

Information (Lehrer):

## Kleiner Führerschein für Datenbanken und ein Programm zur 3D-Visualisierung

Mit der **Bioinformatik** hat sich ein eigener interdisziplinärer Wissenschaftszweig entwickelt, der Sequenzdaten für Nukleinsäuren und Proteine von Organismen sammelt und organisiert, 3D-Strukturdaten verschiedenster makromolekularer Moleküle analysiert und Visualisierungsprogramme für Biomoleküle und biologische Prozesse bereitstellt. Mittlerweile wurden weltweit Abermillionen solcher Datensätze generiert. Um diese Datenflut zu verwalten und auch allen Mitgliedern (vor allem) der Wissenschaftsgemeinschaft zugänglich zu machen, existieren Datenbanken; die umfangreichste befindet sich an den National Institutes of Health (NIH) in den USA (<http://www.ncbi.nlm.nih.gov/sites/gquery>). Diese Art „GOOGLE“-Suchmaschine für Wissenschaftler findet und vergleicht nicht nur Sequenzdaten (grüne Markierungen), sondern ermöglicht mit „PubMed“ auch die Suche nach Originalveröffentlichungen aus der biomedizinischen Forschung (rote Markierung).

The screenshot shows the NCBI Entrez search engine interface. At the top, there is a search bar with the text "Search across databases" and buttons for "GO", "Clear", and "Help". Below the search bar, the text "Welcome to the Entrez cross-database search page" is displayed. The main content area is a grid of database links, each with an icon and a description. A red circle highlights the "PubMed" link, and three green circles highlight the "Protein", "Genome", and "Structure" links.

<b>PubMed:</b> biomedical literature citations and abstracts	<b>Books:</b> online books
<b>PubMed Central:</b> free, full text journal articles	<b>OMIM:</b> online Mendelian Inheritance in Man
<b>Site Search:</b> NCBI web and FTP sites	<b>OMIA:</b> online Mendelian Inheritance in Animals
<b>Nucleotide:</b> Core subset of nucleotide sequence records	<b>dbGaP:</b> genotype and phenotype
<b>EST:</b> Expressed Sequence Tag records	<b>UniGene:</b> gene-oriented clusters of transcript sequences
<b>GSS:</b> Genome Survey Sequence records	<b>CDD:</b> conserved protein domain database
<b>Protein:</b> sequence database	<b>3D Domains:</b> domains from Entrez Structure
<b>Genome:</b> whole genome sequences	<b>UniSTS:</b> markers and mapping data
<b>Structure:</b> three-dimensional macromolecular structures	<b>PopSet:</b> population study data sets
<b>Taxonomy:</b> organisms in GenBank	<b>GEO Profiles:</b> expression and molecular

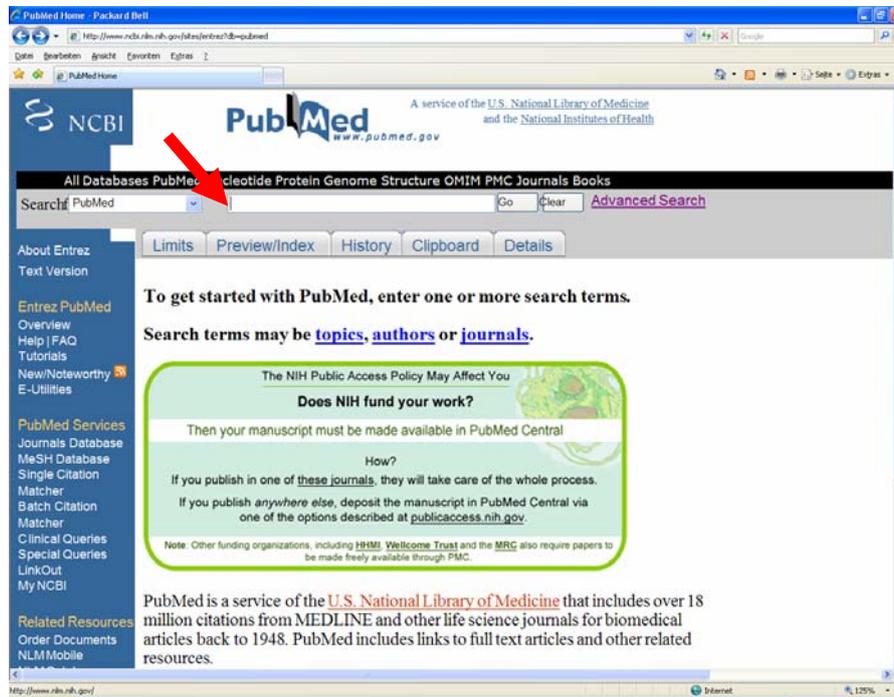
### Inhalt:

- (A) Kurzanleitung zur Suche von wissenschaftlichen Originalpublikationen
- (B) Kurzanleitung zur Suche von Informationen über Proteine
- (C) Kurzanleitung zur Nutzung des Visualisierungsprogramms PyMOL

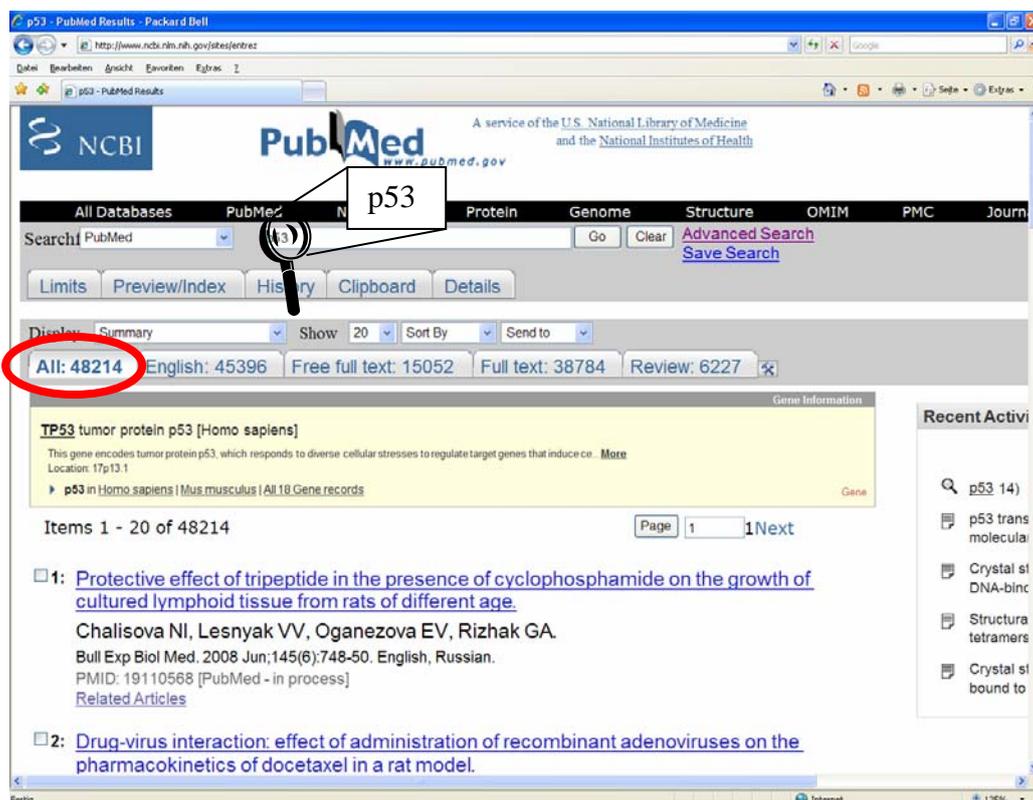


## (A) Kurzanleitung zur Suche von wissenschaftlichen Originalpublikationen:

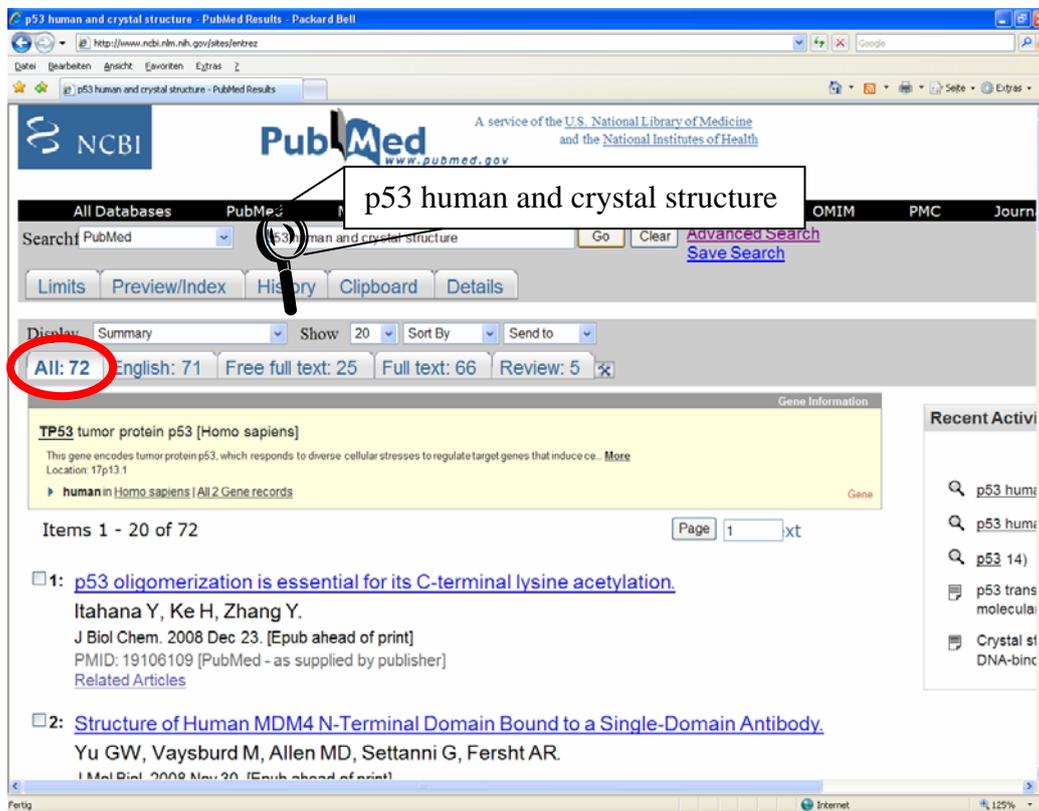
- Mausclick links auf Zeile “**PubMed**: biomedical literature citations and abstracts” (s.o.)
- Suchbegriff eingeben (s. roten Pfeil): z.B. p53



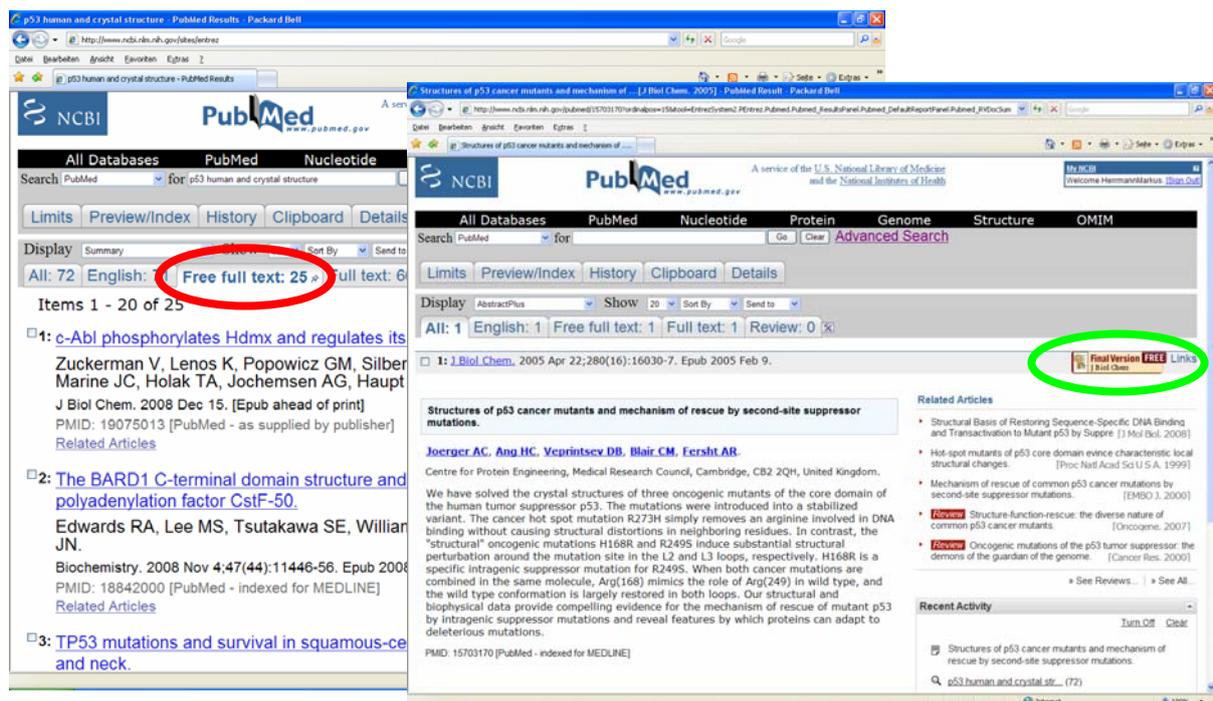
HINWEIS: Man verliert sehr schnell den Überblick und verirrt sich in der Datenflut. So ergibt sich bei der Suche nach dem Tumorsuppressorprotein p53 ein Ergebnis von 48214 Treffern. Auch die Ergänzung von p53 um den Begriff „human“ resultiert in 41011 Treffern.



- Suche durch weitere Begriffe und Operatoren („and“, „or“, „not“) eingrenzen; erst die Eingrenzung mit „p53 human and crystal structure“ ergibt ein akzeptables und überschaubares Ergebnis.



- Mausklick links auf „Free full text“ (s. rote Markierung) zeigt eine Liste von Publikationen, auf deren Zusammenfassungen („abstracts“) direkt zugegriffen werden kann und die als Vollversion gebührenfrei heruntergeladen werden können. Dies ist durch „Final Version FREE“ oder „open access“ oben rechts gekennzeichnet und verlinkt (s. grüne Markierung).



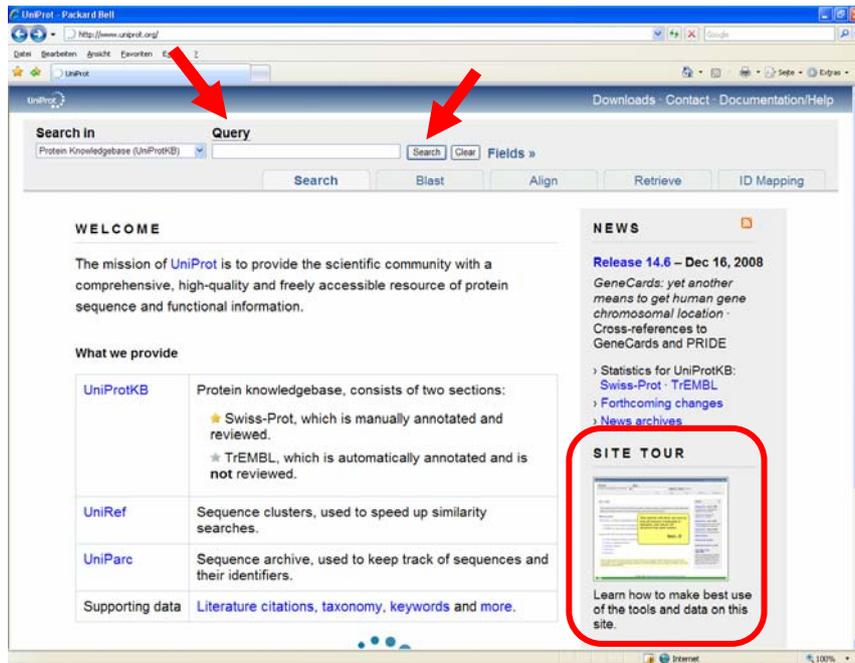
➤ Die Originalpublikation ist als pdf-Datei herunterladbar (s. rote Markierung).

The screenshot shows a web browser window displaying a scientific article. The browser's address bar shows the URL: <http://www.jbc.org/cgi/reprint/280/16/16030>. The article title is "Structures of p53 Cancer Mutants and Mechanism of Rescue by Second-site Suppressor Mutations". The authors listed are Andreas C. Joerger, Hwee Ching Ang, Dmitry B. Veprincev, Caroline M. Blair, and Alan R. Fersht. The article is from the Journal of Biological Chemistry, Vol. 280, No. 16, Issue of April 22, pp. 16030-16037, 2005. The browser's sidebar on the right contains several navigation options, including "Full Text (HTML)", which is circled in red. Other options include "PDF version of: Joerger et al. 280 (16): 16030. (2005)", "Alert me when this article is cited", and "Alert me if a correction is posted".

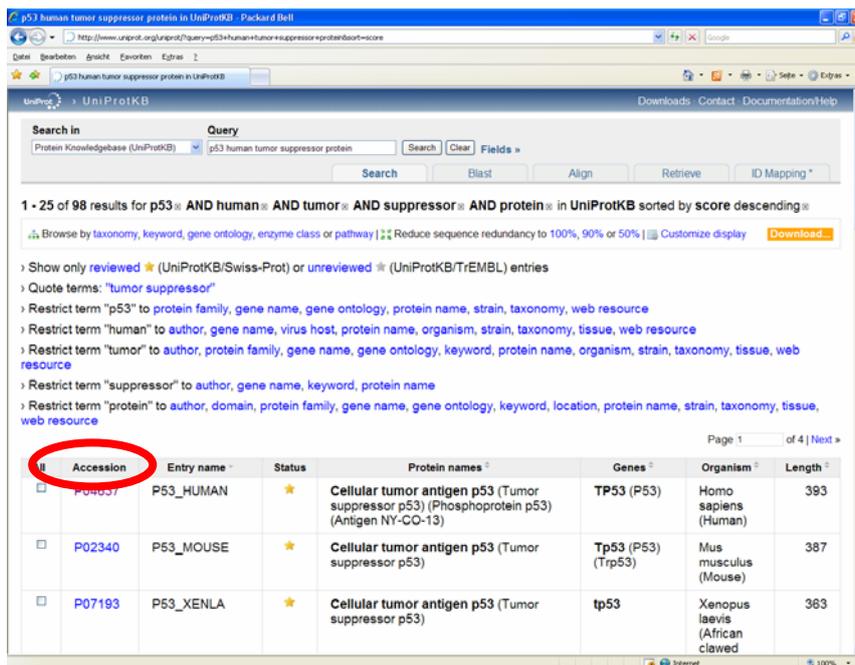
Dies ist nur ein Mikrospektrum der Möglichkeiten, die diese Datenbank der NIH bietet. Es macht sicher viel Freude, die anderen features zu erkunden und auszuprobieren. Jedoch wird schnell die Zeit zum limitierenden Faktor.

**(B) Kurzanleitung zur Suche von Informationen über Proteine:**

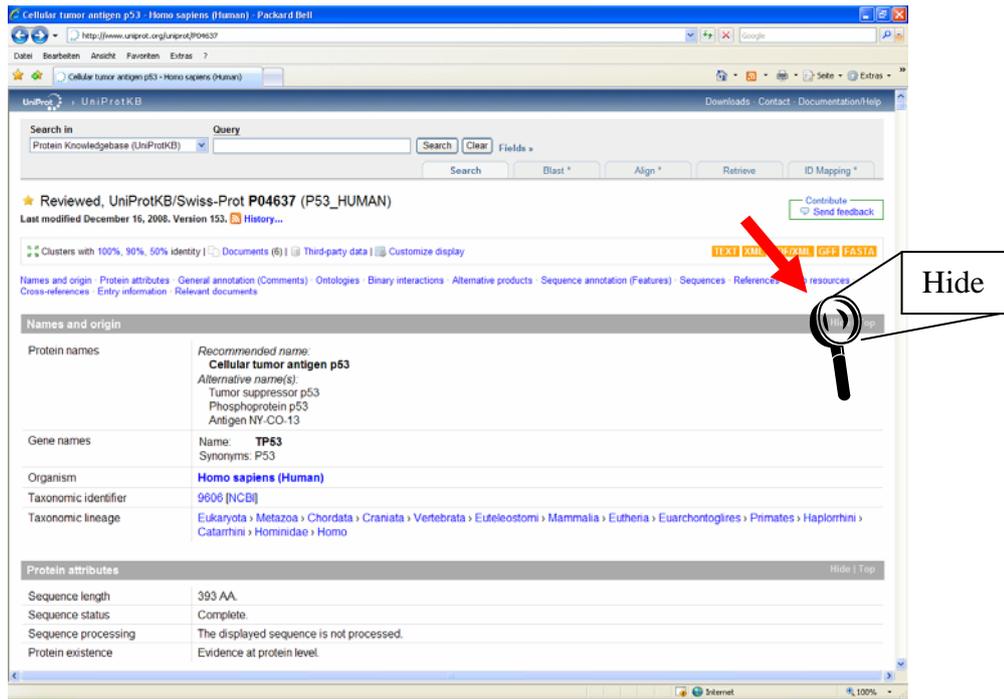
Die Datenbank Universal Protein Resource (kurz: UniProt; <http://www.uniprot.org/>) ist eine umfangreiche Sammlung von Sequenz- und Strukturdaten von Proteinen, die darüber hinaus auch Informationen über Vorkommen, Funktion oder beispielsweise Beteiligung an bestimmten Erkrankungen liefert. Unter „SITE TOUR“ (s. rote Markierung unten rechts) wird ein Lernprogramm für die Nutzung der Datenbank zur Verfügung gestellt.



- Suchbegriff unter „Query“ eingeben und „Search“ klicken (s. rote Pfeile): z.B. p53 human tumor suppressor protein

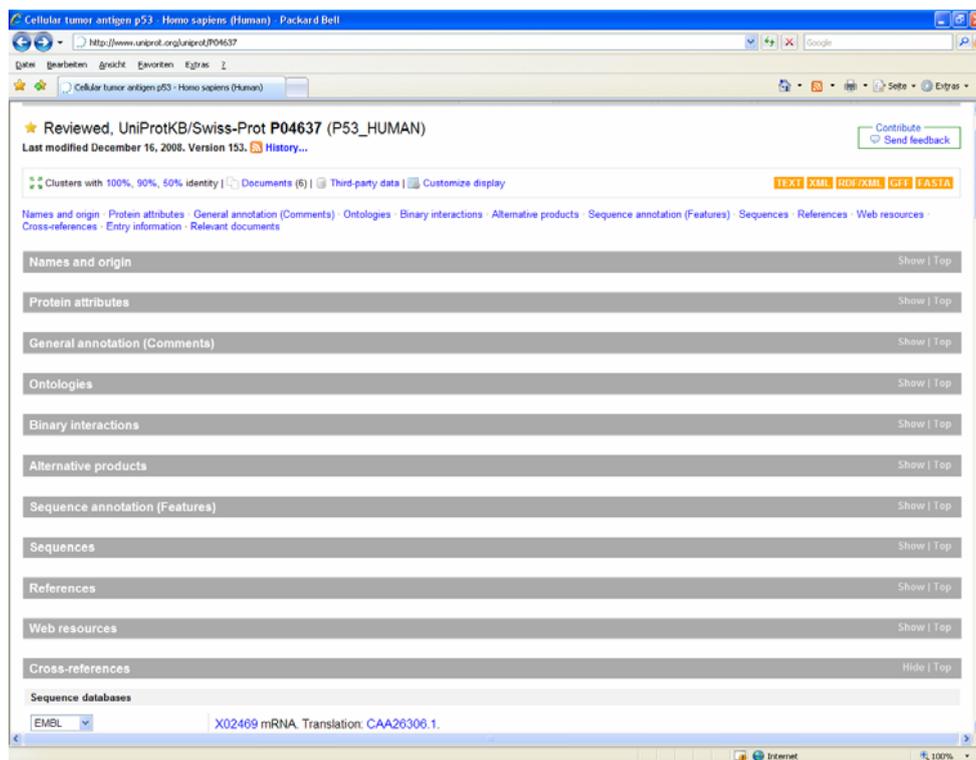


- Zugriffsnummer des gewünschten Proteins in Spalte „Accession“ auswählen (s. rote Markierung)



HINWEIS: Gerade beim menschlichen Tumorsuppressorprotein p53 macht sich die bisher gesammelte Datenflut beim Durchwandern der Seite mal wieder unangenehm bemerkbar. Selbst die Unterteilung durch die grauen Querbalken ist nur bedingt hilfreich, wenn bisher von p53 z.B. über 1.500 Varianten in der Aminosäuresequenz bekannt sind und diese auch alle unter „Sequence annotation (Features)“ aufgelistet werden.

Durch die Funktion „Hide“ am rechten Rand des Balkens steigen Übersichtlichkeit und Bearbeitungsgeschwindigkeit (s. roten Pfeil).



Feature key	Position (s)	Length	Description	Graphical view	Feature identifier
Chain	1 – 393	393	Cellular tumor antigen p53		PRO_0000185703
<b>Regions</b>					
DNA binding	102 – 292	191			
Region	1 – 83	83	Interaction with HRMT1L2		
Region	1 – 44	44	Transcription activation (acidic)		
Region	66 – 110	45	Interaction with WWOX		
Region	100 – 370	271	Interaction with HIPK1 <i>[By similarity]</i>		
Region	116 – 292	177	Interaction with AXIN1 <i>[By similarity]</i>		
Region	241 – 248	8	Interacts with the 53BP2 SH3 domain		
Region	256 – 294	39	Interaction with E4F1		
Region	300 – 393	94	Interaction with CARM1		
Region	319 – 360	42	Interaction with HIPK2		
Region	325 – 356	32	Oligomerization		
Region	359 – 383	5	Interaction with USP7		
Region	368 – 387	20	Basic (repression of DNA-binding)		
Motif	305 – 321	17	Bipartite nuclear localization signal		
Motif	339 – 350	12	Nuclear export signal		
Motif	370 – 372	3	[KR]Q[STAK]K motif		
<b>Sites</b>					
Metal binding	176	1	Zinc		
Metal binding	179	1	Zinc		
Metal binding	238	1	Zinc		
Metal binding	240	1	Zinc		

HINWEIS: Unter „Sequence annotation (Features)“ findet man aber auch sehr interessante Informationen über die Positionen der verschiedenen Domänen innerhalb des Peptids oder Proteins.

Entry	Method	Resolution (Å)	Chain	Positions	PDBsum
1A1U	NMR	-	A/C	324-358	[p]
1A1E	X-ray	1.50	A	326-356	[p]
1C26	X-ray	1.70	A	325-356	[p]
1DT7	NMR	-	X/Y	367-388	[p]
1GZH	X-ray	2.60	A/C	95-292	[p]
1H26	X-ray	2.24	E	376-386	[p]
1HS5	NMR	-	A/B	324-357	[p]
1JSP	NMR	-	A	367-386	[p]
1KZY	X-ray	2.50	A/B	95-289	[p]
1MA3	X-ray	2.00	B	372-389	[p]
1OLG	NMR	-	A/B/C/D	319-360	[p]
1OLH	NMR	-	A/B/C/D	319-360	[p]
1PES	NMR	-	A/B/C/D	325-355	[p]
1PET	NMR	-	A/B/C/D	325-355	[p]
1SAE	NMR	-	A/B/C/D	319-360	[p]
1SAF	NMR	-	A/B/C/D	319-360	[p]
1SAG	NMR	-	A/B/C/D	319-360	[p]
1SAH	NMR	-	A/B/C/D	319-360	[p]
1SAI	NMR	-	A/B/C/D	319-360	[p]
1SAJ	NMR	-	A/B/C/D	319-360	[p]
1SAK	NMR	-	A/B/C/D	319-360	[p]
1SAL	NMR	-	A/B/C/D	319-360	[p]
1TSR	X-ray	2.20	A/B/C	94-312	[p]
1TUP	X-ray	2.20	A/B/C	94-312	[p]
1UOL	X-ray	1.90	A/B	94-312	[p]
1XQH	X-ray	1.75	B/F	369-377	[p]
1YCQ	X-ray	2.30	B	13-29	[p]
1YCR	X-ray	2.60	B	15-29	[p]
1YCS	X-ray	2.20	A	94-312	[p]

➤ unter „Cross-references“ befindet sich eine Auflistung aller bisher ermittelten 3D-Strukturdaten; wenn wie im Falle von p53 keine Strukturdaten der Gesamtsequenz existieren, wählt man bspw. die längste aus (s. rote Markierung)

The screenshot shows the RCSB PDB Structure Explorer interface. The browser address bar displays the URL: <http://www.pdb.org/pdb/cg/structure.cgi?pdbid=1tsr>. The page header includes the RCSB PDB logo and navigation links. A red arrow points to the 'Structure Summary' tab in the navigation menu. The main content area displays the following information for entry 1tsr:

- Title:** P53 CORE DOMAIN IN COMPLEX WITH DNA
- Authors:** Cho, Y., Gorina, S., Jeffrey, P., Pavletich, N.
- Primary Citation:** Cho, Y., Gorina, S., Jeffrey, P.D., Pavletich, N.P. (1994) Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 265: 346-355
- History:** Deposition 1995-07-28 Release 1996-01-29
- Experimental Method:** Type X-RAY DIFFRACTION Data N/A
- Parameters:** Resolution[A] 2.20, R-Value 0.205 (obs.), R-Free n/a, Space Group C 2 (C 1 2 1)
- Unit Cell:** Length [A] a 117.50, b 67.90, c 108.80; Angles [°] alpha 90.00, beta 105.50, gamma 90.00
- Molecular Description Asymmetric Unit:** Polymer: 1 Molecule: DNA (5'-D (\*TP\*TP\*TP\*CP\*CP\*TP\*AP\*GP\*AP\*CP\*TP\*TP\*GP\*CP\*CP\*A P\*AP\*TP\*TP\*A)-3') Chains: E Polymer: 2 Molecule: DNA (5'-D (\*AP\*TP\*AP\*AP\*TP\*TP\*GP\*GP\*CP\*AP\*AP\*GP\*TP\*CP\*TP\*A P\*GP\*GP\*AP\*A)-3')

On the right side, there is an 'Images and Visualization' section with a 3D ribbon diagram of the protein-DNA complex and 'Display Options' such as KING, Jmol, WebMol, and MBT SimpleViewer\*.

- man gelangt direkt auf die Protein-Datenbank von Research Collaboratory for Structural Bioinformatics (RCSB) und kann sich die 3D-Strukturdaten von p53 als pdb-Datei herunterladen (s. roten Pfeil)

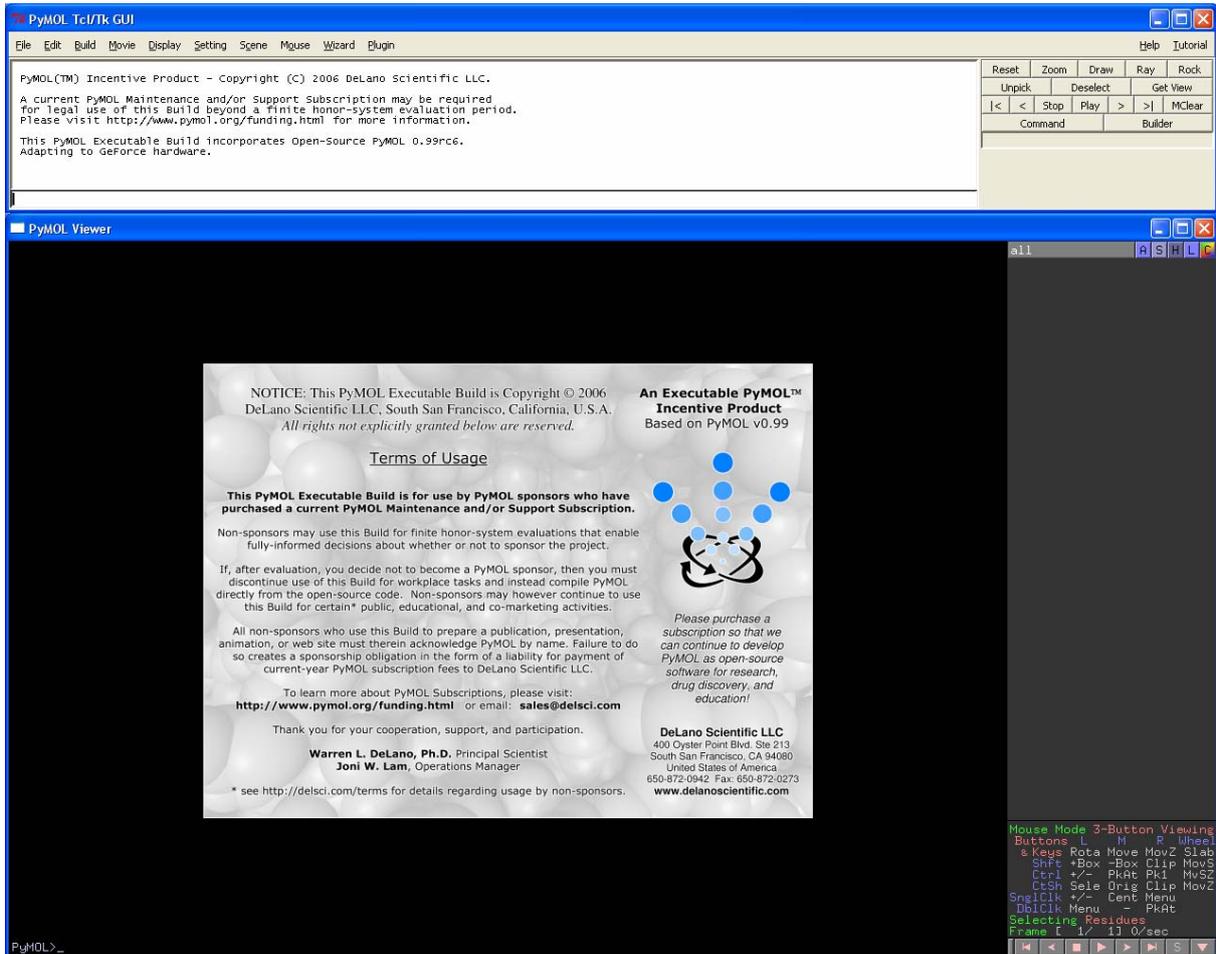
### (C) Kurzanleitung zur Nutzung des Visualisierungsprogramms PyMOL:

Um komplexe Makromoleküle wie Nukleinsäuren oder Proteine am Computerbildschirm dreidimensional darstellen zu können, benötigt man Visualisierungsprogramme wie z.B. PyMOL. DeLano Scientifics ist ein kommerzieller Software-Anbieter, der im Internet eine veraltete Version des Programms kostenlos zum Herunterladen zur Verfügung stellt ( <http://pymol.sourceforge.net/> ).

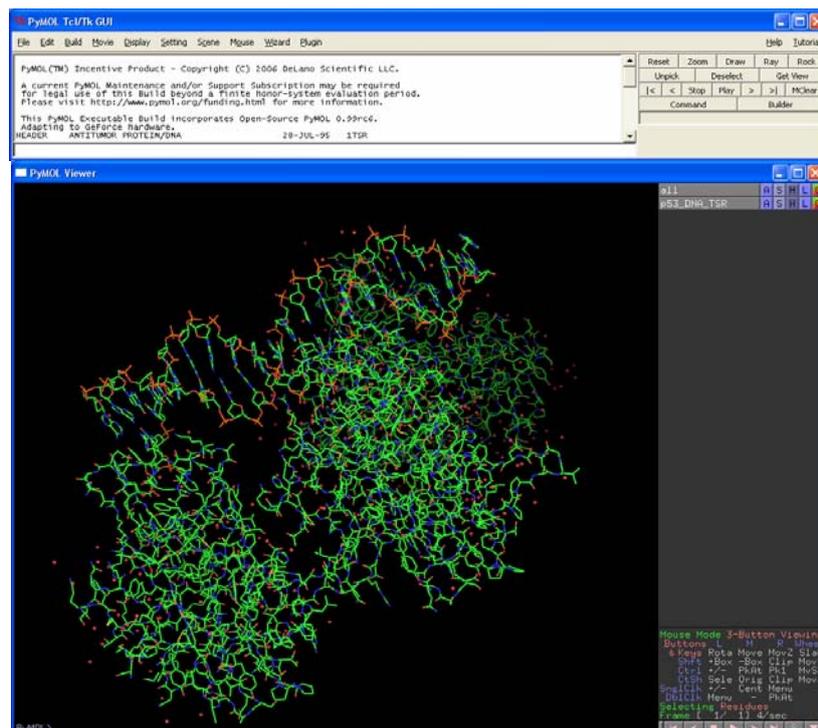
- unter „Download“ kann man unter “Older PyMOL builds (pre-1.0) are still **FREELY accessible** but are no longer maintained” die voll funktionsfähige Version 0.99 herunterladen (Volumen: 11,4 MB) (s. rote Markierung)

The image consists of two screenshots of a web browser. The top screenshot shows the PyMOL Molecular Viewer page on SourceForge. The 'Download' link in the navigation bar is circled in red. The page content includes a list of features, a central molecular model, and a section for 'Older PyMOL builds (pre-1.0) are still FREELY accessible but are no longer maintained.' The bottom screenshot shows the 'Download PyMOL Builds' page, where the same text is highlighted with a red box. The page also includes sections for 'For Current Subscribers', 'For Potential Subscribers', and 'As Open-Source Code'.

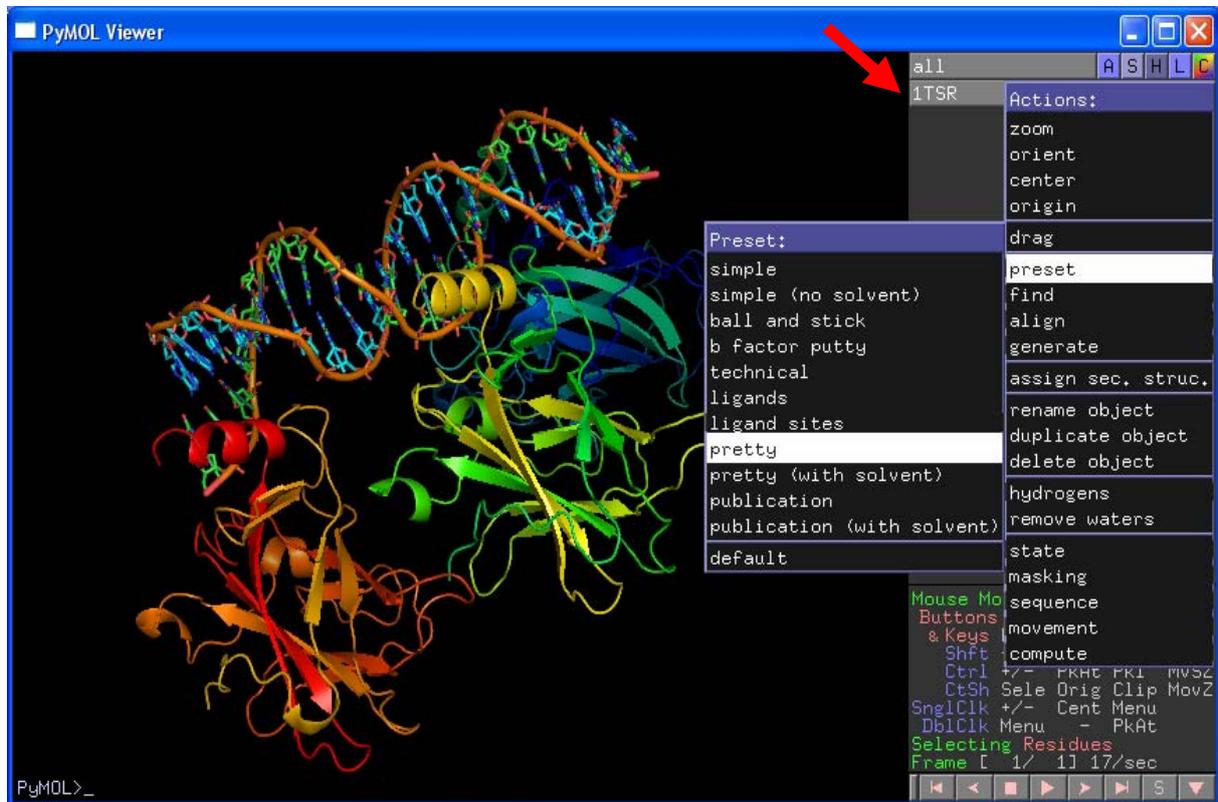
- nach Installation auf dem Rechner das Programm PyMOL starten



- gespeicherte pdb-Datei von p53 (= 1TSR.pdb) öffnen



- in der Zeile 1TSR „A“ für **Actions** anklicken (s. roten Pfeil), aus Liste „preset“ wählen und aus weiterer Liste „pretty“ wählen (andere Möglichkeiten einfach auch mal alle ausprobieren!)



- Bei Gedrückthalten der **linken Maustaste** kann das visualisierte Molekül durch Bewegen der Maus im Raum gedreht werden.
- Bei Gedrückthalten der **rechten Maustaste** kann das visualisierte Molekül durch Ziehen der Maus gezoomt werden.

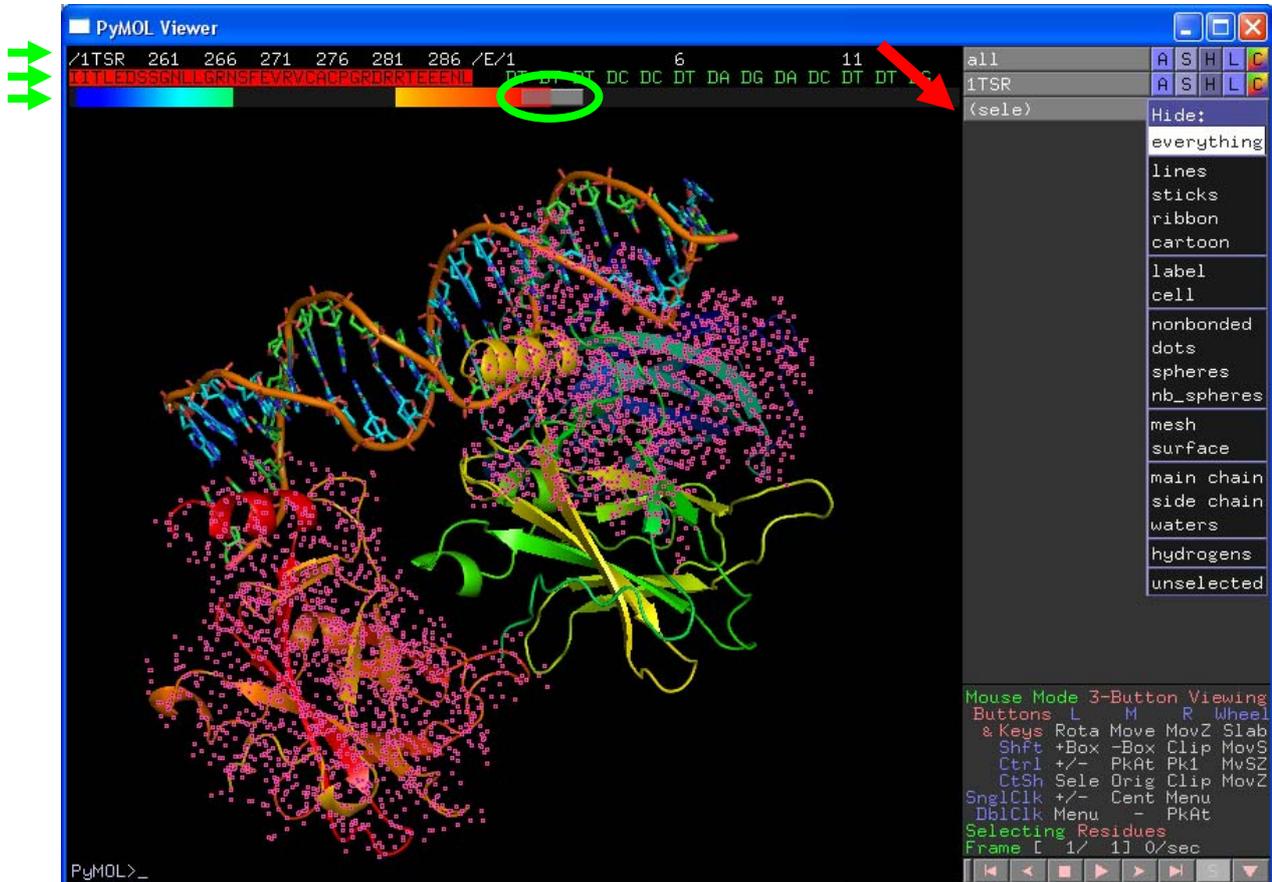
### Molekülsequenzen ausblenden oder entfernen:

- aus Hauptmenüleiste unter „Display“ die Ansicht „Sequence“ auswählen



- es werden oberhalb der Molekülabbildung in 3 Zeilen die jeweiligen Aminosäurepositionen im Peptid/Protein bzw. Nukleotide in der DNA, die Aminosäure im Einbuchstabencode und ein hellgrauer Balken, der als Cursor genutzt werden kann, dargestellt (s. grüne Pfeile, grüne Markierung)
- durch Mausklick links können einzelne Aminosäuren ausgewählt werden; durch Umschalttaste (=shift) + Mausklick links können innerhalb von 2 begrenzenden Aminosäuren ganze Bereiche markiert werden

(die ausgewählten Bereiche werden entsprechend der Farbgebung in der Abbildung in der Cursorzeile wiedergegeben; rechts erscheint unterhalb der Zeile 1TSR eine weitere Zeile (sele), die die Bearbeitung ausgewählter Aminosäuresequenzen steuert (s. roten Pfeil)

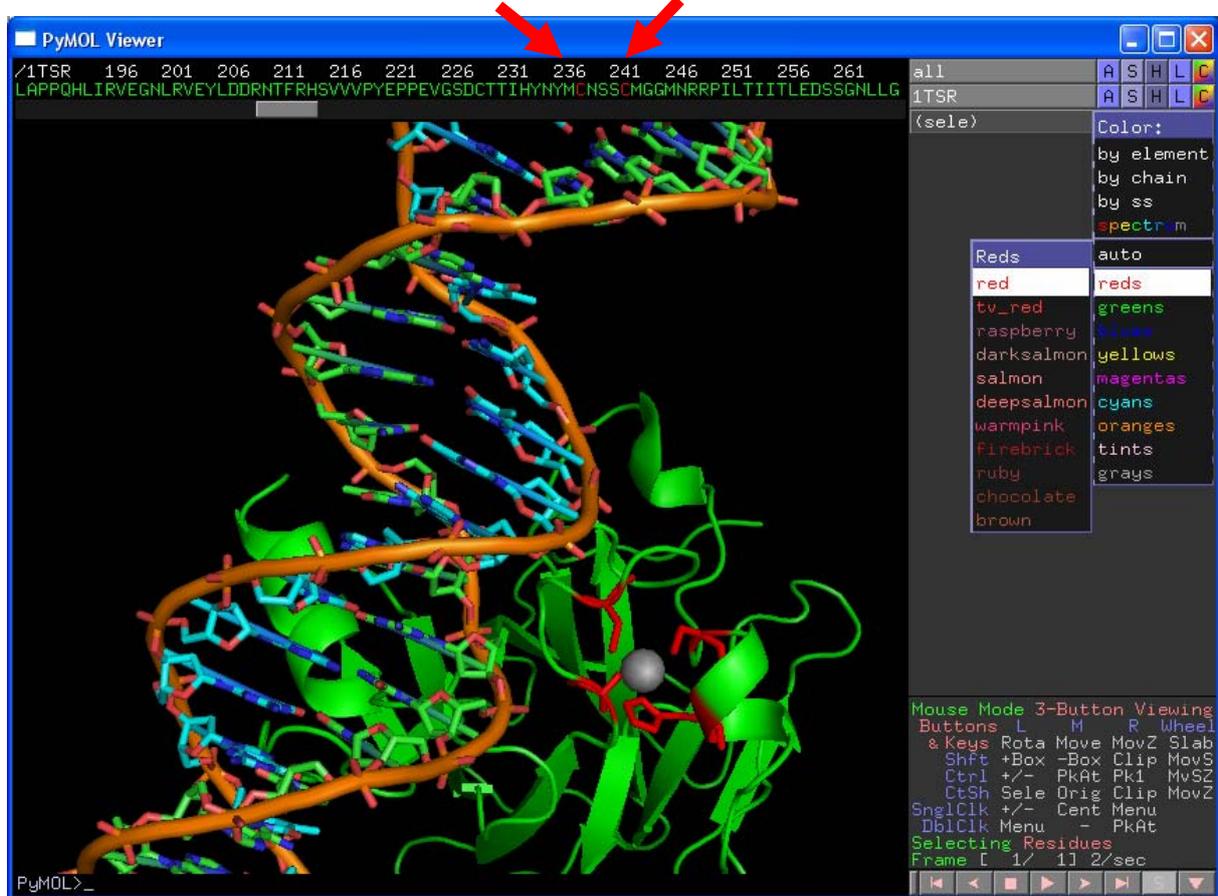


- in der Zeile (sele) „H“ für **Hide** anklicken und aus Liste „everything“ wählen (die ausgewählten Bereiche erscheinen in der Abbildung zunächst mit kleinen pinkfarbenen Quadraten; sie bleiben nach der Bearbeitung erhalten – können aber durch Mausklick links auf den schwarzen Hintergrund entfernt werden; die Molekülteile bleiben in der Sequenzzeile erhalten, werden aus der Molekülabbildung aber ausgeblendet)
- die markierten Molekülanteile können durch Anklicken von „A“ in Zeile (sele) und auswählen von „remove atoms“ auch aus Sequenzzeile entfernt werden

### Farbgebungen / Moleküldarstellungen verändern:

- durch Mausklick links in Sequenzzeile Aminosäuren oder ganze Bereiche markieren
- in der Zeile (sele) „C“ für **Color** anklicken und aus Liste z.B. unter „greens“ weiteren Grün-Ton wählen (gesamte Sequenz des dargestellten p53-Moleküls hellgrün gefärbt; die Zinkfinger-Aminosäuren an Positionen 176, 179, 238 und 242 rot gefärbt (s. rote Pfeile); das zentrale Zink-Ion grau gefärbt; Ansicht gedreht und gezoomt)

- in der Zeile (sele) „S“ für **Show** anklicken und aus Liste z.B. „sticks“ (hier für die Zinkfinger-Aminosäuren) oder „spheres“ (für das Zink-Ion) auswählen



### Hintergrundfarbe variieren:

- in Menüleiste unter „Display“ → „background“ wählen (am Bildschirm ist schwarzer Hintergrund meist optimal; jedoch für Ausdrücke suboptimal)

### Sicherung von Molekülbearbeitungen / als Grafik für Textverarbeitung:

- in Menüleiste unter „File“ → „Save Session“ wählen (Bearbeitungszustand kann parallel zur Original-pdb-Datei als pse-Datei gespeichert werden)
- in Menüleiste unter „File“ → „Save Image“ wählen (Bearbeitungszustand kann parallel zur Original-pdb-Datei als png-Bilddatei gespeichert werden; Einfügen als Grafik durch Textverarbeitungsprogramme möglich)

Auch hierbei handelt es sich nur um eine Kurzanleitung, die rasch erste Schritte im Umgang erlaubt, jedoch die vielfältigen Möglichkeiten, die dieses Visualisierungsprogramm bietet, nur oberflächlich ankratzt. Man sollte es einfach mal spielerisch ausprobieren und sich überraschen lassen, was am Bildschirm so alles passiert 😊.