 1995. Circular permutation of polypeptide chains: implications for protein folding and stability. *Progress in Biophysics & Molecular Biology*, 64: 121-143.
- 27) **Herman A. a Tawfik D. S.** 2007. Incorporating synthetic oligonucleotides via gene reassembly (ISOR): A versatile tool for generating targeted libraries. *Protein Engineering, Design & Selection*, 20(5): 219-226.
- 28) **Hidalgo A., Schliessmann A., Molina R., Hermoso J. a Bornscheuer U. T.** 2008. A one-pot, simple methodology for cassette randomisation and recombination for focused directed evolution. *Protein Engineering, Design & Selection*, 21(9): 567-576.
- 29) **Hughes M. D., Nagel D. A., Santos A. F., Sutherland A. J. a Hine A. V.** 2003. Removing the redundancy from randomised gene libraries. *Journal of Molecular Biology*, 331(5): 973-979.
- 30) **Hutchinson C. A. a Edgell M. H.** 1971. Genetic assay for small fragments of bacteriophage phi X174 deoxyribonucleic acid. *Journal of Virology*, 8(2): 181-189.
- 31) **Chaparro-Riggers J. F., Polizzi K. M. a Bommarius A. S.** 2007. Better library design: Data-driven protein engineering. *Biotechnology Journal*, 2(2): 180-191.
- 32) **Chica R. A., Doucet N. a Pelletier J. N.** 2005. Semi-rational approaches to engineer enzyme activity: Combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology*, 16(4): 378-384.
- 33) **Chovancová E., Pavelka A., Beneš P., Strnad O., Brezovský J., Kozlíková B., Gora A., Šustr V., Klvana M., Medek P., Biedermannová L., Sochor J. a Damborský J.** 2012. CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Computational Biology*, 8(10): e1002708.

- 34) **Chronopoulou E. G. a Labrou N. E.** 2011. Site-saturation Mutagenesis: A Powerful Tool for Structure-Based Design of Combinatorial Mutation Libraries. *Current Protocols in Protein Science*, 26: unit 26.6.
- 35) **Janssen D. B.** 2004. Evolving haloalkane dehalogenases. *Science direct. Current Opinion in Chemical Biology*, 8(2): 150-159.
- 36) **Jesenská A., Bartoš M., Czerneková V., Rychlík I., Pavlík I. a Damborský J.** 2002. Cloning and expression of the haloalkane dehalogenase gene dhmA from *Mycobacterium avium* N85 and preliminary characterization of DhmA. *Applied and Environmental Microbiology*, 68: 3724-3730.
- 37) **Jiang L., Althoff E. A., Clemente F. R., Doyle L., Röthlisberger D., Zanghellini A., Gallaher J. L., Betker J. L., Tanaka F., Barbas C. F., Hilvert D., Houk K. N., Stoddard B. L. a Baker D.** 2008. De novo computational design of retro-aldol enzymes. *Science*, 319(5868): 1387-1391.
- 38) **Klvana M., Pavlova M., Koudelaková T., Chaloupková R., Dvořák P., Prokop Z., Stsiapanava A., Kutý M., Kutá-Smatanova I., Dohnálek J., Kulhánek P. a Damborsky J.** 2009. Pathways and mechanisms for product release in the engineered haloalkane dehalogenases explored using classical and random acceleration molecular dynamics simulations. *Journal of Molecular Biology*, 392(5): 1339-1356.
- 39) **Krauss U., Jaeger K. E. a Eggert T.** 2010. Rapid sequence scanning mutagenesis using in silico oligo design and the Megaprimer PCR of whole plasmid method (MegaWHOP). *Methods in Molecular Biology*, 634:127-135.
- 40) **Kretz K. A., Richardson T. H., Gray K. A., Robertson D. E., Tan X. a Short J. M.** 2004. Gene site saturation mutagenesis: a comprehensive mutagenesis approach. *Methods in Enzymology*, 388: 3-11.
- 41) **Kuipers R. K., Joosten H.-J., Van Berkel W. J. H., Leferink N. G. H., Rooijen E., Ittmann E., van Zimmeren F., Jochens H., Bornscheuer U., Vriend G., Martins dos Santos V. A. P. a Schaap P. J.** 2010. 3DM: systematic analysis of heterogeneous superfamily data to discover protein functionalities. *Proteins*, 78(9): 2101-2113.
- 42) **Kulakova A. N., Larkin M. J. a Kulakov L. A.** 1997. The Plasmid-Located Haloalkane Dehalogenase Gene from *Rhodococcus Rhodochrous* NCIMB 13064. ***Microbiology***, 143(1): 109-115.
- 43) **La D., Esquivel-Rodríguez J., Venkatraman V., Li B., Sael L., Ueng S., Ahrendt S. a Kihara D.** 2009. 3D-SURFER: software for high-throughput protein surface comparison and analysis. *Bioinformatics*, 25(21): 2843-2844.

- 44) **Lahoda M., Chaloupková R., Stsiapanava A., Damborský J. a Kutá-Smatanova I.** 2011. Crystallization and crystallographic analysis of the *Rhodococcus rhodochrous* NCIMB 13064 DhaA mutant DhaA31 and its complex with 1,2,3-trichloropropane. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 67(3): 397-400.
- 45) **Landt O., Grunert H. P. a Hahn U.** 1990. A general method for rapid site-directed mutagenesis using the polymerase chain reaction. *Gene*, 96(1): 125-128.
- 46) **Laurie A. T. R. a Jackson R. M.** 2005. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics*, 21(9): 1908-1916.
- 47) **Lee S. H., Ryu E. J., Kang M. J., Wang E., Piao Z., Choi Y. J., Jung K. H., Jeon J. Y. J. a Shin Y. C.** 2003. A new approach to directed gene evolution by recombined extension on truncated templates (RETT). *Journal of Molecular Catalysis B*, 26: 119-129.
- 48) **Lutz S., Ostermeier M., Moore G. L., Maranas C. D. a Benkovic S. J.** 2001. Creating multiple-crossover DNA libraries independent of sequence identity. *Proceedings of the National Academy of Sciences*, 98: 11,248–11,253.
- 49) **Morley K. L. a Kazlauskas R. J.** 2005. Improving enzyme properties: When are closer mutations better? *Trends in Biotechnology*, 23(5): 231-237.
- 50) **Mullis K., Faloona F., Scharf S., Saiki R., Horn G. a Erlich H.** 1986. Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction Cetus Corporation. *Cold Spring Harbor Symposia on Quantitative Biology*, 51: 263-273.
- 51) **Neylon C.** 2004. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. *Nucleic Acids Research*, 32(4): 1448-1459.
- 52) **Nimrod G., Glaser F., Steinberg D., Ben-Tal N. a Pupko T.** 2005. In silico identification of functional regions in proteins. *Bioinformatics*, 21(1): 1328-1337.
- 53) **Ng P. C. a Henikoff S.** 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research*, 31(13): 3812-3814.
- 54) **Nov Y.** 2012. When Second Best Is Good Enough: Another Probabilistic Look at Saturation Mutagenesis. *Applied and Environmental Microbiology*, 78(1): 258-262.
- 55) **Ochman H., Gerber A. S. a Hartl D. L.** 1988. Genetic applications of an inverse polymerase chain reaction. *Genetics*, 120(3): 621-623.
- 56) **O'Maille P. E., Bakhtina M. a Tsai M. D.** 2002. Structure-based combinatorial protein engineering (SCOPE). *Journal of Molecular Biology*, 321: 677-691.

- 57) **Orencia M. C., Yoon J. S., Ness J. E., Stemmer W. P. a Stevens R. C.** 2001. Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nature Structural & Molecular Biology*, 8(3): 238-242.
- 58) **Ostermeier M., Shim J. H. a Benkovic S. J.** 1999. A combinatorial approach to hybrid enzymes independent of DNA homology. *Nature Biotechnology*, 17(12): 1205-1209.
- 59) **Pace C. N., Fu H., Fryar K. L., Landua J., Trevino S. R., Shrirley B. A., Hendricks M. M., Iimura S., Gajiwala K., Scholtz J. M. a Grimsley G. R.** 2011. Contribution of hydrophobic interactions to protein stability. *Journal of Molecular Biology*, 408(3): 514-28.
- 60) **Parra L. P., Agudo R. a Reetz M. T.** 2013. Directed evolution by using iterative saturation mutagenesis based on multiresidue sites. *ChemBioChem*, 14(17): 2301-2309.
- 61) **Pauling L., Corey R. B. a Branson H. R.** 1951. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*. 37(4): 205-211.
- 62) **Pavelka A., Chovancová E. a Damborský J.** 2009. HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Research*, 37: W376-W383.
- 63) **Pavlova M., Klvana M., Prokop Z., Chaloupková R., Banas P., Otyepka M., Wade R. C., Tsuda M., Nagata Y. a Damborský J.** 2009. Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nature Chemical Biology*, 5(10): 727-733.
- 64) **Reetz M. T., Carballeira, J. D. a Vogel A.** 2006a. Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability. *Angewandte Chemie* 118: 7909-7915.
- 65) **Reetz M. T., Wang L.-W. a Bocola M.** 2006b. Directed evolution of enantioselective enzymes: Iterative cycles of CASTing for probing protein-sequence space. *Angewandte Chemie*, 118: 1258-1263.
- 66) **Reetz M. T. a Carballeira J. D.** 2007. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nature Protocols*, 2(4): 891-903.
- 67) **Reetz M. T.** 2011. Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angewandte Chemie*, 50(1): 138-174.
- 68) **Rockah-Shmuel L., Tawfik D. S. a Goldsmith M.** 2014. Generating Targeted Libraries by the Combinatorial Incorporation of Synthetic Oligonucleotides During Gene Shuffling (ISOR). *Methods in Molecular Biology*, 1179: 129-137.

- 69) **Schier R., Balint R. F., McCall A., Apell G., Larrick J. W. a Marks J. D.** 1996. Identification of functional and structural amino-acid residues by parsimonious mutagenesis. *Gene*, 169:147-155.
- 70) **Sieber V., Martinez C. A. a Arnold F. H.** 2001. Libraries of hybrid proteins from distantly related sequences. *Nature Biotechnology*, 19(5): 456-460.
- 71) **Siloto R. M. P. a Weselake R. J.** 2012. Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatalysis and Agricultural Biotechnology*, 1: 181-189.
- 72) **Song J. K., Chung B., Oh Y. H. a Rhee J. S.** 2002. Construction of DNA-shuffled and incrementally truncated libraries by a mutagenic and unidirectional reassembly method: changing from a substrate specificity of phospholipase to that of lipase. *Applied and Environmental Microbiology*, 68: 6146-6151.
- 73) **Stemmer W. P. C.** 1994. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proceedings of the National Academy of Sciences*, 91(22): 10747-10751.
- 74) **Sullivan B., Walton A. Z. a Stewart J. D.** 2013. Library construction and evaluation for site saturation mutagenesis. *Enzyme and Microbial Technology*, 53(1): 70-77.
- 75) **Sun Z., Wikmark Y., Backvall J.-E. a Reetz M. T.** 2016. New Concepts for Increasing the Efficiency in Directed Evolution of Stereoselective Enzymes. *Chemistry – A European Journal*, 22(15), 5046-5054.
- 76) **Tang L., Gao H., Zhu X., Wang X., Zhou M. a Jiang R.** 2012. Construction of “small-intelligent” focused mutagenesis libraries using well-designed combinatorial degenerate primers. *BioTechniques*, 52(3): 149-158.
- 77) **Valdar W. S. J.** (2002) Scoring residue conservation. *Proteins*, 48(2): 227-241.
- 78) **Walter K. U., Vamvaca K. a Hilvert D.** 2005. An active enzyme constructed from a 9amino acid alphabet. *The Journal of Biological Chemistry*, 280: 37742-37746.
- 79) **Wang W. a Malcolm B. A.** 2002. Two-stage polymerase chain reaction protokol allowing introduction of multiple mutations, deletions and insertions, using QuikChangeTM site-sirected mutagenesis. *In vitro mutagenesis protocols*, 180: 37-43.
- 80) **Warburton M., Ali H. O., Liong W. C., Othusitse A. M., Maddock S. a Wong T. S.** 2015. OneClick: A programme for designing focused mutagenesis experiments. *AIMS Bioengineering*, 2(3): 126-143.

- 81) **Wells J. A., Vasser M. a Powers D. B.** 1985. Cassette mutagenesis: An efficient method for generation of multiple mutations at defined sites. *Gene*, 34(2-3): 315-323.
- 82) **Zhang Z., Li Y., Lin B., Schroeder M. a Huang B.** 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics*, 27(15): 2083-2088.