

Svoluji k zapůjčení své diplomové práce ke studijním účelům a prosím, aby byla vedena přesná evidence vypůjčovatelů. Převzaté údaje je vypůjčovatel povinen řádně ocitovat.

UNIVERZITA KARLOVA V PRAZE  
PŘÍRODOVĚDECKÁ FAKULTA  
PROGRAM: BIOLOGIE  
OBOR: MOLEKULÁRNÍ BIOLOGIE, GENETIKA A VIROLOGIE  
KATEDRA GENETIKY A MIKROBIOLOGIE



# DOMÉNOVÁ A STRUKTURNÍ CHARAKTERIZACE TYROZINOVÝCH FOSFORYLAČNÍCH MÍST PROTEINŮ V NÁDOROVÝCH BUŇKÁCH

DOMAIN AND STRUCTURAL CHARACTERIZATION OF TYROSINE PHOSPHORYLATION SITES IN  
CANCER CELLS

Bc. Dan Vávra

Diplomová práce

Školitel: Mgr. Marian Novotný, Ph.D.

Praha, 2014

**Prohlášení:**

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

V Praze, 28. 7. 2014

Bc. Dan Vávra

## **Poděkování**

Rád bych poděkoval všem, bez kterých by tato práce nemohla vzniknout.

Děkuji především svému školiteli Mgr. Marianu Novotnému, Ph.D. za odborné vedení, cenné rady a trpělivost.

V neposlední řadě také děkuji své rodině za podporu během celého studia.

## **Abstrakt**

Fosforylace je důležitým mechanismem regulace funkce a aktivity proteinů. Fosforylace tyrosinu má zásadní roli v signálních drahách. U mnoha nádorových onemocnění byla zjištěna aberantní fosfotyrosinová signalizace. Má práce se zabývá charakterizací patologických vlastností tyrosinových fosforylačních míst v plicních a kolorektálních nádorech z hlediska aminokyselinové sekvence, sekundární struktury, doménové příslušnosti, exprese, výskytu u modelových organismů. To vše na základě literárních údajů a dat proteinových databází. Zjistil jsem, že v nádorech nevznikla nová fosforylační místa.  $\alpha$ -helixy a  $\beta$ -listy jsou ve srovnání se smyčkami fosforylovány výraznou měrou. Nejvíce fosforylovány jsou kinázové domény a Annexin. V nádorech dochází ke změně exprese fosforylovaných proteinů.

### **Klíčová slova:**

tyrosinová fosforylace, plicní nádory, nádory tlustého střeva, doména, sekundární struktura, změna exprese, mutace

## **Abstract**

Phosphorylation is an important mechanism for regulation of protein function and activity. Tyrosine phosphorylation plays a critical role in signaling pathways. Aberrant tyrosine phosphorylation was observed in many cancer types. My work follows pathological details of tyrosine phosphorylation sites of lung and colorectal cancers. Point of view includes aminoacid sequence, secondary structure, domain localization, expression, model organism ortholog occurrence. The project is based on analysis of literary informations and data from protein databases. There are no new phosphorylation sites in observed cancer types. Regular secondary structures,  $\alpha$ -helices and  $\beta$ -sheets, are significantly phosphorylated in compare with loops. Annexin and Kinase domains are the most phosphorylated. Gene expression change of phosphorylated proteins occurs in observed cancer cells.

### **Key words:**

tyrosin phosphorylation, lung cancer, colon cancer, domain, secondary structure, gene expression change, mutation

## Obsah

<b>Abstrakt</b> .....	5
<b>Abstract</b> .....	6
1. Úvod .....	9
2. Literární přehled .....	11
2.1. Fosforylace proteinů .....	11
2.1.1. Kinázy .....	14
2.1.2. Fosfatázy .....	18
2.1.3. Fosforylace tyrosinu .....	20
2.1.4. Fosforylace v plicních nádorech .....	22
2.1.5. Detekce fosforylace .....	25
2.2. Nádorová onemocnění .....	27
2.2.1. Rakovina plic .....	28
2.2.2. Kolorektální karcinom .....	29
2.2.3. Nádorová onemocnění a fosforylace .....	30
2.3. Databáze .....	32
2.3.1. Databáze fosforylací .....	33
3. Cíle práce .....	35
4. Metody .....	35
4.1. PhosphoSite a aktualizace informací .....	35
4.2. Výchozí série dat .....	36
4.3. Sekvenční analýza .....	41
4.4. Strukturní analýza .....	42
4.5. Expresní analýza .....	43
4.6. Charakteristika orthologů fosforylačních míst .....	44
4.7. Srovnání výskytu fosfoaminokyselin .....	45
5. Výsledky a diskuse .....	46
5.1. Srovnání výskytu fosfotyrosinu u vybraných typů tumorů .....	46
5.2. Vznik nových fosforylačních míst v nádorové tkáni .....	48
5.3. Sekundární struktura fosforylačních míst .....	48
5.4. Doménová analýza fosforylačních míst .....	50
5.5. CATH charakteristika .....	54
5.6. Exprese fosforylovaných proteinů v nádorech .....	57
5.7. Charakteristika orthologů fosforylačních míst .....	60
6. Závěr .....	61
7. Seznam použité literatury .....	62

## Seznam zkratek

aPk atypická protein kináza  
ATP adenosintrifosfát  
CAMK Ca<sup>2+</sup>/kalmodulin dependentní  
proteinkináza  
CDK Cyklindependentní kináza  
CK1 Kaseinkináza I  
EGFR Receptor epidermalního růstového faktoru  
EM electron microscopy  
ePk eukaryotická proteinkináza  
GIST Gastrointestinální stromatární tumor  
GTP guanosintrifosfát  
MAPK Mitogenem aktivovaná proteinkináza  
MEK kináza MAP kinazy  
OMIM Online Mendelian Inheritance in Man  
PDB Protein DataBank  
PGDFR Receptor růstového faktoru odvozeného  
od krevních destiček  
PKA Proteinkináza A  
PKC Proteinkináza C  
PKG Proteinkináza G  
PTM posttranslační modifikace  
PTP Fosfotyrosin fosfatáza  
Rb Retinoblastoma protein  
SNP Jednonukleotidový polymorfismus (Single  
nucleotide polymorphism)  
SNV Jednonukleotidová mutace (Single nucleotide  
variation)



# 1. Úvod

## Posttranslační modifikace

Během posledních několika desetiletí bylo zjištěno, že lidský proteom je mnohem komplexnější než genom. Zatímco genom obsahuje mezi 20000 a 25000 genů (International Human Genome Sequencing Consortium 2004), celkový počet lidských proteinů se odhaduje na více než 1 milion (Jensen 2004). Je proto zřejmé, že jednotlivé geny mohou kódovat více proteinů. Rekombinace, iniciace transkripce na alternativních promotorech a diferenciální terminace i alternativní sestřih jsou mechanismy, které mohou z jednoho genu vytvořit různé mRNA (Ayoubi & Van De Ven 1996).

Proteom je v reakcích na množství podnětů velmi dynamický a zvýšení komplexity i funkční rozmanitosti od genomu směrem k proteomu je dále usnadněno posttranslačními modifikacemi bílkovin (PTM), tedy kovalentním přidáním funkčních skupin nebo polypeptidů, proteolytickým štěpením regulačních podjednotek nebo degradací celých proteinů. Tyto úpravy zahrnují fosforylaci, glykosylaci, ubikvitinaci, nitrosylaci, metylaci, acetylaci, sumoylaci, hydroxylaci, lipidaci a řadu dalších. Jedná se tedy o chemické modifikace, nejčastěji zprostředkované enzymatickou aktivitou, které klíčovým způsobem usměrňují vlastnosti bílkovin a následkem toho téměř všechny aspekty biologie buňky, neboť regulují aktivitu, lokalizaci i interakci s jinými molekulami, jako jsou další proteiny, nukleové kyseliny, lipidy, kofaktory. Mají vliv na buněčný cyklus, regulaci imunitních mechanismů, podílejí se na intercelulární komunikaci (Liu et al. 2011).

PTM lze najít v různých postranních řetězcích aminokyselin nebo peptidových vazbách. Odhaduje se, že 5 % proteomu obsahuje enzymy, které slouží k provádění více než 200 druhů posttranslačních modifikací (Walsh 2006). Mezi tyto enzymy patří kinázy, fosfatázy, transferázy, ligázy, které přidávají či odebírají funkční skupiny, proteiny, lipidy, cukry z postranních řetězců aminokyselin, a proteázy, jež štěpí peptidové vazby za účelem odstranění specifických sekvencí nebo regulačních podjednotek. Mnoho proteinů má také schopnost modifikovat samo sebe pomocí autokatalytických domén, jako jsou například autokinázové a autoproteolytické domény.

Posttranslační modifikace se může objevit v každé fázi " životního cyklu " proteinu. Mnoho

proteinů je například upraveno již krátce po dokončení translace. Je tak zprostředkováno správné sbalení, regulace stability či „adresování“ vznikajícího proteinu do různých buněčných kompartmentů (např. jádro, membrána). Následné další úpravy se mohou vztahovat k aktivaci nebo deaktivaci katalytické aktivity a jiným ovlivněním biologických funkcí dané bílkoviny. Někdy jsou kovalentně připojeny molekuly, které předurčují protein k degradaci (ubikvitinace). Kromě jednotlivých modifikací dochází často ke kombinovaným úpravám pomocí kombinace posttranslačního štěpení s adicí funkčních skupin prostřednictvím postupného mechanismu maturace a aktivace bílkovin (Liu et al. 2011).

Identifikace, analýza a pochopení PTM má velký význam ve studiu buněčné fyziologie či patogeneze, v léčbě nebo prevenci nemocí, studiu srdečních i neurodegenerativních onemocnění, rakoviny, diabetes. Charakterizace PTM poskytuje neocenitelné nahlédnutí do zákonitostí regulace buněčných dějů a chemicko-fyziologických procesů.

Některé PTM, nejčastěji ryze regulačního charakteru, modulující aktivitu, efektivitu, dobu existence, chemickofyzikální vlastnosti (náboj, hydrofobicita) proteinu jsou v závislosti na své povaze reverzibilní. Především chemickofyzikální parametry bílkoviny mění acetylace, metylace, glykosylace, hydroxylace. Například jedna z histonových modifikací, připojení acidických acetylových skupin k lysinovým zbytkům v N-terminální oblasti histonu, odpuzuje na něm navinuté vlákno DNA, dochází k rozvolnění nukleozomu. Účel ireverzibilní acetylace první N-terminální AMK určitých proteinů není s jistotou prokázán, dle dřívějších názorů mělo docházet k prodlužování životnosti dané bílkoviny. Nedávno bylo zjištěno, že protein může být takto označen a zacílen do degradačního procesu (Arnesen 2011). Proteolytické štěpení peptidových vazeb, například odstranění části řetězce trypsinu, je termodynamicky výhodnější reakce, proto peptidové sekvence a regulační domény odstraňuje trvale.

Ve výzkumu posttranslačně modifikovaných proteinů je důležitý hlavně rozvoj specifických metod detekce a purifikace, přičemž z velkého množství různých PTM je největší důraz kladen na několik nejběžnějších, zvláště pak na fosforylace, glykosylace a ubikvitinace (Liu et al. 2011).

Glykosylace je jedna z hlavních posttranslačních modifikací, s významnými dopady na skládání, konformaci, distribuci, stabilitu a aktivitu proteinů. Zahrnuje rozmanitý výběr sacharidových komponent adovaných k bílkovině, od jednoduchých monosacharidových úprav

jaderných transkripčních faktorů až k vysoce komplexním rozvětveným polysacharidovým modifikacím povrchových receptorů buňky. Sacharidy ve formě asparginem vázaných ( N - vázaných ) nebo serin / threonin - vázaných ( O – vázaných) oligosacharidů jsou hlavní strukturní složkou mnoha buněčných povrchů a sekretovaných proteinů (Lodish et al. 2004).

Ubikvitin je 8kDa (kiloDaltonů) polypeptid skládající se z 76 aminokyselin, který se připojuje k (NH<sub>2</sub>)- lysinu v cílových proteinech, pomocí svého C - terminálního glycinu. Po počáteční monoubikvitinaci, může dojít k vytvoření ubikvitinového polymeru, a takto označené proteiny jsou pak rozpoznávány proteazomem 26S, který katalyzuje degradaci ubikvitinovaného proteinu a recyklaci ubikvitinu. To tedy ovlivňuje životnost proteinu. Jako příklad lze uvést ubikvitinační regulaci hladiny nádorového supresoru p53 (Lodish et al. 2004).

## **2. Literární přehled**

### **2.1.Fosforylace proteinů**

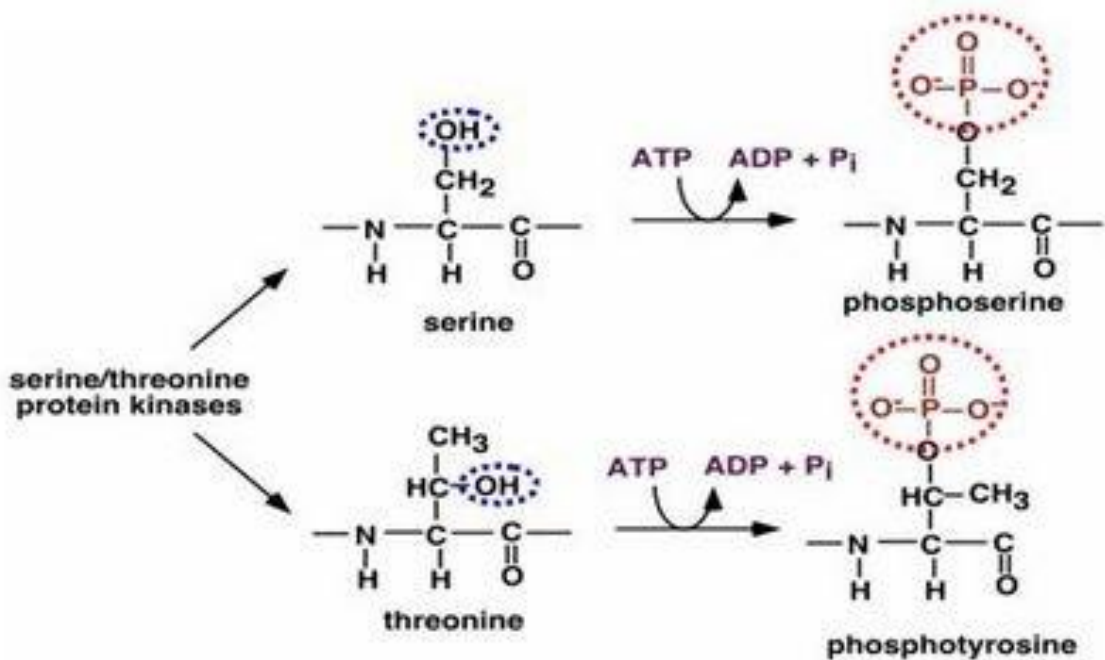
Existuje více substrátů, které mohou být fosforylovány, reverzibilní fosforylace proteinů je však jednou z nejdůležitějších a dobře prostudovaných post-translačních modifikací, dochází k ní během celého životního cyklu buňky eukaryotní i prokaryotní. Jako nejčastější mechanismus modulace funkce bílkovin, hraje fosforylace klíčovou roli v regulaci mnoha procesů včetně buněčného cyklu, dělení, růstu, metabolismu, diferenciaci, transportu, apoptózy nebo signálních drah, kde zajišťuje detekci extra/intracelulárních podnětů i reakce na ně prostřednictvím kináz a fosfatáz (provádějí defosforylaci). Je tedy zřejmé, že defekty ve fosforylaci proteomu, interpretaci a šíření signálu často mají závažné důsledky jak pro jednotlivou buňku, tak pro celý organismus. Odhaduje se, že 30 až 50% lidských proteinů je za určitých podmínek substrátem pro fosforylaci (Pinna & Ruzzene 1996).

Zjištění existence specifických proteinkináz stálo na počátku objasnění mechanismu a účelu fosforylace (Koenig M. & Grabe N. 2004). Roku 1954 Gene Kennedy publikoval objev bílkoviny, katalyzující v játrech fosforylaci kaseinu, byla odhalena úloha proteinkináz. Následně, roku 1955 Fischer s Krebsem a téměř současně rovněž Sutherland s Wosilaitem popsali enzymy, podílející se na metabolismu glykogenu, jejichž funkční vlastnosti jsou modulovány připojením či odstraněním fosfátové skupiny (Philip Cohen 2002). Začátkem osmdesátých let došlo k objevu kinázové aktivity virového (Rousův sarkom) proteinu v-Src,

který zajišťuje tyrosinovou fosforylaci proteinových substrátů. Ve zdravých buňkách byl nejčastěji fosforylován treonin a serin oproti tyrosinu, což znesnadňovalo zjišťování fosforylace prováděné v-Src.

Fosfát, zbytek kyseliny fosforečné, je přidán nejčastěji na kyslíkový atom serinu, treoninu nebo tyrosinu či dusíkový atom histidinu, aspartátu. Fosforylace histidinu, aspartátu probíhá hlavně u rostlin a prokaryot. Dusíkové kinázy živočichů, histidinové i další, jsou nepřibuzné prokaryotním, aspartátkinázy se u živočichů nevyskytují (Klumpp & Krieglstein 2002). Fosforylovány mohou dále být také aminokyseliny arginin, lysin, glutamát (Thomason & Kay 2000).

Aminokyseliny serin, treonin i tyrosin mají nukleofilní (OH)-1 skupinu, která atakuje vazbu mezi fosforem terminálního  $\gamma$ -fosfátového rezidua (PO<sub>3</sub>)-1 a kyslíkem  $\beta$ -fosfátové skupiny na ATP nebo GTP. Z ATP uvolněný  $\gamma$ -fosfátový zbytek se připojuje přes fosforový atom kovalentní esterovou vazbou na místo vodíku této (OH)-1 skupiny postranního řetězce aminokyseliny v proteinu (viz obr. č. 1). Enzymem katalyzovaný přenos protonů z (OH)-1 stimuluje nukleofilní atak  $\gamma$ -fosfátu na ATP, což má tedy za následek přenos fosfátu na AMK za vzniku fosfoAMK a ADP. Při defosforylaci je vazba mezi fosfátem a AMK hydrolyzována, je uvolněn anorganický fosfát. (Pinna & Ruzzene 1996).



**Obr.č.1:** Mechanismus fosforylace - přenos fosfátu z ATP na akceptorová místa serinu a treoninu za vzniku ADP + Pi ([http://www.dolcera.com/wiki/images/thumb/400px-GSK3\\_phosphorylation.jpg](http://www.dolcera.com/wiki/images/thumb/400px-GSK3_phosphorylation.jpg))

Esterové vazby s fosfátovými zbytky jsou v prostředí s neutrálním pH všeobecně velmi stabilní, vazba bílkoviny a fosfátu tedy díky své vysoké energii nepodléhá spontánní hydrolyze, samovolně k defosforylaci nedojde. U kináz i fosfatáz se ale projevuje určité oslabení této vazby, zapříčiněné interakcemi s kovovými ionty v jádře enzymů a postranními řetězci AMK. Pro velkou skupinu proteinů je fosforylace úzce spjata s jejich aktivitou a funkcí, neboť působí konformační změny ovlivňující protein dvěma hlavními způsoby. Protein může být, co se katalytických funkcí týká, aktivován a deaktivován či z hlediska interakcí s jinými bílkovinami identifikován konkrétním enzymem jako vhodný substrát, to vše v důsledku drobných změn terciární nebo kvartérní struktury, iniciovaných v různém rozsahu na okolí působícími chemickofyzikálními vlastnostmi fosfátového rezidua, zejména kyselostí, velkým nábojem i značným prostorem, který zabírá (Hunter 2012). Fosfoproteiny mohou tedy interagovat se strukturně konzervativními doménami okolních proteinů, které se na ně váží se specifitou pro odlišné aminokyseliny. Například SH2 a PTB domény vykazují specifitu pro fosfotyrosin, avšak rozdíly v jejich struktuře zapříčiňují, že každá je specifická pro odlišné fosfotyrosinové motivy (Yaffe 2002). Schopnost fosfoproteinů interagovat s dalšími proteiny má také zásadní význam pro přenos signálu, když jsou efektorové proteiny ovlivňovány signalizujícími fosforylovanými proteiny.

Fosforylace nacházející se ve vazebné doméně retinoblastoma proteinu (Rb) kompletně zamezuje vazbě E2F transkripčního faktoru, jiná fosforylace Rb má vliv na změnu struktury vazebné domény a pouze sníží schopnost E2F vázat. Taková regulace asociace Rb s transkripčními faktory i jinými usměrňovacími proteiny může sloužit jako jeden z mnoha příkladů, že fosforylační modulace funkce bílkoviny často zahrnuje širší škálu možností, než jen alteraci aktivní/inaktivní fáze. Četnost výskytu regulačních fosforylací je nejnižší v uspořádaných strukturách a velmi rychle roste s délkou, neuspořádaností, rozvolněností smyček, protože tam jsou aminokyselinová rezidua čím dál lépe dostupná pro kinázy i fosfatázy (Tyanova et al. 2013).

Vývoj kináz a fosfatáz byl pochopitelně úzce spjat s vývojem fosforylace, která se jako regulační faktor objevila už u prokaryot. Prokaryotních kináz, jež u eukaryot nenalezneme, je sice velké množství, některé se však u určité úzké skupiny přece jen vyskytují nebo vykazují k eukaryotním kinázám značnou homologii.

V důsledku větší velikosti, kompartmentace eukaryotních buněk i vzniku mnohobuněčnosti, došlo k evoluci komplikovanějších způsobů regulace, na rozdíl od prokaryot, kde v mnoha případech postačuje usměrňování prostřednictvím rozpětí hladin koncentrace či dostupnosti substrátu, metabolických meziproduktů atd.

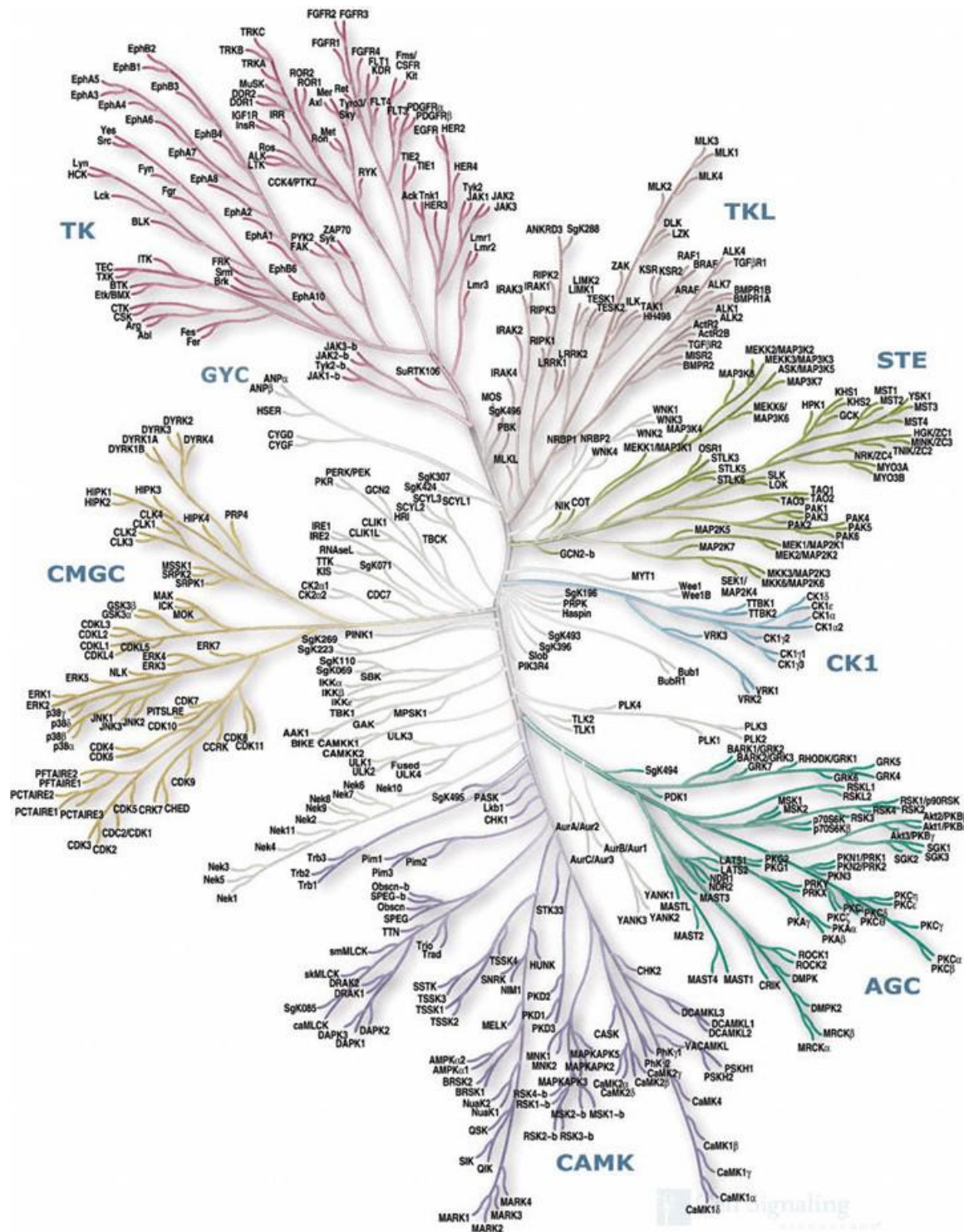
### 2.1.1. Kinázy

Kinázy jsou enzymy, patřící dle názvosloví ustanoveného Mezinárodní biochemickou a molekulárně biologickou unií mezi transferázy, které katalyzují připojení fosfátové skupiny k substrátu, fosforylaci. Substráty pro kinázy jsou rozmanité a zahrnují lipidy, sacharidy, nukleotidy a bílkoviny. V lidském genomu bylo přes sekvenční databáze pomocí skrytých Markovových modelů identifikováno 518 proteinkinázových genů neboli 478 eukaryotických (ePK) a 13 atypických (aPK) proteinkináz, přičemž se podařilo určit 491 ePK a 13 aPK domén. To je tedy podmnožina proteinů reprezentujících lidský kinom. Majoritní množství těchto kináz tedy spadá do nadrodiny determinované výskytem charakteristické katalytické domény ePK a obsahující osm recentně uznávaných kinázových skupin viz obr. č. 2 (Manning et al. 2002):

- 1) TKL (tyrosin like-kináza)
- 2) TK (tyrosin kinázy, konsensuální tyrosinové kinázy): významné členění na receptorové a nereceptorové kinázy; jména rodin podle kináz, jejichž název nezdávka bývá spojen s patologickým stavem, zapříčiněným jejich nesprávnou aktivitou či expresní hladinou; např. rodina Brk (Brk kináza, rakovina prsu)
- 3) CAMK má jen dvě rodiny: Snfl/AMPK (AMP aktivovaná proteinkináza), Ca<sup>2+</sup>/kalmulin dependentní proteinkinázy a jim blízké příbuzné
- 4) STE (homolog kvasinkových kináz - Sterile)
- 5) CK1(kasein kináza)
- 6) RGC (receptor guanylát cykláza)
- 7) CMGC zahrnuje hlavně rodiny: glykogen syntetázy aktivující kinázy 3, MAPK = mitogenem aktivované kinázy (Erk), CLK (Cdc-2 like kináza a příbuzné), GSK3 (glykogen syntáza kináza), CDK a příbuzné (cyklin dependentní kinázy, Cdc = cell division cycle=kinázy buň. cyklu, jejich homology)
- 8) AGC skupina, kam patří např. rodiny: kinázy fosforylující s receptory spřažené G proteiny, kinázy regulované cyklickým nukleotidem (PKG = proteinkináza G, PKA =

proteinkináza A), fosfolipid-dependentní/ diacylglycerol-aktivované kinázy (PKC), kinázy příbuzné PKC a PAK (RAC)

Na rodiny a podrodiny byly skupiny rozděleny analýzou aminokyselinových sekvencí katalytických domén (Hanks & Hunter 1995).



Obr. č. 2: Fylogenetické zobrazení 8 skupin (498 ePK domén) proteinkináz (Manning et al. 2002)

Proteinkinázy lze dle fosforylované aminokyseliny rozdělit do skupin, které mají odlišné katalytické domény: tyrosinkinázy, serin / threonin kinázy, duální kinázy. Přibližně 85-90 % savčího kinomu tvoří serin / threonin kinázy a více než 90% fosfoproteomu tvoří fosfoserin a fosfotreonin. Relativní poměr fosfoserinu, fosfotreoninu, fosfotyrosinu v buňce je 1800:200:1 (Mann et al. 2002). Ačkoli fosfotyrosin není tak převládající jako fosfoserin a fosfotreonin, více ovlivňuje strukturu proteinů (Hunter 2009), má hluboký regulační význam a vztah různých dysregulací fosforylace tyrosinu k lidským onemocněním je významným předmětem biomedicínského výzkumu.

Substrátová specifita protein kináz je založena nejen na cílové aminokyselině, ale také na konsenzuálních sekvencích, které ji obklopují (Pawson & Nash 2003). Tyto sekvence umožňují některým kinázám fosforylovat jednotlivé proteiny a dalším více substrátů (> 300) (Walsh 2006). Kromě toho mohou kinázy fosforylovat jednu nebo více aminokyselin na individuálním proteinu, jsou-li k dispozici specifické konvenční sekvence. Serin/threonin kinázy byly také klasifikovány na základě vlastností aminokyselinových pozic okolí fosforylovaného místa (Pinna & Ruzzene 1996):

A) Kinázy acidofilní nebo kinázy preferující pozice s předešlou fosforylací. Je jich málo, ale jsou velmi běžné. Např. CK1 rodina z CK1 skupiny i CK2 z CMGC skupiny (kaseinové kinázy) a další z klasifikací (Hanks & Hunter 1995).

B) Kinázy bazofilní s preferencí bazických (především lysin) či hydrofobních (někdy větší význam než bazické) reziduí poblíž fosforylované pozice. Např. mimo několika málo celá skupina CAMC, převážné množství z ACG.

C) Kinázy s prolinovou navigací, kdy se na n+1 místě musí nacházet prolin, což u jiných kináz naopak fosforylací zamezuje. V závislosti na povaze určitých kinázových rodin jsou občas ještě vyžadovány zásadité nebo acidické zbytky v blízkosti fosforylovaného místa apod. Rodiny MAPK a CDK ze skupiny CMGC.

Tyrosinkinázy se dělí na receptorové s intra- i extracelulární doménou a nereceptorové (volné), s doménami, které se váží k receptorům postrádajícím enzymatickou aktivitu. Na počátku šíření signálu neseného kaskádovou fosforylací často kináz nebo jiných proteinů, dochází k aktivaci receptorů extracelulárními ligandy (růstový faktor aj.) následované dimerizací a reciprokou fosforylací kináz receptorových či asociovaných s receptorem bez enzymatické aktivity (Pinna & Ruzzene 1996).



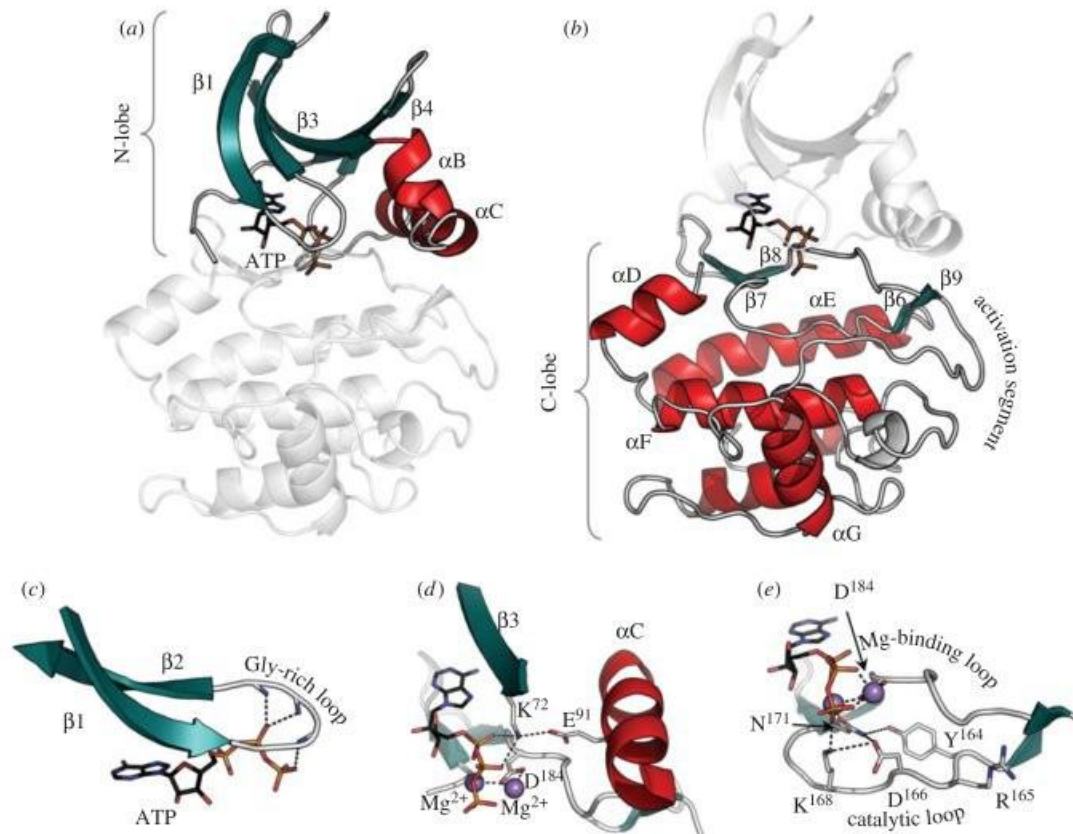
Hodnota vazebné energie pro připojení fosfátu k tyrosinu, serinu, treoninu se liší minimálně, avšak úloha fosfoserinu i fosfotreoninu je ponejvíce strukturní, na rozdíl od fosfotyrosinu a jeho mnohých regulačních interakcí. Tyrosinkinázy jsou aktivovány pouze za silně specifických podmínek, neboť podléhají velmi výrazné negativní regulaci. V případě, že fosfotyrosin není stabilizován vazbou k SH2 doméně, je v důsledku vysoké dynamiky fosfatáz téměř okamžitě odbourán. Přítomnost aromatického kruhu činí vazebnou interakci fosfotyrosinu s SH2 doménou značně stabilní, proto má tyrosinová fosforylace tak významnou regulační roli (Hunter 2009).

Fosforylaci treoninu i tyrosinu ve stabilní sekvenci MAP kinázy provádí MEK, oproti všem ostatním kinázám s dvojnou aktivitou obvyklá a rozšířená. Funkcí duálních kináz mělo být dle dřívějších představ šíření signálu z tyrosinkináz na serin/treonin kinázy. Dvojná aktivita kináz, schopnost fosforylace tyrosinu i treoninu či serinu je však v mnoha případech autofosforylační nebo se projevuje jen výjimečně, někdy pouze za spolupůsobení silně specifických faktorů a tyto kinázy tedy příliš běžné nejsou (Pinna & Ruzzene 1996).

Kinázy mají regulační podjednotky, které mají různé substráty a fungují jako aktivátory nebo inhibitory ostatních domén. Jejich fosforylací dochází k regulaci kinázové aktivity (Johnson & Lewis 2001). Většina proteinkináz je v bazálním stavu defosforylovaná a neaktivní, a jsou aktivovány fosforylací. Malé množství kináz je konstitutivně aktivní, a jsou fosforylací naopak deaktivovány (CK2 serin/treonin kináza).

ATP je kosubstrátem pro téměř všechny proteinové kinázy, i když malý počet kináz používá guanosin trifosfát. ATP je ideální struktura pro přenos  $\gamma$  - fosfátových skupin, fosforyltransfer (Walsh 1979). Zatímco substrátové specifity kináz se liší, ATP - vazebné místo je obecně zachováno (Walsh 2006). Dosti konzervovaná kinázová doména eukaryotních proteinkináz váže substrát, váže fosfátový donor (ATP nebo GTP) pomocí komplexu s hořčnatými kationty, u obou moderuje správnou prostorovou orientaci a následné přenesení gama fosfátu z ATP na hydroxylový zbytek substrátu (Hanks & Hunter 1995). Typická kinázová doména (viz obr. č.3) je tvořena N koncovým a C koncovým lalokem. Fosfát je orientován k substrátu katalyticky aktivní smyčkou, umístěnou mezi  $\beta_6$  a  $\beta_7$  ( $\beta$ -hřeben) C laloku a umístění hořčíkových iontů určuje smyčka mezi  $\beta_8$  a  $\beta_9$ . C koncový lalok se však skládá vyjma čtyř  $\beta$ -hřebenů hlavně z  $\alpha$ -helixů. N terminální lalok obsahuje pět antiparalelních  $\beta$ -hřebenů a dva  $\alpha$ -helixy. Na glycin bohatá smyčka určující orientaci ATP se nachází mezi prvním a druhým

$\beta$ -hřebenem, mezi třetím a čtvrtým  $\beta$ -hřebenem jsou dva  $\alpha$ -helixy, někdy jen jeden (nazýván  $\alpha$ C), ale nepostradatelný (Taylor et al. 2012).

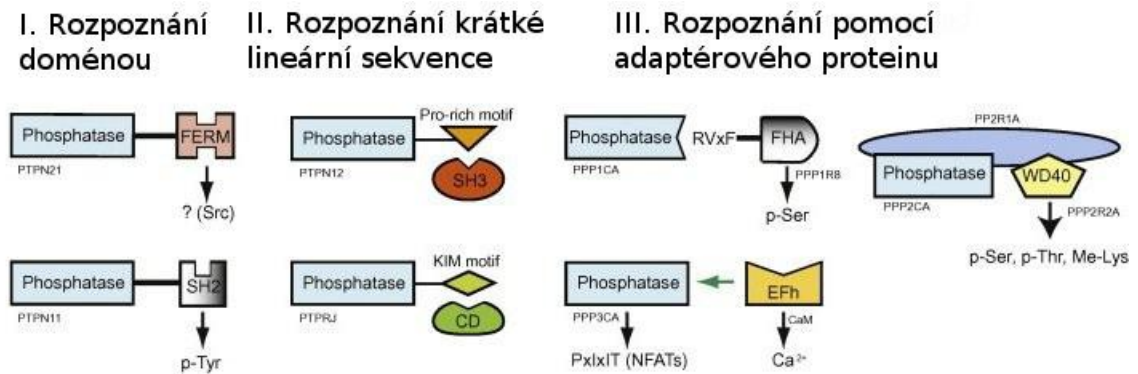


Obr. č.3: Stavba kinazové domény eukaryotické proteinkinázy (PKA), N koncový lalok (a) C koncový lalok (b) a důležité funkční motivy: glycin bohatá smyčka (c), konzervované zbytky v  $\beta$ 3 a  $\alpha$ C (d), katalytická smyčka (e). (Taylor et al. 2012)

### 2.1.2. Fosfatázy

Fosfoproteiny jsou v případě potřeby zpětně defosforylovány pomocí enzymatické aktivity fosfatáz. Lidský proteom obsahuje přibližně 150 fosfatáz, které vykazují specifitu pro fosfoserinová/fosfotreoninová a fosfotyrosinová rezidua (Kennelly 2001). Přestože defosforylace je konečným cílem obou těchto skupin fosfatáz, každá tuto reakci realizuje prostřednictvím odlišných mechanismů. Serin/threonin fosfatázy zprostředkovávají přímou hydrolýzu fosfátové skupiny pomocí bimetalového (Fe/Zn) centra, zatímco tyrosin fosfatázy tvoří kovalentní meziprodukt thiofosforyl, který usnadňuje odstranění tyrosinového zbytku. Substrát je každou fosfatázou rozpoznán jedním ze tří známých způsobů. Doménou (např. SH2) schopnou určit cílové proteiny či adaptorovým proteinem nekovalentně vázaným k fosfatáze a

nebo může rovněž také dojít k identifikaci krátkého lineárního motivu fosfatázy i jeho následnému navázání cílovým proteinem, viz obr. č.4 (Sacco et al. 2012).



**Obr. č. 4:** Možnosti identifikace substrátu fosfatázou (Sacco et al. 2012)  
 1) prostřednictvím domény SH2; 2) rozpoznáním krátkého lineárního motivu;  
 3) identifikace adaptérovým proteinem

Substrátová specifita fosfatáz je in vivo vysoká, in vitro se vytrácí, pravděpodobně kvůli přítomnosti jiných složek a substrátů, in vivo nedostupných. Genom *Saccharomyces cerevisiae* obsahuje 1,9% proteinkináz a jen 0,23% proteinfosfatáz (Hunter & Plowman 1997). Organismy tedy prokazatelně obsahují daleko nižší množství fosfatáz než kináz, fosfatázy však mají velký regulační význam a nepřítomnost, nedostatek či nadbytek různých typů těchto enzymů dokáže buněčné funkce vážně ohrozit. Např. v případě absence fosfatázy deaktivující defosforylací glykogen syntázu by byly energetické zásoby buňky velice rychle eliminovány. Navzdory tomu, že fosfatáz není obecně tolik jako kináz, tyrosinfosfatázy jsou naopak početnější než tyrosinkinázy, což poukazuje na významnou úlohu tyrosinové fosforylace v regulačních pochodech a interakcích (Mustelin 2007). V tabulce č. 1 je přehled eukaryotických fosfatáz.

Fosforyl. AK	rodina	podrodina	zástupce
S/T	PPM		PP2C
	FPC		FPC
	PPP		PP1, PP2A, kalcineurin
Y,S	HAD		Eya, CTD, cronophin
Y	Třída I Cys- založených PTP	klasické PTP	CD45, PTP1B
		VH1-like a duální fosfatázy	MKP
	Třída II Cys- založených PTP		CDC25A, CDC25B, CDC25C
	Třída III Cys- založených PTP		LMPTP

**Tabulka č. 1:** Tř. I Cys-založ. PTP: Klasické PTP se dělí na receptorové (transmembránové) tyrosinfosfatázy, nereceptorové fosfatázy. VH1-like a duální fosfatázy - fosfatáza MAP kinázy.

Tř. II Cys-založ. PTP sice nemá strukturní příbuznost s Tř. I, ale funkční podobnost je zde značná.

Tř. III Cys-založ. PTP: LMPTP = jediný lidský zástupce, tyrosinfosfatáza o nízké molekulové hmotnosti.

PPM i FPC vyžadují Mg<sup>2+</sup>.

HAD (Asp-založ.): Patří sem některé serin- a tyrosinfosfatázy. Různé substráty – fosforylované bílkoviny, tuky, sacharidy, nukleotidy (Mustelin 2007).

### 2.1.3. Fosforylace tyrosinu

Během studií polyomaviru roku 1979 vedla náhodná pozorování kinázové aktivity v-Src k objevu fosforylace tyrosinu jako nového typu modifikace proteinů. Od té doby byl učiněn velký pokrok v objasňování mechanismů, jakými tyrosinová fosforylace ovlivňuje vlastnosti a funkce bílkovin (Hunter 2009).

Fosforylace tyrosinu je obvykle obstarávána specifickými tyrosinkinázami. Tyrosin kinázy se dělí na receptorové a nereceptorové.

Receptor tyrosin kinázy (RTK) zaujímají 58 z 90 tyrosin kináz v lidském genomu (Manning et al. 2002). Zjištění, že navázáním ligandu vyvolaná autofosforylace RTK je následována rychlým zvýšením koncentrace růstových faktorů (EGF, PDGF), odhalilo důležitou roli

tyrosinové fosforylace v růstových signálních drahách, proliferaci a rovněž tedy i v onkogenezi. Fosfotyrosinový zbytek aktivované RTK je dle AMK sekvencí blízko (downstream) fosfotyrosinu rozpoznán a navázán doménou SH2, což je klíčové pro start a šíření signálu. Specifita tyrosinkináz je kromě rozeznávané sekvence upřesněna SH2 doménou (Pawson 2004). Proteiny s SH2 doménou zastávají celou řadu funkcí, mohou to být adaptérové bílkoviny schopné rekrutovat další signalizační proteiny, enzymy působící na membránové molekuly (např. fosfolipázy), cytoplazmatické tyrosin kinázy, které šíří signál dál, E3 ubiquitin ligázy nebo transkripční faktory. Mezi fosfotyrosin vázící domény také patří PTB, jež se rovněž podílí na RTK signalizaci a ještě několik dalších. RTK mají N-terminální ligandy vázící extracelulární doménu, transmembránovou a C-terminální cytoplazmatickou doménu, která obsahuje také katalytickou doménu (Manning et al. 2002). Vazba ligandu (růstové faktory, cytokiny) vyvolá oligomeraci RTK, dojde k aktivaci katalytické cytoplazmatické domény transfosforylací Tyr v aktivační smyčce nebo juxtamembránové doméně, což má za následek tyrosinovou fosforylací rekrutovaných cytoplazmatických proteinů, extracelulární signál je tak přenesen přes plasmatickou membránu.

Nereceptorové tyrosin kinázy (nRTK) nemají transmembránovou doménu, většinou jsou to rozpustné intracelulární proteiny, některé však mohou asociovat s plazmatickou membránou prostřednictvím PTM, jako je N-terminální myristoylace a zastávat funkci katalytických podjednotek receptorů postrádajících vlastní katalytickou doménu. K aktivaci nRTK dochází vazbou ligandu iniciovanou oligomerací nebo konformační reorganizací receptoru, na který jsou vázány. Integrované systémy signálních drah, počínající aktivací RTK fosforylací pozic se specifickými sekvencemi, se signálem dále neseným proudem různých SH2/PTB proteinů, zahrnují cytoplazmatické i nukleární reakce. Patří sem MAP kinázové kaskády, Ras/ERK, PI-3 kinázy aj. Fosforylace tyrosinu je zapojena do procesů buněčné adheze pomocí integrinové signalizace, metabolické regulace inzulinovými RTK, kontroly buněčného cyklu prostřednictvím inhibice tyrosinové fosforylace CDK, transkripční aktivace tyrosinovou fosforylací latentních cytoplazmatických transkripčních faktorů rodiny STAT, nervového přenosu, stárnutí přes IGF-1R (Hunter 2009).

Negativní regulace většiny tyrosinkináz je velmi přesná a účinná, důsledkem čehož jsou aktivní jen za velmi konkrétních podmínek. Fosfotyrosin mění strukturu proteinů mnohem výrazněji než fosfotreonin s fosfoserinem. Tyto poznatky pomohly vysvětlit, proč je zastoupení fosfotyrosinu vedle ostatních fosfoAMK přes úctyhodné množství nově popisovaných

tyrosinkináz tak nízké (<1%). K defosforylaci dojde velmi rychle, jakmile fosfotyrosin není stabilizován navázáním k SH2 doméně spojující kinázu se substrátem.

Regulace, založená na tyrosinové fosforylaci je typická pro kmen Metazoa (mnohobuněční heterotrofní živočichové). V říši hub (Fungi) se u hlenek (*Dictyostelium*) vyskytuje protein obsahující SH2 doménu (viz níže) i fosfotyrosin-dependentní přenos signálu, kvasinky (*Saccharomyces*) mají v MAP kinázové dráze kinázy s dvojnou aktivitou, src-like (nereceptorové) ani receptorové kinázy se však u nich, stejně jako u rostlin (říše *Plantae*), dle dosavadních poznatků nevyskytují (Darnell 1997). Trubénka *Monosiga brevicollis* má vyšší množství tyrosinkináz a málo kódovaného tyrosinu. Trubénky (Choanoflagellata) jsou považovány za vývojovou větev, která se odloučila od kmene Metazoa (Tan et al. 2009). Množství kódovaného tyrosinu a tedy i fosfotyrosinu vykazuje u kmene Metazoa negativní korelaci se vzrůstající komplexitou organismu (měřeno počtem buněčných typů), což je vysvětlováno zdokonalením systému signálních kaskád postupnou evoluční eliminací fosforylačních míst, která mohou způsobovat aberantní a patologická chování buněk. Co nejjednodušší a nejpřesnější signalizační systém organismům poskytuje výhodu z hlediska fitness. Například mnoho lidských proteinů neobsahuje žádný fosfotyrosin.

#### **2.1.4. Fosforylace v plicních nádorech**

Výchozími literárními zdroji pro mou práci byly následující publikace výzkumných projektů, zabývajících se aberantní fosfotyrosinovou signalizací v nádorových buňkách. Z nich jsem získal seznamy fosforylačních míst rakovinných nádorů.

První projekt (Rikova et al. 2007) se zabývá identifikací onkogenních kináz v plicních nádorech globální analýzou fosfotyrosinové signalizace. Tato studie představuje první rozsáhlý rozbor aberantní činnosti tyrosin kináz v plicních nádorech. Tumory jsou komplexní tkáně, obsahující 50-90% rakovinných buněk. Do projektu vybrané nádory z pacientů obsahují vždy více než 50% rakovinných buněk a kontrola získaných výsledků probíhala zejména prostřednictvím srovnávání s čistými NSCLC buněčnými liniemi (obsahují 100% rakovinných buněk). Užití metody zahrnují hlavně imunoafinitní purifikaci vzorků pomocí fosfotyrosin-specifických protilátek za účelem zvýšení koncentrace fosfotyrosinu a následnou analýzu tandemovou hmotnostní spektrometrií. Zesílená exprese RTK a s ní související zvýšený výskyt fosfotyrosinu, zjištěný u 30 % tumorů výraznou měrou podporuje tvrzení, že

aktivita RTK hraje roli ve vývoji NSCLC. V získaných profilech fosfotyrosinové signalizace byly rozpoznány známé, v tumorech i buněčných kulturách nadexprimované, fosforylační dočinnosti uvedené (EGFR, c-Met) onkogenní RTK kinázy i některé další, jejichž souvislost s vývojem NSCLC dosud nebyla zjištěna či ověřena. Biochemická, funkční a klinická analýza umožnila rozlišit nové podtypy NSCLC, exprimující aktivní formy RTK, jako ALK a ROS (fúzní proteiny) nebo PDGFR $\alpha$ , DDR1. Zjištění, že zhruba v polovině buněčných linií a tumorů se zvýšenou hladinou EGFR kinázy nese tento enzym aktivační mutace, vedlo k alternativě klasifikace typů rakoviny na základě kinázové saturace, snaze o selekci nádorů exprimujících kinázy aktivované nad přirozenou míru. Je známo, že naprostá většina fosforylačních pozic na těle tyrosin kinázy má pozitivní vztah k její aktivitě. Některé RTK (Met, ALK, DDR1, ROS, VEGFR-2, IGF1R, PDGFR $\alpha$ , EGFR, Axl) a nRTK (FAK, LYN, FYN, HCK, FRK, BRK) jsou v NSCLC velmi silně fosforylované. Role těchto enzymů v patogenezi NSCLC není dostatečně objasněna, nedávno prokázané mutace aktivačních smyček mnohých z nich však evokují myšlenku, že mají v nádorovém onemocnění úlohu základních řídicích faktorů, jejichž identifikace je nezbytná pro zefektivnění léčby rakoviny kinázovými inhibitory. Důležitým poznatkem je fakt, že rakovinné buňky exprimují mnoho různých aktivovaných kináz interagujících v propojených sítích odlišných hierarchických úrovní nebo nezávisle na sobě v paralelních separátních drahách. Celkem 4551 zaznamenaných tyrosinových fosforylačních míst, na množství přesahujícím 2700 různých proteinů, obsahuje dle databáze PhosphoSite více než 85 % nově objevených, s funkcí zatím nedostatečně známou. Seznam vybraných 674 fosfotyrosinových pozic je součástí publikace (Rikova et al. 2007).

Cílem další studie bylo stanovení predikčních fosforylačních profilů pro rakovinu plic. V návaznosti na publikovaný projekt Rikova et al. 2007, bylo provedeno srovnání fosforylačních dat z normální a nádorové plicní tkáně. Odhalilo stovky míst, která jsou v tumorech fosforylována odlišně. Hlavním výsledkem výzkumu tedy je, že metoda rozlišení zdravé tkáně od rakovinného nádoru pomocí analýzy fosforylace tyrosinových zbytků je velmi citlivá, účinná dokonce též za použití relativně malého souboru fosforylovaných peptidů. Jedná se o způsob vysoce přesné klasifikace plicních nádorů na základě konkrétních fosforylačních profilů nebo identifikace disproporčně aktivovaných signálních drah (dle globálního stavu jejich fosforylace) a to vše prostřednictvím statistického in silico rozboru hmotnostních spekter, umožňujícího v nádorových proteinech determinovat pozice hyper- i hypofosforylované. Hlavní datový soubor obsahuje 151 NSCLC a 48 vzorků zdravé tkáně. Ze 151 NSCLC bylo 52 spinocelulárních a 42 adenokarcinomů s dostupnými klinickými

informacemi použito v klasifikační analýze. Mezi těmito 94 vzorky NSCLC bylo identifikováno oproti zdravé tkáni silně rozdílně fosforylovaných 129 AMK, z toho 77 hyper- a 52 hypofosforylací. Ze 4551 fosforylačních pozic byly do projektu vybrány pouze ty, které se vyskytovaly minimálně v 10% posuzovaných vzorků. K nezávislému ověření získaných výsledků sloužil kontrolní datový soubor, sestávající z 16 NSCLC. Například, množství fosfotyrosinu detekované ve dvou proteinových souborech z EGFR dráhy (základní skupina 11 a rozšířená skupina 47 proteinů), stačilo k rozpoznání zdravé a nádorové tkáňe s velmi vysokou, u široké 47- členné skupiny však výrazně vyšší přesností. EGFR kináza je receptor tyrosin kináza, podílející se na mnoha biologických procesech včetně apoptózy, buněčné adheze a růstu. Cytoplazmatická část EGFR obsahuje tyrosiny Y1172 a Y1197, o nichž se ví, že jejich fosforylací je regulována proliferační aktivita. Oba jsou v NSCLC mnohem více fosforylovány. Y1172 je hyperfosforylován jen v adenokarcinomech (AD), Y1197 v AD i spinocelulárních karcinomech (SCC), ale v AD o dost silněji. V celkovém množství získaných dat se jinak téměř nepodařilo najít tyrosiny a signální dráhy, které by byly dostatečně relevantně odlišně fosforylovány u AD oproti SCC, což může být zapříčiněno nedostatečnou citlivostí užitých metod pro drobné rozdíly. Největší konzistenci rozdílné četnosti fosforylací mezi tumorem a zdravou tkání vykazuje ADH1B (alkoholdehydrogenáza 1B), účastníci se glykolýzy, glukoneogeneze i metabolismu aminokyseli, jejíž tyrosin Y34 je v rakovinných buňkách fosforylován méně. ERK1/2 se mohou podílet na zastavení buněčného cyklu, jejich nízká hladina v tumorech by proto mohla naopak urychlovat jeho průběh. Tyrosiny Y186 a Y204 na ERK1, ERK2 jsou v NSCLC fosforylovány méně.

V rozeznání zdravé tkáňe od nádorové je klasifikace na základě fosfotyrosinových profilů stejně účinná jako srovnání hladin mRNA transkriptů, při rozdělení tumorů na typy je fosforylační hledisko účinné, ale mRNA klasifikace výrazně přesnější a v odlišení ranných stádií vývoje nádoru od pozdních jsou obě metody neúčinné. Mapování stavu fosforylace vybraných aminokyselinových zbytků vypovídá o funkční aktivitě asociovaných genových produktů více než mRNA transkripční a proteinové expresní hladiny. To je důležité v případě tumorů, kde přítomnost nebo absence různých receptorových proteinů nemusí reflektovat aktivitu následujících členů signálních kaskád (Chang-Jiun Wu et al. 2009).

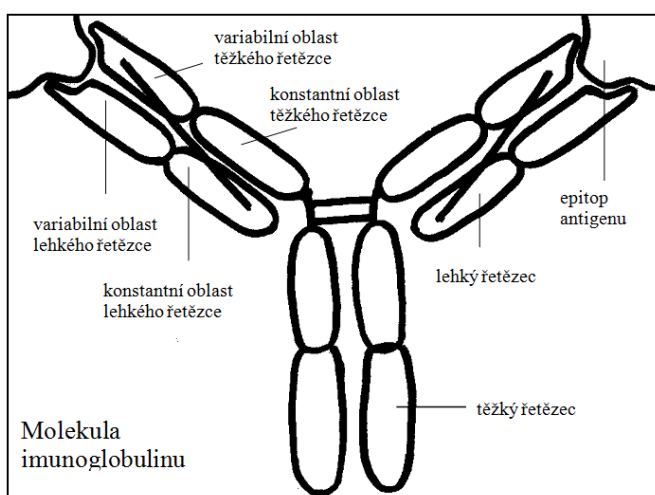


## 2.1.5. Detekce fosforylace

Postranslační modifikace lze experimentálně určovat. V současné době se používají především dva základní způsoby detekce fosforylovaných aminokyselin, které jsou při výzkumu fosforylace proteinů užívány.

### 2.1.5.1. Imunodetekce

Opsonizace (viz obr. č.5) protilátkami (např. radioznačené  $^{32}\text{P}$  ortofosfátem) zaměřenými proti konkrétním fosfoepitopům aminokyselin je základním nástrojem pro studium místně specifické fosforylace proteinů. Byly vyvinuty pro detekci fosforylace jednotlivých aminokyselin (fosfoserin, fosfotreonin, fosfotyrosin), jsou využívány při Western blot, imunoprecipitaci (IP), imunohistochemii (IHC), ELISA, průtokové cytometrii a v poslední době i při imobilizaci na poli s pevným nosičem. Nejefektivnější však jsou imunoglobuliny se specifitou pro fosfotyrosin. Fosfoserin i fosfotreonin jsou menších rozměrů než fosfotyrosin a nebyly by identifikovány, je proto užíváno protilátek proti motivu nebo většímu epitopu (např. u MAP kináz a cyklin dependentních kináz charakteristický fosfotreonin s prolinem na +1 místě). Přestože jsou dnes fosfospecifické protilátky velmi populární, jen málo z nich oplývá vysokou specifitou pro cílové epitopy a imunodetekce fosforylací je časově i finančně náročná (Kovarova et al. 2003).

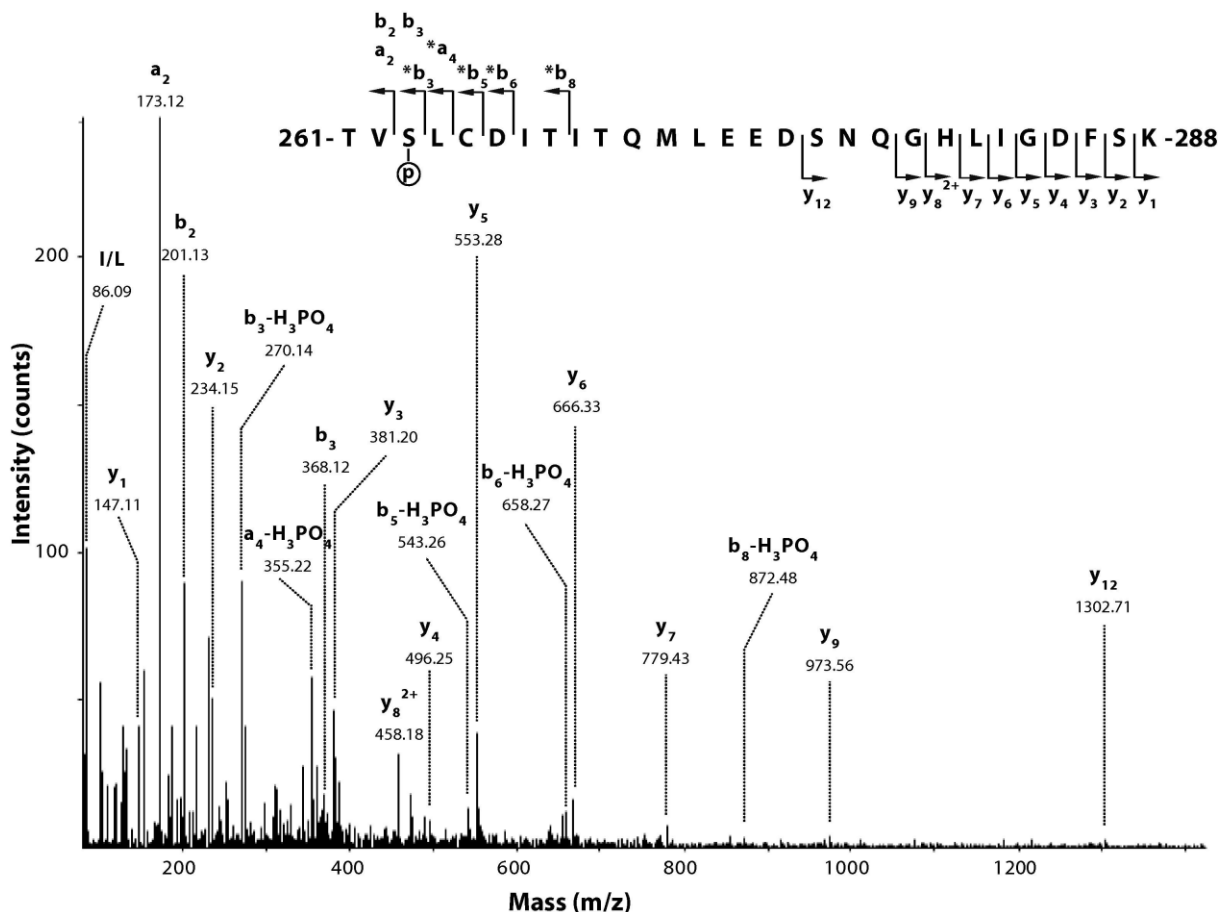


**Obr. č.5:** Molekula imunoglobulinu – těžké a lehké řetězce s variabilními i konstantními oblastmi, vazba k epitopu (Zídková 2003)

### 2.1.5.2. Hmotnostní spektrometrie

Podstatou hmotnostní spektrometrie je fragmentace separovaných fosfoproteinových vzorků na ionty, ty jsou uspořádány dle hmotnosti, identifikovány. Pro analýzu fosforylací jsou užívány hmotnostní spektrometrie druhého nebo třetího stupně (Zhou et al. 2001). Tandemová hmotnostní spektrometrie spojená s kapalinovou chromatografií (lc-ms/ms) v hmotnostním analyzátoru selektuje dané ionty, které jsou znovu štěpeny, potom vstupují do druhého hmotnostního analyzátoru, následuje detekce fosforylovaných aminokyselinových zbytků a korelace se sekvenčními daty (Witze et al. 2007). Výsledná spektra (viz obr. č.6) reprezentují absolutní či relativní saturaci iontů s konkrétním poměrem hmotnosti a náboje ( $m/Q$ ). Občas však mohou být získané informace o zastoupení proteinů zkreslené, omezená je také možnost analýzy nízkých množství proteinů, není důrazná selekce specifických fosforylačních míst, přičemž specifická kinázová místa je náročné identifikovat experimentálně (Gao et al. 2010).

Současné metody hmotnostní spektrometrie využívají značení aminokyselin v buněčné kultuře stabilním izotopem (SILAC) nebo značení peptidů in vitro s tandemovou hmotnostní spektrometrií, následuje relativní stanovení změn fosforylace. Některé kvantifikační postupy zahrnují také peptidové standardy s těžkými isotopy a umožňují pochopit globální změny fosfoproteomu v reakci na různé podněty nebo chorobné stavy (Ham 2012).



**Obr. č.6:** Spektrum z detekce fosfoserinu 263 pomocí řad iontů, tandemová hmotnostní spektrometrie (Esmenjaud-Mailhat et al. 2007)

## 2.2. Nádorová onemocnění

Podstatou nádorových onemocnění je nekontrolované buněčné dělení, k němuž může dojít u jakéhokoli buněčného typu. Vznik rakoviny může být podmíněn genetickými dispozicemi, dále faktory chemickými (karcinogeny), fyzikálními (gama záření) a biologickými (viry, poruchy hormonální regulace). Důležitou vlastností nádorových buněk je, s výjimkou poruchy regulace dělení, také necitlivost vůči apoptotickému signálu, který může přicházet od okolní tkáně, buněk imunitního systému nebo například v důsledku závažného poškození DNA často z buňky samotné. Ve zdravé buňce jsou signály apoptotické v rovnováze s antagonistickými. U defektních buněk dojde k porušení této stability, apoptotický signál je umlčen či oslaben, což je potenciálem pro možný rozvoj nádoru, metastáz atd. Někdy však může během vícestupňové karcinogeneze docházet i k posílení apoptózy (Lowe & Lin 2000).

Nepočítaje nádory kůže, ročně vzroste počet pacientů s různými druhy rakoviny o cca 12 milionů (Jemal et al. 2011). Nádorová onemocnění zahrnují širokou škálu typů s odlišným zdravotním rizikem, od snadno léčitelných nezhoubných kožních nádorů až po silně nebezpečné, například kolorektální nádory. Jako se rakovinné buňky původem z epitelu svými charakteristikami většinou velmi liší od buňek kostních nádorů, stejně vzdáleně se často mohou jevit i nádory vzniklé u různých pacientů ze stejné tkáně. Dle schopností invazivního růstu i tvorby metastáz lze rozeznávat nádory zhoubné (maligní, silně invazivní, metastáze) a nezhoubné (benigní, pomalý růst nebo stagnace, nemetastázuji). Různé typy rakoviny lze také dělit podle míry podobnosti s mateřskou tkání či podle toho, zda se jedná o primární nádor nebo metastázu (Lowe & Lin 2000).

Nádorová onemocnění jsou jednou z nejčastějších příčin úmrtí na celém světě. Ve své práci se zabývám plicním a kolorektálním karcinomem. Oba typy nádorů patří u nás i ve světě mezi nejfrekventovanější zhoubná onemocnění.

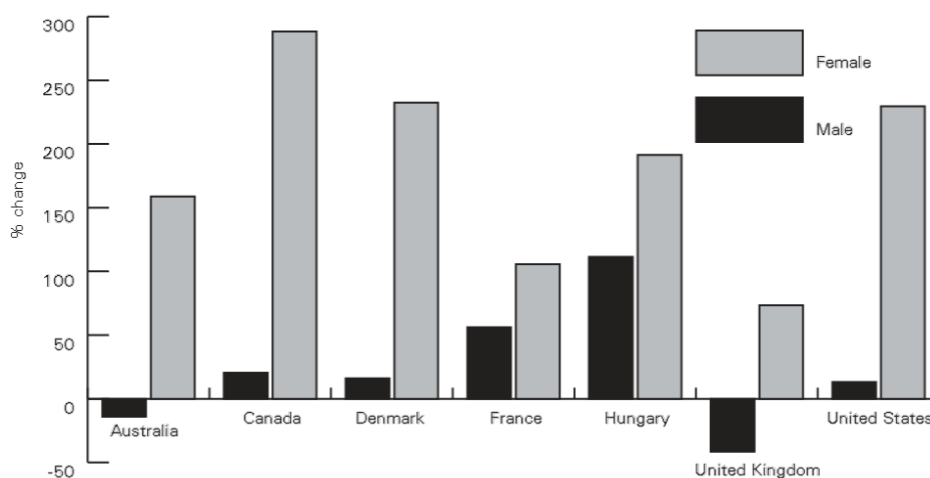
### **2.2.1. Rakovina plic**

Průměrně 1,35 milionů nových pacientů a 1,18 milionů mrtvých (viz obr. č. 7) každý rok s těžištěm výskytu v Evropě a Severní Americe dělá z rakoviny plic nejběžnější a nejzhoubnější nádorové onemocnění. V severní Evropě a USA je nejvyšší úmrtnost na karcinom plic u žen, ve východní Evropě u mužů (*Gender in lung cancer and smoking research*. World Health Organization, 2007). Naprostá většina případů karcinomu plic je epitelální povahy. Majoritní příčinou kolem 90 % výskytu plicních nádorů v rozvinutých zemích je kouření tabákových výrobků (Peto et al. 2006), nezanedbatelné riziko rovněž představuje kouření pasivní, dále se zde uplatňuje kontaminace ovzduší radonem, azbestem, exhalacemi a určitý podíl nesou i virové infekce.

V klinické praxi nejužívanější je kategorizace plicních tumorů na malobuněčný (Small cell lung cancer – SCLC) a nemalobuněčný (non-Small cell lung cancer – NSCLC) karcinom plic. Malobuněčný karcinom tvoří asi 20% plicních nádorů. Rychle metastázuje nejčastěji do uzlin, mozku, kostní dřeně, jater. Nemalobuněčný karcinom je typický pomalejším růstem a menší pravděpodobností vzniku metastáz. Dělí se na tři hlavní podtypy: adenokarcinom, spinocelulární karcinom, velkobuněčný plicní karcinom. Adenokarcinom zaujímá přibližně

40% plicních nádorů (Lu et al. 2010). Při pozdním diagnostikování je průměrná doba přežití pacientů 20 měsíců (*Lung Carcinoma: Tumors of the Lungs* Merck Manual Professional Edition 2007).

Mutace protoonkogenu K-ras jsou zodpovědné za až 30% adenokarcinomů (Aviel-Ronen et al. 2006) V asi 4% NSCLC se projevuje EML4-ALK tyrosinkinázový fúzní gen (Kumar et al. 2013). Receptor epidermálního růstového faktoru (EGFR) má regulační vliv na proliferaci, apoptózu, angiogenezi. Mutace a amplifikace EGFR jsou běžné v NSCLC. Dalším mutovaným či amplifikovaným genem plicních nádorů je například c-Met (Herbst et al. 2008).



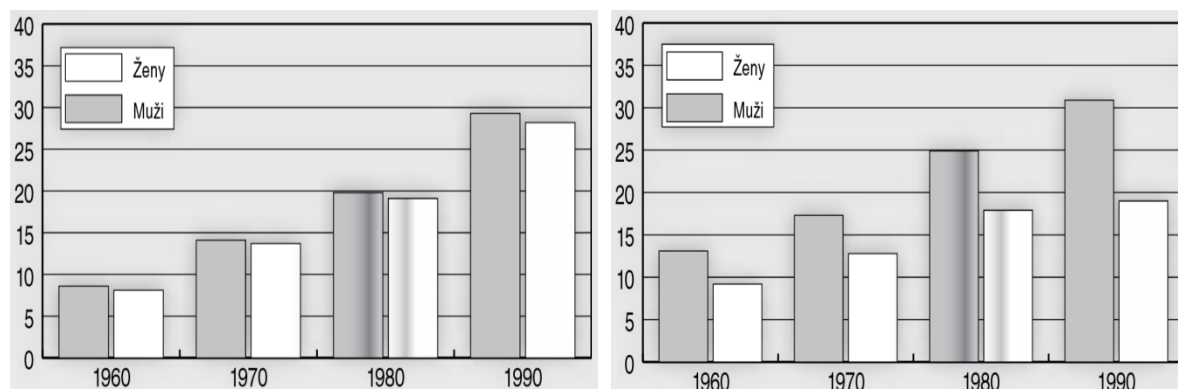
Obr. č. 7: Procentuální změna úmrtnosti na rakovinu plic ve vybraných zemích v průběhu 30 let (1968-1998), (GLOBOCAN 2000. Cancer incidence, mortality and prevalence worldwide.)

### 2.2.2. Kolorektální karcinom

Jedním z nejběžněji se vyskytujících druhů rakoviny je také kolorektální karcinom, méně přesně nazývaný také rakovina tlustého střeva, působící smrt kolem 655 000 lidí ročně (<http://info.cancerresearchuk.org/cancerstats/types/bowel/incidence/>), přičemž mezi nejrizikovější země patří Česká republika, dále Slovensko a Maďarsko. Nejvíce postiženi jsou muži (60,3 oproti 31,5 případů/100000 obyvatel), hlavně ve věku nad 50 let viz obr. č. 8 (Musil 2003). Za příčinu tohoto onemocnění je považována nadměrná konzumace masa, smažené potraviny, tuků, uzenin, s tím spojený vysoký příjem v lipidech rozpustných a v tukové tkáni ukládaných škodlivin i fyziologicky nevhodných variant mastných kyselin. Nezanedbatelný je také vliv pití alkoholu, kouření, nedostatku vlákniny v potravě (Pecorino 2006). Kolorektální karcinom se většinou vyvíjí pozvolna z benigního polypu střevní sliznice, jehož výskyt může někdy být i geneticky podmíněn (Gardnerův syndrom, familiární adenomatózní polypóza aj. mají poměrně vysoké

riziko vzniku kolorektálního karcinomu).

Více než 95% karcinomů jsou adenokarcinomy. Nejčastěji metastázuji do jater, následně do plic. Další typy kolorektálních karcinomů se vyskytují zřídka. Patří mezi ně např. lymfomy, sarkomy, spinocelulární karcinomy. Na molekulární úrovni lze kolorektální karcinomy dělit do řady podskupin. V praxi nejužívanější je dělení na K-ras dependentní a K-ras independentní nádory, dle přítomnosti mutace K-ras. Tato mutace není přítomna asi u 60 % pacientů. U těch je předpoklad pro kladnou reakci na imunoterapii proti EGFR (Vogelstein & Kinzler 2004). Pokud je kolorektální karcinom diagnostikován včas, prognóza přežití je 80 %. Při diagnostikování pokročilých stádií je průměrná doba přežití pacientů 11 měsíců (Musil 2003).



Obr. č. 8: Incidence kolorektálního karcinomu v ČR na 100 000 obyvatel/rok (Musil 2003)

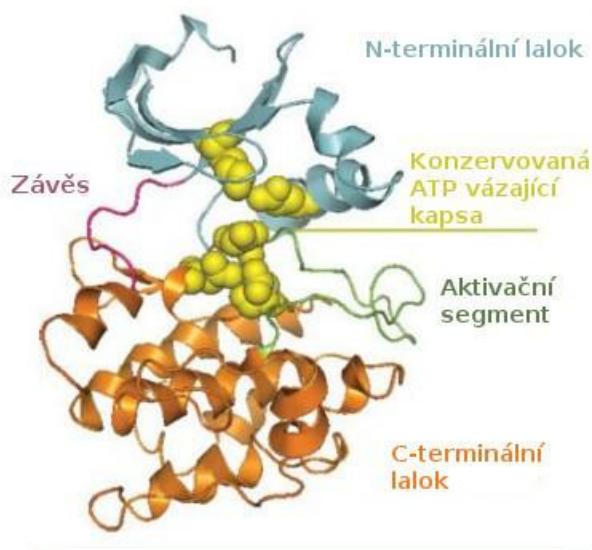
### 2.2.3. Nádorová onemocnění a fosforylace

V buňkách plicních nádorů je míra fosforylace proteinů vysoce zvýšená, mimo jiné na EGFR (receptor epidermálního růstového faktoru) (Wu et al. 2009), následkem čehož dochází také k abnormálnímu nárůstu fosforylace tyrosinu (Guha et al. 2008). Nadbytečné množství na tyrosinu fosforylovaných proteinů se projevuje tvarovými odlišnostmi a nadměrnou proliferací rakovinných buněk, atypickou morfologií takové tkáně jako celku. S rostoucí hladinou fosfotyrosinových bílkovin klesá pravděpodobnost přežití pacienta, nevyjímaje ani případy bez jiných detekovatelných chorobných vlastností nádorové tkáně (Gong et al. 2002). Pro bujení a invazi karcinomu je nepostradatelná nadexprese či hyperaktivita src-like kináz i uvedení EGFR do činnosti (Wu et al. 2010). Nezávisle na množství zkoumaných fosforylačních pozic, lze s přesností odpovídající microarray analýze spolehlivě rozlišit rakovinné buňky od normálních, čistě na základě odlišností ve fosforylaci tyrosinů (Wu et al. 2009).

Některé virové onkoproteiny jsou rovněž konstitutivně aktivní tyrosinkinázy, například adenovirový analog MEK1 (Hunter 2007).

Prekurzory buněk pojivové tkáně gastrointestinálního traktu obsahují c-kit, s tyrosinkinázovou aktivitou spřažený, protoonkogenem kódovaný receptor růstového faktoru. Mutací tohoto protoonkogenu zapříčiněná konstitutivní aktivita tyrosinkinázy je základem vzniku gastrointestinálních stromálních tumorů. Dochází k neregulované fosforylaci c-kit, porušení některých signálních drah a hlavně extrémní tyrosinkinázovou aktivitou podmíněné proliferaci buněk (Joensuu et al. 2001).

Příčinou aberantní tyrosinové signalizace v nádorových buňkách nejčastěji bývají mutace v DNA. U nebezpečných řídicích mutací DNA (drivers), determinovaných místy silného zastoupení nesynonymních změn oproti synonymním, je na rozdíl od cestujících mutací (passenger) velká pravděpodobnost, že iniciují mechanismy, v jejichž důsledku se buňka začne chovat jako buňka rakovinná (Torkamani et al. 2009). Největší množství řídicích mutací je lokalizováno poblíž hydrofilních AMK při povrchu proteinu a v mnoha případech určitým způsobem ovlivňují fosforylaci nezávisle na tom, zda se nacházejí v kinázové doméně či substrátu. (Talavera et al. 2010). Mutace substrátových bílkovin postihují fosforylovatelné AMK či AMK poblíž tyrosinu, serinu nebo treoninu, čímž je často znemožněna identifikace fosforylačního místa. V kinázové doméně jsou řídicí mutace ve srovnání s cestujícími blíže k místu vázajícímu ATP, viz obr. č. 9 (Izarzugaza et al. 2009).



**Obr. č.9:** Kináza, odvozeno od MAP3K1; 2 laloky propojené závěsem, mezi laloky se naváže ATP, řídicí mutace často ovlivňují peptidovou sekvenci poblíž tohoto místa (Izarzugaza et al. 2009)

Určité fosforylačně konkrétní SNV (jednonukleotidová mutace) proteinu TP53 (tumorsupresor) však mohou dokonce prodloužit dobu života pacientů s glioblastomem nebo zvýšit pravděpodobnost přežití žen s rakovinou vaječníků.

Rozborem 793 odběrů genetického materiálu z tumorů osmi druhů rakoviny bylo identifikováno 775 genů obsahujících SNV, která se nachází v kinázové doméně nebo fosforylačním místě (7 AMK zbytků před i za fosfoAMK) a ovlivňuje tedy fosforylaci. Celkem 949 v daném genovém souboru určených fosforylačně konkrétních SNV se vyskytuje minimálně v jednom druhu nádorového onemocnění. Pro výzkum a objasňování rolí řídicích mutací DNA buněk rakovinných tumorů je přínosné zjištění, že zhruba třetina genů majících SNV ve fosforylační pozici se neobjevuje jen u jednoho druhu rakoviny (Reimand & Bader 2013).

Vývoj kinázových inhibitorů za účelem terapeutického tlumení zvýšené nebo disproporčně načasované fosforylační aktivity v rakovinných buňkách je znesnadňován redundantní funkcí některých kináz i novými mutacemi v katalytické doméně vzniklou inhibiční rezistencí (Hunter 2009).

### **2.3.Databáze**

V návaznosti na expanzi proteomiky a jiných s datovými soubory operujících odvětví výzkumu, je zde velká nutnost co nejpřehlednější systematizace shromažďování a třídění získaných dat. Kromě této funkce jsou však databáze rovněž prostředím umožňujícím aplikovat další výzkumné metody vedoucí k novým poznatkům. Některé proteinové databáze selektují informace o proteinech obecně (UniProt/TrEMBL, UniProt/Swiss-Prot), ostatní jsou více specializované. Databáze enzymů obsahují mimo jiné data týkající se kináz (Kinbase). Mezi databáze zabývající se chorobami souvisejícími s různými patologiemi proteomu patří OMIM, orientovaná na chyby v posttranslačních modifikacích ve vztahu ke konkrétním onemocněním.

V důsledku časové a finanční náročnosti praktického určování posttranslačních úprav je kladen důraz na pečlivé uchování dat, k čemuž je využíváno několik databází posttranslačních modifikací, které jsou obecné nebo cílené na konkrétní skupinu organismů (např. PTM virů). Data v nich uložená slouží ke srovnávání a posuzování výsledků nových postupů, konstrukci predikčních modelů (Gnad et al. 2011) atp.



### 2.3.1. Databáze fosforylací

Databáze postranlačních modifikací jsou dosud roztržštěné, zatím si mezi nimi žádná nedobyla podobného statutu standardu jako Uniprot mezi proteinovými databázemi. Stručně tedy zmíním hlavní databáze zaměřené na uchování dat o fosforylacích.

Databáze Phosida (Gnad et al. 2011) obsahuje data týkající se nejen fosforylace, rovněž však také N – glykosylace a acetylce. Hlavními objekty zájmu jsou člověk, myš, octomilka, ale i mnoho dalších modelových organismů, včetně prokaryot. Databáze HPRD (Human Protein Reference Database) nabízí kategorizaci lidských proteinů z hlediska lokalizace, funkce, motivu, domény, PTM (Prasad et al. 2009). Databáze Phospho.ELM (Dinkel et al. 2011) shromažďuje literární a experimentální data o fosforylaci serinů, treoninů, tyrosinů lidských proteinů a proteinů modelových organismů (octomilka).

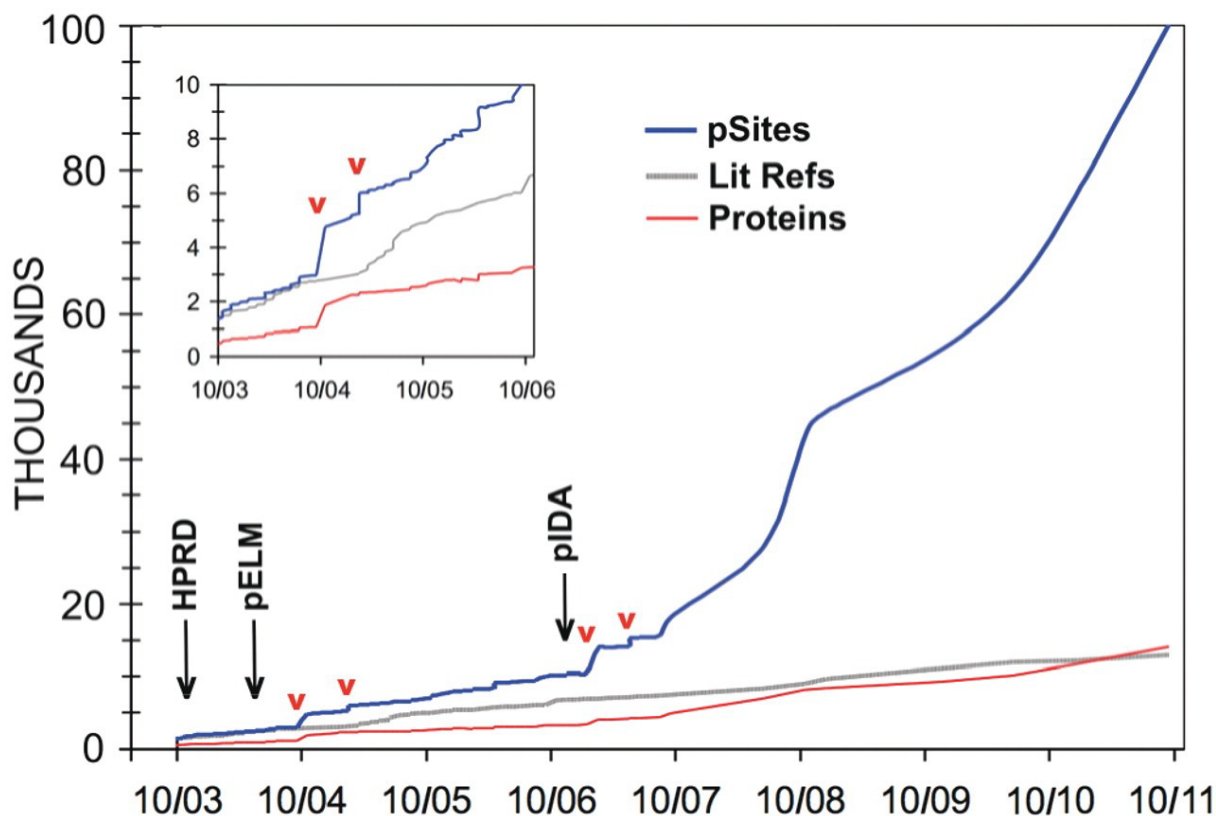
PTM databáze zaměřené na fosforylaci proteinů (viz tabulka č.2) jsou neúplné, liší se velikostí i množstvím proteinů a identifikovaných fosforylačních míst, což je dáno užitím rozdílných metod rozpoznávání fosforylovaných pozic i jiných modelových organismů (Gnad et al. 2011).

Databáze	Organismy	Web	
PhosphoSitePlus	<i>H. sapiens</i> , <i>M. musculus</i> , <i>D. melanogaster</i>	<a href="http://www.phosphosite.org">http://www.phosphosite.org</a>	Hornbeck et al. 2012
PHOSIDA	<i>H. sapiens</i> , <i>M. musculus</i> , <i>D. melanogaster</i>	<a href="http://www.phosida.com/">http://www.phosida.com/</a>	Gnad et al., 2011
Phospho.ELM	<i>H. sapiens</i> , <i>M. musculus</i> , <i>D. melanogaster</i>	<a href="http://phospho.elm.eu.org">http://phospho.elm.eu.org</a>	Diella et al., 2008
HPDR	<i>H. sapiens</i>	<a href="http://www.hprd.org">http://www.hprd.org</a>	Prasad et al., 2009

Tabulka č.2: Stručný popis vybraných PTM databází

Nejobsahlejší databází fosforylací je však PhosphoSitePlus. Databáze PhosphoSitePlus vznikla ze starší Phosphosite, zaměřené na fosforylace, včetně rámcového popisu příslušných

kináz i fosfatáz lidských a myších bílkovin. PhosphoSitePlus obsahuje data o PTM, hlavně fosforylaci, acetylaci, ubiquitinaci pro více než 19 tisíc zejména lidských a myších, ale částečně i jiných proteinů, s celkovým počtem fosforylačních míst přesahujícím 200 tisíc. Zahrnuje dva typy dat, která se liší věrohodností. Data kontrolovaná a publikovaná z detailně orientovaných výzkumných projektů i výsledky, na nichž se podílela tandemová hmotnostní spektrometrie. Databáze PhosphoSitePlus neustále narůstá, její rozšiřování do roku 2011 ukazuje obr. č. 10. Informace jsou manuálně selektovány dle více kritérií, lze přehledněji vyhledávat v závislosti na konkrétní buněčné linii, fosforylované pozici, druhu onemocnění (Hornbeck et al. 2012).



Obr. č. 10: Rostoucí rozsah Phosphosite (2003-2011), červená linie = počet proteinů, modrá linie = počet fosforylačních míst, šedá linie = reference, červené šipky = publikace výsledků MS, černé šipky = vznik jiných databází (PHOSIDA, Phospho.ELM, Human Protein Reference Database) (Hornbeck et al. 2012).

### 3. Cíle práce

Obrovský rozvoj high-throughput přístupů v nedávné době otvírá nové možnosti při výzkumu nádorových onemocnění. Jednou z těchto možností je i průzkum nashromážděných údajů o fosforylacích v nádorových liniích.

Cílem mé práce je charakterizace fosfotyrosinových profilů dvou významných typů nádorů – nádorů plicních, které jsou celosvětově jedním z nejnebezpečnějších a nejagresivnějších nádorů a nádorů kolorektálních, které mají vysokou četnost v České republice a které patří spíše mezi nádory pomalu se rozvíjející. Rozhodl jsem se provést sekvenční, doménovou, strukturní a expresní charakterizaci obou typů nádorů a jejich srovnání za použití dostupných bioinformatických nástrojů a databází.

Zabýval jsem se ověřením změny exprese fosforylovaných proteinů v rakovinných buňkách. Zajímalo mě, jak se liší fosforylační místa tumorů od zdravé tkáně na úrovni sekundární struktury peptidového řetězce a primární struktury ve smyslu možného vzniku nových fosforylačních míst, prostřednictvím změny aminokyselinové sekvence stávajících. Dalším důležitým bodem bylo také shromáždění dat týkajících se lokalizace fosforylačních míst ve známých proteinových doménách, respektive zjištění, které domény jsou v daných typech tumorů nejvíce fosforylovány i případná identifikace rozdílů mezi nimi. Poslední z hlavních úkolů představuje podpoření či vyvrácení domněnky, že potenciální patologické tyrosinové fosforylace mohou být v datasetech identifikovány na základě absence fosforylace u orthologů modelových organismů.

### 4. Metody

#### 4.1. PhosphoSite a aktualizace informací

Proces výběru a třídění dat nově zahrnovaných do databáze PhosphoSite se skládá ze tří důsledně kontrolovaných hlavních kroků obsahujících mnoho dílčích způsobů ověřování informací.

##### 1) Výběr vhodných záznamů

Informace jsou získávány z literatury a dalších zdrojů. K vyhledávání publikací dochází prostřednictvím poloautomatických algoritmů identifikujících data, jejichž vyhodnocením byl

stanoven předpoklad, že se potenciálně mohou týkat popisu fosforylačních míst lidských, myších i jiných proteinů. Výstupní informace jsou dále selektovány kvalifikovanými vědeckými pracovníky, kteří vyberou pouze seriózní zprávy týkající se nových fosforylačních míst nebo popisu jejich biologické funkce. Fosforylační místa objevená in vitro jsou považována za relevantní jen v případě, pokud je k dispozici důkaz o výskytu in vivo. Shromážděné záznamy postupují do organizační fáze.

## **2) Organizační fáze**

Je vybráno přístupové číslo ze Swissprot nebo NCBI, jež bude sloužit jako referenční kód nového proteinu ve PhosphoSite, přidána sekvence, alternativní názvy, některé odkazy na další zdroje informací. Někdy jsou podobná data shromážděna i pro orthology a isoformy. K zahrnutí popisu isoformem dojde pouze pokud v konkrétním článku hraje roli jejich specifické značení a nerespektování tohoto rozlišení by bylo pro uživatele PhosphoSite matoucí. Další přidávané záznamy obsahují metody užití k podpoře tvrzení, že je pozice fosforylována in vivo, popis domnělých příslušících kináz a fosfatáz i definici případné signální či konformační funkce dané fosforylace.

## **3) Editace**

Finální strukturovaná data jsou kontrolována z hlediska přesnosti i konzistence, editor ověřuje informace ze všech článků a zajistí konečné zařazení položek do systému.

## **4.2. Výchozí série dat**

V předkládané práci jsem pracoval se čtyřmi základními soubory dat:

Datový soubor LUNG obsahuje 674 tyrosinových fosforylačních míst determinovaných UniProt kódem daného proteinu, genovým symbolem, aminokyselinovou sekvencí části blízkého okolí příslušného tyrosinu a jeho sekvenčním číslem. Jedná se o fosforylovatelné pozice proteinů plicních nádorů, jež jsou součástí výše zmíněné publikace Rikova et al. 2007, jako vybraná skupina z celkového počtu fosfotyrosinů detekovaných při tomto projektu. Pět fosforylačních míst muselo být vyřazeno pro nedostatek dohledatelných informací z různých důvodů. LUNG je rozdělen na dvě části, LUNG/all a LUNG/only (580 fosforylačních míst, která se nevyskytují současně v následujícím souboru HCT116).

Datový soubor HCT116 zahrnuje 172 tyrosinových fosforylovatelných pozic proteinů kolorektálních nádorů, determinovaných UniProt kódem daného polypeptidu, genovým symbolem, aminokyselinovou sekvencí fosforylačního místa – od sedmé aminokyseliny před fosfotyrosinem po sedmou aminokyselinu za fosfotyrosinem, sekvenčním číslem fosfotyrosinu. HCT116 pochází z databáze PhosphoSitePlus,. Byla selektována pouze jedna buněčná linie (hct116) s nejvíce daty, aby nebyl počet fosforylací oproti LUNG/all příliš vysoký. Také jsem nechtěl míchat různá stádia nemoci (před a po léčení). Fosforylační místa byla tedy vybrána z buněčné linie hct116 pod kritériem absence léčebných postupů, což však spolu s ignorací serinových i treoninových fosforylací nakonec skupinu zúžilo velmi výrazně. Ani po komunikaci se správci databáze ale nevíme, zda data byla někdy publikována v časopise. Sekce HCT116/all obsahuje všech 172 tyrosinfosforylačních míst, sekci HCT116/only tvoří pouze 79 míst, která se nevyskytují současně v LUNG.

Přílohové materiály již diskutované publikace Chang-Jiun Wu et al. 2009 poskytly možnost separace datasetu S1 čítajícího 161 fosfotyrosinových míst proteinů buněk plicních nádorů. Fosforylační místa jsou determinována genovými symboly a sekvenčními čísly fosfotyrosinů, na základě čehož jsem pro další potřeby dodatečně dohledal aminokyselinové sekvence okolí i UniProt kódy. Uvedené údaje však neumožnily 17 míst jednoznačně identifikovat, ta tedy musela být vyřazena a proto zbylo 144 fosfotyrosinových pozic. S1 obsahuje 33 specifických fosforylačních míst a 111 míst vyskytujících se současně v LUNG. UniProt kódy proteinů z datasetů LUNG/all, LUNG/only, HCT116/all, HCT116/only, S1 ukazuje tabulka č.3.

Datová série STY je tvořena patnácti úplnými seznamy všech fosforylovatelných pozic v databázi PhosphoSite Plus (tyrosinových, serinových i treoninových) - jedním kompletním soupisem pro celou databázi PhosphoSitePlus a čtrnácti menšími (rovněž z PhosphoSitePlus) pro vybrané typy nádorových onemocnění. Jedná se o rakovinu plic, kolorekta, prsu, jater, žaludku, jícnu, slinivky, prostaty, štítné žlázy, dále o chronickou i akutní lymfoblastickou a myeloidní leukémii, glioblastom. Každý fosfotyrosin je (kromě UniProt ID, sekvenčního čísla, okolní sekvence, genového symbolu) opatřen také výčtem literárních referencí, publikací výzkumných projektů, kterými byl v konkrétním typu nádoru identifikován. Sekce STY/all u každého nádorového onemocnění zahrnuje všechna fosforylační místa, STY/10% je omezena pouze na místa zmíněná alespoň u 10% z příslušného celkového seznamu datasetů pro daný typ nádoru.

<b>HCT116/only</b>
P00519 P42684 P54760 P08631 P08069 Q12866 P07947 Q06418 P12931 Q9NR20 P53779 P49841 Q9H2X6 Q00526 Q00526 Q9Y463 O00571 Q9H9T3 P54753 P54760 P62807 P62807 P58876 P58876 Q93079 Q93079 O60814 O60814 Q99880 Q99880 Q99879 Q99879 Q99877 Q99877 Q5QNW6 Q5QNW6 P57053 P57053 P34932 P08238 O95678 P15259 Q8N0Y7 Q6NZI2 Q6NZI2 P49023-2 Q71UM5 Q9NRC6 P02786 Q05639 Q05639 Q9H792 P07947 Q8IZP0 BAA76823 Q8TEW0 Q8IWW6 Q9H5V8 P46108 P46109 Q96PD2 Q96PD2 Q96RT1 Q96PD2 Q9C0H9 P09651 P22626 P51991 Q15464 Q96J84 Q9Y2X7 Q9NZM3 Q99569 Q15464 Q15464 Q15464 Q9Y446 Q99569 O43353
<b>HCT116/all</b>
P00519 P42684 P54760 P08631 P08069 Q12866 P07947 Q06418 P12931 P30530 P06213 P07948 P29317 Q05397 Q05397 P06241 P06239 P08581 Q9NR20 P53779 P49841 Q9H2X6 Q00526 Q00526 Q9Y463 Q00535 Q92630 Q9UPZ9 P45983 P45984 O15264 P53778 P24941 Q13627 P27361 Q9UPZ9 Q16539 P06493 P06493 P24941 P28482 P49840 Q86Z02 Q9H422 Q13523 Q13164 Q92796 NP_001734 NP_005175 NP_001734
NP_005175 O00571 Q9H9T3 P54753 P54760 P62807 P62807 P58876 P58876 Q93079 Q93079 O60814 O60814 Q99880 Q99880 Q99879 Q99879 Q99877 Q99877 Q5QNW6 Q5QNW6 P57053 P57053 P34932 P08238 P08238 O95678 P05787 P15259 Q8N0Y7 P18669 Q6NZI2 Q6NZI2 P49023-2 Q71UM5 P42677 P05388 Q9NRC6 P46783 P02786 Q05639 Q05639 P68104 P68104 Q9H792 P07947 P07948 P06241 Q06124 P23528
O00401 Q8TEW0 P05023 P05556 P16144 P07195 P18206 Q8IZP0 BAA76823 Q8TEW0 Q8IWW6 Q9H5V8 P46108 P46109 Q96PD2 Q96PD2 Q96RT1 Q96PD2 Q9C0H9 P09651 P22626 P51991 Q15464 Q96J84 P15529-4 Q13480 Q9NZM3 P08581 Q86YV5 Q13627 P00533 Q9Y624 O00459 Q9Y2X7 Q9NZM3 Q99569 Q15464 Q15464 Q15464 Q9Y446 Q99569 O43353 Q07912 O60716 O60716 P29317 P29317 Q05397 Q05397 P35568
Q9Y4H2 Q9Y4H2 Q96A22 P56945 P56945 Q99569 O95297 Q06124 P29597 Q8IZP0 P29317 P56945 Q05655 P18433 Q13671 P40763 P40763-2 Q9UDY2 P07355 Q9NRY4 P29353
<b>LUNG/only</b>
O43603 P08559 P08559 P13796 P60660 P06733 P06733 P08727 P05787 P08670 P08729 P02538 P08670 P08670 P08670 P07954 P04899 P08754 P22102 P50395 P20592 P69905 P69905 P02042 P04075 P68104 Q12959 P11413 P11413 P11413 Q9HC35 P18669 P18669 P18669 P62805 P62805 Q9UKY1 Q14247 Q9UJY1 P08107 P11142 P38646 P14625 Q9UIA9 P48739 P20963 P07766 P30273 P30273 P63252 P24071 P12314 P12314 O43914 P53990 P20936 P05129 Q9NZM1 Q9NZM1 Q9BSJ8 Q9BSJ8 P33151 P04040 P04040 Q99439
Q15417 Q01995 P37802 Q9NVF5 Q00610 Q00610 Q9NQR4 P35606 P17302 P61604 Q9Y6Y1 P12814 O43707 P13796 Q8IZD4 O14786 Q8N353 Q9ULV4 Q8NCT2 A1X283 P62333 Q99758 Q99758 P07108 P60709 P60709 P60709 P60709 P61158 P61158 P61158 Q8N142 Q7Z4W1 Q9BY49 P30038 P63010 O00519 O00519 P12532 Q96GR2 Q96GR2 Q96GR2 P17174 P00505 Q12955 Q9Y566 Q6PKA2 P04083 P50995 P50995 P07355 P07355 P07355 P07355 P07355 P08758 P08133 Q06481 Q06481 P05067 P05067 Q9H6X2 P26358 P63010 P17844
P20591 P20591 P59665 P36543 O14964 Q9Y490 Q16526 O00757 Q86UX7 O14964 O15145 P35568 Q9Y4H2 O15268 Q9UED8 Q9UED8 Q9NYT0 Q8N4X5 Q8N4X5 P21333 P21333 Q14315 P22105 Q00341 Q13753 Q13753 Q13753 P56597 O96000 O96000 Q9Y316 P12235 P05141 P05141 P35580 P35579 P31939 Q86VI4 Q15012 P49023 P49023 Q14152 P51512 P11216 Q08357 O00329 P19174 O43781 Q9Y4K4 Q15759 Q00536 Q05655 P00519-2 Q06187 P16591 P09769 P42685 P06241 P23458 P23458 P07948 P07948 Q14289 Q14289 P43405
Q9UM73 Q9UM73 Q9UM73 Q9UM73 P30530 Q08345 Q08345 Q08345 Q16832 Q16832 P00533 P00533 P29320 P11362 P17948 P17948 P35916 P35916 P06213

P06213 P29376 P29376 P29376 P08581 P08581 P16234 P16234 P16234 P47712 P15924 Q6S376 P21796 P07737 P35080 P25787 P25787 P60900 Q9Y606 Q9UPS7 NP_001017995 Q9BVF7 P55196 P62826 P62826 P46940 Q86VI3 P39023 P08865 Q15056 P26599 Q15427 P08621 P00338 Q8WWI1 Q9UGI8 Q9UGI8 Q9UGI8 P24666 Q86X29 Q86X29 Q8TDW0 O00453 P30040 P30040 Q13813 P42768
Q15431 P02768 P02768 Q15019 O95487 O00459 Q06124 P09769 P06241 P08631 P06239 P07948 Q3SX64 Q3SX64 Q13122 P11686 O15056 P60903 P31949 P31949 Q7Z3A1 P47897 P26639 Q9Y3B8 O75663 P02787 P29401 Q8NCT0 Q9Y2L5 P06753 Q9BZV1 P35221 P35221 P18206 Q13347 P50552 Q86UW9 Q9C0B5 Q13480 Q13480 Q9NPY3 Q9H251 Q9BZF9 Q16181 Q9BSJ2 P27986 P27986 P27986 P35580 P35579 Q92835 O15357 P00533 P00533 P16234 P15529-4 P12318 P12318 Q86X29 Q86X29 Q86X29 Q99650 P31942 P62995 P62995
P02671 Q8WZ73 Q96JD3 P78364 O60496 Q9UQC2 Q9Y5V3 Q8TEW8 O00459 P16591 P56945 Q03135 Q13191 O60716 Q9UHF2 Q92796 Q9Y2H0 Q99704 Q99704 Q12929 Q12929 O75955 O15117 Q13322 Q9NQ75 Q9NQ75 P35568 Q9Y4H2 Q9Y4H2 O14654 Q93052 O95297 Q14511 Q14511 Q6VY07 Q96HC4 O00560 O00560 Q8N5H7 O43609 Q9UDY2 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q8IZW7 Q8IZW7 P12110 Q14467 P15924 P15924 P26006 O00188 P15941 P15941 P15941 P15941 P15941 P15941 P15941
Q13835 Q13835 Q99959 P18827 P31431 P04083 P07355 P07355 Q9UHD8 Q99816 P31689 Q9NZJ4 Q9NRL2 O15551 O15551 O00501 O00501 Q9UJU6 Q96JJ3 O43491 P21333 O75369 P14923 P05783 P13647 P13647 Q15149 Q99426 Q9UPS7 Q9Y490 Q9Y490 Q9Y490 Q9Y490 Q9Y4G6 Q99424 P35573 P04075 P04075 P16152 P00450 P00450 Q92830 Q8N1G2 P43490 P29474 Q92831 P16885 P06737 P06737 Q9NP81 P21980 P21980 Q16881 P11172 P26640 Q6ZML7 P46940 P21359 Q9UJM3 Q9UJM3 P23743 P27986 P48426 P11216 Q9Y2H5
P22307 P22307 Q92835 P62140 Q99952 P29350 P29350 Q15256 Q13443 O75976 O75976 P11274 Q13131 Q05655 P78527 P00519-2 P09769 P09769 P42685 P42685 P06239 P07948 Q05397 Q13882 P43405 P43405 P43405 P43405 Q07912 P07947 P07947 Q9UM73 Q9UM73 Q9UM73 P54753 P10721 P54284 Q9BZW8 O15430 O15430 P20023 P17302 Q8NFI5 Q8NFI5 Q8NFI5 Q8NFI5 Q9NQ84 P68871 P02042 O43613 Q12791 Q96MP8 Q96MP8 Q8NHJ6 Q86X29 Q86X29 Q9BRK3 Q13291 Q96DU3 Q9HAV3 P02730 P02730 Q92499 Q13151 P09651
Q9Y2Z2 Q9Y2Z2 Q14764 Q9NS98 Q9NRY4 Q99623 P42229 Q9UKY1 P82933 O00541 Q9BVM7 Q8WXE0 Q96B33 Q96B33 P56747 P56747 O95484 P52943 Q5VTJ3 Q5VTJ3 Q8NBF2 Q8TEA1 O94886 Q9H0K5 Q9BV38 A1YPR0 Q9H7R5 Q9H7R5 Q9H9E3 P16284 Q02818 Q9UHB9 Q99873 Q9Y5P8 Q9NX69 Q9NWQ8 Q9NWQ8 Q9NWQ8 Q9NWQ8 Q6ZUJ8 Q96QU7 Q9HAP4 Q5SXH7 Q9UKY7 Q9H7V7 P53801 P53801 XP_370878 XP_113914 BAA76790 Q8WV15 Q8WV15 P35663 Q9NWT0 Q96TA1 Q17RB0 Q9H7U1 Q9ULK3 XP_373030
<b>LUNG/all</b>
O43603 P05023 P08559 P08559 P13796 P60660 P62158 P06733 P06733 P08727 P05787 P05787 P08670 P08729 P02538 P08670 P08670 P08670 P07954 P04899 P08754 P22102 P50395 P20592 P69905 P69905 P02042 P04075 P68104 P68104 P68104 Q12959 Q92796 P11413 P11413 P11413 Q9HC35 P18669 P18669 P18669 P18669 P62805 P62805 Q9UKY1 Q14247 Q9UJY1 P08107 P11142 P38646 P14625 P08238 Q9UIA9 P48739 P20963 P07766 P30273 P30273 P63252 P24071 P12314 P12314 O43914 P53990 P20936 P05129
Q9NZM1 Q9NZM1 Q9BSJ8 Q9BSJ8 P33151 P04040 P04040 Q99439 Q15417 Q01995 P37802 Q9NVF5 Q00610 Q00610 Q9NQR4 P35606 P23528 P17302 P61604 Q9Y6Y1 P12814 O43707 P13796 Q8IZD4 O14786 Q8N353 Q9ULV4 Q8NCT2 A1X283 P62333 Q99758 Q99758 P07108 P60709 P60709 P60709 P60709 P61158 P61158 P61158 Q8N142 Q7Z4W1 Q9BY49 P30038 P63010 O00519 O00519 P12532 Q96GR2 Q96GR2 Q96GR2 P17174 P00505 Q12955 Q9Y566 Q6PKA2 P04083 P50995 P50995 P07355 P07355 P07355 P07355 P07355 P08758 P08133 Q06481 Q06481 P05067 P05067 Q9H6X2 P26358 P63010 P17844 P20591

P20591 P59665 P36543 O14964 Q9Y490 Q16526 O00757 Q86UX7 O14964 O15145 P35568 Q9Y4H2 O15268 Q9UED8 Q9UED8 Q9NYT0 Q8N4X5 Q8N4X5 P21333 P21333 Q14315 P16144 P22105 P05556 Q00341 Q13753 Q13753 Q13753 P56597 O96000 O96000 Q9Y316 P12235 P05141 P05141 P35580 P35579 P31939 Q86VI4 Q15012 P49023 P49023 Q14152 P51512 P11216 Q08357 O00329 P19174 Q13627 Q92630 O43781 P06493 P06493 P24941 P24941
Q00535 P49840 Q86Z02 Q9H422 Q9UPZ9 Q9UPZ9 Q9Y4K4 P28482 Q15759 P53778 O15264 Q16539 P27361 Q13164 P45983 P45984 Q00536 Q05655 Q13523 P00519-2 Q06187 P16591 P09769 P42685 P06241 P06241 P23458 P23458 P06239 P07948 P07948 P07948 Q05397 Q05397 Q05397 Q14289 Q14289 P43405 Q9UM73 Q9UM73 Q9UM73 Q9UM73 P30530 P30530 Q08345 Q08345 Q08345 Q16832 Q16832 P00533 P00533 P29317 P29320 P11362 P17948 P17948 P35916 P35916 P06213 P06213 P06213 P29376 P29376 P29376 P08581
P08581 P08581 P16234 P16234 P16234 P47712 P15924 Q6S376 P46783 P21796 P07737 P35080 P25787 P25787 P60900 Q9Y606 Q9UPS7 NP_001017995 Q9BVF7 P55196 P62826 P62826 P46940 Q86VI3 P39023 P05388 P08865 P42677 Q15056 P26599 Q15427 P08621 P00338 P07195 Q8WWI1 Q9UGI8 Q9UGI8 Q9UGI8 P24666 Q86X29 Q86X29 Q8TDW0 O00453 Q8TEW0 P30040 P30040 Q13813 P42768 O00401 Q15431 P02768 P02768 Q15019 O95487 O00459 Q06124 Q06124 P09769 P06241 P06241 P08631 P06239 P07948 P07948 Q3SX64 Q3SX64
Q13122 P11686 O15056 P60903 P31949 P31949 Q7Z3A1 P47897 P26639 Q9Y3B8 O75663 P02787 P29401 Q8NCT0 Q9Y2L5 P06753 Q9BZV1 P35221 P35221 P18206 P18206 Q13347 P50552 Q86UW9 Q9C0B5 Q13480 Q13480 Q13480 Q9NZM3 Q9NPY3 Q9H251 Q9Y624 Q9BZF9 Q16181 Q9BSJ2 P27986 P27986 P27986 P35580 P35579 Q92835 O15357 Q13627 P00533 P00533 P00533 P08581 P16234 P15529-4 P15529-4 P12318 P12318 Q86X29 Q86X29 Q86X29 Q99650 P31942 P62995 P62995 P02671 Q8WZ73 Q96JD3 P78364 O60496 Q9UQC2
Q9Y5V3 Q8TEW8 O00459 O00459 Q86YV5 P16591Q8IZP0 P56945 P56945 P56945 P56945 Q03135 Q13191 O60716 O60716 O60716 Q9UHF2 Q92796 Q9Y2H0 Q99704 Q99704 Q12929 Q12929 O75955 O15117 Q13322 Q9NQ75 Q9NQ75 P35568 P35568 Q9Y4H2 Q9Y4H2 Q9Y4H2 Q9Y4H2 O14654 Q93052 O95297 O95297 Q14511 Q14511 Q6VY07 Q96HC4 O00560 O00560 Q8N5H7 P29353 O43609 Q9UDY2 Q9UDY2 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q8IZW7 Q8IZW7 P12110 Q14467 P15924 P15924 P26006
O00188 P15941 P15941 P15941 P15941 P15941 P15941 P15941 Q13835 Q13835 Q99959 Q99569 P18827 P31431 P04083 P07355 P07355 P07355 Q9UHD8 Q99816 P31689 Q9NZJ4 Q9NRL2 O15551 O15551 O00501 O00501 Q9UJU6 Q96JJ3 O43491 P21333 O75369 P14923 P05783 P13647 P13647 Q15149 Q99426 Q9UPS7 Q9Y490 Q9Y490 Q9Y490 Q9Y490 Q9Y4G6 Q99424 P35573 P04075 P04075 P16152 P00450 P00450 Q92830 Q8N1G2 P43490 P29474 Q92831 P16885 P06737 P06737 Q9NP81 P21980 P21980 Q16881 P11172 P26640
Q6ZML7 P46940 P21359 Q13671 Q9UJM3 Q9UJM3 P23743 P27986 P48426 P11216 Q9Y2H5 P22307 P22307 Q92835 P62140 Q06124 Q99952 P29350 P29350 P18433 Q15256 Q13443 O75976 O75976 P11274 Q13131 Q05655 Q05655 P78527 P00519-2 P09769 P09769 P42685 P42685 P06239 P07948 Q05397 Q05397 Q13882 P43405 P43405 P43405 P43405 Q07912 Q07912 P29597 P07947 P07947 Q9UM73 Q9UM73 Q9UM73 P29317 P29317 P29317 P54753 P10721 P54284 Q9BZW8 O15430 O15430 P20023 P17302 Q8NFI5 Q8NFI5 Q8NFI5
Q8NFI5 Q9NQ84 P68871 P02042 O43613 Q12791 Q96MP8 Q96MP8 Q8NHJ6 Q86X29 Q86X29 Q9BRK3 Q13291 Q96DU3 Q9HAV3 P02730 P02730 Q92499 Q13151 P09651 Q9Y2Z2 Q9Y2Z2 Q14764 Q9NS98 Q9NRY4 Q9NRY4 Q99623 P40763-2 P40763 P42229 Q9UKY1 P82933 O00541 Q9BVM7 Q8WXE0 Q96B33 Q96B33 P56747 P56747 O95484 P52943 Q5VTJ3 Q5VTJ3 Q8NBF2 Q8TEA1 O94886 Q9H0K5 Q9BV38 A1YPR0 Q9H7R5 Q9H7R5 Q9H9E3 P16284 Q02818 Q9UHB9 Q99873 Q9Y5P8 Q9NX69 Q9NWQ8 Q9NWQ8 Q9NWQ8 Q9NWQ8 Q6ZUJ8



Q96QU7 Q9HAP4
Q5SXH7 Q9UKY7 Q9H7V7 P53801 P53801 XP_370878 XP_113914 BAA76790 Q8WV15 Q8WV15 P35663 Q96A22Q9NWT0 Q96TA1 Q17RB0 Q9H7U1 Q9ULK3 XP_373030 EAW84943 EAW96536 Q6P0M1 XP_930908 Q8WWZ8
<b>S1</b>
P06493 P06493 O15264 P24941 P24941 P27361 Q92630 Q16539 Q9UPZ9 Q9UPZ9 P28482 O43781 Q00526 P06239 P07948 Q08345 Q08345 Q08345 P00519-2 P29320 Q05397 Q05397 P17948 P17948 P08631 P00519 P00519 P42684 P05067 Q06481 P05067 P00325 Q96A22 P42677 Q7Z6J2 P05067 Q9H792 O95817 Q05397 Q6NZI2 P07355 P04083 P08758 P35579 P60709 P60709 P60709
P63261 P63261 P35221 P49023 P49023 Q00610 P08670 P21333 P08670 P68104 P68104 Q05639 Q05639 P02768 P02768 P02042 P30273 P05023 P08559 P29803 P13796 P33151 O14964 P08238 P11413 P11413 P50552 O00329 P17302 Q92796 Q12959 O00757 P09467 P46940 P19174 P59665 P59666 P05141 Q9BXI2 P06733 P29401 Q13480 Q13480 Q13480 O00459 P35579 P27986 Q13627 P00533 P00533 P15529-4 Q03135 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q9HBL0 Q96A22 P07355 P07355 Q9UDY2 P53801 Q06124 Q9NQ75 Q9NWQ8 Q8NFJ5 Q8NFJ5 Q8NFJ5 Q07912 Q9NRY4 Q13291 P42229 P68871 P43405 P13647 P15941 P15941 P15941 P15941 O60716 Q12929 P16284 Q05655 P56945 Q9UJM3 Q16881 O00560 O15389 Q92569 P51692 Q15464 O60716 Q14511 Q14511 Q8IX19 Q06124 P42224

Tabulka č.3: UniProt kódy proteinů z datasetů LUNG/all, LUNG/only, HCT116/all, HCT116/only, S1

### 4.3. Sekvenční analýza

UniProt (The UniProt Consortium 2014) - The Universal Protein Resource je komplexní databáze proteinových sekvencí a funkčních informací. Jedna z dílčích součástí, UniProtKB (Magrane et al. 2011), obsahuje sekci Swiss-Prot (Boeckmann et al. 2005), zahrnující soubor manuálně anotovaných, tříděných, kontrolovaných a neredundantních proteinových sekvencí. Z literatury shromážděné informace nejdříve projdou řadou in silico selekčních operací. Proces manuální anotace se skládá z rozboru, srovnání, sloučení všech dostupných sekvencí daného proteinu, kritického přezkoumání souvisejících experimentálních i predikčních dat. Swiss-Prot má za cíl poskytnout všechny známé relevantní informace o konkrétním proteinu. Každý záznam individuálního polypeptidu popisuje odlišné produkty pocházející z podobných genů různých organismů, včetně každé varianty vzniklé alternativním sestřihem, polymorfismy nebo posttranslačními modifikacemi. Proteinové rodiny a skupiny jsou pravidelně přezkoumávány a doplňovány aktuálními vědeckými poznatky.

Ke každému proteinu prvního (lung) a druhého (HCT-116) datasetu, jsem v databázi UniProtKB/ Swiss-Prot dle příslušného ID vyhledal stránku obsahující souhrn informací a

numerickým pořadím opatřený zápis aminokyselinové sekvence daného polypeptidu izolovaného ze zdravých buněk. Odpovídající místa takto získaných standardních sekvencí jsem porovnal se sekvenčními detaily okolí fosfotyrosinů (uvedeny ve výchozích datových souborech) stejných proteinů získaných z buněk rakovinných.

#### **4.4. Strukturní analýza**

PDBe (Gutmanas et al. 2014) – Protein Data Bank of Europe, je evropská databáze pro shromažďování, třídění a šíření informací o prostorové struktuře bílkovin. Jako jeden z členů wwPDB (Berman et al. 2003) ve spolupráci s ostatními členy i dalšími databázemi integruje informace z biomedicínských portálů po celém světě a bioinformatických, krystalografických, NMR, EM výzkumných projektů kde udržuje krok s technickým i metodickým rozvojem těchto oblastí. PDBe obsahuje komplexní data ohledně 3D uspořádání, primárních, sekundárních, terciárních, kvartérních struktur, sekvencí a rodin proteinů, umožňuje vyhledávání na základě jednoduchých textových požadavků i složitějších 3D parametrů. Tým PDBe vyvinul mnoho sofistikovaných nástrojů pro strukturní analýzu makromolekul, zasloužil se o inovace prostředků ukládání a rozboru strukturních dat.

Prostřednictvím všech ID z prvního až třetího datasetu jsem, bylo-li to možné, v databázi UniProt ke každému proteinu zjistil, zda má známou experimentálně určenou 3D strukturu v databázi PDBe, obsahujícím soubory s popisem 3D struktury a grafické vyobrazení sekundární struktury daného sekvenčního úseku. Následovala selekce vhodného intervalu, pokud byl k dispozici, do kterého spadalo sekvenční číslo konkrétního fosfotyrosinu. V případě existence většího množství struktur studovaného fosforylačního místa proteinu měly přednost intervaly se strukturou detekovanou pomocí X-ray před méně přesnou NMR metodou, žádoucí byla rovněž co nejnižší rozlišovaná vzdálenost (Resolution, Å) i eventuální dobrá pozice fosfotyrosinu (spíše blíž ke středu než okraji intervalu). Pomocí kódu vybraného nejvhodnějšího úseku zobrazený odkaz databáze PDBe umožnil získání PDB souboru s popisem 3D struktury i určení sekundární struktury příslušného fosforylačního místa přes sekvencí doplněné grafické schéma. Tímto způsobem tedy byly shromážděny informace ohledně strukturních vlastností fosforylovatelných pozic, zjištěno poměrné zastoupení fosforylovaných sekundárních struktur v nádorech plic a kolorekta.

## 4.5. Expresní analýza

Oncomine je databáze zabývající se integrací informací ohledně genové exprese v rakovinných nádorech (Rhodes et al. 2004). Jedná se o rozbor, standardizaci, třídění, shromažďování publikovaných expresních dat z microarray analýz, jejich srovnání s normálním stavem ve zdravých buňkách příslušného typu a systematizovanou prezentaci odborné veřejnosti. Selektuje expresní profily odlišných typů rakoviny i anotací některých genových skupin, poskytuje možnosti identifikace genů, procesů, signálních drah, k jejichž dysregulaci u jednotlivých druhů tumorů dochází. Databáze aktuálně obsahuje 715 datových kolekcí, zahrnujících celkem 86733 sérií výsledků jednotlivých výzkumných projektů s aplikací microarray metod, z toho 12764 týkajících se zdravých tkání. Ke své práci jsem využil zdarma dostupnou část databáze, která zaujímá 33% kompletního množství informací, je tvořena převážně daty z nádorové histologie i analýz porovnávajících tumory s čistou linií rakovinných buněk (neboť tumory neobsahují 100% rakovinných buněk) a zdravou tkání.

Přes genové symboly z datových souborů LUNG, HCT116, S1 jsem v databázi Oncomine (v sekci Gene Summary View) vyhledal expresní hladiny konkrétních fosfoproteinů zjištěné různými výzkumnými projekty u jednotlivých typů tumorů. Zajímaly mě profily z plicních nádorů pro datasey LUNG a S1, z kolorektálních nádorů pro HCT116. Expresní charakteristika zaujímala vždy jednu ze čtyř variant:

- a) protein ve srovnání s normální hladinou exprimovaný ve zvýšené míře (Over)
- b) snížené míře (Under)
- c) ve více výzkumných projektech zjištěna pod- i nadexprese (Over/Under)
- d) data o změně exprese nejsou k dispozici (pracovně značeno „//“)

Pro každý dataset i případnou dílčí sekci (LUNG/all, LUNG/only, HCT116/all, HCT116/only, S1) byla zvlášť ověřena celková četnost změny exprese daných proteinů v nádorové tkáni a stanovena procentuální zastoupení fosforylačních míst, vyskytujících se na proteinech s expresní charakteristikou a), b), c) nebo d).

## 4.6. Charakteristika orthologů fosforylačních míst

S využitím databáze PhosphoSitePlus, jedné z nejkomplexnějších současných databází posttranslačních modifikací, jsem provedl zmapování výskytu orthologů daných lidských tyrosinových pozic i jejich fosforylace u modelových organismů.

Z nabídky způsobů vyhledávání ve PhosphoSitePlus jsem využil možnost orientace na základě ID proteinu (sekce „Protein, Sequence, or Reference Search“), z následně zobrazeného seznamu organismů jsem zvolil člověka (human). Pro každý polypeptid je k dispozici sloupec s aminokyselinovými sekvencemi fosforylačních míst i sekvenčními čísly modifikovaných aminokyselin u člověka. Další sloupce obsahují stejné detaily fosforylovatelných pozic orthologů daného proteinu u některých modelových organismů, nejčastěji hlavně myši a krysy, v mnoha případech také dále králíka, prasete, kuřete nebo tura, zřídka eventuálně ještě psa, křečka, ovce. Pokud je aminokyselina u zvířete fosforylována, je v sekvenci zapsána malým písmenem, velká písmena značí aminokyseliny bez modifikace.

Konkrétní fosforylovatelné pozice jsem ve sloupcích vyhledával podle sekvenčních čísel a úseků aminokyselinových sekvencí přítomných ve výchozích datasetech. Získaná data jsem zapisoval do tabulky v programu Microsoft Office Excel 2007. Tyrosinfosforylační místo se buď u určitého organismu vůbec nevyskytuje (při práci značeno symbolem „x“) nebo je přítomno, ale není fosforylováno (symbol „O“) či fosforylováno je („p“). V datasetu LUNG/all byly pro každý protein zaznamenány informace o všech uvedených ortholozích (všech zmíněných organismů) jeho tyrosinfosforylačního místa (míst), soubory LUNG/only, HCT116/all, HCT116/only, S1 jsem zdokumentoval z hlediska výskytu i stavu fosforylace orthologů pouze u myši, krysy a stavu fosforylace v bílkovinách zdravé lidské tkáně. Výsledky byly pro snadnější vyvození závěrů sumarizovány u individuálních organismů v jednotlivých datasetech a procentuálně vyjádřeny. Sumarizace probíhala pomocí funkce COUNTIF do jejíhož vzorce byly postupně vepisovány souřadnice vymezující sčítané oblasti a také symboly, jejichž počet v ohraničeném poli jsem chtěl vyjádřit (x, O, p).

## 4.7. Srovnání výskytu fosfoaminokyselin

V datové sérii STY jsem u každé z patnácti dílčích částí, čítající tisíce fosforylovatelných pozic, pomocí programu Microsoft Office Excel 2007 zjistil procentuální zastoupení tří variant fosforylačních míst, rozlišovaných podle fosforylované aminokyseliny (serin, treonin, tyrosin). K tomuto účelu jsem využil maticový vzorec, složený z funkcí DÉLKA, DOSADIT, SUMA.

Údaj o sekvenčním čísle fosforylované aminokyseliny v souboru STY vždy zahrnuje také její jednopísmennou zkratku (serin = S, treonin = T, tyrosin = Y), počty jednotlivých typů fosforylačních míst bylo proto možné sumarizovat následujícím způsobem. Pokud se v některém údaji se sekvenčním číslem vyskytovala zkratka konkrétní fosforylovatelné aminokyseliny, pro kterou byl právě určován počet míst ji obsahujících, funkce DOSADIT se stejným písmenem vepsaným do svého vzorce tento znak mezi ostatními eliminovala dosazením „žádného znaku“ (ve vzorci symbolizováno ""). Eliminací v této souvislosti rozumějme neviditelnost znaku pro funkci DÉLKA. Funkce DÉLKA stanovila počet znaků takto zkráceného textového výrazu. Toto číslo bylo odečteno od hodnoty získané funkcí DÉLKA bez předchozí úpravy textového výrazu funkcí DOSADIT (výraz byl tedy v tomto případě o jeden znak delší). Výsledek rozdílu těchto dvou funkcí DÉLKA byl roven jedné. V údaji, kde se jednopísmenná zkratka příslušné aminokyseliny nevyskytovala, nedošlo ke zkrácení textu funkcí DOSADIT o jeden znak, výsledek rozdílu dvou funkcí DÉLKA o shodných hodnotách byl roven nule. Funkce SUMA provedla součet výsledků diskutovaných rozdílů funkcí DÉLKA pro všechna fosforylační místa. Finální hodnota odpovídá počtu fosforylovatelných pozic obsahujících danou fosfoaminokyselinu.

U čtrnácti souborů fosforylačních míst vybraných nádorových onemocnění, jsem pro každý určil procentuální zastoupení typů fosforylovatelných pozic zvlášť v seznamu selektovaném pouze na místa s minimálně 10% deklarovaných referencí (sekce STY/10%) a v seznamu celkovém (sekce STY/all). Následně byla provedena konfrontace výsledků s poměrem výskytu jednotlivých fosfoaminokyselin v úplném soupisu všech fosforylací zahrnutých do PhosphoSitePlus i mezi všemi typy tumorů navzájem. Nejdůležitější vzhledem k zaměření mé práce byla hlavně rozdílná množství zastoupení fosfotyrosinu.

## 5. Výsledky a diskuse

### 5.1. Srovnání výskytu fosfotyrosinu u vybraných typů tumorů

Databáze	STY/all					
	Treonin		Serin		Tyrosin	
PhosphoSitePlus	19,70%		63,60%		16,70%	
Nádorové onemocnění	STY/all			STY/10%		
	Treonin	Serin	Tyrosin	Treonin	Serin	Tyrosin
Rakovina prsu	15,02%	55,25%	29,73%	0,52%	0,00%	99,48%
Rakovina plic	15,41%	44,63%	39,96%	1,86%	0,93%	97,21%
Rakovina jícnu	6,58%	15,43%	78,00%	2,51%	7,71%	89,78%
Rakovina žaludku	18,86%	40,17%	40,96%	4,08%	20,99%	74,94%
Rakovina kolorekta	8,20%	20,59%	71,20%	1,92%	3,19%	94,89%
Rakovina slinivky	0,85%	0,57%	98,58%	0,00%	0,00%	100,00%
Rakovina jater	12,83%	26,56%	60,62%	4,17%	10,93%	84,89%
Rakovina štítné žlázy	0,00%	0,00%	100,00%	0,00%	0,00%	100,00%
Rakovina prostaty	17,60%	76,06%	6,33%	9,15%	71,92%	18,93%
Chron. lymfoblast. leuk.	9,21%	39,20%	51,59%	0,00%	30,00%	70,00%
Chron. myeloid. leuk.	15,10%	47,16%	37,74%	1,86%	5,33%	92,81%
Akut. lymfoblast. leuk.	0,05%	0,00%	99,95%	0,16%	0,00%	99,84%
Akut. myeloid. leuk.	0,67%	1,73%	97,60%	0,00%	0,00%	100,00%
Glioblastom	9,52%	31,03%	59,45%	3,13%	23,61%	73,25%

**Tabulka č.4:** Procenta výskytu fosfoaminokyselin v datasetu STY zahrnujícím čtrnáct typů rakoviny a celou databázi PhosphoSitePlus. STY/all obsahuje pro každý dílčí soubor všechny fosforylace, STY/10% pouze fosforylace s alespoň 10% publikovaných referencí. Barevné rozlišení: treonin = bílá, serin = světle zelená, tyrosin = světle červená, STY/all = žlutá sekce, STY/10% = tyrkysová sekce, dílčí datové série = světle modrá.

Mým cílem bylo potvrdit, že tyrosin je v nádorových tkáních vysoce nadprůměrně fosforylován.

Datová série STY/all je tvořena patnácti úplnými seznamy všech fosforylovatelných pozic (tyrosinových, serinových, treoninových) - jedním kompletním soupisem pro celou databázi PhosphoSitePlus a čtrnácti menšími (rovněž z PhosphoSitePlus) pro vybrané typy nádorových onemocnění. Série STY/10% je omezena pouze na místa zmíněná alespoň u 10% z příslušného celkového seznamu referencí.

V tabulce č.4 je vidět, že zastoupení tyrosinových fosforylací je oproti běžnému výskytu 16,70% mezi všemi proteiny PhosphoSitePlus téměř u každého nádorového onemocnění silně zvýšené. To je v souladu s publikovanými názory ohledně role aberantní fosfotyrosinové signalizace v onkogenezi, ze kterých má práce vycházela. Rakovina štítné žlázy má z detekovaných fosforylací 100% fosfotyrosinu v obou souborech, STY/all i STY/10%. Akutní lymfoblastická i myeloidní leukémie a rakovina slinivky mají v datasetu STY/all hodnoty velmi blízké 100% fosfotyrosinu, ve STY/10% dosáhnou 100% s výjimkou akutní lymfoblastické leukémie, kde dojde k nárůstu o zlomek množství fosfotreoninu, vzhledem k možným nepřesnostem získávání dat a prezentace výsledků zanedbatelnému. U rakoviny prsu, plic, kolorekta, jícnu, žaludku, jater, chronické myeloidní i myeloblastické leukémie a glioblastomu je zastoupení tyrosinových fosforylací ve STY/all různou měrou zvýšené nad normál, o zhruba 13% u rakoviny prsu, o několikanásobek standardní hladiny u ostatních. Dataset STY/10% vykazuje množství fosfotyrosinu těchto nádorových onemocnění ještě silněji vzrostlé. Celkově lze o kolekci vybraných typů rakoviny říci, že podíly fosfotreoninu i fosfoserinu jsou ve STY/all oproti běžnému výskytu snižené, ve STY/10% dále značně poklesnou ku prospěchu zastoupení tyrosinových fosforylací. To lze interpretovat menším počtem projektů zabývajících se detekcí fosfoserinu a fosfotreoninu, jejich poměr vůči fosfotyrosinu tedy může být mírně zkreslený malým počtem referencí. Pro rakovinu prostaty byly získány výsledky velmi odlišné od všech ostatních onemocnění, neboť obsah tyrosinových fosforylací je zde ve STY/all podprůměrný, ve STY/10% lehce nadprůměrný a na zdaleka nejvyšší hodnotě se u obou STY sekcí drží fosfoserin.

## 5.2. Vznik nových fosforylačních míst v nádorové tkáni

Jedním ze způsobů, jak může dojít k transformaci buňky na buňku rakovinnou, je vznik nových tyrosinových fosforylačních míst, prostřednictvím jednonukleotidových mutací (SNV), majících vliv na změnu původní sekvence fosforylačního místa (Reimand, Bader 2013; Radivojac et al. 2008). Z tohoto důvodu mě zajímalo, zda v mnou analyzovaných typech tumorů nějaká nová fosforylační místa skutečně vznikla.

Dataset LUNG/all obsahuje 669 tyrosinových fosforylačních míst proteinů plicních nádorů. Každá fosforylační pozice je zde vymezena sekvenčním číslem fosforylovatelného tyrosinu a sekvenčními detaily jeho okolí v proteinech nádorových buněk. Sekvenčními detaily v tomto případě rozumějme buď přímo peptidovou sekvenci fosforylačního místa nebo určitý interval, nezahrnující alespoň na jedné straně od fosforylovatelného rezidua přesně sedm aminokyselinových zbytků, nýbrž počet vyšší či nižší. Dataset HCT116/all obsahuje 171 tyrosinových fosforylačních míst proteinů vybrané buněčné linie kolorektálních nádorů. Každá fosforylační pozice je vymezena sekvenčním číslem fosforylovatelného tyrosinu a aminokyselinovou sekvencí celého fosforylačního místa, získanou z proteinů nádorových buněk.

Při srovnání sekvenčních determinant všech fosforylovatelných pozic obou souborů LUNG/all i HCT116/all s příslušnými úseky sekvencí proteinů zdravých buněk v databázi UniProt, jsem nenalezl žádné rozdíly. Z mých výsledků vyplývá, že alespoň u některých typů nádorů nemusí docházet ke vzniku nových fosforylačních míst.

## 5.3. Sekundární struktura fosforylačních míst

Neboť bylo prokázáno, že k fosforylacím proteinů zdravých buněk dochází převážně v neuspořádaných strukturách, zajímalo mě, jaká je situace v buňkách rakovinných.

Z 669 tyrosinových fosforylačních míst datasetu LUNG/all se jedna třetina (223 míst; 33,33%) nachází na proteinech známé sekundární struktury. Pro 86,10% (192 míst) z těchto 223 fosforylovatelných pozic je známa sekundární struktura rovněž pro fosforylační místo samotné. Zbylých téměř 14% zaujímají pozice, které nejsou ve struktuře vidět a pravděpodobně se jedná o neuspořádané struktury. Soubor LUNG/only, odvozený od LUNG/all, obsahuje pouze 580 fosforylací, které rakovina plic nemá společné s nádory kolorekta. Toto kritérium



snížilo počet fosforylací lokalizovaných na bílkovinách identifikované sekundární struktury z 223 na 183. LUNG/only tedy zaujímá 33,55% míst na proteinech s pojmenovanou sekundární strukturou. Mezi těmito 183 fosfotyrosinovými pozicemi je 161 míst (87,98% ze 183) známé sekundární struktury. Podrobné zastoupení sekundárních struktur v datasetech LUNG/only, LUNG/all ukazuje tabulka č. 5.

Sekundární struktura	LUNG/all		LUNG/only	
	Proteinů	%	Proteinů	%
$\alpha$ -helix	79	35,43%	72	39,34%
$\beta$ -list	49	21,97%	42	22,95%
Smyčka	64	28,70%	47	25,68%
Známa	192	86,10%	161	87,98%

**Tabulka č.5:** Zastoupení sekundárních struktur v datasetech LUNG/all, LUNG/only. Procenta LUNG/all jsou odvozena z celku 223 fosforylačních míst proteinů identifikované sekundární struktury, u LUNG/only z celku 183 fosforylačních míst proteinů identifikované sekundární struktury. V levém sloupci jsou jednotlivé sekundární struktury, řádek „Známa“ popisuje kompletní množství fosforylačních pozic se známou strukturou. Následuje vždy počet proteinů dané struktury a jeho procentuální vyjádření pro každý dataset.

Dataset HCT116/all a jeho 171 tyrosinových fosforylačních pozic rakoviny kolorekta zahrnuje 54 míst (31,58%) na proteinech definované sekundární struktury. Z těchto 54 míst je 75,93% (41 míst) se známou sekundární strukturou. Soubor HCT116/only, čítá 79 fosforylací z HCT116/all, které jsou ve srovnání s LUNG/all pro HCT116/all specifické. Mezi nimi se však vyskytuje jen 17,72% (14 pozic) lokalizovaných na bílkovinách identifikované sekundární struktury. Struktura fosforylačního místa je pak známa pouze pro 10 z těchto 14 míst. Podrobné zastoupení sekundárních struktur v datasetech HCT116/only, HCT116/all ukazuje tabulka č. 6.

Sekundární struktura	HCT116/all		HCT116/only	
	Proteinů	%	Proteinů	%
$\alpha$ -helix	11	20,37%	4	28,57%
$\beta$ -list	9	16,67%	2	14,29%
Smyčka	21	38,89%	4	28,57%
Známa	41	75,93%	10	71,43%

**Tabulka č.6:** Zastoupení sekundárních struktur v datasetech HCT116/all, HCT116/only. Procenta jsou odvozena z celku 54 fosforylačních míst proteinů identifikované sekundární struktury, u HCT116/only z celku 14 fosforylačních míst proteinů identifikované sekundární struktury. V levém sloupci jsou jednotlivé sekundární struktury, řádek „Známa“ popisuje kompletní množství fosforylačních pozic se známou strukturou. Následuje vždy počet proteinů dané struktury a jeho procentuální vyjádření pro každý dataset.

Soubor S1 zahrnuje celkem 144 fosfotyrosinových pozic plicních nádorů, z nichž 59 míst, tj. 40,97% náleží proteinům identifikované sekundární struktury. Konkrétní sekundární struktura

fosforylačního místa je ale určena jen u 51 míst (86,44% z 59 pozic). Podrobné zastoupení sekundárních struktur v datasetu S1 ukazuje tabulka č. 7.

Sekundární struktura	S1	
	Proteinů	%
$\alpha$ -helix	20	33,90%
$\beta$ -list	13	22,03%
Smyčka	18	30,51%
Známa	51	86,44%

**Tabulka č.7:** Zastoupení sekundárních struktur v datasetu S1. Procenta jsou odvozena z celku 59 fosforylačních míst proteinů identifikované sekundární struktury. V levém sloupci jsou jednotlivé sekundární struktury, řádek „Známa“ popisuje kompletní množství fosforylačních pozic se známou strukturou. Následuje počet proteinů dané struktury a jeho procentuální vyjádření.

Pořadí četnosti výskytu jednotlivých sekundárních struktur v S1 koresponduje s datasetem LUNG/all i LUNG/only. Nejčastější je  $\alpha$ -helix, potom smyčka. V souboru HCT/116/all je tomu naopak, smyčka je první,  $\alpha$ -helix druhý. HCT116/only má jejich poměr vyrovnán. Zde sice může docházet ke zkreslení již značně nízkým počtem proteinů definované struktury v HCT116/only, nicméně tendenci k poklesu četnosti smyček a nárůstu  $\alpha$ -helixů při pomyslném přechodu ze sekce -all do -only mají datasety LUNG i HCT116 společnou.  $\beta$ -list je strukturou v každém z diskutovaných datasetů (LUNG/all, LUNG/only, HCT116/all, HCT116/only, S1) zastoupenou nejméně.

Významná současná studie přinesla výsledky, které ukazují, že četnost výskytu regulačních tyrosinových fosforylací proteinů zdravých buněk je nejnižší v uspořádaných strukturách a velmi rychle roste s délkou, neuspořádaností, rozvolněností smyček, protože tam jsou aminokyselinová rezidua čím dál lépe dostupná pro kinázy i fosfatázy. Přes 90% fosforylací leží v neuspořádaných strukturách (Tyanova et al. 2013). V souborech tyrosin fosforylačních míst proteinů plicních a kolorektálních tumorů, analyzovaných během mé práce jsou však výraznou měrou fosforylovány  $\alpha$ -helixy a  $\beta$ -listy. V LUNG/all, LUNG/only, S1 do oblastí jedné z těchto dvou sekundárních struktur spadá převážná většina fosforylací se známou strukturou fosforylačního místa, v HCT116/all, HCT116/only je jejich podíl také značný.

#### 5.4.Doménová analýza fosforylačních míst

Zajímalo nás, zda jsou některé domény v uvedených typech tumorů výrazně více fosforylovány i jaké jsou mezi nimi případné rozdíly.

V souboru LUNG/all, obsahujícím 669 tyrosinových fosforylovatelných pozic proteinů plicních nádorů, vykazuje 351 (52,47%) fosforylačních míst lokalizaci na známých proteinových doménách. Celkem je v LUNG/all zastoupeno 161 domén. Z toho je 106 domén (65,84%) v datasetu reprezentováno pouze jedinou fosforylací, 26 domén (16,15%) dvěma fosforylacemi a 15 domén (9,32%) třemi fosforylacemi. Podrobnější zastoupení domén se čtyřmi a více přiřazenými fosforylačními místy ukazuje tabulka č. 8.

LUNG/all			LUNG/only		
Doména	Proteinů	%	Doména	Proteinů	%
PF07714 - Pkinase Tyr	51	7,62%	PF07714 - Pkinase Tyr	41	7,07%
PF00069 - Pkinase	26	3,89%	PF00191 - Annexin	10	1,72%
PF00191 - Annexin	10	1,49%	PF00169 - PH	8	1,38%
PF00017 - SH2	10	1,49%	PF00017 - SH2	7	1,21%
PF00169 - PH	8	1,20%	PF00022 - Actin	7	1,21%
PF00022 - Actin	7	1,05%	PF00168 - C2	6	1,03%
PF00038 - Filament	6	0,90%	PF00069 - Pkinase	5	0,86%
PF00168 - C2	6	0,90%	PF00038 - Filament	5	0,86%
PF00300 - His Phos 1	4	0,60%	PF02189 - ITAM	4	0,69%
PF02189 - ITAM	4	0,60%	PF00402 - Calponin	4	0,69%
PF00402 - Calponin	4	0,60%	PF10515 - APP amyloid	4	0,69%
PF10515 - APP amyloid	4	0,60%	PF00412 - LIM	4	0,69%
PF00412 - LIM	4	0,60%	PF00300 - His Phos 1	3	0,52%
PF01044 - Vinculin	4	0,60%	PF01044 - Vinculin	3	0,52%

**Tabulka č.8:** Domény nejvíce reprezentované v datasetech LUNG/all, LUNG/only. Procenta LUNG/all jsou odvozena z celku 669 fosforylačních míst proteinů plicních nádorů, u LUNG/only z celku 580 fosforylačních míst proteinů plicních nádorů (která nemá LUNG/all společná s HCT116/all). V levém sloupci jsou Pfam kódy a názvy domén. Následuje vždy počet fosforylačních míst spadajících do dané domény a jeho procentuální vyjádření pro každý dataset.

V souboru S1, obsahujícím 144 tyrosinových fosforylovatelných pozic proteinů plicních nádorů, vykazuje 88 (61,11%) fosforylačních míst lokalizaci na známých proteinových doménách. Celkem je v S1 zastoupeno 44 domén. Z toho je 30 domén (68,18%) v datasetu reprezentováno pouze jedinou fosforylací, 8 domén (18,18%) dvěma fosforylacemi.

Podrobnější zastoupení domén se třemi a více přiřazenými fosforylačními místy ukazuje tabulka č. 9.

S1		
Doména	Proteinů	%
PF07714 - Pkinase Tyr	14	9,72%
PF00069 - Pkinase	13	9,03%
PF00022 - Actin	5	3,47%
PF00009 - GTP EFTU	4	2,78%
PF10515 - APP amyloid	3	2,08%
PF00191 - Annexin	3	2,08%

**Tabulka č.9:** Domény nejvíce reprezentované v datasetu S1. Procenta S1 jsou odvozena z celku 144 fosforylačních míst proteinů plicních nádorů. V levém sloupci jsou Pfam kódy a názvy domén. Následuje počet fosforylačních míst spadajících do dané domény a jeho procentuální vyjádření.

Dataset S1 je o mnoho menší než LUNG/all, vyskytují se zde podobné domény, zastoupení opět vede tyrosin kinázová a kinázová. GTP EFTU je reprezentována v LUNG/all třemi, v LUNG/only jedním fosforylačním místem. U nádorových buněk se mohou projevat alterace exprese elongačních faktorů domény GTP EFTU (Gopalkrishnan et al. 1999).

Častý výskyt tyrosin kinázové (Pkinase Tyr) a kinázové (Pkinase) domény je v souladu s publikovanými výsledky výzkumů, které byly výchozími zdroji informací mé práce. Rakovinné buňky se dle autorů často vyznačují abnormálně zesílenou fosfotyrosinovou signalizací. V datasetu LUNG/only, zúženém pouze na 580 fosforylačních pozic, která nemá LUNG/all společná s HCT116/all, zůstala většina domén zastoupena stejným nebo nižším množstvím míst, které však příliš zásadně jejich celkový podíl neovlivnilo. Výraznou výjimku tvoří kinázová doména (Pkinase), u níž počet reprezentujících fosforylačních míst poklesl velmi významně. Fosforylovatelné pozice kinázové domény, které ubyly, se mohou nacházet přímo na určitých tyrosin kinázách, jež mají plicní i kolorektální tumory společné. Tato místa bývají fosforylována při autofosforylační aktivaci tyrosin kináz, k níž v tumorech ve zvýšené míře také dochází. Doména SH2 má jako součást tyrosin kinázy významnou roli při fosforylaci tyrosinu, její častější výskyt proto není překvapivý. PH (Pleckstrin homology domain) váže membránové lipidy, podjednotky G proteinů, protein kinázu C a je někdy obsažena v serin/treonin i tyrosin kinázách (Yu et al. 2004). Annexiny tvoří skupinu proteinů s membránovou lokalizací na povrchu buněk, zastávají funkce v protizánětlivých a proapoptotických mechanismech, membránovém lešení, koagulaci i fibrinolýze. U části byl zaznamenán zvýšený výskyt v plicích i jiných solidních typech tumorů a jsou využívány při imunoterapiích, jako cíle protilátkové opsonizace (Oh et al. 2004). Co se týká domény Actin, aktinová mikrofilamenta hrají mimo jiné důležitou roli v buněčné migraci a tedy i v invazivních schopnostech rakovinných buněk, tvorbě metastáz (Alberts et al. 2002). Doména Filament

obsahuje proteiny, dávající vznik různým typům cytoskeletálních útvarů intermediárních filament, metylace genů této skupiny se často vyskytují v některých typech nádorů (Ulirsch 2013). C2 doména, schopna vazby k lipidovým membránám, bývá součástí s membránami interagujících proteinů, rovněž tedy jistých protein kináz (Hata et al. 1995). U části histidin fosfatáz domény His Phos 1 je známa jejich role v invazi buněk plicních nádorů (Xu et al. 2010). ITAM (Immunoreceptor Tyrosin-based Activation Motif) jsou proteiny mnoha receptorů imunitních buněk, aktivované tyrosinovou fosforylací, aberantní signalizace jejich prostřednictvím se projevuje v buňkách některých nádorů (Kim et al. 2014). Calponin, kalcium vázající protein s funkcí v hladké svalovině bývá také ve zvýšené míře přítomen v určitých typech tumorů (Debald et al. 2014). APP amyloid je transmembránový protein koncentrovaný v synapsích neuronů, určitých typech nádorů prsu, štítné žlázy aj. (Yang et al. 2012, Takagi et al. 2013). Proteiny s LIM doménou se uplatňují v cytoskeletální organizaci a migraci rakovinných buněk (Johannessen et al. 2006). Vinculin je protein podílející se na organizaci cytoskeletu, deregulace parametrů jeho výskytu také může mít vliv při tvorbě metastáz (Goldmann et al. 2013).

V souboru HCT116/all, obsahujícím 171 tyrosinových fosforylovatelných pozic proteinů plicních nádorů, vykazuje 107 (62,57%) fosforylačních míst lokalizaci na známých proteinových doménách. Celkem je v HCT116/all zastoupeno 31 domén. Z toho je 18 domén (58,06%) v datasetu reprezentováno pouze jedinou fosforylací, 7 domén (22,58%) dvěma fosforylacemi. Podrobnější zastoupení domén se třemi a více přiřazenými fosforylačními místy ukazuje tabulka č. 10.

HCT116/all			HCT116/only		
Doména	Proteinů	%	Doména	Proteinů	%
PF00069 - Pkinase	28	16,37%	PF00125 - Histone	18	22,78%
PF07714 - Pkinase Tyr	18	10,53%	PF07714 - Pkinase Tyr	9	11,39%
PF00125 - Histone	18	10,53%	PF00069 - Pkinase	7	8,86%
PF00009 - GTP EFTU	4	2,34%	PF00009 - GTP EFTU	2	2,53%
PF00017 - SH2	4	2,34%	PF00300 - His Phos 1	2	2,53%
PF00300 - His Phos 1	3	1,75%	PF00017 - SH2	1	1,27%

**Tabulka č.10:** Domény nejvíce reprezentované v datasetech HCT116/all, HCT116/only. Procenta HCT116/all jsou odvozena z celku 171 fosforylačních míst proteinů kolorektálních nádorů, u HCT116/only z celku 79 fosforylačních míst proteinů kolorektálních nádorů (která nemá HCT116/all společná s LUNG/all). V levém sloupci jsou Pfam kódy a názvy domén. Následuje vždy počet fosforylačních míst spadajících do dané domény a jeho procentuální vyjádření pro každý dataset.

HCT116/all obsahuje jako ostatní datasety nejvyšší zastoupení domén kinázové a tyrosin kinázové. I další častěji reprezentované domény jsou společné se soubory LUNG/all, LUNG/only, S1. Významným rozdílem mezi doménovým zastoupením u plicních a kolorektálních nádorů je histonová doména se silným výskytem v HCT116/all. V HCT116/only množstvím míst dokonce značně převyšující tyrosin kinázovou doménu. Během onkogeneze se mohou uplatňovat modifikace histonů ve specifických kontextech, např. byla zjištěna role histonové metylace při umlčování tumorsupresorového genu v kolorektálních nádorech (Kondo et al. 2003).

Je velmi zajímavé, že ve všech popisovaných datových sériích náleží do dobře definovaných domén relativně malý počet tyrosinových fosforylačních míst.

## 5.5. CATH charakteristika

Tímto strukturním srovnáním jsem chtěl identifikovat sekvenčně i funkčně nepřibuzné proteiny, které jsou však fosforylovány se srovnatelnou frekvencí, díky podobné struktuře.

V souboru LUNG/all, obsahujícím 669 tyrosinových fosforylovatelných pozic proteinů plicních nádorů, se nachází 223 (33,33%) fosforylačních míst na proteinech se známou sekundární strukturou. Z této podmnožiny vykazuje 137 (61,43%) míst příslušnost k ustanoveným CATH nadrodinám (CATH Superfamily). Celkem je v LUNG/all zastoupeno 56 CATH nadrodin. Mezi nimi je 33 nadrodin (58,93%) v datasetu reprezentováno pouze jedinou fosforylací, 12 nadrodin (21,43%) dvěma fosforylacemi. Podrobnější zastoupení CATH nadrodin se třemi a více přiřazenými fosforylačními místy ukazuje tabulka č. 11, kde je vidět, že největší počet, 38 fosforylací náleží strukturálním uspořádáním s převažujícím  $\alpha$ -helixem (první pozice CATH kódu je 1), 31 fosforylací spadá do struktur s přibližně vyváženým podílem  $\alpha$ -helixu i  $\beta$ -listu (první pozice CATH kódu je 3), 11 fosforylací se nalézá na útvech majoritního výskytu  $\beta$ -listu (první pozice CATH kódu je 2). Doménové zastoupení CATH nadrodin, určených pro zmíněných 137 z 669 fosforylací LUNG/all, zahrnuje některé z domén nejběžnějších v celém LUNG/all (tyrosin kinázová, kinázová, Annexin, SH2, PH, His Phos 1, Filamin) i osm jiných.

V souboru LUNG/only, omezeném na 580 fosforylačních pozic proteinů plicních nádorů, která nemá LUNG/all společná s HCT116/all, se počet míst definované sekundární struktury snížil ke 183 (31,55% z 580). Z toho 114 (62,30%) fosforylačních míst spadá pod celkový počet 55 CATH nadrodin, jichž je 34 (61,82%) v datasetu zastoupeno jednou fosforylací, 10 (18,18%) dvěma fosforylacemi. Podrobné zastoupení nadrodin se třemi a více přiřazenými fosforylačními místy ukazuje tabulka č. 12, kde je vidět, že k formacím s převažujícím  $\alpha$ -helixem patří 26 fosforylací, 24 fosforylací spadá do struktur s přibližně vyváženým podílem  $\alpha$ -helixu i  $\beta$ -listu, 10 fosforylací je lokalizováno na útvarech majoritního výskytu  $\beta$ -listu.

LUNG/all				
CATH	Míst	%	Doména	Míst
1.10.510.10	28	12,56%	PF07714 - Pkinase Tyr	20
			PF00069 - Pkinase	8
1.10.220.10	10	4,48%	PF00191 - Annexin	7
			Neznáma	3
3.30.200.20	8	3,59%	PF00069 – Pkinase	4
			PF07714 - Pkinase Tyr	2
			Neznáma	2
3.30.505.10	7	3,14%	PF00017 - SH2	6
			Neznáma	1
2.30.29.30	6	2,69%	PF00169 - PH	5
			PF00568 - WH1	1
2.60.40.10	5	2,24%	PF13895 - Ig 2	3
			PF00630 – Filamin	1
			PF00041 - fn3	1
3.40.50.1240	4	1,79%	PF00300 - His Phos 1	4
3.20.20.70	3	1,35%	PF00274 – Glycolytic	1
			Neznáma	2
3.30.360.10	3	1,35%	PF02781 - G6PD C	3
3.40.50.300	3	1,35%	PF00071 - Ras	2
			PF00270 – DEAD	1
3.40.50.720	3	1,35%	PF13561 - adh short C2	2
			Neznáma	1

**Tabulka č.11:** CATH nadrodiny nejvíce reprezentované v datasetu LUNG/all. Procenta jsou odvozena z celku 223 fosforylačních míst nádorových proteinů identifikované sekundární struktury. V levém sloupci jsou CATH kódy, následuje počet fosforylačních míst strukturně spadajících do dané CATH nadrodiny a jeho procentuální vyjádření. Pravá část tabulky poskytuje rozdělení CATH nadrodin na jednotlivé přítomné domény a počty fosforylací jim náležejících.

LUNG/only				
CATH	Míst	%	Doména	Míst
1.10.510.10	17	9,29%	PF07714 - Pkinase Tyr	16
			PF00069 - Pkinase	1
1.10.220.10	9	4,92%	PF00191 - Annexin	7
			Neznáma	2
2.30.29.30	6	3,28%	PF00169 - PH	5
			PF00568 - WH1	1
3.30.505.10	5	2,73%	PF00017 - SH2	4
			Neznáma	1
2.60.40.10	4	2,19%	PF13895 - Ig 2	3
			PF00630 – Filamin	1
3.30.200.20	4	2,19%	PF07714 - Pkinase Tyr	2
			Neznáma	2
3.20.20.70	3	1,64%	PF00274 – Glycolytic	1
			Neznáma	2
3.30.360.10	3	1,64%	PF02781 - G6PD C	3
3.40.50.1240	3	1,64%	PF00300 - His Phos 1	3
3.40.50.300	3	1,64%	PF00071 – Ras	2
			PF00270 – DEAD	1
3.40.50.720	3	1,64%	PF13561 - adh short C2	2
			Neznáma	1

**Tabulka č. 12:** CATH nadrodiny nejvíce reprezentované v datasetu LUNG/only. Procenta jsou odvozena z celku 183 fosforylačních míst nádorových proteinů identifikované sekundární struktury. V levém sloupci jsou CATH kódy, následuje počet fosforylačních míst strukturně spadajících do dané CATH nadrodiny a jeho procentuální vyjádření. Pravá část tabulky poskytuje rozdělení CATH nadrodin na jednotlivé přítomné domény a počty fosforylací jim náležejících.

Při pomyslném přechodu z LUNG/all do zúženého LUNG/only si většina CATH nadrodin zachovává stejný či o jeden menší počet fosforylačních míst. Pouze 3.30.200.20 a 3.30.505.10 přijdou o dvě fosforylace, jediná 1.10.510.10 však ztratí výrazné množství, čítající 11 fosforylací, přesto zůstane nejvýznamněji reprezentovanou, což není překvapivé, neboť pokrývá kinázové domény, které jsou nejčastěji fosforylovanými doménami v našem datasetu. V LUNG/all, LUNG/only i HCT116/all jsou nejvíce fosforylovanými terciárními strukturami ortogonální svazek  $\alpha$ -helixů (začátek CATH kódu 1.10), dvou- a třívrstvé  $\alpha$   $\beta$  sendviče (CATH kódy začínají 3.30 a 3.40). Nejčastěji zastoupeným CATH foldem je 3.40.50. Zde se jedná o funkčně i sekvenčně nepříbuzné proteiny, které jsou však fosforylovány v mezích určitého intervalu srovnatelné frekvence, díky jejich podobné struktuře.

HCT116/all, datová série čítající 171 tyrosinových fosforylovatelných míst bílkovin kolorektálních nádorů, zahrnuje 54 pozic (31,58%), pro něž je známa sekundární struktura.



Skupina 28 míst (51,85%) z oněch 54 náleží kompletnímu počtu 9 CATH nadrodin. Šesti nadrodinám (66,67%) náleží každé jedna fosforylace. Podrobné zastoupení nadrodin se dvěma a více přiřazenými fosforylačními místy ukazuje tabulka č. 13.

Dataset HCT116/only, obsahující 79 tyrosinových fosforylačních míst, jež jsou souboru HCT116/all ve vztahu k LUNG/all specifické, zahrnuje malý počet 14 (17,72%) pozic proteinů identifikované sekundární struktury, mezi nimiž pouhých 5 (pro úplnost 35,71%) spadá do CATH kategorizace, přičemž se jedná jen o jedinou zde reprezentovanou CATH nadrodinu. V HCT116/only je 1.10.510.10 zastoupena oproti HCT116/all také výrazně nižším počtem členů, obdobně jako v případě srovnání LUNG/all, LUNG/only.

HCT116/all				
CATH	Míst	%	Doména	Míst
1.10.510.10	16	29,63%	PF07714 - Pkinase Tyr	8
			PF00069 - Pkinase	8
3.30.200.20	4	7,41%	PF00069 - Pkinase	3
			Neznáma	1
3.30.505.10	2	3,70%	PF00017 - SH2	2
HCT116/only				
CATH	Míst	%	Doména	Míst
1.10.510.10	5	35,71%	PF07714 - Pkinase Tyr	4
			PF00069 - Pkinase	1

*Tabulka č. 13: CATH nadrodiny nejvíce reprezentované v datasetech HCT116/all (vrchní část tabulky), HCT116/only (spodní část tabulky). Procenta HCT116/all jsou odvozena z celku 54 fosforylačních míst nádorových proteinů identifikované sekundární struktury, u HCT116/only z celku 14 nádorových proteinů identifikované sekundární struktury. V levém sloupci jsou CATH kódy, následuje počet fosforylačních míst strukturně spadajících do dané CATH nadrodiny a jeho procentuální vyjádření. Pravá část tabulky poskytuje rozdělení CATH nadrodin na jednotlivé přítomné domény a počty fosforylací jim náležejících.*

## 5.6. Expresí fosforylovaných proteinů v nádorech

Databáze Oncomine uchovává mnoho sérií výsledků expresních microarray analýz, které potvrzují, že v nádorech často dochází ke změnám exprese různých proteinů. Z dřívějších publikovaných prací na téma změny exprese proteinů v nádorových buňkách můžeme jmenovat například Ross et al. 2000. S ohledem na tyto skutečnosti jsem tedy provedl ověření změn exprese fosforylovaných proteinů analyzovaných v mé práci.

V LUNG/all, souboru 669 tyrosinfosforylačních pozic plicních nádorů dochází u 59,94% ke změně exprese proteinů, na nichž se nacházejí. Z toho největší množství fosforylačních míst je exprimováno méně než ve zdravých buňkách, nižší počet je nadexprimován a pro nejméně pozic jsou známy výsledky studií prokazující v plicních nádorech pod- i nadexpresi. V LUNG/only, sérii čítající 580 tyrosinfosforylačních míst plicních nádorů, které nemá rakovina plic společně s vybranou buněčnou linií rakoviny kolorekta souboru HCT116/all, jsou procentuální výsledky podobné LUNG/all. Ke změně exprese zde dochází u 61,38% míst, největší množství je podexprimováno, o mnoho nižší počet nadexprimován. Nejméně zastoupena jsou místa lokalizovaná na proteinech, pro něž byla různými studiemi v plicních nádorech zaznamenána pod- i nadexpresi. Názornější přehled výskytu fosforylačních míst LUNG/all a LUNG/only na proteinech se změnou expresní hladiny ukazuje tabulka č. 14.

Expresní hladina	LUNG/all		LUNG/only	
	Míst	%	Míst	%
Over	104	15,55%	89	15,34%
Under	242	36,17%	218	37,59%
Over/Under	55	8,22%	49	8,45%
//	268	40,06%	224	38,62%

**Tabulka č. 14:** Zastoupení tyrosinových fosforylačních míst se změnou exprese v datasetech LUNG/all, LUNG/only. Procenta LUNG/all jsou odvozena z celku 669 fosforylačních míst proteinů plicních nádorů, u LUNG/only z celku 580 fosforylačních míst proteinů plicních nádorů (která nemá LUNG/all společně s HCT116/all). V levém sloupci jsou jednotlivé varianty změny exprese. „Over“ vyjadřuje nadexpresi; „Under“ podexpresi; pro fosforylační místa ze skupiny „Over/Under“ byly v různých projektech zjištěny obě varianty; „//“ značí místa lokalizovaná na proteinech, u nichž nebyla změna exprese v plicních nádorech zaznamenána. Následuje vždy počet fosforylačních míst dané expresní varianty a jeho procentuální vyjádření pro každý dataset.

Dataset S1 obsahuje 144 tyrosinfosforylačních míst plicních nádorů. Výsledky jsou do určité míry podobné, jako u LUNG/all, LUNG/only. Výskyt na proteinech se změnou exprese vykazuje 63,19% fosforylačních míst. Největší množství míst je podexprimováno. Dále jsou zde rovněž místa lokalizovaná na proteinech v plicních nádorech nadexprimovaných nebo na proteinech, u nichž byla ve více výzkumných projektech detekována pod- i nadexpresi. Oba tyto případy vykazují v S1 podobné zastoupení. Přehled výskytu fosforylačních míst S1 na proteinech se změnou expresní hladiny ukazuje tabulka č. 15.

Expresní hladina	S1	
	Míst	%
Over	17	11,81%
Under	58	40,28%
Over/Under	16	11,11%
//	53	36,81%

**Tabulka č. 15:** Zastoupení tyrosinových fosforylačních míst se změnou exprese v datasetu S1. Procenta jsou odvozena z celku 144 fosforylačních míst proteinů plicních nádorů. V levém sloupci jsou jednotlivé varianty změny exprese. „Over“ vyjadřuje nadexpresi; „Under“ podexpresi; pro fosforylační místa ze skupiny „Over/Under“ byly v různých projektech zjištěny obě varianty; „//“ značí místa lokalizovaná na proteinech, u nichž nebyla změna exprese v plicních nádorech zaznamenána. Následuje počet fosforylačních míst dané expresní varianty a jeho procentuální vyjádření.

Ze 171 tyrosinových fosforylačních míst vybrané buněčné linie rakoviny kolorekta tvořících soubor HCT116/all se 48,81% míst nachází na proteinech, u nichž byla v kolorektálních tumorech zaznamenána změna exprese oproti normálnímu stavu. Nejvíce fosforylačních míst je nadexprimováno, mírně nižší množství podexprimováno. Nízké je zastoupení tyrosinových pozic, pro které různé zdroje uvádějí pod- i nadexpresi. V HCT116/only, sérii 79 fosforylačních míst, které nemá HCT116/all společné s LUNG/all, dochází u 46,84% míst ke změně exprese. Zdaleka nejvyšší je podíl míst nadexprimovaných, výrazně méně je míst podexprimovaných, velmi nízké zastoupení vykazují místa se záznamy o obou těchto variantách změny exprese. Přehled výskytu fosforylačních míst HCT116/all, HCT116/only na proteinech se změnou expresní hladiny ukazuje tabulka č. 16.

Expresní hladina	HCT116/all		HCT116/only	
	Míst	%	Míst	%
Over	39	23,21%	24	30,38%
Under	35	20,83%	11	13,92%
Over/Under	8	4,76%	2	2,53%
//	86	51,19%	42	53,16%

**Tabulka č. 16:** Zastoupení tyrosinových fosforylačních míst se změnou exprese v datasetech HCT116/all, HCT116/only. Procenta HCT116/all jsou odvozena z celku 171 fosforylačních míst proteinů kolorektálních nádorů, u HCT116/only z celku 79 fosforylačních míst proteinů kolorektálních nádorů (která nemá HCT116/all společná s LUNG/all). V levém sloupci jsou jednotlivé varianty změny exprese. „Over“ vyjadřuje nadexpresi; „Under“ podexpresi; pro fosforylační místa ze skupiny „Over/Under“ byly v různých projektech zjištěny obě varianty; „//“ značí místa lokalizovaná na proteinech, u nichž nebyla změna exprese v kolorektálních nádorech

*zaznamenána. Následuje vždy počet fosforylačních míst dané expresní varianty a jeho procentuální vyjádření pro každý dataset.*

V datasetech LUNG/all, LUNG/only, S1 je většina fosforylačních míst lokalizována na proteinech v buňkách rakoviny plic vykazujících změnu exprese ve srovnání s normálním stavem, nejvýraznější je podíl míst na proteinech podexprimovaných. V datasetech HCT116/all, HCT116/only dochází u velké části fosforylačních míst ke změně exprese v kolorektálních nádorech, převažují fosforylační místa nadexprimovaná. Celkově všechny analyzované datasety vykazují výraznou změnu exprese fosforylovaných proteinů. Z dostupných dat však nelze určit, zda jde o příčinou souvislost a fosforylace proteinu ovlivňuje expresi či změna exprese ovlivňuje fosforylaci nebo jde o dva zcela nezávislé jevy.

## **5.7. Charakteristika orthologů fosforylačních míst**

S využitím databáze PhosphoSitePlus, jedné z nejkomplexnějších současných databází posttranslačních modifikací, jsem provedl zmapování výskytu orthologů daných lidských tyrosinových pozic i jejich fosforylace u modelových organismů.

Z nabídky způsobů vyhledávání ve PhosphoSitePlus jsem využil možnost orientace na základě ID proteinu (sekce „Protein, Sequence, or Reference Search“), z následně zobrazeného seznamu organismů jsem zvolil člověka (human). Pro každý polypeptid je k dispozici sloupec s aminokyselinovými sekvencemi fosforylačních míst i sekvenčními čísly modifikovaných aminokyselin u člověka. Další sloupce obsahují stejné detaily fosforylovatelných pozic orthologů daného proteinu u některých modelových organismů, nejčastěji hlavně myši a krysy, v mnoha případech také dále králíka, prasete, kuřete nebo tura, zřídka eventuálně ještě psa, křečka, ovce. Pokud je aminokyselina u zvířete fosforylována, je v sekvenci zapsána malým písmenem, velká písmena značí aminokyseliny bez modifikace.

Konkrétní fosforylovatelné pozice jsem ve sloupcích vyhledával podle sekvenčních čísel a úseků aminokyselinových sekvencí přítomných ve výchozích datasetech. Získaná data jsem zapisoval do tabulky v programu Microsoft Office Excel 2007. Tyrosinfosforylační místo se buď u určitého organismu vůbec nevyskytuje (při práci značeno symbolem „x“) nebo je přítomno, ale není fosforylováno (symbol „O“) či fosforylováno je („p“). V datasetu

LUNG/all byly pro každý protein zaznamenány informace o všech uvedených orthologích (všech zmíněných organismů) jeho tyrosinfosforylačního místa (míst), soubory LUNG/only, HCT116/all, HCT116/only, S1 jsem zdokumentoval z hlediska výskytu i stavu fosforylace orthologů pouze u myši, krysy a stavu fosforylace v bílkovinách zdravé lidské tkáně. Výsledky byly pro snadnější vyvození závěrů sumarizovány u individuálních organismů v jednotlivých datasetech a procentuálně vyjádřeny. Sumarizace probíhala pomocí funkce COUNTIF do jejíhož vzorce byly postupně vepisovány souřadnice vymezující sčítané oblasti a také symboly, jejichž počet v ohraničeném poli jsem chtěl vyjádřit (x, O, p).

## 6. Závěr

Shromáždil jsem data charakterizující sekvenční a strukturní vlastnosti sérií tyrosinfosforylačních míst bílkovin plicních i kolorektálních nádorů, detaily expresních hladin, zastoupení proteinových domén i CATH nadrodin.

Analyzované soubory nevykazují vznik nových fosforylačních míst. Zdá se tedy, že alespoň v některých nádorech ke vzniku nových tyrosinfosforylačních míst nedochází. Zaznamenal jsem výraznou míru fosforylace  $\alpha$ -helixů a  $\beta$ -listů, oproti smyčkám, které jsou často fosforylovány dle dostupné literatury. Podařilo se ověřit, že v nádorech dochází ve velké míře ke změnám exprese fosforylovaných proteinů. Nejvíce fosforylovanými doménami jsou kinázové domény a Annexin, jimž odpovídají CATH nadrodiny 1.10.510.10 a 1.10.220.10. Velmi zajímavé je i to, že hned tři CATH nadrodiny spadají do jednoho foldu a to 3.40.50, což je velmi častý tzv. Rossmann fold.. Došlo ke zmapování výskytu orthologů tyrosinfosforylačních míst u modelových organismů myši a krysy. Hypotéza, že by mohla být možná identifikace patologických tyrosinových fosforylací na základě nepřítomnosti orthologa u modelových organismů, se ukázala jako nepoužitelná, pravděpodobně především proto, že fosforylační data jsou stále velmi neúplná.

Byl určen poměr fosfoserinu, fosfotreoninu, fosfotyrosinu u proteinů vybraných nádorových onemocnění. Všechny tumory mají výrazně zvýšený obsah fosfotyrosinu, kromě rakoviny prostaty, kde je abnormálně vysoké zastoupení fosfoserinu. Tyto výsledky jsou však pravděpodobně zavádějící, neboť řada datasetů, uložených v databázi PhosphoSitePlus, byla získána pouze za použití fosfotyrosinových protilátek a fosfoseriny či fosfotreoniny nemohly

proto vůbec detekovat. Bohužel se zdá, že v současné době není možné na dostupných datech srovnání zastoupení fosforylací provádět. Dostupná data jsou neúplná, nekonzistentní mezi jednotlivými datasey a navíc současné experimentální metody upřednostňují detekci fosfotyrosinu před ostatními fosforylacemi.

## 7. Seznam použité literatury

ARNESEN, Thomas, 2011. Towards a functional understanding of protein N-terminal acetylation. PLoS biology [online]. vol. 9, no. 5, s. e1001074, ISSN 1545-7885. Dostupne z: doi:10.1371/journal.pbio.1001074

Aviel-Ronen, S; Blackhall FH; Shepherd FA; Tsao MS (July 2006). "K-ras mutations in non-small-cell lung carcinoma: a review". Clinical Lung Cancer (Cancer Information Group) 8 (1): 30–38. doi:10.3816/CLC.2006.n.030. PMID 16870043

Ayoubi T. A. and Van De Ven W. J. (1996) Regulation of gene expression by alternative promoters. FASEB J. 10, 453-60.

Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res. 2007 Jan;35(Database issue):D301-3. Epub 2006 Nov 16. PMID:17142228

Boeckmann B, Blatter MC, Famiglietti L, Hinz U, Lane L, Roehert B, Bairoch A. Protein variety and functional diversity: Swiss-Prot annotation in its biological context. C R Biol. 2005 Oct-Nov;328(10-11):882-99. Epub 2005 Jul 28. Review. PMID:16286078

Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walte. Molecular biology of the cell. 2004 ISBN:9780815341055

Cancer. World Health Organization, February 2006

Cohen, P. (2002) The origins of protein phosphorylation, Nature Cell Biology, 4, E127-E130.

DARNELL, James E., 1997. Phosphotyrosine signaling and the single cell: metazoan boundary. Proceedings of the National Academy of ... [online]. vol. 94, no. October, s. 11767–11769,. Dostupne z: <http://www.pnas.org/content/94/22/11767.short>

Debald M, Jin JP, Linke A, Walgenbach KJ, Rauch P, Zellmer A, Fimmers R, Kuhn W, Hartmann G, Walgenbach-Brünagel G. Calponin-h2: a potential serum marker for the early detection of human breast cancer? Tumour Biol. 2014 Aug 7. [Epub ahead of print] PMID:25099617

DINKEL, Holger, Claudia CHICA, Allegra VIA, Cathryn M GOULD, Lars J JENSEN, Toby J GIBSON and Francesca DIELLA, 2011. Phospho.ELM: a database of phosphorylation sites--update 2011. Nucleic acids research [online]. vol. 39, no. Database issue, s. D261–7, ISSN 1362-4962. Dostupne z: doi:10.1093/nar/gkq1104

ESMENJAUD-MAILHAT, Charlotte, Valerie LOBJOIS, Carine FROMENT, Roy M GOLSTEYN, Bernard MONSARRAT and Bernard DUCOMMUN, 2007. Phosphorylation of

- CDC25C at S263 controls its intracellular localisation. FEBS letters [online]. vol. 581, no. 21, s. 3979–85, ISSN 0014-5793. Dostupne z: doi:10.1016/j.febslet.2007.07.024
- Gao, J.J., Thelen, J.J., Dunker, A.K., Xu, D. (2010) Musite, a Tool for Global Prediction of General and Kinasespecific Phosphorylation Sites, *Molecular & Cellular Proteomics*, 9, 2586-2600.
- GLOBOCAN 2000. Cancer incidence, mortality and prevalence worldwide  
<http://info.cancerresearchuk.org/cancerstats/types/bowel/incidence/> MUSIL, Dalibor. Populační screening kolorektálního karcinomu. *Interni Med.* 2003, čís. 3, s. 126-130.
- GNAD, Florian, Jeremy GUNAWARDENA and Matthias MANN, 2011. PHOSIDA 2011: the posttranslational modification database. *Nucleic acids research* [online]. vol. 39, no. Database issue, s. D253–60, ISSN 1362-4962. Dostupne z: doi:10.1093/nar/gkq1159
- Goldmann WH, Auernheimer V, Thievensen I, Fabry B. Vinculin, cell mechanics and tumour cell invasion. *Cell Biol Int.* 2013 May;37(5):397-405. doi: 10.1002/cbin.10064. Epub 2013 Mar 13. Review.PMID:23494917
- GONG, Y, T HIRANO, Y KATO, K YOSHIDA, Y SHOU, T OHIRA, N IKEDA, Y EBIHARA and H KATO, 2002. Phosphorylated tyrosine-containing proteins in primary lung cancer correlates with proliferation and prognosis. *British journal of cancer* [online]. vol. 86, no. 12, s. 1893–8, ISSN 0007-0920. Dostupne z: doi:10.1038/sj.bjc.6600327
- Gopalkrishnan RV, Su ZZ, Goldstein NI, Fisher PB. Translational infidelity and human cancer: role of the PTI-1 oncogene *Int J Biochem Cell Biol.* 1999 Jan;31(1):151-62. Review.PMID:10216950
- GUHA, Udayan, Raghothama CHAERKADY, Arivusudar MARIMUTHU, a Scott PATTERSON, Manoj K KASHYAP, H C HARSHA, Mitsuo SATO, Joel S BADER, Alex E LASH, John D MINNA, Akhilesh PANDEY and Harold E VARMUS, 2008. Comparisons of tyrosine phosphorylated proteins in cells expressing lung cancer-specific alleles of EGFR and KRAS. *Proceedings of the National Academy of Sciences of the United States of America* [online]. vol. 105, no. 37, s. 14112–7., ISSN 1091-6490. Dostupne z: doi:10.1073/pnas.0806158105
- Gutmanas A, Alhroub Y, Battle GM, Berrisford JM, Bochet E, Conroy MJ, Dana JM, Fernandez Montecelo MA, van Ginkel G, Gore SP, Haslam P, Hatherley R, Hendrickx PM, Hirshberg M, Lagerstedt I, Mir S, Mukhopadhyay A, Oldfield TJ, Patwardhan A, Rinaldi L, Sahni G, Sanz-García E, Sen S, Slowley RA, Velankar S, Wainwright ME, Kleywegt GJ. *PDBe: Protein Data Bank in Europe Nucleic Acids Res.* 2014 Jan;42(Database issue):D285-91. doi: 10.1093/nar/gkt1180. Epub 2013 Nov 27.PMID: 24288376
- Ham B. M. (2012) *Proteomics of biological systems : Protein phosphorylation using mass spectrometry techniques.* Hoboken, N.J.: John Wiley & Sons.
- HANKS, SK and T HUNTER, 1995. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *The FASEB Journal* [online]. Dostupne z: <http://www.fasebj.org/content/9/8/576.short>
- Hata Y, Hofmann K, Sudhof TC, Brose N (1995). "Mammalian homologues of *Caenorhabditis elegans unc-13* gene define novel family of C2-domain proteins". *J. Biol. Chem.* 270 (42): 25273–80. doi:10.1074/jbc.270.42.25273. PMID 7559667

- Herbst, RS; Heymach JV; Lippman SM (September 2008). "Lung cancer". *New England Journal of Medicine* 359 (13): 1367–1380. doi:10.1056/NEJMra0802714. PMID 18815398
- HORNBECK, Peter V, Jon M KORNHAUSER, Sasha TKACHEV, Bin ZHANG, Elzbieta SKRZYPEK, Beth MURRAY, Vaughan LATHAM and Michael SULLIVAN, 2012. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic acids research* [online]. vol. 40, no. Database issue, s. D261–70, ISSN 1362-4962. Dostupne z: doi:10.1093/nar/gkr1122
- [http://www.dolcera.com/wiki/images/thumb/400px-GSK3\\_phosphorylation.jpg](http://www.dolcera.com/wiki/images/thumb/400px-GSK3_phosphorylation.jpg)
- Hunter T. The age of crosstalk: phosphorylation, ubiquitination, and beyond. 2007 Dec 14;28(5):730-8. *Mol Cell Joensuu H, Roberts PJ, Sarlomo-Rikala M, Andersson LC, Tervahartiala P, Tuveson D, Silberman S, Capdeville R, Dimitrijevic S, Druker B, Demetri GD* (April 2001). "Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor". *N. Engl. J. Med.* 344 (14): 1052–6.
- HUNTER, T and G D PLOWMAN, 1997. The protein kinases of budding yeast: six score and more. *Trends in biochemical sciences* [online]. vol. 22, no. 1, s. 18–22, ISSN 0968-0004. Dostupne z: <http://www.ncbi.nlm.nih.gov/pubmed/9020587>
- HUNTER, Tony, 2009. Tyrosine phosphorylation: thirty years and counting. *Current opinion in cell biology* [online]. vol. 21, no. 2, s. 140–6, [vid. 14. November 2012]. ISSN 1879-0410. Dostupne z: doi:10.1016/j.ceb.2009.01.028
- HUNTER, Tony, 2012. Why nature chose phosphate to modify proteins. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* [online]. vol. 367, no. 1602, s. 2513–6, ISSN 1471-2970. Dostupne z: doi:10.1098/rstb.2012.0013
- International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature.* 431, 931-45.
- IZARZUGAZA, Jose M G, Oliver C REDFERN, Christine a ORENGO and Alfonso VALENCIA, 2009. Cancer-associated mutations are preferentially distributed in protein kinase functional sites. *Proteins* [online]. vol. 77, no. 4, s. 892–903, ISSN 1097-0134. Dostupne z: doi:10.1002/prot.22512
- JEMAL, Ahmedin, Freddie BRAY, Jacques FERLAY, Elizabeth WARD and David FORMAN, 2011. *Global Cancer*
- Jensen O. N. (2004) Modification-specific proteomics: Characterization of post-translational modifications by mass spectrometry. *Curr Opin Chem Biol.* 8, 33-41.
- Johannessen M1, Møller S, Hansen T, Moens U, Van Ghelue M. The multifunctional roles of the four-and-a-half-LIM only protein FHL2. *Cell Mol Life Sci.* 2006 Feb;63(3):268-84
- Johnson L. N. and Lewis R. J. (2001) Structural basis for control by phosphorylation. *Chem Rev.* 101, 2209-42.
- Kennelly P. J. (2001) Protein phosphatases--a phylogenetic perspective. *Chem Rev.* 101, 2291-312.



- KESHAVA PRASAD, T S, Renu GOEL, Kumaran KANDASAMY, Shivakumar  
KEERTHIKUMAR, Sameer KUMAR, Suresh MATHIVANAN, Deepthi TELIKICHERLA,  
Rajesh RAJU, Beema SHAFREEN, Abhilash VENUGOPAL, Lavanya BALAKRISHNAN,  
Arivusudar MARIMUTHU, Sutopa BANERJEE, Devi S SOMANATHAN, Aimy  
SEBASTIAN, Sandhya RANI, Somak RAY, C J HARRYS KISHORE, Sashi KANTH,  
Mukhtar AHMED, Manoj K KASHYAP, Riaz MOHMOOD, Y L RAMACHANDRA, V  
KRISHNA, B Abdul RAHIMAN, Sujatha MOHAN, Prathibha RANGANATHAN, Subhashri  
RAMABADRAN, Raghothama CHAERKADY and Akhilesh PANDEY, 2009. Human Protein  
Reference Database--2009 update. Nucleic acids research [online]. vol. 37, no. Database issue,  
s. D767–72, ISSN 1362-4962. Dostupne z: doi:10.1093/nar/gkn892
- Kim Y1, Ju H2, Kim DH3, Yoo HY4, Kim SJ5, Kim WS5, Ko YH6. CD79B and MYD88  
mutations in diffuse large B-cell lymphoma. *Hum Pathol.* 2014 Mar;45(3):556-64. doi:  
10.1016/j.humpath.2013.10.023. Epub 2013 Oct 31.
- KLUMPP, Susanne and Josef KRIEGLSTEIN, 2002. Phosphorylation and dephosphorylation of  
histidine residues in proteins. *European journal of biochemistry / FEBS* [online]. vol. 269, no. 4,  
s. 1067–71., ISSN 0014-2956. Dostupne z: <http://www.ncbi.nlm.nih.gov/pubmed/11856347>
- Koenig, M. and Grabe, N. (2004) Highly specific prediction of phosphorylation sites in proteins,  
*Bioinformatics*, 20, 3620-3627.
- Kondo Y1, Shen L, Issa JP. Critical role of histone methylation in tumor suppressor gene silencing  
in colorectal cancer. *Mol Cell Biol.* 2003 Jan;23(1):206-15.
- KOVAROVA, H, M HAJDUCH, M LIVINGSTONE, P DZUBAK and I LEFKOVITS, 2003.  
Analysis of state-specific phosphorylation of proteins by two-dimensional gel electrophoresis  
approach. *Journal of Chromatography B* [online]. vol. 787, no. 1, s. 53–61, ISSN 15700232.  
Dostupne z: doi:10.1016/S1570-0232(02)00729-8
- Kumar, V; Abbas AK; Aster JC (2013). "Chapter 5". *Robbins Basic Pathology* (9th ed.). Elsevier  
Saunders. p. 212. ISBN 978-1-4377-1781-5.
- Lauren Pecorino, *Molecular biology of cancer*, 2006
- Liu C., Li H. (2011) In silico predictions of post-translational modifications. *Methods Mol. Biology*  
Volume 760, 325-40.
- Lodish, H. et al.: *Molecular Cell Biology*, 5th edition. W. H. Freeman & Co. New York, 2004,  
ISBN 0-7167-4366-3
- LOWE, SW and AW LIN, 2000. Apoptosis in cancer. *Carcinogenesis* [online]. vol. 21, no. 3, s.  
485–495, ISSN 14602180. Dostupne z: doi:10.1093/carcin/21.3.485
- Lung Carcinoma: Tumors of the Lungs [online]. Merck Manual Professional Edition, Online  
edition
- Magrane M, Consortium U.Database (Oxford). UniProt Knowledgebase: a hub of integrated  
protein data.2011 Mar 29;2011:bar009. doi: 10.1093/database/bar009. Print  
2011.PMID:21447597
- Mann M. et al. (2002) Analysis of protein phosphorylation using mass spectrometry: Deciphering

the phosphoproteome. *Trends Biotechnol.* 20, 261-8.

MANNING, G, D B WHYTE, R MARTINEZ, T HUNTER and S SUDARSANAM, 2002. The protein kinase complement of the human genome. *Science (New York, N.Y.)* [online]. vol. 298, no. 5600, s. 1912–34, ISSN 1095-9203. Dostupne z: doi:10.1126/science.1075762

MOONEY, 2008. Gain and loss of phosphorylation sites in human cancer. *Bioinformatics (Oxford, England)* [online]. vol. 24, no. 16, s. i241–7, ISSN 1367-4811. Dostupne z: doi:10.1093/bioinformatics/btn267

MUSTELIN, Tomas, 2007. A brief introduction to the protein phosphatase families. *Methods in molecular biology (Clifton, N.J.)* [online]. vol. 365, s. 9–22., ISSN 1064-3745. Dostupne z: doi:10.1385/1-59745-267-X:9

Oh P, Li Y, Yu J, Durr E, Krasinska KM, Carver LA, Testa JE, Schnitzer JE.. Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy. *Nature* 2004 Jun 10;429(6992):629-35.PMID:15190345

Pawson T. and Nash P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science.* 300, 445-52.

Pawson T. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. 2004 *Cell.* Jan 23;116(2):191-203 Lu, C; Onn A, Vaporciyan AA et al. (2010). "78: Cancer of the Lung". *Holland-Frei Cancer Medicine (8th ed.)*. People's Medical Publishing House.

PETO, R, Lopez AD, Boreham J et al. Mortality from smoking in developed countries 1950–2000: Indirect estimates from National Vital Statistics. [s.l.] : Oxford University Press, 2006

PINNA, L a and M RUZZENE, 1996. How do protein kinases recognize their substrates? *Biochimica et biophysica acta* [online]. vol. 1314, no. 3, s. 191–225., ISSN 0006-3002. Dostupne z: <http://www.ncbi.nlm.nih.gov/pubmed/8982275>

RADIOJAC, Predrag, Peter H BAENZIGER, Maricel G KANN, Matthew E MORT, Matthew W HAHN and Sean D

REIMAND, Juri and Gary D BADER, 2013. Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers. *Molecular systems biology* [online]. vol. 9, no. 637, s. 637, ISSN 1744-4292. Dostupne z: doi:10.1038/msb.2012.68

Rhodes DR1, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D, Barrette T, Pandey A, Chinnaiyan AM. ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia.* 2004 Jan-Feb;6(1):1-6.

Rikova K1, Guo A, Zeng Q, Possemato A, Yu J, Haack H, Nardone J, Lee K, Reeves C, Li Y, Hu Y, Tan Z, Stokes M, Sullivan L, Mitchell J, Wetzell R, Macneill J, Ren JM, Yuan J, Bakalarski CE, Villen J, Kornhauser JM, Smith B, Li D, Zhou X, Gygi SP, Gu TL, Polakiewicz RD, Rush J, Comb MJ. Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell.* 2007 Dec 14;131(6):1190-203.

Ross D.T. et al. (2000). "Systematic variation in gene expression patterns in human cancer cell lines". *Nature Genetics* 24 (3): 227–235. doi:10.1038/73432. PMID 10700174.

- SACCO, Francesca, Livia PERFETTO, Luisa CASTAGNOLI and Gianni CESARENI, 2012. The human phosphatase interactome: An intricate family portrait. *FEBS letters* [online]. . S.l.: Federation of European Biochemical Societies, vol. 586, no. 17, s. 2732–9, ISSN 1873-3468. Dostupne z: doi:10.1016/j.febslet.2012.05.008
- Statistics. [online]. vol. 61, no. 2, s. 69–90., Dostupne z: doi:10.3322/caac.20107.Available
- Takagi K1, Ito S, Miyazaki T, Miki Y, Shibahara Y, Ishida T, Watanabe M, Inoue S, Sasano H, Suzuki T Amyloid precursor protein in human breast cancer: an androgen-induced gene associated with cell proliferation. *Cancer Sci.* 2013 Nov;104(11):1532-8. doi: 10.1111/cas.12239. Epub 2013 Aug 26.
- TALAVERA, David, Martin S TAYLOR and Janet M THORNTON, 2010. The (non)malignancy of cancerous amino acidic substitutions. *Proteins* [online]. vol. 78, no. 3, s. 518–29, ISSN 1097-0134. Dostupne z: doi:10.1002/prot.22574
- Tan CS, Pasculescu A, Lim WA, Pawson T, Bader GD, Linding R. *Science.* 2009 Sep 25;325(5948):1686-8. doi: 10.1126/science.1174301. Epub 2009 Jul 9. PMID:19589966
- TAYLOR, Susan S, Malik M KESHWANI, Jon M STEICHEN and Alexandr P KORNEV, 2012. Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* [online]. vol. 367, no. 1602, s. 2517–28,. ISSN 1471-2970. Dostupne z: doi:10.1098/rstb.2012.0054
- Thomason P., Kay R. (2000) Eukaryotic signal transduction via histidine-aspartate phosphorelay. *Journal of Cell Science* 113, 3141-3150.
- TORKAMANI, Ali, Gennady VERKHIVKER and Nicholas J SCHORK, 2009. Cancer driver mutations in protein kinase genes. *Cancer letters* [online]. . S.l.: Elsevier Ireland Ltd, vol. 281, no. 2, s. 117–27, ISSN 1872-7980. Dostupne z: doi:10.1016/j.canlet.2008.11.008
- TYANOVA, Stefka, Jurgen COX, Jesper OLSEN, Matthias MANN and Dmitrij FRISHMAN, 2013. Phosphorylation variation during the cell cycle scales with structural propensities of proteins. *PLoS computational biology* [online]. vol. 9, no. 1, s. e1002842, ISSN 1553-7358. Dostupne z: doi:10.1371/journal.pcbi.1002842
- Ulirsch J, Fan C, Knafel G, Wu MJ, Coleman B, Perou CM, Vimentin DNA methylation predicts survival in breast cancer. *Breast Cancer Res Treat.* 2013 Jan;137(2):383-96. doi: 10.1007/s10549-012-2353-5. Epub 2012 Dec 13. PMID:23239149
- UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2014 Jan;42(Database issue):D191-8. doi: 10.1093/nar/gkt1140. Epub 2013 Nov 18
- Vogelstein, B; Kinzler, KW (2004). "Cancer genes and the pathways they control". *Nature Medicine* 10 (8): 789–99. doi:10.1038/nm1087. PMID 1528678
- Votýpka, J., V. Varga, M. Varga, L. Zídková, 2003. Parazitismus (Biologická olympiáda 2003/2004, přípravný text pro kategorie A, B). Institut dětí a mládeže MŠMT ČR, Praha.
- Walsh C. (1979) *Enzymatic reaction mechanisms*. San Francisco: W. H. Freeman. xv, 978 p.
- Walsh C. (2006) *Posttranslational modification of proteins : Expanding nature's inventory*.

Englewood, Colo.: Roberts and Co. Publishers. xxi, 490 p.

- WITZE, Eric S, William M OLD, Katheryn A RESING and Natalie G AHN, 2007. Mapping protein post-translational modifications with mass spectrometry. *Nature methods* [online]. vol. 4, no. 10, s. 798–806, ISSN 1548-7091. Dostupne z: doi:10.1038/nmeth1100
- WU, Hsin-Yi, Vincent S TSENG, Lien-Chin CHEN, Hui-Yin CHANG, I-Chi CHUANG, Yeou-Guang TSAY and Pao-Chi LIAO, 2010. Identification of tyrosine-phosphorylated proteins associated with lung cancer metastasis using label-free quantitative analyses. *Journal of proteome research* [online]. vol. 9, no. 8, s. 4102–12., ISSN 1535-3907. Dostupne z: doi:10.1021/pr1006153
- WU, Chang-Jiun, Tianxi CAI, Klarisa RIKOVA, David MERBERG, Simon KASIF and Martin STEFFEN, 2009. A predictive phosphorylation signature of lung cancer. *PloS one* [online]. vol. 4, no. 11, s. e7994, ISSN 1932-6203. Dostupne z: doi:10.1371/journal.pone.0007994
- Xu A, Hao J, Zhang Z, Tian T, Jiang S, Hao J, Liu C, Huang L, Xiao X, He D. 14-kDa phosphohistidine phosphatase and its role in human lung cancer cell migration and invasion. *Lung Cancer*. 2010 Jan;67(1):48-56. doi: 10.1016/j.lungcan.2009.03.005. Epub .PMID:19344975
- Yaffe M. B. (2002) Phosphotyrosine-binding domains in signal transduction. *Nat Rev Mol Cell Biol*. 3, 177-86.
- Yang Z1, Fan Y, Deng Z, Wu B, Zheng Q Amyloid precursor protein as a potential marker of malignancy and prognosis in papillary thyroid carcinoma. *Oncol Lett*. 2012 Jun;3(6):1227-1230. Epub 2012 Mar 8.
- Yu JW, Mendrola JM, Audhya A, Singh S, Keleti D, DeWald DB, Murray D, Emr SD, Lemmon MA. Genome-wide analysis of membrane targeting by *S. cerevisiae* pleckstrin homology domains. *Mol Cell*. 2004 Mar 12;13(5):677-88
- ZHOU, H, J D WATTS and R AEBERSOLD, 2001. A systematic approach to the analysis of protein phosphorylation. *Nature biotechnology* [online]. vol. 19, no. 4, s. 375–8., ISSN 1087-0156. Dostupne z: doi:10.1038/86777