

Resistenz gegen HIV (Computer- und Internet-basierte Aufgabe)

In den frühen 1990er Jahren zeigten verschiedene Untersuchungen, dass einige Menschen trotz wiederholten Kontaktes mit dem HI-Virus nicht zu Virus-Trägern wurden oder der Krankheitsverlauf trotz einer nachweisbaren Infektion deutlich langsamer als üblich ausfielen (Verzögerungen von mehreren Jahren wurden beobachtet). Erste Erklärungsversuche dieses Phänomens gab es wenige Jahre später, als Wissenschaftler wichtige Co-Rezeptor-Moleküle auf der Oberfläche der Wirtszellen entdeckten, die für die Anheftung des HIV an die Zelle und die Infektion der Zelle wichtig sind. Die Forscher vermuteten, dass die resistenten Personen eine ungewöhnliche Form des Co-Rezeptor-Moleküls besitzen könnten, die dem Virus den Eintritt in die Wirtszelle unmöglich macht. Ein solcher Rezeptor ist der Co-Rezeptor CCR5, der an der Immunantwort des Wirtes beteiligt ist (Dean & O'Brien, 1998).

Um ihre Hypothese zu testen, sequenzierten die Forscher die Gene, die für den Co-Rezeptor CCR5 codieren. Sie untersuchten Proben von über 700 HIV-infizierten Patienten, und verglichen diese mit den Sequenzen von über 700 gesunden Personen. Die Ergebnisse der Sequenzierung zeigten, dass sowohl bei Proben von einigen infizierten Patienten, deren Krankheitsverlauf nur sehr langsam fortschritt, wie auch bei einzelnen Proben der gesunden Personen Mutationen im CCR5-Gen auftraten (Samson et al., 1996).

Aufgaben

1. Recherchieren Sie selbstständig auf der Internetseite des NCBI (www.ncbi.nlm.nih.gov) nach vollständigen Sequenzen des CCR5-Gens und Sequenzen des mutierten CCR5-Gens (siehe Kurzanleitung im Anhang).
 - a. Wählen Sie dazu bei SEARCH die Datenbank-Option ‚Nucleotide‘ und suchen Sie nach den Akzessionsnummern AY463215.1 und U66285.1!
 - b. Laden Sie die Sequenzen jeweils im FASTA-Format herunter und speichern sie Sie an geeigneter Stelle ab. (Rechts oben auf der Seite wird die Option ‚Download‘ angezeigt und durch Anwählen des Formats *FASTA* kann man die gewünschte Sequenz für die weitere Bearbeitung im FASTA-Format abspeichern werden).

2. Öffnen Sie das Programm GeneDoc und importiere die beiden ausgewählten Sequenzen. Versuchen Sie nun, die beiden Sequenzen miteinander abzugleichen, indem Sie nach Übereinstimmungen der Sequenzen suchen (Alignierung; siehe Kurzanleitung im Anhang). Stellen Sie fest,
 - a. wo die Veränderung liegt und
 - b. um welche Mutation es sich handelt.

3. Welche Konsequenzen hat eine solche Mutation für die anschließende Translation in der Proteinbiosynthese? Stellen Sie Vermutungen an!

4. Überprüfen Sie Ihre Vermutungen, indem Sie die Sequenzen wieder in ihren Ausgangszustand versetzt (keine Lücken o.ä.) und die DNA-Sequenz in eine Aminosäure-Sequenz übersetzen lassen (siehe Kurzanleitung im Anhang).
Worin unterscheiden sich die beiden resultierenden Proteine?

Option (bei ausreichend verbleibender Zeit oder anstelle von Aufgabe 1a):

5. Recherchieren Sie selbstständig auf der Internetseite des NCBI (www.ncbi.nlm.nih.gov) nach 2 (oder mehr) weiteren vollständigen Sequenzen („complete cds“) des CCR5-Gens und Sequenzen des mutierten CCR5-Gens, indem Sie geeignete Schlagworte für die Suche eingeben (z.B. „CCR5 chemokine receptor“). Verfahren Sie beim Download der Sequenzen wie in 1b angegeben und vergrößern Sie Ihr Alignment (siehe Kurzanleitung im Anhang).

Zusatzinformation:

Das Gen für den Co-Rezeptor CCR5 liegt auf dem dritten menschlichen Chromosomenpaar und es gibt zwei Allele für diesen Genort. Sind diese Allele gleich, ist der Träger homozygot für dieses Allel, bestehen jedoch Unterschiede, wird der Träger als heterozygot für dieses Allel bezeichnet. Ist ein Mensch homozygot für das „normale“ Allel CCR5 (CCR5/CCR5), besitzt er keine Resistenz gegen AIDS; ist er heterozygot, besitzt also ein normales Allel CCR5 und ein mutiertes Allel CCR5 Δ 32 (CCR5/CCR5 Δ 32), so wird der Krankheitsverlauf stark verzögert. Nur diejenigen, die homozygot für das mutierte Allel sind (CCR5 CCR5 Δ 32), sind tatsächlich resistent gegen eine Infektion mit HIV.

6. Erklären Sie, warum nur diejenigen Menschen, die homozygot für das mutierte Allel sind, resistent gegen eine HIV-Infektion sind, während die heterozygoten Träger der Mutation einen stark verzögerten Krankheitsverlauf aufweisen?

Kurzanleitung

GenBank® – die Suche nach DNA-Sequenzen

Um authentische Sequenzen zu vergleichen, kann man sich die Gendatenbanken im Internet zu Nutze machen. Die Sequenzdaten aller Gen-Datenbanken werden öffentlich verwaltet und stehen jedem zur Nutzung und Analyse über die www-Seiten der Organisationen zur Verfügung. Eine dieser Datenbanken, die DNA-Sequenzen sammelt, verwaltet und nutzbar macht, ist die GenBank® des NIH (National Institute of Health) in den USA, die vom NCBI, dem National Center for Biotechnology Information verwaltet wird (www.ncbi.nlm.nih.gov). Die Gendatenbanken sind das Resultat wissenschaftlicher Publikationen, denn die meisten Zeitschriften fordern, dass zeitgleich mit der schriftlichen Veröffentlichung von Forschungsarbeiten auch die Sequenzen in zentralen öffentlichen Datenbanken gespeichert werden müssen. Im traditionellen Zweig von GenBank® werden zurzeit über 106 Milliarden Basen in über 108 Millionen Sequenzen verwaltet (Stand August 2009).

Die Startseite des NCBI bietet einen einfachen Einstieg in die Suche nach Sequenzdaten. Ein Ausklappenmenü im Kopf der Seite erlaubt es, die Suche gezielt auf bestimmte Daten einzuschränken, so kann man z.B. gezielt nach DNA-Sequenzen suchen, indem man bei *Search: „Nucleotide“* anwählt. Da die Suche textbasiert abläuft, ist es möglich, als Suchbegriffe Artnamen, Gen-Bezeichnungen, Autoren, Akzessionsnummern o.ä. anzugeben, z.B. *Search „Nucleotide“ for „CCR5 mutant chemokine receptor“*.

Die Suche liefert dann eine Auflistung aller Einträge, die als Hyperlinks direkt zum vollständigen Datenbankeintrag weiterleiten. Unterhalb der Bezeichnung findet man die Akzessionsnummer, die der eindeutigen Identifizierung einer Sequenz dient und immer als Referenz angegeben werden sollte, da insbesondere Gen-Bezeichnungen keinem universellen Benennungsmuster unterliegen. Die Auflistung ändert sich ständig, da die Datenbank stetig aktualisiert und erweitert wird. Aus dieser Liste wählt man einen für die jeweilige Fragestellung geeigneten Eintrag aus. Im aktuellen Beispiel ‚Resistenz gegen HIV‘ wird nach Sequenzen im menschlichen Genom gesucht, daher könnte man z.B. *„Homo sapiens mutant CCR5 chemokine receptor (CCR5) gene, complete cds“* (Akzessionsnummer AY744141.1) anwählen. Ein solcher Datenbankeintrag enthält dann detaillierte Angaben zu der gesuchten Sequenz (z.B. Anzahl der Basen, Vollständigkeit der Sequenz, etc.), der biologischen Art des betreffenden Organismus, der zugehörigen Publikation, weiteren Literaturangaben sowie die abgeleitete Aminosäureübersetzung der DNA-Sequenz. Einige dieser Angaben sind zusätzlich mit Verlinkungen versehen, die zu weiteren relevanten Datenbanken (z.B. mit Informationen zur Systematik etc.) führen. Am Ende des Eintrags findet man die gesuchte DNA-Sequenz. Bei dieser Darstellung handelt es sich um das GenBank-Format, weitere Darstellungs-Optionen findet man unterhalb der fettgedruckten Gen-Bezeichnung: *FASTA* oder *Graphics*. Das GenBank-Dateiformat ist jedoch für den Transfer zwischen den Analyseprogrammen wenig geeignet. Besser gelingt dieses mit dem FASTA-Format (.aa), da dieses kompatibel mit zahlreichen Analyse-Programmen ist. Dies gilt für Nukleotid- wie auch Aminosäuresequenzen gleichermaßen. Rechts oben auf der Seite wird die Option *Send* angezeigt. Durch Anwählen dieses Links öffnet sich ein Menü, in dem empfohlen wird *„Complete Record“* und unter *„Choose Destination“*: *„File“* zu markieren. Daraufhin öffnet sich ein kleines Untermenü, in dem man das Datei-Format anwählen kann. Durch Auswahl des *Formats FASTA* kann man die gewünschte Sequenz für die weitere Bearbeitung in einem Analyseprogramm im FASTA-Format an einem geeigneten Ort abspeichern.

Kurzanleitung GeneDoc – PC-Programm zur Sequenz-Analyse

GeneDoc

Multiple Sequence Alignment Editor & Shading Utility Version 2.6.002

Copyright © 2000 by Karl Nicholas

Dieses Tutorial stellt erste grundlegende Schritte eines Sequenzalignments mit dem Programm GeneDoc dar. Es soll einen Einblick in die wissenschaftliche Arbeit bei Sequenzanalysen geben und erhebt keinen Anspruch auf Vollständigkeit.

Das Programm wird von den Herstellern zum kostenlosen Download angeboten (www.nrbsc.org/gfx/genedoc/).

Das Erstellen des Datensatzes

Import von Sequenzen:

Wählt man in der Menüleiste File (Datei) > Import (Importieren) an, öffnet sich ein kleines separates Fenster ‚Import Type‘ (Art des Imports), in dem das Dateiformat der Sequenz erfragt wird. I.d.R. passt die Voreinstellung und der Import der Sequenz kann durch Anklicken von ‚Import‘ veranlasst werden. (Das Dateiformat FASTA (Pearson) ist mit vielen Programmen kompatibel und zu Beginn wurde bereits empfohlen, Sequenzen aus GenBank® im FASTA-Format abzuspeichern (Dateiname.aa)). Die gewünschte Datei kann nun ausgewählt und geöffnet werden. Nach jedem Öffnen einer Sequenzdatei erscheint das ‚Import Type‘ Fenster erneut und es können weitere Sequenzen ergänzt werden. Ist der Datensatz vollständig, wird im Fenster ‚Import Type‘ die Option ‚Done‘ (Erledigt) gewählt.

Das Programm zeigt nun eine Auflistung der Sequenzen. Dabei erscheint links die Bezeichnung der jeweiligen Sequenz und nach einem Doppelpunkt wird die vollständige Sequenz in der Zeile angezeigt.

Möchte man ein bereits bestehendes Alignment zur weiteren Bearbeitung öffnen, wählt man das entsprechende Dateiformat an (dieses kann z.B. ebenfalls bereits als FASTA (Pearson), also Dateiname.aa vorliegen oder in anderen Formaten, wie z.B. im Clustal-Format als Dateiname.aln).

Bearbeitung der Sequenzen / des Alignments

Umbenennen der Sequenzen

Die etwas unpraktische Bezeichnung der Gene kann vom Nutzer geändert werden. Zu diesem Zweck wählt man in der Menü-Leiste ‚Project‘ > ‚Edit Sequence List‘ (Bearbeiten der Sequenz-Liste). Es erscheint ein Fenster, in dem alle Sequenzen des Datensatzes aufgelistet sind. Durch einen Doppelklick auf die gewünschte Sequenz erscheint ein neues Fenster mit dem aktuellen Namen der Sequenz. Dieser kann durch eine andere Bezeichnung ersetzt und mit ‚Ok‘ bestätigt werden (im Alignment angezeigt werden die ersten 10 Zeichen (Voreinstellung)).

Grad der Übereinstimmung der Sequenzen

Die Nukleotide sind mit unterschiedlichen Graustufen hinterlegt, die den Grad der Übereinstimmung zwischen den verschiedenen Sequenzen deutlich machen. Dabei ist es möglich, über Knöpfe (Buttons) in der Menü-Leiste unterschiedliche Level der Schattierung anzuwählen: 0, 2, 3 oder 4 Graustufen (Level 4 im Beispiel unten). Nach jeder Verschiebung der Sequenzen sollte die Schattierung durch Anklicken des Buttons aktualisiert werden.




540 * 560 * 580
TCTCATTTTCCATACAGTCAAGTATCAATTCTGGAAGAATTTCAGAC
TCTCATTTTCCATACATTAAGATAGTCATCTTGGGGCTGGTCTCTGC
TCTCATTTTCCATACA T A AT G T C C

Verschieben der Sequenzen

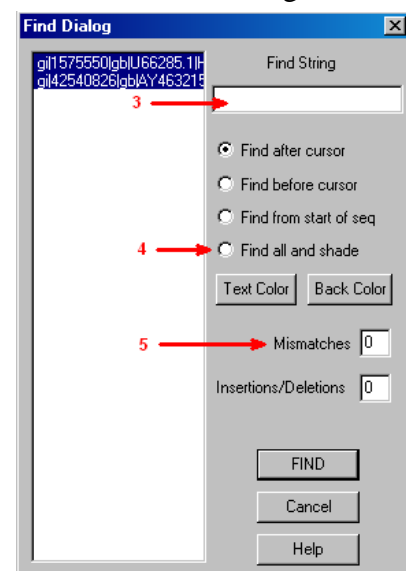
In der Menü-Leiste sind zwei Knöpfe (Buttons) zu finden, die die Verschiebung einzelner Nukleotide (Grab and Slide →1) bzw. ganzer Sequenzabschnitte (Grab and Drag →2) ermöglichen.



Suche nach Übereinstimmungen

Um die Sequenzen miteinander abzugleichen, ist es notwendig, nach Übereinstimmungen zu suchen. Dies kann manuell erfolgen oder aber mit Hilfe der Suchfunktion des Programms. Dazu wählt man in der Menü-Leiste ‚Edit‘ (Bearbeiten) > ‚Find‘ (Finden) Eine „Abkürzung“ besteht in dem Knopf (Button): . Es öffnet sich ein kleines Fenster ‚Find Dialog‘, in dem man in das Feld ‚Find String‘ (Finde Sequenz) (→3) eine Abfolge von Nukleotiden eingibt, die in einer der vorliegenden Sequenzen vorhanden ist. Es sollte die Option ‚Find all and Shade‘ (Finden und Schattieren) (→4) aktiviert werden (die Schattierungs-Farbe kann bei ‚Text Color‘ verändert werden). Durch Anklicken von ‚Find‘ durchsucht das Programm alle Sequenzen auf eine Übereinstimmung mit der angegebenen Nukleotid-Sequenz und markiert diese farbig.

620 * 640 * 660
GCTGCTTGTCATGGTCATCTGCTACTCGGGAATCCTAAAACT
TCCTAAAAACTCTGCTTCGGTGTGCGAAATGAGAAGAAGAGGCA
C C G T GA A A C



Dadurch wird die Alignierung der Sequenzen erheblich erleichtert. Zur Überprüfung der Übereinstimmungen kann immer wieder die Graustufen-Schattierung aller Sequenzen aktualisiert werden (farbige Markierungen gehen dann verloren). Werden keine Treffer angezeigt, kann man entweder nach neuen Abschnitten suchen oder im Fenster ‚Find Dialog‘ auch eine Anzahl von Fehltreffern (Mismatches oder Insertion/Deletion) (→5) zulassen.

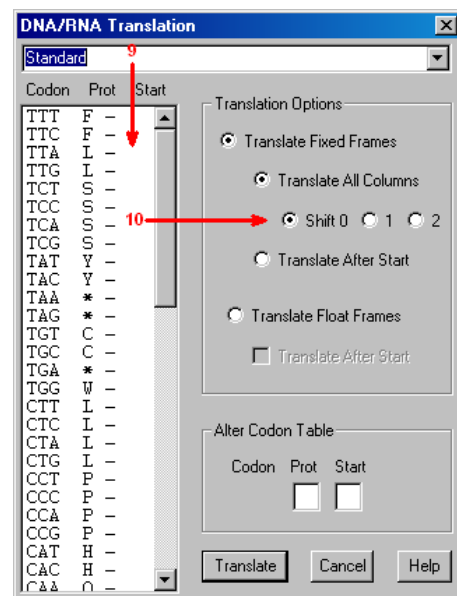
Speichern des Alignments

Ein Alignment kann für die weitere Bearbeitung in GeneDoc oder weitere Analysen in anderen Programmen gespeichert werden. In diesem Fall kann man entweder über ‚File’ > ‚Export’ die Datei im FASTA-Format oder beispielsweise im Clustal-Format speichern. (Zu dem gleichen Ergebnis kommt man durch Anwahl von ‚Project’ > ‚Edit Sequence List’ > ‚Export’).

Translation der Sequenzen (DNA-Sequenz in Aminosäure-Sequenz)

In einem Alignment können nicht nur die Nukleotide einer DNA-Sequenz miteinander verglichen werden, sondern auch die daraus resultierenden Aminosäure-Sequenzen. Diese Translation übernimmt das Programm. Zunächst muss man jedoch alle Lücken (gaps), die zuvor für ein erfolgreiches Alignment eingefügt wurden, wieder entfernt werden (sie dienen lediglich der Veranschaulichung von Insertionen / Deletionen).

Um nun die Translation durchzuführen wählt man ‚Project’ > ‚Trans DNA to Protein’. Es öffnet sich ein kleines Fenster ‚DNA/RNA Translation’, in dem die Codierung der Aminosäuren aufgelistet ist (→9). An dieser Stelle hat man die Möglichkeit, eine Verschiebung des Leserasters („Shift”) um 1 oder 2 Nukleotide festzulegen, falls z.B. die Sequenz nicht mit einem vollständigen Triplet beginnt (→10) (u.U. müssen mehrere Versionen ausgetestet werden). Die Aminosäure-Sequenz erscheint in einem neuen Fenster, so dass durch Schließen oder Verkleinern dieses Fensters wieder die DNA-Sequenz sichtbar wird.



Lösungsvorschläge

2a.

Ab Base 554 fehlen in der mutierten Sequenz 32 Basen.

Ausschnitt aus GeneDoc:

```

standard : * 540 * 560 * 580 * 600 * 620 *
          ACCTGCAGCTCTCATTTCATACAGTCAAGTATCAATTCTGGAAGAATTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCTGCGCTGCTTGTTCATGGT
mutiert  : ACCTGCAGCTCTCATTTCATACAC-----TTAAAGATAGTCATCTTGGGGCTGGTCTGCGCTGCTTGTTCATGGT
          ACCTGCAGCTCTCATTTCATACAC-----TTAAAGATAGTCATCTTGGGGCTGGTCTGCGCTGCTTGTTCATGGT
  
```

2b. Deletion

3.

- Leserasterverschiebung und damit andere Aminosäuresequenz (Primärstruktur) und dementsprechend andere Sekundär- und Tertiärstruktur des Proteins
- Leserasterverschiebung und damit verfrühtes Entstehen eines Stopp-Codons, so dass das Protein deutlich verkürzt wird.
- keine Leserasterverschiebung, Anzahl der deletierten Basen ein Vielfaches von 3.

4.

Durch die Mutation entsteht eine Leserasterverschiebung. Die Aminosäure-Sequenz wird dadurch verändert und es kommt zu einem verfrühten Entstehen eines Stopp-Codons. Das Protein wird deutlich kürzer und damit fehlen dem CCR5-Rezeptor 3 Transmembranabschnitte, die für die Anheftung des Virus notwendig wären (siehe „Illustration der Veränderung“).

Standard-Allel (Sequenzausschnitt von Aminosäure 170 bis 216): ... Gln, Lys, Glu, Gly, Leu, His, Tyr, Thr, Cys, Ser, Ser, His, Phe, Pro, Tyr, Ser, Gln, Tyr, Gln, Phe, Trp, Lys, Asn, Phe, Gln, Thr, Leu, Lys, Ile, Val, Ile, Leu, Gly, Leu, Val, Leu, Pro, Leu, Leu, Val, Met, Val, Ile, Cys, Tyr, Ser, Gly, ...

Mutierte Allel (Sequenzausschnitt von Aminosäure 170 bis 216): ... Gln, Lys, Glu, Gly, Leu, His, Tyr, Thr, Cys, Ser, Ser, His, Phe, Pro, Tyr, Ile, Lys, Asp, Ser, His, Leu, Gly, Ala, Gly, Pro, Ala, Ala, Cys, His, Gly, His, Leu, Leu, Leu, Gly, Asn, Pro, Lys, Asn, Ser, Ala, Ser, Val, Ser, Lys, **STOP**

Ausschnitt aus GeneDoc:

```

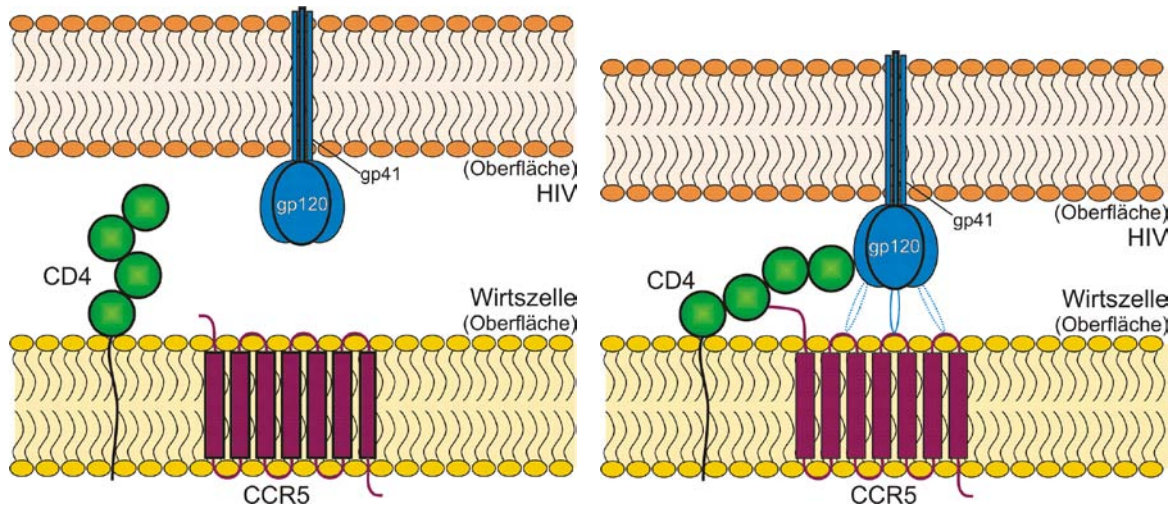
          140 * 160 * 180 * 200 * 220 *
standard : ALKARTVTFGVVTSVITWVAVFASLPGIIFTRSQEGLHYTCSSHPFYISQYQFWKNFQTLKIVILGLVLPFLVMVICYSILKTLRLCRNEKKRHRRAVR
mutiert  : ALKARTVTFGVVTSVITWVAVFASLPGIIFTRSQEGLHYTCSSHPFYIKDSHLGAGPAAACHGHLLENPKNSASVSK*EEAOGCEAYLHHHDCLES
          ALKARTVTFGVVTSVITWVAVFASLPGIIFTRSQEGLHYTCSSHPFY L6
  
```

6.

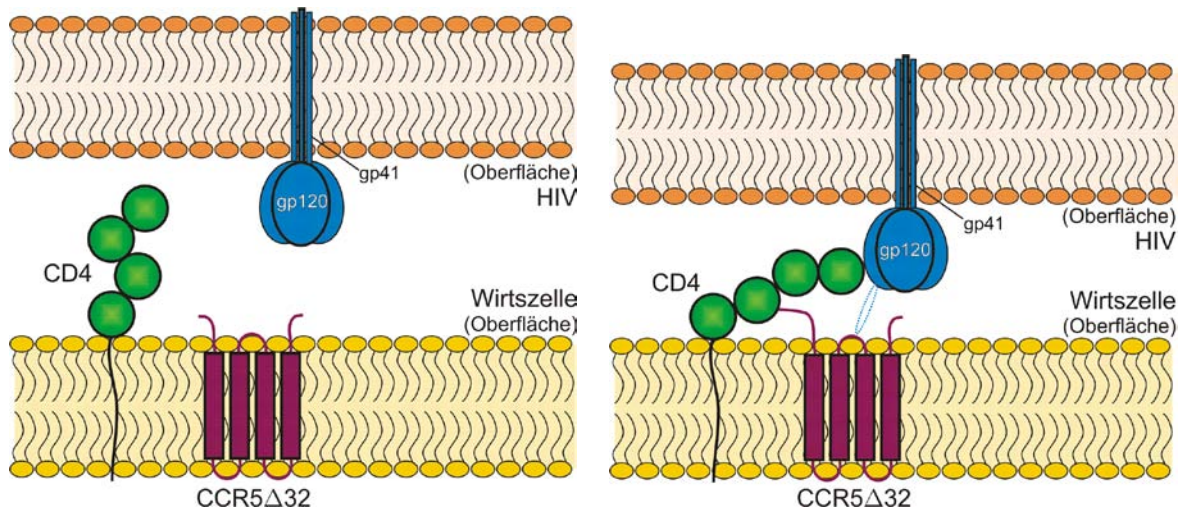
Bei homozygoten Trägern der Mutation wird der Rezeptor ausschließlich in stark abgewandelter Form exprimiert, HIV kann die Zelle nicht infizieren. Bei heterozygoten Trägern der Mutation wird der funktionsfähige Rezeptor in stark verringertem Maß exprimiert und somit schreitet die HIV-Infektion deutlich verlangsamt voran.

Illustration der Veränderung:

Intakter Co-Rezeptor CCR5 und Anheftung des HIV



Mutierter Co-Rezeptor CCR5 Δ 32 und Anheftung des HIV



Durch den vorzeitigen Abbruch der Proteinbiosynthese fehlen dem mutierten Protein die letzten drei Transmembran-Segmente. Dadurch kann eine Anheftung des HI-Virus nicht mehr vollständig erfolgen, eine Membranverschmelzung wird nicht mehr initiiert.

Literatur & links:

Dean, M. and S. O'Brien (1998). "Die Suche nach Resistenz-Genen gegen AIDS." Spektrum der Wissenschaft 2: 38-45.

Samson, M., F. Libert, et al. (1996). "Resistance to HIV-1 infections in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene." Nature **382**: 722-725.

HIV: the ultimate evolver (www.evolution.berkeley.edu/evolibrary/article/_0_0/medicine_04) (letzter Zugriff: Juli 2009)

Ghost of epidemic past (www.evolution.berkeley.edu/evolibrary/news/081001_hivmalaria) (letzter Zugriff: Juli 2009)

HIV immunity (www.pbs.org/wgbh/evolution/library/10/4/1_104_06.html) (letzter Zugriff: Juli 2009)

NCBI: www.ncbi.nlm.nih.gov (letzter Zugriff: Juli 2010)

GeneDoc: www.nrbsc.org/gfx/genedoc/ (letzter Zugriff: Juli 2010)