

Protokoll Versuch D2

Bioinformatik

Gruppe 8
Susanne Duncker und Friedrich Hahn

Identifizierung des zu untersuchenden Proteins

Unsere Gruppe untersuchte in diesem Versuch die Proteinsequenz GENVKIPVAIKEL. Um herauszufinden, welche Proteine sich dahinter verbergen könnten, führten wir eine BLAST-Suche¹ (Untermenü „search for short, nearly exact matches“, da diese Methode für kurze Sequenzen besser geeignet ist als BLASTP) unter Verwendung der SWISSPROT-Datenbank² durch.

Sequences producing significant alignments:			Score (Bits)	E Value	
gi 6175066 sp P04412 EGFR DROME	Epidermal growth factor recep...		43.5	6e-05	
gi 2811086 sp P00533 EGFR HUMAN	Epidermal growth factor recep...		39.7	8e-04	G
gi 1352359 sp Q01279 EGFR MOUSE	Epidermal growth factor receptor		39.7	8e-04	G
gi 125346 sp P00535 ERBB AVIER	Tyrosine-protein kinase transform		39.7	8e-04	
gi 6016434 sp P00534 ERBB ALV	Tyrosine-protein kinase transformi		39.7	8e-04	
gi 52782769 sp O18735 ERBB2 CANFA	Receptor tyrosine-protein k...		37.5	0.003	G
gi 119533 sp P04626 ERBB2 HUMAN	Receptor tyrosine-protein kin...		37.5	0.003	G
gi 3913591 sp Q60553 ERBB2 MESAU	Receptor tyrosine-protein ki...		37.5	0.003	

Abb. 1: die ersten acht Zeilen des von BLAST gezeigten Ergebnisses der Proteinsuche

Das Ergebnis (s. Abb. 1) legt nahe, dass es sich bei dem gesuchten Protein um den Epidermal growth factor receptor (EGFR) precursor aus *Drosophila melanogaster* handeln könnte, weswegen wir uns für unsere weitere Analyse auf dieses Protein festlegten und für die weiteren Analysen die bei NCBI³ hinterlegte DNS-Sequenz (s. S. 9) dieses Proteins verwendeten.

Informationen zum Protein

Da das von uns untersuchte Protein aus *Drosophila* stammt, informierten wir uns zunächst auf FLYBASE⁴ über die Funktion des EGFR. Wie dort nachzulesen ist, hat der auf dem Chromosom 2R liegende EGFR eine Fülle von Aufgaben, die uns am ehesten unter dem Stichwort „Entwicklung“ zusammenfassbar zu sein schienen (z.B. Bildung des anterior-posterior-Musters, Gehirnentwicklung, Blastodermsegmentierung uvm.). Des Weiteren fanden wir dort heraus, dass EGFR eine Transmembranrezeptor- sowie eine Protein-Tyrosin-Aktivität aufweist und ATP bindet.

Um zu vergleichen, wie verwandt die von uns in Abbildung 1 gefunden Proteine sind, führten wir mit CLUSTALW⁵ ein multiples Alignment der Sequenzen der EGFRs von Mensch, Maus und Fruchtfliege sowie der erbB2-Sequenzen von Mensch und Maus durch. Das Ergebnis (s. Abb. 2) zeigte, dass sich sowohl die erbB2-Sequenzen als auch die EGFR-Sequenzen von Mensch und Maus untereinander stark ähneln, der EGFR von *Drosophila* jedoch strukturell am weitesten von den vier anderen Sequenzen abweicht. Dies war zu erwarten, da sich Mensch und Maus evolutiv betrachtet als Säugetiere viel näher stehen als der Fliege als Arthropode. Dass sich jedoch die vertebreten Sequenzen von erbB2 und EGFR untereinander wohl näher stehen als alle drei EGFR-Sequenzen, war interessant.

¹ <http://www.ncbi.nlm.nih.gov/BLAST/>

² Auf der unter 1 genannten Homepage kann man zwischen mehreren Datenbanken auswählen – swissprot war die einzige von uns ausprobierte Datenbank, die gute Ergebnisse lieferte

³ <http://www.ncbi.nlm.nih.gov/>

⁴ <http://www.flybase.org>

⁵ <http://www.ebi.ac.uk/clustalw/>

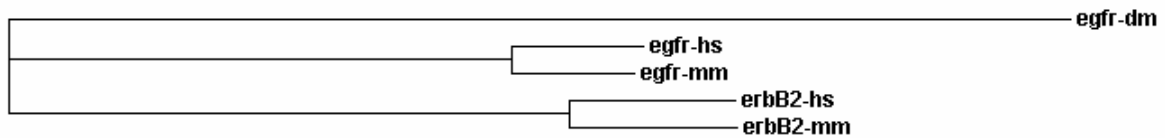


Abb. 2: multiples Alignment der oben beschriebenen Sequenzen

Sekundärstruktur

Wir verwendeten SIGNALP⁶, um uns vermutlich vorhandene Signalpeptidsequenzen anzeigen zu lassen, da es sich bei unserem Protein um einen Precursor, den Vorläufer eines reifen Proteins, handelt.

Das Ergebnis (s. Abb. 3) zeigt eine Signalpeptidsequenz (grün) für ein sekretorisches Protein an, die mit hoher Wahrscheinlichkeit (blau) zwischen der 32 und 33 Aminosäure des Vorläuferproteins abgespalten wird.

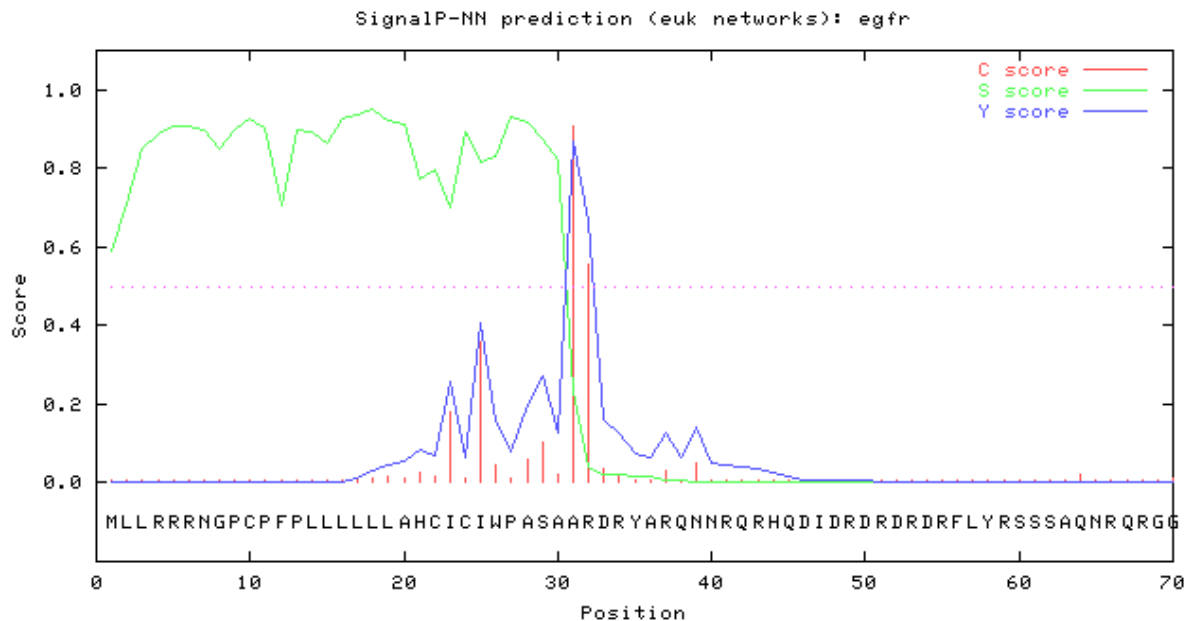


Abb. 3: Ergebnis der Signalpeptidanalyse von SIGNALP

Bei der Suche nach der Sequenz des reifen EGFR sahen wir uns einem unerwarteten Problem gegenüber: Vom EGFR aus *Drosophila melanogaster* gibt es zwei Isoformen, A und B (siehe Anhang 1, S. 10). *Drosophila simulans* hingegen hat nur eine Form, wie uns beim Betrachten der Ergebnisse unserer Suche nach „EGFR drosophila“ bei NCBI auffiel. Ein Vergleich der drei Sequenzen zeigte, dass der EGFR aus *Drosophila simulans* sowohl Isoform A als auch Isoform B des EGFR aus *Drosophila melanogaster* sehr ähnlich ist (siehe Alignment durch CLUSTALW⁷, Anhang 2, S. 11f), sogar ähnlicher als die Isoformen untereinander. Das lässt sich dadurch erklären, dass sich die beiden Isoformen an ihrem N-Terminus stark unterscheiden – möglicherweise handelt es sich hierbei um weitere Signalpeptid-, Rezeptorsequenzen oder ähnliches. Für unsere weiteren Analysen verwendeten wir den EGFR aus *Drosophila simulans*, da wir hierbei nur eine Sequenz verwenden mussten.

⁶ <http://www.cbs.dtu.dk/services/SignalP>

⁷ <http://www.ebi.ac.uk/clustalw/index.html>

Da wir unser Wissen, dass es sich bei EGFR um ein Transmembranprotein handelt, lediglich aus der NCBI-Datenbank hatten, gaben wir unsere Aminosäuresequenz zur Transmembrananalyse in PHOBIUS⁸ ein.

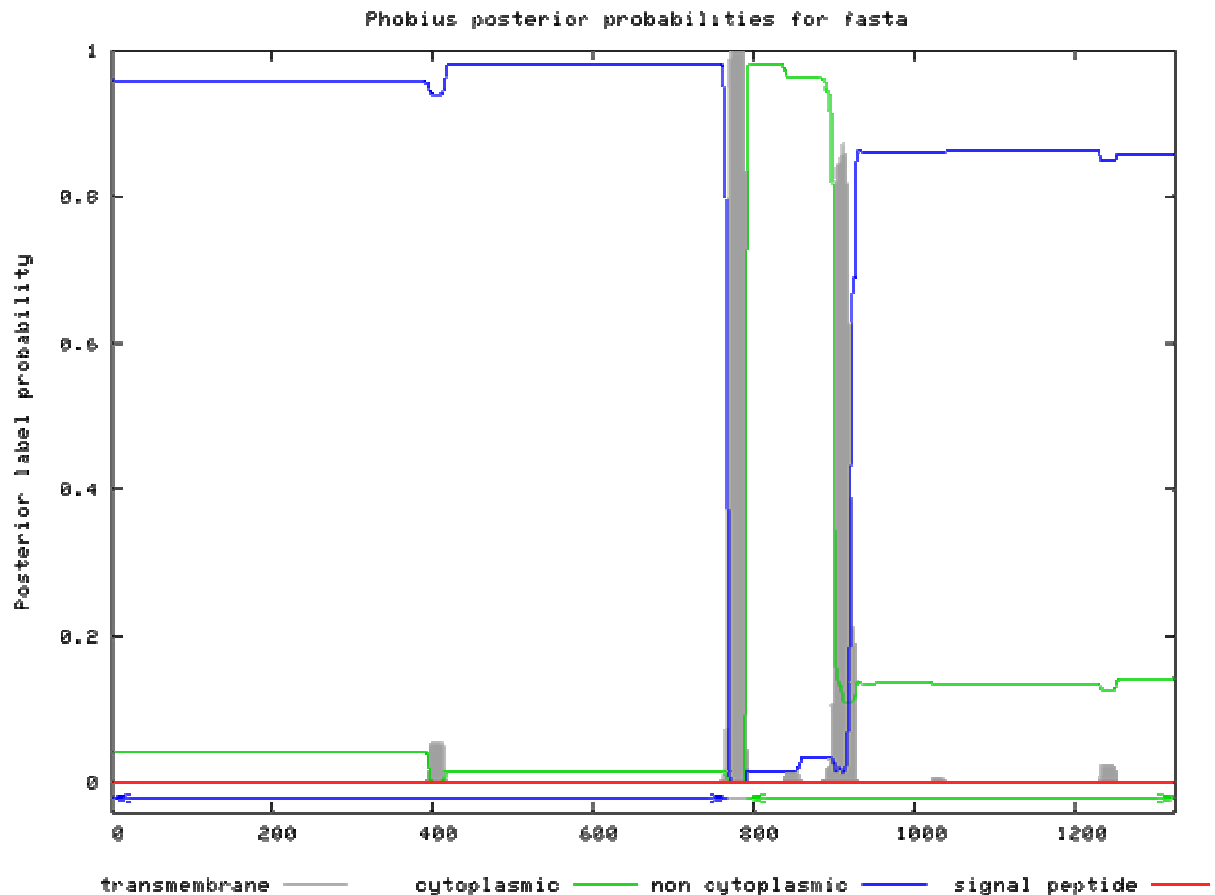


Abb. 4: Ergebnis der Transmembrandomänenanalyse mit Phobius

Das Ergebnis (s. Abb. 4) zeigt uns zwei mögliche Transmembrandomänen an, wovon jedoch nur eine (die erste) von PHOBIUS als gesichert erachtet und als zwischen Aminosäure (AS) 768 und 791 befindlich angezeigt wird.

Zur Sicherung der hier gezeigten Daten führten wir eine weitere Transmembrananalyse mit SOSUI⁹ durch.

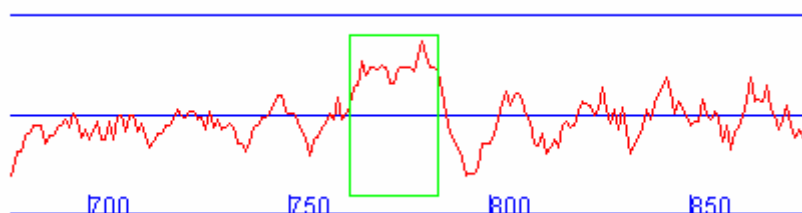


Abb. 5: Anzeige der Transmembrandomäne durch SOSUI

Auch SOSUI zeigt uns nur eine Transmembrandomäne an (s. Abb. 5), und zwar zwischen AS 765 und 787. Die beiden vorhergesagten Transmembrandomänen sind also ziemlich genau gleich lang (24 und 23 AS) und ungefähr an derselben Position (768 – 791 und 765 – 787 AS, d.h. 19 von 23 AS Überlappung).

⁸ <http://phobius.cgb.ki.se/>

⁹ http://sosui.proteome.bio.tuat.ac.jp/sosui_submit.html

Konservierte Domänen (s. Abb. 6a und 6b), die im EGFR enthalten sind, fanden wir, wie die Sequenz selbst, auf der NCBI-Homepage.

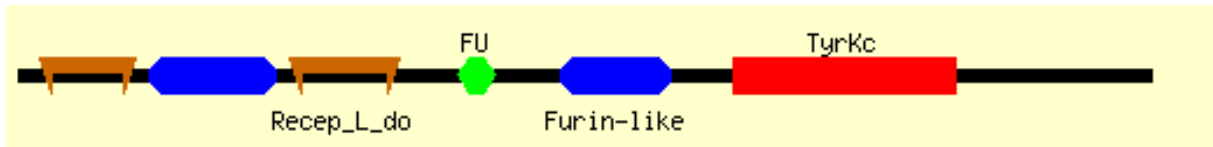


Abb. 6a: konservierte Domänen im EGFR

	smart00261	Furin-like repeats
	pfam00757	Furin-like cysteine rich region
	pfam01030	Receptor L domain
	cd00192	Tyrosine kinase, catalytic domain

Abb. 6b: Legende zu Abb. 6a

Um die Sekundärstrukturelemente des EGFR von *Drosophila simulans* vorherzusagen, verwendeten wir GOR IV¹⁰, welches uns folgende Sekundärstrukturelementanordnung (s. Abb. 7; rot = Faltblätter, blau = Helices) vorschlug:

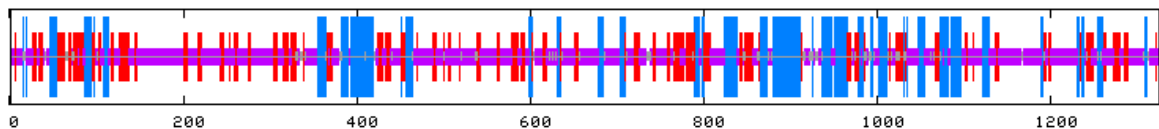


Abb.7: von GOR IV vorhergesagte Sekundärstrukturelemente des EGFR; blau: Helices; rot: Faltblätter

Man erkennt zwischen AS 800 und 1100 einen Bereich, der stark von Helices geprägt ist. Kurz vor AS 800, sprich dort, wo die Transmembrandomäne liegen sollte, befinden sich laut der Sekundärstrukturvorhersage Faltblattstrukturen.

Tertiärstruktur

Die Vorhersage der Tertiärstruktur ließen wir von SWISSMODEL¹¹ im „First Approach Mode“ durchführen. Dieses sandte uns eine Tertiärstruktur zu, die in Abbildung 8 zu sehen ist.

¹⁰ http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html

¹¹ <http://swissmodel.expasy.org/>



Abb. 8: von SwissModel vorgeschlagene Sekundärstruktur

Die vorgeschlagene Struktur wollten wir gerne mit der echten Struktur vergleichen. Leider ist der EGFR aus *Drosophila simulans* noch nicht gelöst und in die PROTEIN DATA BANK¹² (PDB) eingegeben, daher verwendeten wird die Struktur der Aminosäuren 1 – 501 des menschlichen EGFR im Komplex mit TGF α ., die unter der ID 1MOX bei der PDB gespeichert ist (s. Abb. 9).

¹² <http://www.rcsb.org/pdb/>

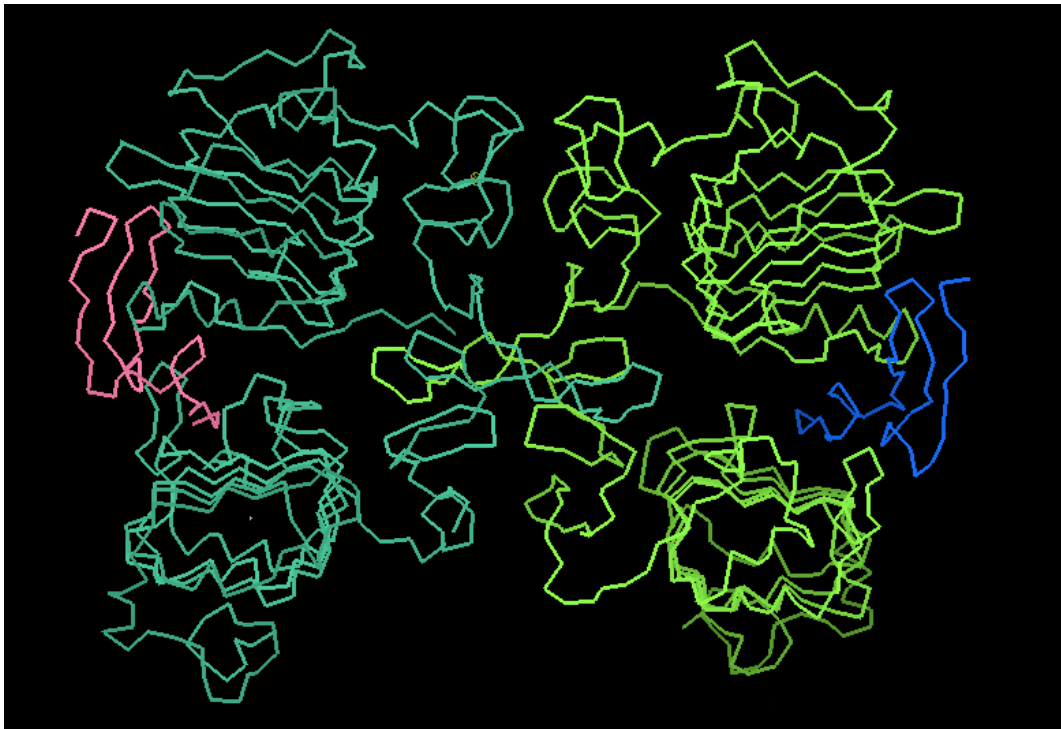


Abb. 9: die unter 1MOX gespeicherte Struktur des humanen EGFR

Zum Vergleich überlagerten wir die beiden Strukturen, wie in Abbildung 10 zu sehen ist (die Farbgebung wurde zwecks Übersichtlichkeit beibehalten).

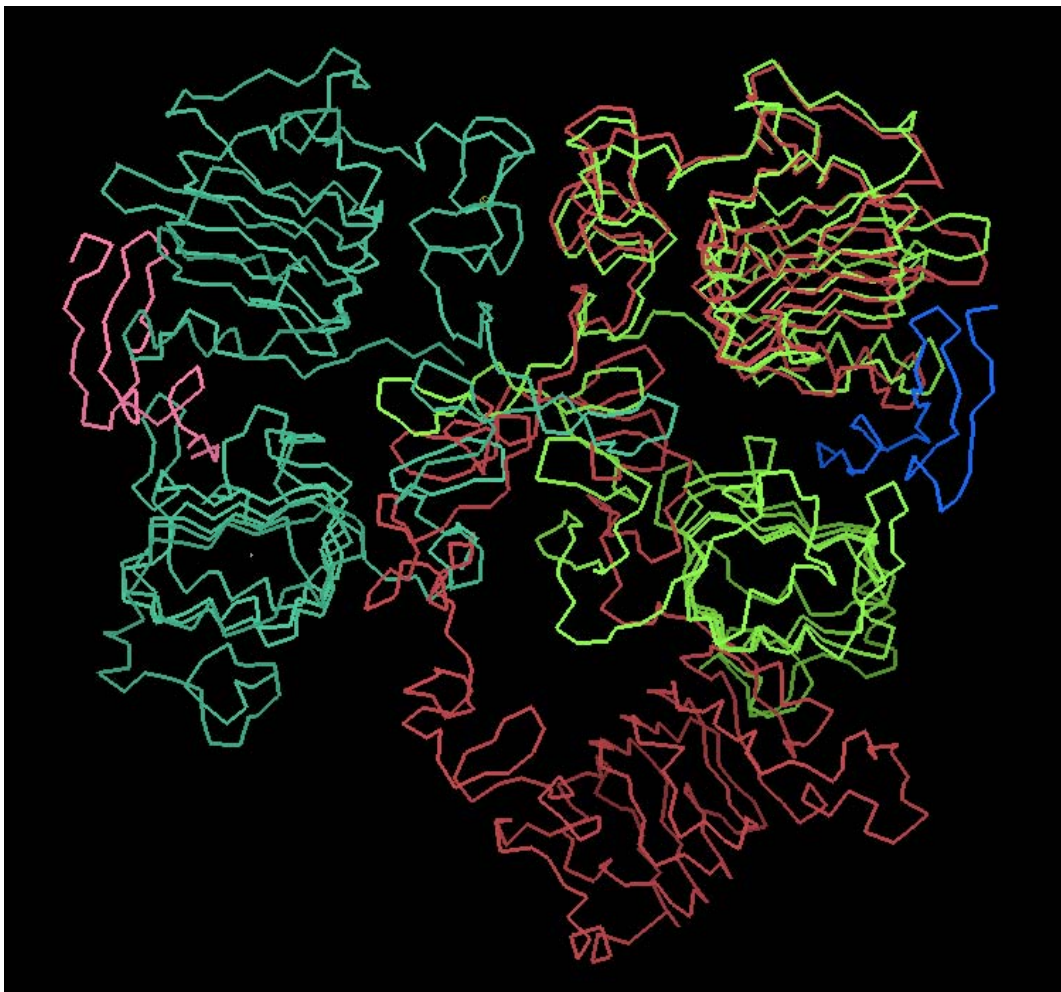


Abb. 10: Überlagerung der vorgeschlagenen mit der gelösten Struktur

Man sieht, dass die vorgeschlagene Struktur an und für sich gut passt, in der Mitte des Moleküls jedoch eine Drehung beschreibt, die verhindert, dass auch die zweite Hälfte des Moleküls mit der türkis eingefärbten Kette des humanen EGFR zur Deckung gebracht werden kann.

Fazit

Abschließend kann man sagen, dass die Strukturvorhersage recht gut geklappt hat, jedoch wurde besonders die Tertiärstruktur ja nicht de novo bestimmt, sondern aufgrund von Strukturvergleich (SWISSMODEL arbeitet auf dieser Basis). Die Übereinstimmung ist also nicht unbedingt überraschend.

Anhang 1: Aminosäuresequenzen

Aminosäuresequenz des EGFR-Precursors aus *Drosophila melanogaster*

1	mllrrrngpc	pfplllllla	hciciwpasa	ardryarqnn	rqrhqdidrd	rdrdrflyrs
61	ssaqrnrqrgg	anfalglgan	gvtiptsled	knknefvkkg	icigtksrls	vpsnkehhyr
121	nldr drytnct	yvdgnlklw	lpnenldlsf	ldnirevtgy	ilishvdkk	vvfpklqiir
181	grtlfslsve	eekyalftvy	skmytleipd	lrdvlnqvgv	fhnnynlchm	rtiqwseivs
241	ngtdaynyd	ftaperecpk	chescthgcv	gegpkncqkf	skltcspqca	ggrcygpkipr
301	ecchlfscagg	ctgptqkdc	acknffdeav	skeecppmrk	ynpttyvlet	npegkyayga
361	tcvkecpghl	lrdngacvrs	cpqdkmdkkg	ecvpcngpcp	ktcpgvtvlh	agnidsfrnc
421	tvidgniril	dqtfsgfqdv	yanytmgpry	ipldperrev	fstvkeitgy	lniegthpqf
481	rnlsyfrnle	tihrqlmes	mfaalaivks	slyslemrnl	kqissgsuvi	qhnrldcyvs
541	nirwpaiqke	pekvwvnen	lradlcekng	ticsdqcned	gcwgagtdqc	ltcknfnfng
601	tciadcgyis	naykfdnrtc	kichpecrtc	ngagadhqc	cvhvrddgqh	vsecpknkyn
661	drvgcrecha	tcdgctgpkd	tigigacttc	nlainndat	vkrcllkddk	cpdgyfweyv
721	hpqeqgslkp	lagravcrkc	hplcelctny	gyheqvcskc	thykrreque	tecpadhytd
781	eeqrecfqrh	pecngctgpg	addckscrnf	klfdanetgp	yvnstmfnc	skcplemrhv
841	nyqytaigpy	caaspprsk	itanldvnm	fiitgavlv	ticilcvvty	icrqkqkakk
901	etvkmmtals	gcedseplr	sniganlckl	rivkdaelrk	ggvlmgafg	rvykgvwvpe
961	genkipvai	kellkstgae	sseeflreay	imaseehvnl	klavcmss	qmmllitqlmp
1021	lgclldyvrn	nrdkigskal	lnwstqiakg	msyleekrlv	hrdlaarnvl	vqtpslvkit
1081	dflakllss	dsneykaagg	kmpikwale	cinnrvftsk	sdwafgvti	welltfgqrp
1141	henipakdip	dlievglkle	qpeicsldiy	ctllscwhld	aamrptfkql	ttvfaefard
1201	pgrylaipgd	kftrlpayts	qdekdlirkl	aptdgseai	akpddylqpk	aapgpshrtd
1261	ctdempklr	yckdpsnkns	stgdderdss	arevgvgnlr	ldlpvdeddy	lmpctqpgpn
1321	nnnnmnnpnq	nnmaavgvaa	gymdligvpv	svdnpeylln	aqlgvgesp	iptqtigipv
1381	mggpgtmevk	vpmpgsepts	sdheyndtq	relqplhrnr	ntetr	

Aminosäuresequenz des EGFR aus *Drosophila simulans*

1	cigtksrlsv	psnkehhyrn	ldr drytncty	vdgnleltwl	pnenldlsfl	dnirevtgyi
61	lishvdkkv	vfplqiirg	rtlfslsvee	ekyalftys	kmytleipdl	rdvlnqvgvf
121	hnnynlchmr	tiqwseivsn	gtdaynydf	taperecpkc	hescthgcv	egpkncqkfs
181	klctspqag	grcygpkipre	cchlfscagg	tgptqkdcia	cknffdegvc	keecppmrky
241	npttyvletn	pegkyaygat	cvkecpghll	rdngacvrsc	pqdkmdkge	cvpcngpcpk
301	tcpgvtvlha	gnidsfrnct	vidgnirild	qtfsgfqdv	anytmgpryi	pldperlevf
361	stvkeitgyl	niegthpqfr	nlsyfrnlet	ihgrqlmesm	faalaivkss	lyslemrnlk
421	qissgsvviq	hnrldcyvs	irwpaiqkep	eqkvwvnenl	radlcekngt	icsdqcnedg
481	cwagagtdqcl	tcknfnfngt	ciscdgcfin	aykfdnrtck	ichpecrtcn	gagadhqc
541	vhvrddgqhcv	secpknkynd	rgvcrechat	cdgctgpkdt	igigacttcn	laiinndatv
601	krcllkddkc	pdgyfweyv	pqeqgslkpl	agravcrkch	plcelctnyg	yheqvcskct
661	hykrreqcet	ecpadhytde	eqrecfqchp	ecngctgpga	ddckacrnfk	lfadanetgpy
721	vnstmfncs	kcplemrhvn	yqytaigpyc	aaspprski	tanldvnmif	iitgavlvpt
781	icilcvvtyi	crqkqkakke	tkvmtmaltg	cedseplrps	niganlcklr	ivkdaelrk
841	gvlmgafgr	vykgvwvpeg	envkipvaik	ellkstgaes	seeflreayi	masvehvnl
901	klavcmssq	mmllitqlmpl	gclldyvrnn	rdkigskall	nwstqiakgm	syleekrlvh
961	rdlaarnvlv	qtpslvkitd	fglakllssd	sneykaaggk	mpikwalec	innrvftsk
1021	dwafgvtiw	elltfgqrph	enipakdipd	lievglkleq	peicsldiyc	tllscwhlda
1081	amrptfkqlt	tvfaefardp	grylaipgd	fttrlpaytsq	dekdlirkl	pttdgseai
1141	epddylqpk	apgpshrtdc	tdeipklrny	ckdpsnkns	tgddetdssa	revgvgnlrl
1201	dlpvdeddy	mptcqpnpn	nnnmnnpnqn	nmaavgvaag	ymdligvpvs	vdnpeyllna
1261	qtlgvgespi	ptqtigipvm	gvpgtmevk	pmpgseptss	dheyndtqr	elqplhrnr
1321	tetr					

Aminosäuresequenz des **EGFR Isoform B**, *Drosophila melanogaster*

1	mllrrrngpc	pfplllllla	hciciwpasa	ardryarqnn	rqrhqdidrd	rdrdrflyrs
61	ssaqrnrqrgg	anfalglgan	gvtiptsled	knknefvkkg	icigtksrls	vpsnkehhyr
121	nldr drytnc	yvdgnleltw	lpnenldlsf	ldnirevtgy	ilishvdkk	vvfpklqiir
181	grtlfslsve	eekyalfvty	skmytleipd	lrdvlnqvgv	fhnnynlchm	rtiqwseivs
241	ngtdayyny	ftaperecpk	chescthgcv	gegpkncqkf	skltcspqca	ggrcygpkipr
301	ecchlfscag	ctgptqkdc	acknffdeg	ckeecppmrk	ynpttyvlet	npegkyayga
361	tcvkecpghl	lrdngacvrs	cpqdkmdkg	ecvpcngpcp	ktcpgvtvlh	agnidsfrnc
421	tvidgniril	dqtfsgfqdv	yanytmgpry	ipldperlev	fstvkeitgy	lniegthpqf
481	rnlsyfrnle	tihrgrqlmes	mfaalaivks	slyslemrnl	kqissgsuvi	qhnrldcyvs
541	nirwpaikqe	peqkvwnen	lradlcekng	ticsdqcned	gcwgagtdqc	ltcknfnfng
601	tciadcgyis	naykfdnrtc	kichpecrtc	ngagadhqc	cvhvrddgqh	vsecpknkyn
661	drgvcrecha	tcdgctgpkd	tigigacttc	nlainndat	vkrcllkddk	cpdgyfweyv
721	hpqeqgslkp	lagravcrkc	hplcelctny	gyheqvcskc	thykrreque	tecpadhytd
781	eeqrecfcch	pecngctgpg	addckscrnf	klfdanetgp	yvnstmfnc	skcplemrhv
841	nyqytaigpy	caaspprsk	itanldvnm	fiitgavlv	ticilcvvty	icrqkqkakk
901	etvkmmtals	gcedseplr	sniganlckl	rivkdaelrk	ggvlmgafg	rvykgvwppe
961	genvkipvai	kellkstgae	sseeflreay	imasvehvnl	lklavcmss	qmmllitqlmp
1021	lgclldyvrn	nrdkigskal	lnwstqiakg	msyleekrlv	hrdlaarnvl	vqtpslvkit
1081	dflgallkss	dsneykaagg	kmpikwlale	cirnrvtfsk	sdwafgvti	welltfgqrp
1141	henipakdip	dlievglkle	qpeicsldiy	ctllscwhld	aamrptfkql	ttvfaefard
1201	pgrylaipgd	kftrlpayts	qdekdlirkl	apttdgseai	aepddylqpk	aapgpshrtd
1261	ctdeipklr	yckdpsnkns	stgddetds	arevgvgnlr	ldlpvdeddy	lmpctqpgpn
1321	nnnninnpnq	nnmaavgvaa	gymdligvpv	svdnpeylln	aqlgvgesp	iptqtigipv
1381	mgvpgtmevk	vpmpgsepts	sdheyndtq	relqplhrnr	ntetr	

Aminosäuresequenz des **EGFR Isoform A**, *Drosophila melanogaster*

1	mmiismwmsi	srglwdsssi	wsvllilacm	asittsssvs	nagyvdngnm	kvcigtksrl
61	svpsnkehhy	rnldr drytnc	tyvdgnlelt	wlpnenldls	fldnirevtg	yilishvdkk
121	kvvfpklqii	rgrtlfsls	eeekyalfv	yskmytleip	dlrdvlnqgv	gfhnnynlch
181	mrtiqwseiv	sngtdayyny	dftaperecp	kchescthg	wgegpkncqk	fsklcspqc
241	agrcygpkip	recchlfscag	gctgptqkdc	iacknffdeg	vckeecppmr	kynpttyvle
301	tpegkyayg	atcvkecpgh	llrdngacvr	scpqdkmdkg	gecvpcngpc	pktcpgvtvl
361	hagnidsfrn	ctvidgniri	ldqtfsgfqd	vyanytmgpr	yipldperle	vfstvkeitg
421	yniegthpq	frnlsyfrnl	etihgrqlme	smfaalaivk	sslyslemrnl	lkqissgsuv
481	iqhnrldcyv	snirwpaikq	epeqkvwnen	nradlcekng	gticsdqcned	dgcwgagtdq
541	cltcknfnfn	gtciadcgyi	snaykfdnrt	ckichpecr	cngagadhqc	ecvhvrddgqh
601	cvsecpknky	ndrgvcrech	atcdgctgpk	dtigigactt	cnlainnda	tvkrcllkdd
661	kcpdgyfwey	vhpqeqgslk	plagravcrk	chplcelctn	gyheqvcskc	cthykrreqc
721	etecpadhyt	deeqrecfc	hpecngctgp	gaddckscrn	klfdanetgp	pyvnstmfnc
781	tskcplemrh	vnyqytaigp	ycaaspprsk	kitanldvnm	ifiitgavlv	pticilcvvt
841	yicrqkqkak	ketvkmmtal	sgcedseplr	psniganlck	lrivkdaelr	kggvlmgaf
901	grvykgvwp	egenvkipva	ikellkstga	esseeflrea	yimasvehvnl	llklavcms
961	sqmmlitqlm	plgclldyvr	nnrdkigska	llnwstqiak	gmsyleekrl	vhrdlaarnv
1021	lvqtpslvki	tdflgallks	sdsneykaag	gkmpikwlal	ecirnrvtfs	ksdwafgvt
1081	iwelltfgqr	phenipakdi	pdlievglkl	eqpeicsldi	yclllscwhl	daamrptfkq
1141	lttvfaefar	dpgrylaipg	dkftrlpayt	sqdekdlirk	lapttdgsea	iaepddylqp
1201	kaapgpshr	dctdeipklr	ryckdpsnk	sstgddetds	sarevgvgnl	rlldpvdedd
1261	ylmpctqpgp	nnnninnpn	qnnmaavgva	agymdligvp	svdnpeylln	naqlgvges
1321	piptqtigip	vmgvpgtmev	kvpmpgsept	ssdheyndt	qrelqplhrn	rntetr

Anhang 2: Vergleich des EGFR von *Drosophila simulans* mit den EGFR Typ A und B, durchgeführt von <http://www.ebi.ac.uk/clustalw/>

ClustalW Results

Scores Table

SeqA Name	Len(aa)	SeqB Name	Len(aa)	Score
1 egfr-ds	1325	2 egfr-dm-a	1377	99
1 egfr-ds	1325	3 egfr-dm-b	1426	99
2 egfr-dm-a	1377	3 egfr-dm-b	1426	96

Alignment

CLUSTAL W (1.82) multiple sequence alignment

```

egfr-ds      -----
egfr-dm-b    MLLRRRNGPCFPPLLLLLLAHCICIWPAASAARDRYARQNNRQRHQDIDRDRDRFLYRS 60
egfr-dm-a    -----MMIISMWMS-----ISRG-----LWDS 17

egfr-ds      -----CIGTKSRLSVSPSNKEHHYR 19
egfr-dm-b    SSAQNRQRGGANFALGLGANGVTIPTSLLEDKNKNEFVKGKICIGTKSRLSVSPSNKEHHYR 120
egfr-dm-a    SSIWS-----VLLILACMAS-ITTSSVSNAGYVDNGNMKVCIGTKSRLSVSPSNKEHHYR 71
                      *****

egfr-ds      NLRDRYTNCTYVDGNLELTWLPNENLDLSFLDNIREVTGYILISHVDVKKVVFVKLQIIR 79
egfr-dm-b    NLRDRYTNCTYVDGNLELTWLPNENLDLSFLDNIREVTGYILISHVDVKKVVFVKLQIIR 180
egfr-dm-a    NLRDRYTNCTYVDGNLELTWLPNENLDLSFLDNIREVTGYILISHVDVKKVVFVKLQIIR 131
                      *****

egfr-ds      GRTLFSLSVEEEKYALFVTYSKMYTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVS 139
egfr-dm-b    GRTLFSLSVEEEKYALFVTYSKMYTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVS 240
egfr-dm-a    GRTLFSLSVEEEKYALFVTYSKMYTLEIPDLRDVLNGQVGFHNNYNLCHMRTIQWSEIVS 191
                      *****

egfr-ds      NGTDAYYNYDFTAPERECPKCHESCTHGCWGEGPKNCQKFSKLTCSPPCAGGRCYGPCKPR 199
egfr-dm-b    NGTDAYYNYDFTAPERECPKCHESCTHGCWGEGPKNCQKFSKLTCSPPCAGGRCYGPCKPR 300
egfr-dm-a    NGTDAYYNYDFTAPERECPKCHESCTHGCWGEGPKNCQKFSKLTCSPPCAGGRCYGPCKPR 251
                      *****

egfr-ds      ECCHLFCAGGCTGPTQKDCIACKNFFDEGVCKEECPMRKYNPTTYVLETNPEGKYAYGA 259
egfr-dm-b    ECCHLFCAGGCTGPTQKDCIACKNFFDEGVCKEECPMRKYNPTTYVLETNPEGKYAYGA 360
egfr-dm-a    ECCHLFCAGGCTGPTQKDCIACKNFFDEGVCKEECPMRKYNPTTYVLETNPEGKYAYGA 311
                      *****

egfr-ds      TCVKECPGHLLRDNGACVRSCPQDKMDKGGEVPCNGPCPKTCPGVTVLHAGNIDSFRNC 319
egfr-dm-b    TCVKECPGHLLRDNGACVRSCPQDKMDKGGEVPCNGPCPKTCPGVTVLHAGNIDSFRNC 420
egfr-dm-a    TCVKECPGHLLRDNGACVRSCPQDKMDKGGEVPCNGPCPKTCPGVTVLHAGNIDSFRNC 371
                      *****

egfr-ds      TVIDGNIRILDQTFSGFQDVYANYTMGPRIPLDPERLEVSTVKEITGYLNIEGTHPQF 379
egfr-dm-b    TVIDGNIRILDQTFSGFQDVYANYTMGPRIPLDPERLEVSTVKEITGYLNIEGTHPQF 480
egfr-dm-a    TVIDGNIRILDQTFSGFQDVYANYTMGPRIPLDPERLEVSTVKEITGYLNIEGTHPQF 431
                      *****

egfr-ds      RNLSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRLCYVS 439
egfr-dm-b    RNLSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRLCYVS 540
egfr-dm-a    RNLSYFRNLETIHGRQLMESMFAALAIVKSSLYSLEMRNLKQISSGSVVIQHNRLCYVS 491
                      *****

egfr-ds      NIRWPAIQKEPEQKVWVNNENLRADLCEKNGTICSDQCNEGCGWAGTDQCLTCKNFNFNG 499

```

egfr-dm-b	NIRWPAIQKEPEQKVWVNNENLRADLCEKNGTICSDQCNEGDCWGAGTDQCLTCKNFNFG	600
egfr-dm-a	NIRWPAIQKEPEQKVWVNNENLRADLCEKNGTICSDQCNEGDCWGAGTDQCLTCKNFNFG	551
egfr-ds	TCISDCGFISNAYKFDNRTCKICHPECRTCNGAGADHCQECVHVRDGGHCVSECPKNKYN	559
egfr-dm-b	TCIADCGYISNAYKFDNRTCKICHPECRTCNGAGADHCQECVHVRDGGHCVSECPKNKYN	660
egfr-dm-a	TCIADCGYISNAYKFDNRTCKICHPECRTCNGAGADHCQECVHVRDGGHCVSECPKNKYN	611
egfr-ds	DRGVCRECHATCDGCTGPKDTIGIGACTTCNLAIINNDATVKRCLLKDDKCPDGYFWEYV	619
egfr-dm-b	DRGVCRECHATCDGCTGPKDTIGIGACTTCNLAIINNDATVKRCLLKDDKCPDGYFWEYV	720
egfr-dm-a	DRGVCRECHATCDGCTGPKDTIGIGACTTCNLAIINNDATVKRCLLKDDKCPDGYFWEYV	671
egfr-ds	HPQEQGSLKPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRREQCETECPADHYTD	679
egfr-dm-b	HPQEQGSLKPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRREQCETECPADHYTD	780
egfr-dm-a	HPQEQGSLKPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRREQCETECPADHYTD	731
egfr-ds	EEQRECFQCHPECNGCTGPGADDCKACRNFKLFDANETGPYVNSTMFNCTSKCPLMRHV	739
egfr-dm-b	EEQRECFQCHPECNGCTGPGADDCKSCRNFKLFDANETGPYVNSTMFNCTSKCPLMRHV	840
egfr-dm-a	EEQRECFQCHPECNGCTGPGADDCKSCRNFKLFDANETGPYVNSTMFNCTSKCPLMRHV	791
egfr-ds	NYQYTAIGPYCAASPPRSSKITANLDVNMIFIITGAVLVPTICILCVVTYICRQKQKAKK	799
egfr-dm-b	NYQYTAIGPYCAASPPRSSKITANLDVNMIFIITGAVLVPTICILCVVTYICRQKQKAKK	900
egfr-dm-a	NYQYTAIGPYCAASPPRSSKITANLDVNMIFIITGAVLVPTICILCVVTYICRQKQKAKK	851
egfr-ds	ETVKMTMALSGCEDSEPLRPSNIGANLCKLRIVKDAELRKGGVLGMGAFGRVYKGVWVPE	859
egfr-dm-b	ETVKMTMALSGCEDSEPLRPSNIGANLCKLRIVKDAELRKGGVLGMGAFGRVYKGVWVPE	960
egfr-dm-a	ETVKMTMALSGCEDSEPLRPSNIGANLCKLRIVKDAELRKGGVLGMGAFGRVYKGVWVPE	911
egfr-ds	GENVKIPVAIKELLKSTGAESSEEFLEAYIMASVEHVNLLKLLAVCMSSQMMLITQLMP	919
egfr-dm-b	GENVKIPVAIKELLKSTGAESSEEFLEAYIMASVEHVNLLKLLAVCMSSQMMLITQLMP	1020
egfr-dm-a	GENVKIPVAIKELLKSTGAESSEEFLEAYIMASVEHVNLLKLLAVCMSSQMMLITQLMP	971
egfr-ds	LGCLLDYVRNRRDKIGSKALLNWSTQIAKGMSYLEEKRLVHRDLAARNVLVQTPSLVKIT	979
egfr-dm-b	LGCLLDYVRNRRDKIGSKALLNWSTQIAKGMSYLEEKRLVHRDLAARNVLVQTPSLVKIT	1080
egfr-dm-a	LGCLLDYVRNRRDKIGSKALLNWSTQIAKGMSYLEEKRLVHRDLAARNVLVQTPSLVKIT	1031
egfr-ds	DFGLAKLLSSDSNEYKAAGGKMPIKWLALEICIRNRVFTSKSDVWAFGVTIWELLTFGQRP	1039
egfr-dm-b	DFGLAKLLSSDSNEYKAAGGKMPIKWLALEICIRNRVFTSKSDVWAFGVTIWELLTFGQRP	1140
egfr-dm-a	DFGLAKLLSSDSNEYKAAGGKMPIKWLALEICIRNRVFTSKSDVWAFGVTIWELLTFGQRP	1091
egfr-ds	HENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCWHLDAAMRPTFKQLTTVFAEFARD	1099
egfr-dm-b	HENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCWHLDAAMRPTFKQLTTVFAEFARD	1200
egfr-dm-a	HENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCWHLDAAMRPTFKQLTTVFAEFARD	1151
egfr-ds	PGRYLAIPGDKFTRLPAYTSQDEKDLIRKLAPTTDGSEAI ¹ VEPDDYLQPKAAPGPSHRTD	1159
egfr-dm-b	PGRYLAIPGDKFTRLPAYTSQDEKDLIRKLAPTTDGSEAI ¹ VEPDDYLQPKAAPGPSHRTD	1260
egfr-dm-a	PGRYLAIPGDKFTRLPAYTSQDEKDLIRKLAPTTDGSEAI ¹ VEPDDYLQPKAAPGPSHRTD	1211
egfr-ds	CTDEIPKLNRYCKDPSNKNSSGTGDETSSAREVGVGNLRLDLPVDEDDYLMPTCQPGPN	1219
egfr-dm-b	CTDEIPKLNRYCKDPSNKNSSGTGDETSSAREVGVGNLRLDLPVDEDDYLMPTCQPGPN	1320
egfr-dm-a	CTDEIPKLNRYCKDPSNKNSSGTGDETSSAREVGVGNLRLDLPVDEDDYLMPTCQPGPN	1271
egfr-ds	NNNN ¹ NNPNQNNMAAVGVAAGYMDLIGVPVSVDNPEYLLNAQTLGVGESPIPTQTIGIPV	1279
egfr-dm-b	NNNN ¹ NNPNQNNMAAVGVAAGYMDLIGVPVSVDNPEYLLNAQTLGVGESPIPTQTIGIPV	1380
egfr-dm-a	NNNN ¹ NNPNQNNMAAVGVAAGYMDLIGVPVSVDNPEYLLNAQTLGVGESPIPTQTIGIPV	1331
egfr-ds	MGVPGTMEVKVPMGPSEPTSSDHEYNDTQRELQPLHRNRNTETRV	1325

```
egfr-dm-b      MGVPGTMEVKVPMPGSEPTSSDHEYNDTQRELQPLHRNRNTETRV 1426
egfr-dm-a      MGVPGTMEVKVPMPGSEPTSSDHEYNDTQRELQPLHRNRNTETRV 1377
                *****
```

Sich unterscheidende Aminosäuren wurden orange hervorgehoben, in einigen Sequenzen fehlende Aminosäuren rot.