

V7 Modellierung von biomolekularen Komplexen

- **Protein-Protein-Docking**
- **Protein-DNA-Komplexe**

In Vorlesung #7 beschäftigen wir uns mit biomolekularen Wechselwirkungen zwischen zwei oder mehreren Proteinen und mit Protein:DNA-Komplexen.

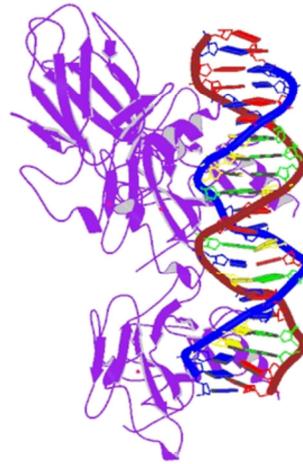
Beispiel eines Protein-DNA-Komplexes

PDB-Struktur 1TUP: Tumor Suppressor **p53**

Durch Röntgenkristallographie bestimmter Protein-DNA-Komplex in der PDB-Datenbank.

Lila (links): p53-Protein

Blau/roter DNA-Doppelstrang (rechts)



The protective action of the wild-type *p53* gene helps to suppress tumors in humans. The *p53* gene is the most commonly mutated gene in human cancer, and these mutations may actively promote tumor growth.



www.sciencemag.org (1993)

7. Vorlesung WS 19/20

Softwarewerkzeuge

www.rcsb.org

Wir beginnen mit Protein-DNA-Komplexen. Dies können z.B. Komplexe von DNA mit Transkriptionsfaktoren oder mit Enzymen zur DNA-Replikation und zur DNA-Reparatur sein.

In der Abbildung ist die Struktur von doppelsträngiger DNA mit dem Transkriptionsfaktor p53 gezeigt. p53 ist das Gen, das am häufigsten in den Genomen von Tumorkranken mutiert ist. Im Jahr 1993 wurde es zum Molecule of the Year in der Zeitschrift Science benannt. p53 wird als „guardian of the cell“ bezeichnet, also der Wächter der Zelle.

p53 ist besonders wichtig bei der Erkennung, ob DNA-Reparaturen erforderlich sind und bei der Einleitung der dafür erforderlichen Schritte.

spezifische Bindung / Kristallkontakt



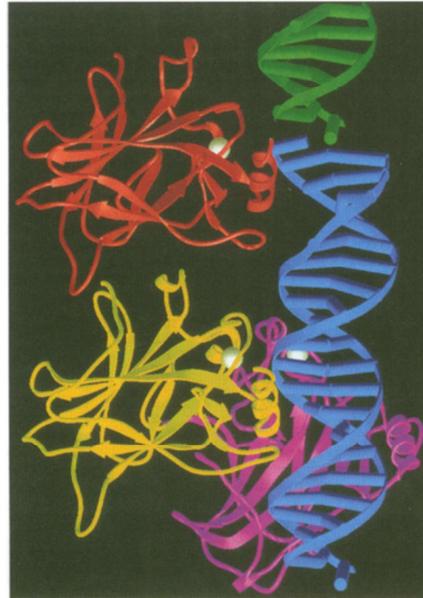
Nikola Pavletich,
Sloan Kettering
Cancer Center

RESEARCH ARTICLE

Crystal Structure of a p53 Tumor Suppressor-DNA Complex: Understanding Tumorigenic Mutations

Yunje Cho, Svetlana Gorina, Philip D. Jeffrey, Nikola P. Pavletich

Fig. 3. Schematic ribbon drawing of the asymmetric unit, which contains three p53 core domain molecules and one DNA duplex. Two of the core domains bind DNA (blue); one (yellow) interacts extensively with a consensus binding site, and the other (red) binds at a nonconsensus site at the interface of DNA fragments related by crystallographic symmetry (a portion of the symmetry-related DNA fragment is shown in green). The third core domain molecule (purple) does not bind DNA, but makes protein-protein contacts stabilizing crystal packing. The zinc atoms are shown as white spheres.



Science 265, 346-355 (1994)

7. Vorlesung WS 19/20

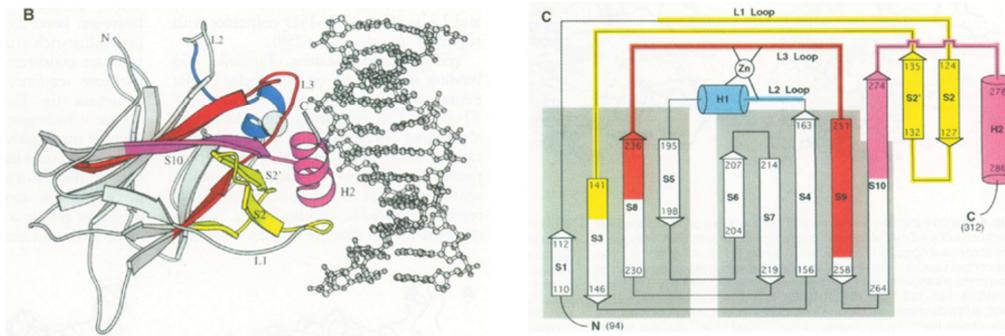
Softwarewerkzeuge

3

Die rechte Abbildung zeigt eine Kristallstruktur von doppelsträngiger DNA (blau und grün) mit 3 Kopien (rot, gelb, lila) der Kerndomäne von p53. Nur die rote und gelbe Domäne machen Kontakte mit DNA, die lila Domäne ist nur an Protein-Protein-Wechselwirkungen im Kristall beteiligt. Sie hat also keine biologische Relevanz. Man bezeichnet solche Kontakte als „Kristallpackung“. Die rote Domäne bindet an einer nicht-Konsensus-Stelle der DNA, wo diese also nicht die übliche p53-bindende Sequenz hat.

Daher ist nur die gelbe Domäne (welche an eine Konsensus-Sequenz bindet) von biologischer Relevanz.

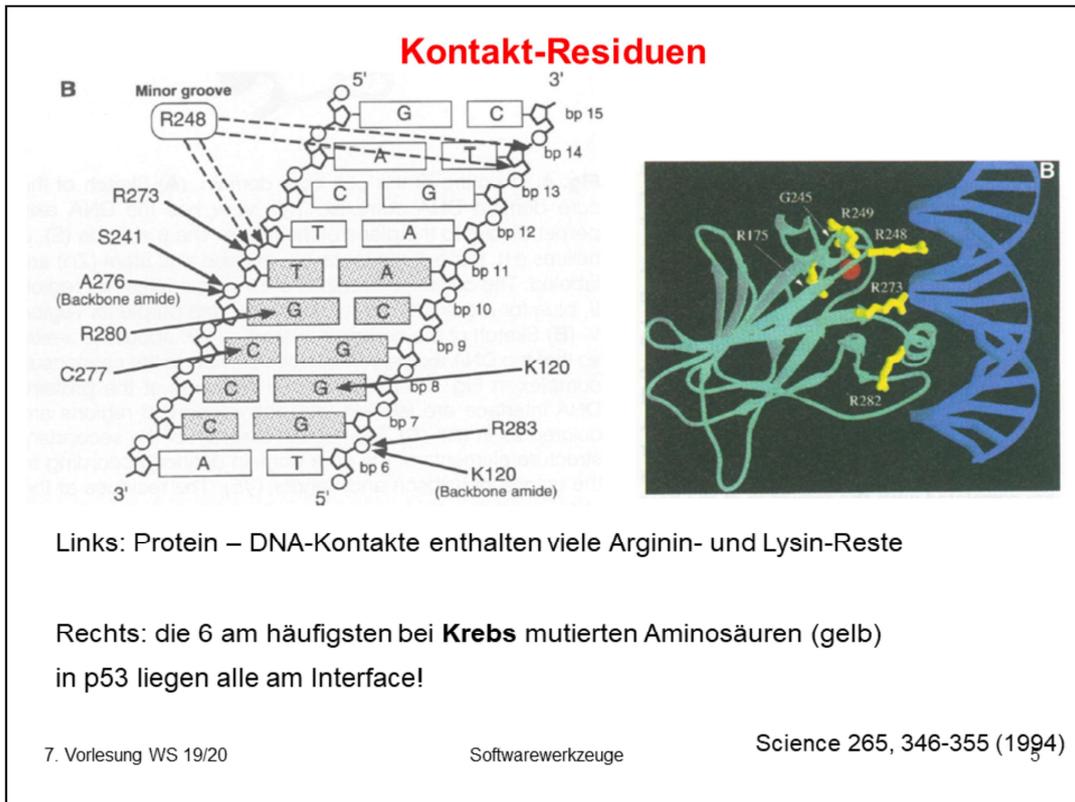
Konservierte Bereiche sind am Interface angereichert



colored as in (A). (C) Topological diagram of the secondary structure elements of the core domain defined according to the criteria of Kabsch and Sander (45). The residues at the start and the end of each secondary structure element are indicated. The conserved regions are colored according to (A), and the boundaries of the two β sheets that make up the β sandwich are shaded.

Science 265, 346-355 (1994)

Die rechte Abbildung zeigt die innere Beta-Strang-Architektur dieser p53-Kerndomäne. Die linke Abbildung zeigt deutlich die Kontakte zwischen p53 und DNA. Die pinke Helix lagert sich an die DNA an, jedoch nicht in deren große Furche (major groove) hinein. Nahe an der DNA liegt außerdem der rechte und der gelbe Loop.



Links ist die DNA-Sequenz gezeigt und die Kontakte mit 9 Proteinresiduen. Darunter sind drei positiv geladene Arginine (R) und zwei positive geladene Lysine (K). Rechts sind die 6 am häufigsten bei Krebs mutierten Aminosäuren in gelb markiert. Bis auf Gly245 sind dies alles Arginine. 3 liegen nahe an der DNA (248, 273 und 282), zwei weiter entfernt (249 und 175).

Da elektrostatische Wechselwirkungen als langreichweitig bekannt sind, können alle diesen positiven Aminosäuren die Bindung an das negativ geladene Phosphatrückgrat der DNA stabilisieren.

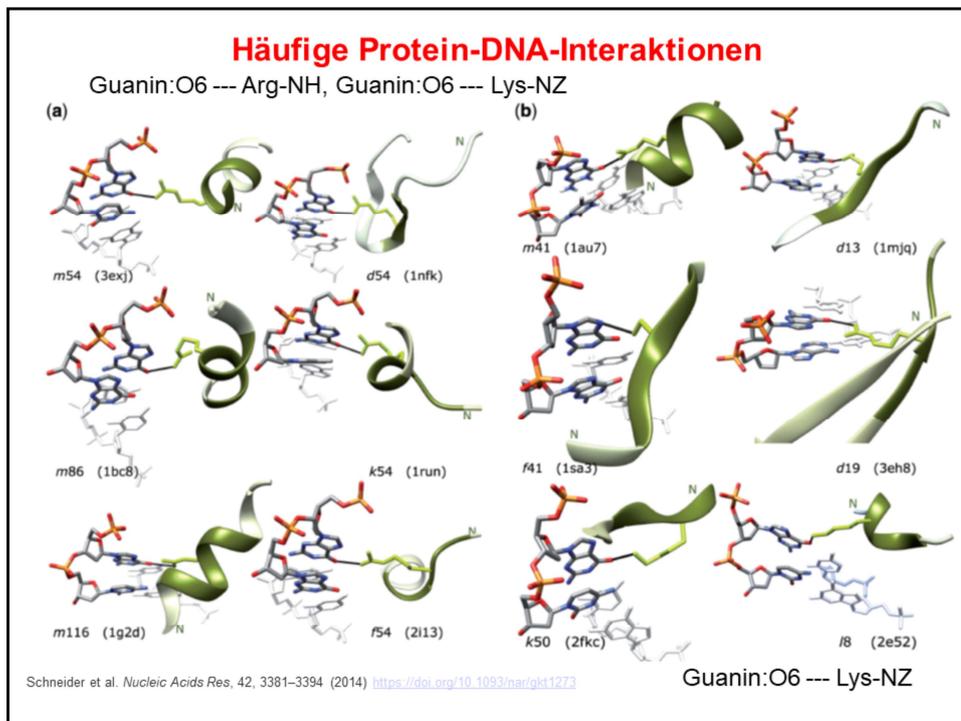


Figure 3. Examples of the common protein/DNA interactions. Interacting motifs are labeled by the codes of the interacting PB and ntC and by PDB id of structures in which they were identified. Interacting PBs are drawn as green cartoon with atoms of the central amino acid in light green and the nucleotide step as a stick model using commonly used ‘chemical’ colors; the contacts (black sticks) are directed to the major groove edge of guanines in the right-handed double helical DNA. The 5'-end phosphates are on the left top of each motif. The N-ends of the PBs are labeled; the complementary DNA strand and amino acids adjacent to the depicted PB are in light gray. (a) Motifs common to all types of structures approximately in order of their occurrence in the group of all 1018 structures. All contacts shown are between the **guanine atom O6** and the **arginine NH** observed in crystal structures 3exj (65), 1nfk (66), 1bc8 (67), 1run (68), 1g2d (69) and 2i13 (70). (b) Motifs m41 and d13 are highly populated in transcription factors (TrF-R2) and underrepresented in nucleases (Nuc-R2), motifs f41, d19, k50 and l8 are highly populated in Nuc-R2 and less in TrF-R2. They appear in crystal structures 1au7 (71), 1mjg (72), 1sa3 (73), 3eh8 (74), 2fkc (75) and 2e52 (76). The motifs m41, d13 and d19 show interaction between the guanine O6 and arginine NH, k50 and l8 between the guanine O6 and lysine NZ, and f41 between the guanine N7 and serine OG.

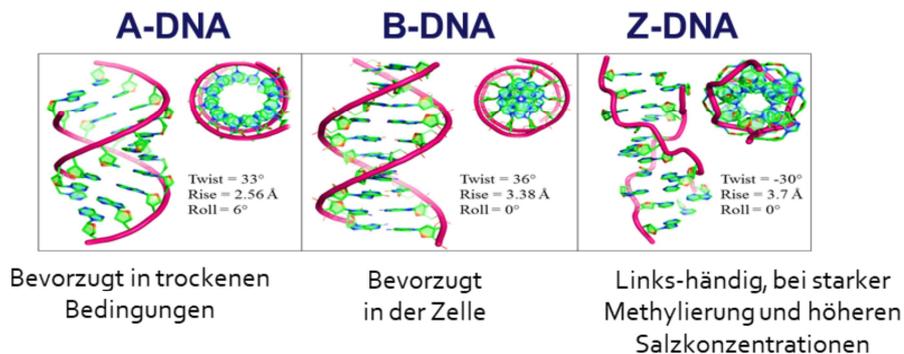
© The Author(s) 2013. Published by Oxford University Press. This is an Open

Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Prinzipien von Protein-DNA-Komplexen

Das wichtigste Prinzip für eine gute Interaktion ist elektrostatische Komplementarität. DNA ist stark negativ geladen, die Proteinoberfläche muss entsprechend positiv geladen sein

Protein und DNA haben eine flexible Konformation. Die DNA-Konformation bleibt meist nicht in der kanonischen B-DNA-Konformation, sondern deformiert sich etwas, zum Teil in die A-Konformation.



7. Vorlesung WS 19/20

Softwarewerkzeuge

7

Die kanonische Konformation von zellulärer DNA ist die sogenannte B-Konformation (Mitte). Die A-DNA-Konformation (links) kann bei trockenen Bedingungen eingenommen werden, also wenn der Wassergehalt sehr reduziert wird. Hier sind die Windungen deutlich weiter aufgeweitet. Die Basenpaare liegen außerdem nicht wie bei B-DNA in der Mitte der Helix übereinander, sondern am Rand des Helixinneren. Eine weitere Konformation ist Z-DNA (rechts). Hier ist die DNA anders herum (linkshändig) gewunden. Die Basen liegen nur teilweise im Inneren und stehen deutlich mehr für Interaktionen zur Verfügung.

Protein-Protein Interaktionen

Sind **wichtig** für viele **zelluläre Prozesse**

Eine biologische Zelle enthält zu jedem Zeitpunkt ca.
 10^9 Proteinkopien aus etwa 5000 verschiedenen Proteinen.

Im Mittel bildet jedes Protein etwa 6 Interaktionen.

Etwa die Hälfte aller Proteine bildet **stabile** oder **transiente Komplexe**

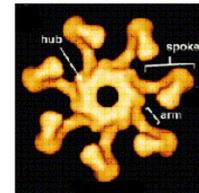
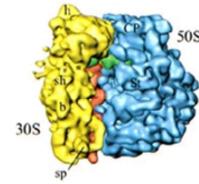
Charakterisierung der **Protein-Protein-Assoziation**:

thermodynamische Stabilität : Bindungskonstante k_D

Kinetische Ratenkonstante für Assoziation: k_{on}

driving forces der Protein-Protein-Wechselwirkung:

- lang-reichweitige elektrostatische Anziehung
- Feste Assoziation durch hydrophoben Effekt



Kein Kommentar.

Eigenschaften von Protein-Protein-Interfaces

Table 2. Properties of protein-protein interfaces

Parameter	Protein-protein complexes ^a	Homodimers ^b		Weak dimers ^c	Crystal packing ^d
		Bahadur	Dey		
Number in dataset	70	122	276	19	188
BSA (Å ²)	1910	3900	3700	1620	570, 1510
(S.D.)	(760)	(2200)	(2160)	(670)	(520)
Amino acids per interface	57	104	100	50	48
BSA (Å ²) per amino acid	34	38	37	32	32
Composition (BSA %)					
Non-polar	58	65	65	62	58
Neutral polar	28	23	22	25	25
Charged	14	12	13	13	17
Chain segments ^f	5-6	3-4	3-2	5-8	6-3
H bonds					
<i>n</i> _{IB} (number per interface)	10	19	18	7	5
BSA per bond (Å ²)	190	210	209	230	280
Water molecules ^g					
Number per interface	20	44			23
Number per 1000 Å ²	10	11			15
Bridging H bonds	6	13			6
Residue conservation ^h					
% in core	55	60			40
<i>z</i> in core and rim	0.65 and 0.80	0.63 and 0.77			0.98 and 0.99

Homodimere sind meist permanente Komplexe
-> sehr große Schnittstellen, großer Anteil an unpolaren Aminosäuren

BSA: buried surface area
S.D. standard deviation

Artifizielle Kristallkontakte enthalten einen wesentlich kleineren Anteil an konservierten Residuen
(40% gegenüber 55-60%).

Dies ist eine Statistik über die Eigenschaften von Protein-Protein-Schnittstellen. BSA misst die Fläche der Interfaces. Die Interfaces von permanenten Homodimeren (3900 Å²) sind mehr als zwei mal so groß wie die Interfaces von schwach gebundenen Dimeren (weak dimers, 1620 Å²) und Kontakten, die lediglich im Kristall existieren (crystal packing, 570/1510 Å²).

Die Zusammensetzung (%BSA) ist ebenfalls sehr unterschiedlich. Homodimere haben besonders viele hydrophobe Aminosäuren am Interface (65%), schwache Dimere nur 62% und Kristallkontakte 58%. Umgekehrt gibt es in Kristallkontakte mehr Kontakte, die geladene Aminosäuren beinhalten (17%), als für schwache Dimere und Homodimere.

Fläche von Protein-Protein-Interfaces

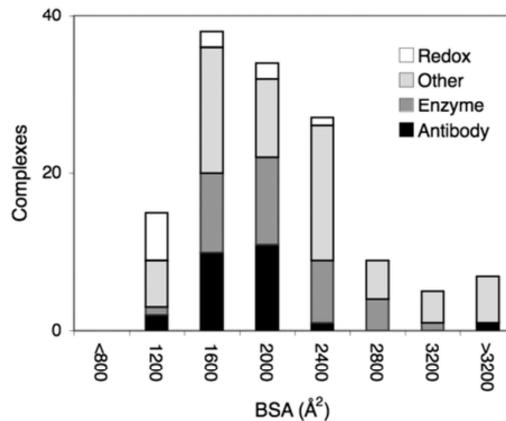
Redox-Komplexe dienen z.B. zur Übertragung von Elektronen (Cytochrom c).



Cytochrom c aus
Pferdeherz
1hrc.pdb;
www.rcsb.org

Beobachtung: Redox-Komplexe haben recht kleine Schnittstellen.
Ihre Funktion benötigt nur eine transiente = kurzlebige Bindung.

Dies wird durch die kleine Größe der Schnittstelle unterstützt.



In dieser Statistik wurden die Protein-Schnittstellen nach ihrer Funktion aufgeteilt. Redox-Komplexe wie z.B. mit dem Elektronen-Überträger-Protein Cytochrom c haben eher kleine Interfaces. Hier muss lediglich eine Redox-Reaktion stattfinden (z.B. ein Elektronentransfer zwischen den beiden Proteinen). Danach lösen sie sich wieder voneinander. Antikörper müssen möglichst dauerhaft an ihre Bindungspartner binden und haben daher größere Interfaces (1600-2000 Å²).

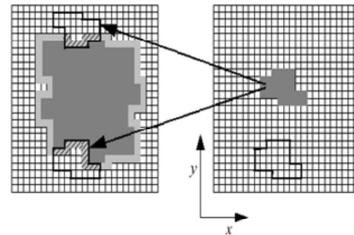
Protein-Protein-Komplexe: Komplementäre Oberflächen

Die bekannten Strukturen von Protein-Protein-Interaktionen des Menschen (bekannte experimentelle Strukturen plus Homologie-Modelle) decken nur etwa 4% der geschätzten Anzahl von etwa 300.000 Protein-Protein-Interaktionen zwischen menschlichen Proteinen ab.

Quelle: Proteins, 81, 2192–2200 (2013)

Idee: versuche, die Strukturen von PP-Komplexen durch **Docking** zu modellieren.

Berechne die Komplementarität der Oberflächen zwischen beiden starren Proteinen in allen möglichen Orientierungen (alle erlaubten Translationen + Rotationen),
Suche in 6 Freiheitsgraden



Docking ist eine populäre Methode, um Modelle für Protein-Protein-Komplexe zu generieren. Meist werden beide Proteine als starr betrachtet. Dann braucht man nur die relative Orientierung der beiden Proteine zueinander zu berechnen. In der Abbildung sind zwei mögliche Translationen des rechten Proteins gezeigt.

Workflow für Protein-Protein-Docking

(1) Docking von starren Proteinen
mit Katchalski-Kazir-Algorithmus
z.B. FTDock-Program, verwendet FFT,
Optimale Lösung hat maximales $a \times b$

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 2192-2199, March 1992
Biophysics

Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques

(protein-protein interaction/surface complementarity/macromolecular complex prediction/molecular docking)
EPHRAIM KATCHALSKI-KATZIR^{1,2}, ISAAC SHARIV³, MIRIAM EISENSTEIN⁴, ASHER A. FRIESEM⁵,
CLAUDE AFLALO³, AND ILYA A. VAKSER¹

Departments of ¹Membrane Research and Biophysics, ²Electronics, ³Structural Biology, and ⁴Biochemistry, Weizmann Institute of Science,
Rehovot 76100, Israel

Next, to distinguish between the surface and the interior of each molecule, we retain the value of 1 for the grid points along a thin surface layer only and assign other values to the internal grid points. The resulting functions thus become

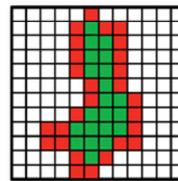
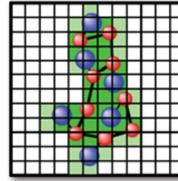
$$\bar{a}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \rho & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases} \quad [2a]$$

and

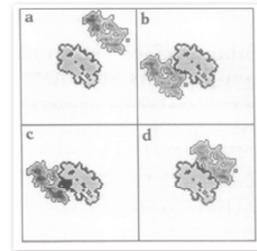
$$\bar{b}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \delta & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases} \quad [2b]$$

where the surface is defined here as a boundary layer of finite width between the inside and the outside of the molecule. The parameters ρ and δ describe the value of the points inside the molecules, and all points outside are set to zero. Two-

Typical values: $\rho = -15$, $\delta = 1$
=> penalty for overlap of volumes



2D cross sections at $l = 46$ ($N = 90$)



- a) no contact
- b) limited contact
- c) overlap (black area)
- d) good geometric match

7. Vorlesung WS 19/20

Softwarewerkzeuge

12
Proteins, 81, 2192-2200 (2013)

Der Katchalski-Kazir-Algorithmus ist weit verbreitet und basiert auf einer möglichst großen Komplementarität der Oberflächen beider Proteine.

Wie links gezeigt, werden beide Proteinstrukturen auf ein Gitter abgebildet. Beim Docking muss vermieden werden, dass sich beide Proteine überlappen. Deshalb wird rho sehr negativ gewählt.

Man positioniert das zweite Protein irgendwo um das erste Protein herum (tatsächlich probiert man alle möglichen Positionen aus). Danach multipliziert man in jeder Gitterzelle (l,m,n) die entsprechenden Werte der Proteine $\bar{a}_{l,m,n}$ und $\bar{b}_{l,m,n}$ und bildet dann die Summe über alle Gitterpositionen. Das verbirgt sich hinter $a \times b$. Man sucht die Lösung mit maximalem $a \times b$. Dies wird erreicht, wenn sich über einen möglichst großen Bereich die Proteine an ihren Rändern berühren. In der Abbildung ganz rechts ist das für Panel d) der Fall. Zur effizienten Berechnung dieser Summe für alle möglichen Translationen und Rotationen wird wieder die Fast Fourier Transformation (FFT) benutzt, siehe MAFFT in V3.

Workflow für Protein-Protein-Docking

(2) Wende Zdock-Scoring-Funktion auf die FTDock-Ergebnisse an
-> wähle 1000-2000 Kandidaten-Modelle mit dem besten Score aus.

Zdock Scoring-Funktion: statistisches Paar-Potential zwischen Kontaktresiduen
(→ DOPE in V6), Oberflächenkomplementarität und Elektrostatik

(3) Bewerte Zdock-Lösungen noch einmal mit pyDOCK-Scoring-Funktion:

- **Desolvationsenergie** proportional zur Abnahme der gesamten
Solvent-accessible surface area (-> Hydrophober Effekt in V5)

- **Coulomb-Elektrostatik** und **van-der-Waals-Wechselwirkungen** zwischen
Kontakt-Residuen

Hier ist ein modernes Beispiel gezeigt, in dem FFT-Docking als erster Schritt verwendet wurde, gefolgt von einem ReScoring mit einer anderen, genaueren Energiefunktion. Die Details davon sind hier nicht wichtig.

Ergebnis in CAPRI5-Wettbewerb

Table I
Results of Our pyDock Protocol for All Protein-Protein Targets of the Last CAPRI Edition

Target	Type	Predictors			Scorers		
		Submission rank ^a	Quality ^b	Successful groups ^c	Submission rank ^a	Quality ^b	Successful groups ^c
T46	HH	—	—	2 (40)	—	—	8 (16)
T47	HU	1	***	25 (29)	2 ^d	***	13 (14)
T48	UU	3	*	14 (32)	No scorers	No scorers	No scorers
T49	UU	4	*	14 (33)	6	*	7 (13)
T50	UH	1	**	18 (40)	4	**	12 (17)
T51	DHD	—	—	3 (46)	—	—	5 (13)
T53	UH	3 ^e	**	20 (42)	1	**	11 (13)
T54	UH	—	—	4 (41)	—	—	0 (13)
T58	UU	5	**	11 (23)	No scorers	No scorers	No scorers

U, unbound; H, homology-based model; D, domain.

^aRank of the best model within our submission to CAPRI.

^bQuality of our best model according to CAPRI criteria.

^cNumber of successful groups for each target; in brackets, total number of participants.

^dModel Rank 1 had medium accuracy (**).

^eModel Rank 1 had acceptable accuracy (*).

HH: Docking von 2 Homologie-Modellen für die beiden Proteine

UU: Docking der Kristallstrukturen der ungebundenen (freien / apo) Strukturen der Proteine
(ist natürlich zuverlässiger als HH)

In dem CAPRI-Wettbewerb versuchen Bioinformatik-Arbeitsgruppen, Docking-Lösungen für Proteinkomplexe zu erzeugen, die gerade erst aufgeklärt wurden (vgl. der CAMEO-Wettbewerb für Homologie Modelling-Server in V6). Die Predictor-Server (links) versuchen sowohl, günstige Strukturen der Komplexe zu generieren als auch ein Ranking unter ihnen zu erstellen. Die Scoring-Server (rechts) erstellen „nur“ für eine Anzahl an gegebenen Strukturen ein Ranking. Sie erzeugen also keine Konformationen.

Docking von Homologie-Modellen funktioniert üblicherweise schlechter als wenn die Strukturen der Bindungspartner experimentell bestimmt wurden. Das Docking der Proteinstrukturen im ungebundenen Zustand (UU) ist das große Ziel. Für manche Systeme (T47 und T50) funktionierte das Docking mit pyDock exzellent. Bei diesen Systemen (UH) wurde eine ungebundene Struktur an ein Homologie-Modell gedockt. Allerdings waren für diese Systeme auch viele andere Arbeitsgruppen erfolgreich.

Protein-Protein-Komplexe

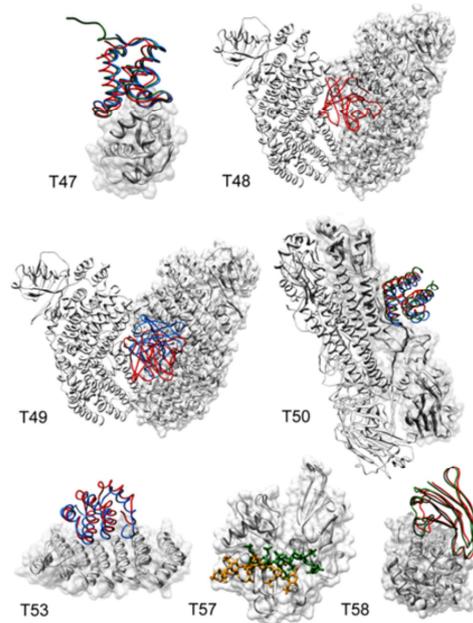
Beste Docking-Modelle für die Targets
T47, T48, T49, T50, T53, T57, und T58.

Rezeptorstruktur ist jeweils überlagert und
weiss gefärbt.

Rot/blau: beste Dockingpositionen für das
zweite Protein.

In grün ist (falls bekannt) die
experimentelle Struktur gezeigt.

Dies sind die Erfolgsfälle.



7. Vorlesung WS 19/20

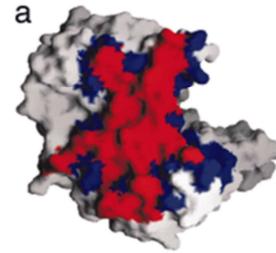
Softwarewerkzeuge

Proteins, 81, 2192–2200 (2013)
15

Dies sind Beispiele für erfolgreiche Docking-Lösungen. Es gibt natürlich auch (viel mehr) Fälle, in denen die Docking-Modelle weit entfernt von der experimentellen Struktur liegen.

Komposition von Protein-Protein-Bindungsschnittstellen Ent

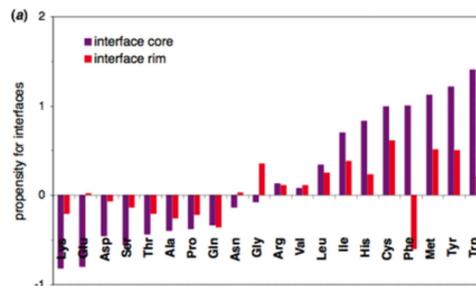
Bindungsschnittstelle einer Untereinheit
des homodimerischen Proteins
2-Phospho-d-Glycerat Hydrolase (PDB ID
code 1ebh).
"core" (rot) und "rim" (blau) Regionen



(a) Verglichen mit der gesamten
Proteinoberfläche enthält die
zentrale Region der Schnittstelle
(Interface-core) deutlich mehr
aromatische Aminosäuren,
aber deutlich weniger geladene
Aminosäuren.

Der Rand („rim“) der Schnittstelle ist
recht ähnlich zur Gesamtoberfläche.

<http://www.pnas.org/content/102/43/15447.full>



Janin et al. Quart Rev Biophys 41, 133-180 (2008)

7. Vorlesung WS 19/20

Softwarewerkzeuge

16

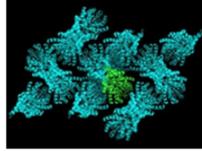
Nun schauen wir uns eine Statistik über die Häufigkeit der 20 Aminosäuren an Bindungs-Schnittstellen an. In der Abbildung oben rechts ist auf einer Untereinheit eines Homo-Dimers die Bindungsschnittstelle farblich eingefärbt. Man unterscheidet den „core“ (Kern) – dies sind alle Aminosäuren, die im gebundenen Zustand im Kontakt mit dem anderen Protein sind, aber nicht mit dem Lösungsmittel – und das „rim“ (Kuppe) – das sind die Kontakt-Aminosäuren am Rand der Schnittstelle, die außerdem noch Kontakt mit Wasser haben.

In der Abbildung unten sieht man die logarithmierten Häufigkeiten in den beiden Regionen (ungeschickterweise sind die Farbcodes gegenüber der Abbildung vertauscht) relativ zur restlichen Proteinoberfläche. Am rim gibt es z.B. viel weniger Phenylalanin, aber mehr Glycin. Im Core gibt es sehr viele aromatische Aminosäuren Trp und Tyr und sehr wenige geladene Aminosäuren Lys, Glu und Asp. Bloß Arg ist eine Ausnahme davon.

Komposition von Bindungsschnittstellen

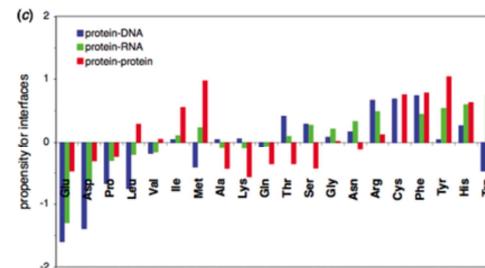
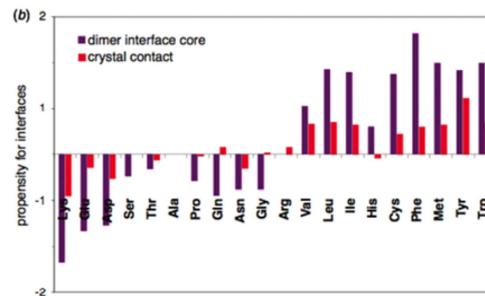
Kristallkontakt: Bindungsschnittstelle zweier Proteine im Kristall, ist oft nur eine artifizielle Interaktion durch Kristallisationsprozess.

<http://pymolwiki.org>



(b) Artifizielle Kristallkontakte haben (zu) wenig hydrophobe Aminosäuren am Interface.

(c) An Interfaces mit negativ geladener DNA gibt es viel weniger negative Glu- und Asp-Residuen als bei PP-Kontakten. Dafür mehr positive Arg-Residuen.



Janin et al. Quart Rev Biophys 41, 133-180 (2008)
Softwarewerkzeuge

17

Auf dieser Folie gibt es analoge Statistiken für den Vergleich von Protein-Protein-Schnittstellen (oben links) aufgeteilt in tatsächliche Dimer-Schnittstellen und artifizielle Kristallkontakte. Die Kristallkontakt-Häufigkeiten sind recht ähnlich zur restlichen Proteinoberfläche. Die Dimere haben viele aromatische und hydrophobe AS und sehr wenig geladene Glu, Lys und Asp. Durch Betrachtung der AS-Häufigkeiten kann man daher recht gut entscheiden, ob es sich bei einem Proteinkomplex um eine „guten“ Kontakt mit typischen AS-Häufigkeiten handelt.

Die untere Abb. vergleicht Protein-Protein-Schnittstellen mit Protein-DNA/RNA-Schnittstellen. Zwischen DNA und RNA-Kontakt gibt es relativ wenig Unterschiede, bis auf Trp ganz rechts. Trp wechselwirkt anscheinend relativ gut mit einzelsträngiger RNA (pi-Stacking), aber sehr schlecht mit doppelsträngiger DNA (Basen liegen nach innen, Trp kommt da nicht ran). Die negativ geladenen Glu/Asp sind an Schnittstellen mit negativ geladener DNA / RNA sehr ungünstig.

Protein-Nukleinsäure-Komplexe

Table 3. *Properties of protein–nucleic acid interfaces*

Average value	Protein/RNA ^a	Protein/DNA ^b	Protein/protein ^c
Number of complexes	81	75	70
BSA (Å ²)			
Mean	2530	3100	1910
S.D.	(1210)	(1050)	(760)
Protein/nucleic acid	1210/1320	1540/1560	–
Number of amino acids/nucleotides	43/18	48/18	57
BSA (Å ²) per amino acid/nucleotide	28/75	33/72	34
Composition (protein/nucleic acid, BSA %)			
Non-polar	55/33	52/41	58
Neutral polar	21/41	24/16	28
Charged (negative)	4/26	2/43	5
Charged (positive)	20/0	23/0	9
f_{bur} (% buried atoms, protein/nucleic acid)	29/29	24/28	34
L_D (packing index, protein/nucleic acid)	37/43	39/46	42
H bonds			
n_{HB} (number per interface)	20	22	10
BSA per bond (Å ²)	125	145	190
Water molecules			
Number per interface	32	21	20
Number per 1000 Å ²	13	7	10
Bridging H bonds	11		6

Eher größere Kontaktflächen.

Hoher Anteil an positiv geladenen Aminosäuren.

Janin et al. *Quart Rev Biophys* 41, 133-180 (2008)

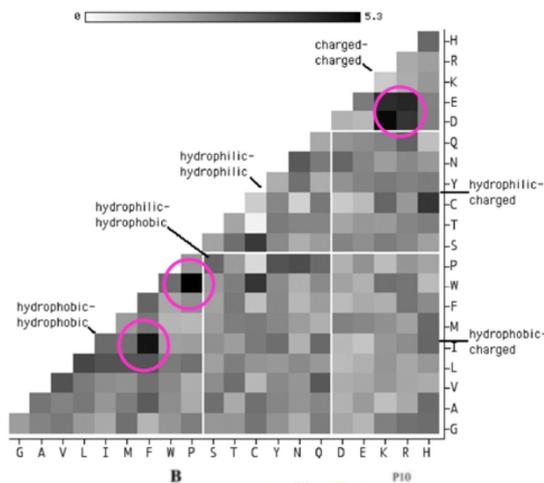
7. Vorlesung WS 19/20

Softwarewerkzeuge

18

Diese Statistik vergleicht die Oberflächen, Zusammensetzung und Kontakte an den 3 Schnittstellen.

Paarungspräferenzen an PP-Interfaces



Häufige Kontakte (schwarz):

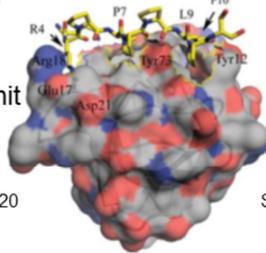
(1) Trp - Pro.

Man findet sie oft an Interfaces von prolinreichen Peptiden und Adapterdomänen wie SH3-Domänen.

(2) die beiden hydrophoben AS Phe und Ile.

(3) Salzbrücken wie Lys – Asp.

PI3-Kinase-SH3 Domäne mit Peptidligand



7. Vorlesung WS 19/20

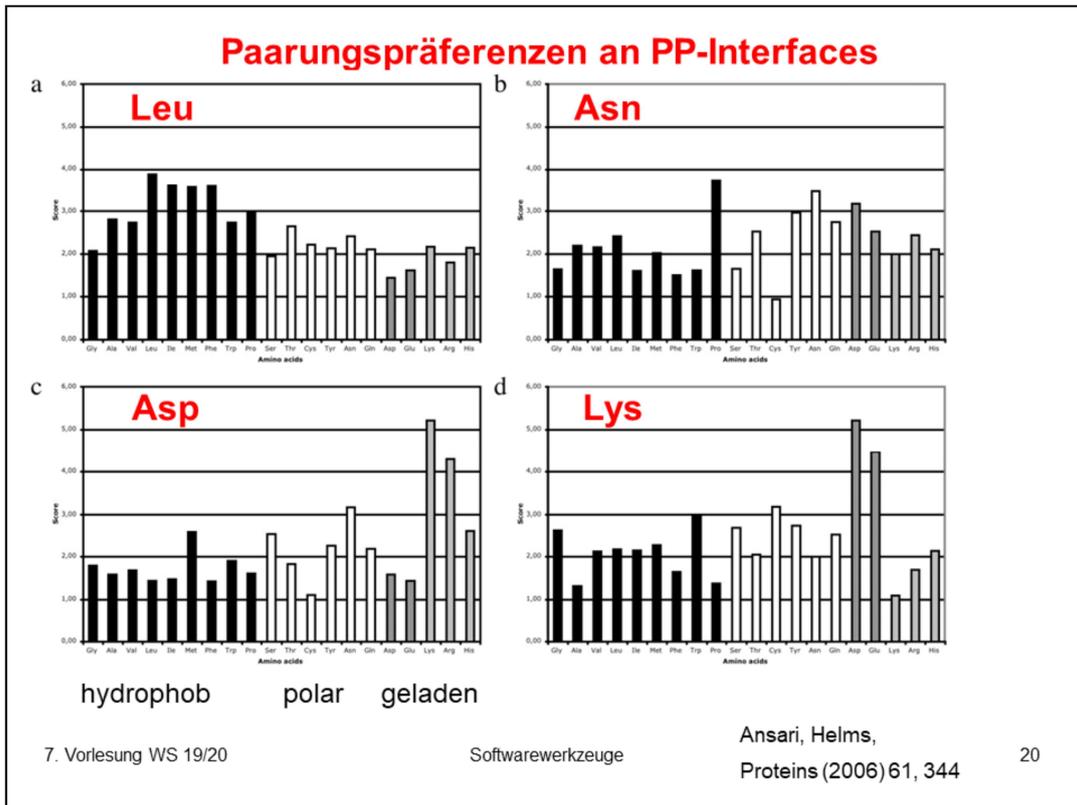
Softwarewerkzeuge

Ansari, Helms, Proteins (2006) 61, 344

Batra-Safferling et al. Biol. Chem., 391, 33 (2010)

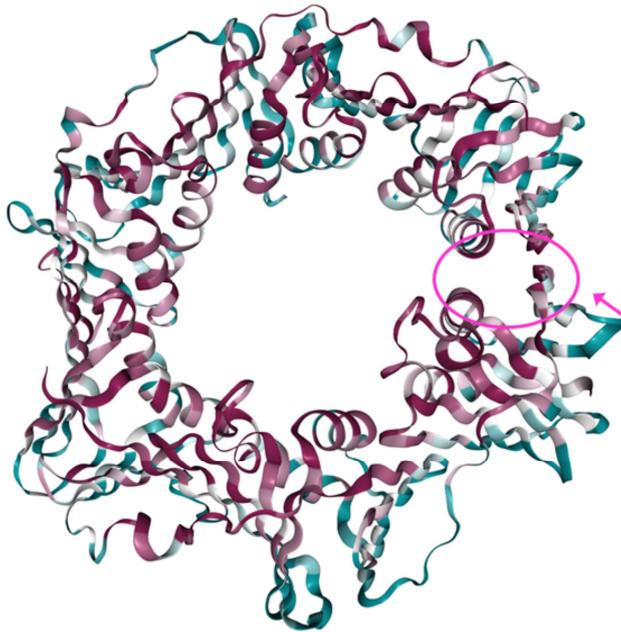
19

Nachdem wir zunächst untersucht hatten, wie häufig die 20 AS an Interfaces vorkommen, kann man ebenfalls analysieren, wie häufig sich bestimmte AS-Kontakte dort bilden. Hier sind 3 häufige Kontakte pink markiert. Natürlich gibt es oft Salzbrücken zwischen Lys/Arg und Asp/Glu. Kontakte zwischen Trp (W) und Pro (P) gibt es oft an den Interfaces von SH3-Domänen. Die Pro-Ringe eines Proteins stecken dann in Taschen des anderen Proteins. Die Abb. unten zeigt solch einen Komplex (vgl. auch Folie 17 in V5). Hier liegen die Prolinringe von P7 und P10 in Taschen auf der SH3-Oberfläche. Das Peptid hat eine PPII-Konformation mit einer Periodizität von 3 (nicht 3,6 wie in einer alpha-Helix). Deshalb zeigen die beiden Proline im Abstand 3 in dieselbe Richtung.



Dies ist die Kontakt-Häufigkeit von Leu, Asn, Asp und Lys an Interfaces. Asp macht natürlich viele Kontakte mit entgegengesetzt geladenen Lys und Arg, Lys umgekehrt mit Asp und Glu. Diese Graphik ist gewissermaßen jeweils ein Schnitt durch die Matrix auf der vorigen Folie.

Sequenzkonservierung an PP-Interfaces



Konservierung am Dimer-Interface der β Untereinheit der DNA Polymerase III von *E. coli* (Ashkenazy et al. 2016).

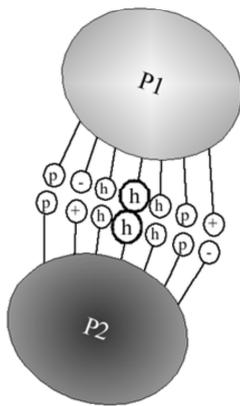
Rot: starke Konservierung.
Blau: schwache Konservierung.

Am Interface zwischen den beiden Untereinheiten sind die Residuen stark konserviert.

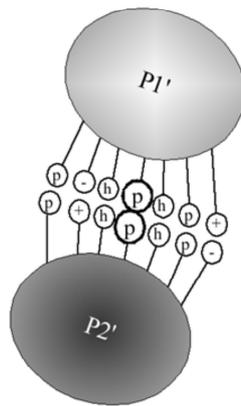
Man kann ebenfalls die Konservierung an Bindungsschnittstellen analysieren. Im Allgemeinen ist die Konservierung dort im Vergleich zur restlichen Proteinoberfläche leicht erhöht. Ein bequemes Tool für so etwas ist der Webserver consurf (<https://consurf.tau.ac.il/>) von Nir Ben-Tal.

Dort wird basierend auf einem Multiplen Sequenzalignment der Proteinfamilie die Konservierung an jeder Position bestimmt. In diesem Beispiel ist die Konservierung am pink umkreisten Dimer-Interface erhöht (rot eingefärbt). Der relative Vergleich zur restlichen Proteinoberfläche kann aber irreführend sein, da andere Teile der Oberfläche Interaktionen mit anderen Proteinen eingehen können.

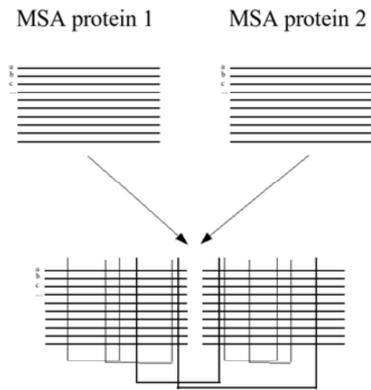
Korrelierte Mutationen an PP-Interfaces



Mitte:
Kontakt zwischen
zwei hydrophoben
AS



Beide Residuen des
Kontakts sind in
einem anderen
Organismus' durch
polare AS ersetzt.



Man kann in MSA-Alignments
zweier Proteine gezielt nach
solchen Korrelationen suchen
→ Vorhersage

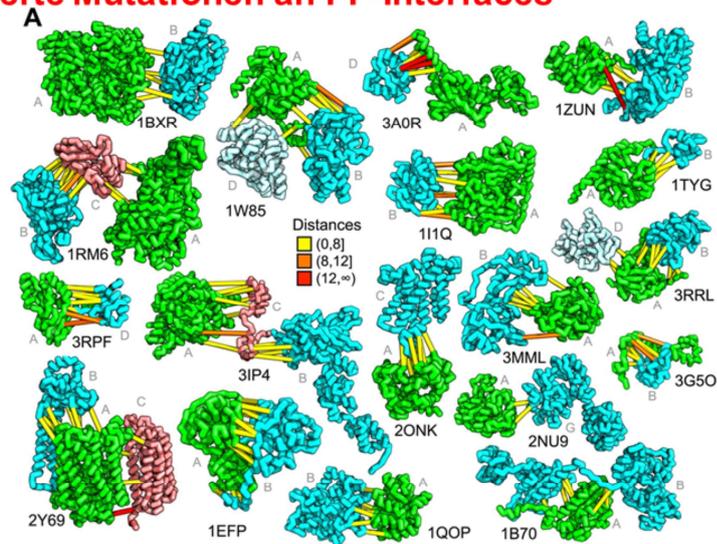
(dabei enthält jede Reihe Proteine
desselben Organismus)

Ein wichtiges Indiz für das Vorliegen von Kontakten sind jedoch korrelierte Mutationen von gegenüberliegenden Aminosäuren an Interfaces (linke Abb.). Generell wäre eine Mutation von einer hydrophoben AS in eine polare AS ungünstig, falsch die Kontakt-AS unverändert bleibt. Wenn diese in einem Organismus gemeinsam in den anderen Typ austauschen, dann ist das ok.

Wie rechts gezeigt, kann man nun in den multiplen Sequenzalignments von zwei Proteinfamilien gezielt nach solchen Korrelationen suchen. Diese sind dann Indiz für mögliche Kontakte.

Korrelierte Mutationen an PP-Interfaces

Korrelierte Mutationen wurden an den Interfaces vieler bekannter Protein-Kontakte gefunden.



Ovchinniko et al. (2014)
Elife 3, e02030.

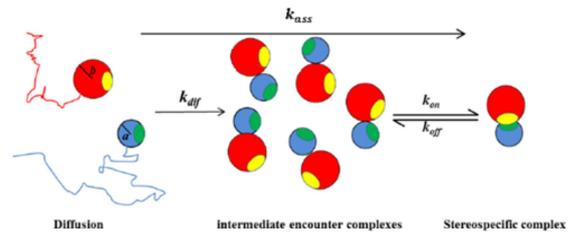
Dies ist ein link zu der Publikation: <https://elifesciences.org/articles/02030>

Mit der sogenannten Direct Coupling Analysis (DCA) kann man solche Korrelationen aufspüren. Eine DCA-ähnliche Methode ist in dem Tool GREMLIN aus der Gruppe von David Baker implementiert. In diesem Beispiel sind alle Korrelationen in bekannten Komplexen gezeigt. Die beiden Bindungspartner sind jeweils auseinandergezogen, damit man die Korrelationen besser einzeichnen kann. Die farbliche Einfärbung bezieht sich auf die Distanz im gebundenen Zustand.

Assoziationspfad für PP-Bindung

Schritte bei Protein-Protein-Assoziation:

- Zufällige Diffusion (1)
- Elektrostatische Anziehung (2)
- Bildung von Encounter-Komplex (3)
- Dissoziation oder Bildung des endgültigen Komplexes via Übergangszustand (TS) (4)

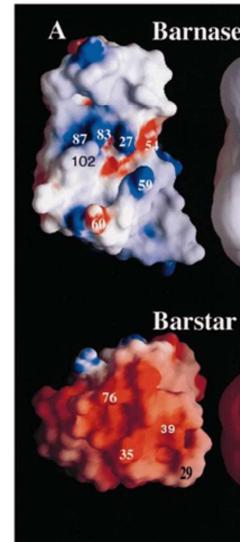


Bei der Bindung müssen die Schnittstellen desolvatisiert werden.
Dies könnte energetisch aufwändig sein.

Nun betrachten wir die Energielandschaft, auf der Protein-Protein-Assoziation stattfindet.

Barnase: Barstar sind elektrostatisch komplementär

- Barnase: eine **Ribonuklease**, die außerhalb der Zelle aktiv ist.
Barstar: ihr intrazellulärer Inhibitor;
beide Proteine haben Durchmesser $\sim 30 \text{ \AA}$
- Sehr gut charakterisiertes Modellsystem für **elektrostatisch gesteuerte** diffusive Annäherung zweier Proteine
- Interaktion zwischen Barnase und Barstar gehört zu den **stärksten** bekannten **Interaktionen** zwischen Proteinen
sehr **schnelle Assoziationsrate**:
 $10^8 - 10^9 \text{ M}^{-1} \text{ s}^{-1}$ bei 50 mM ionischer Stärke



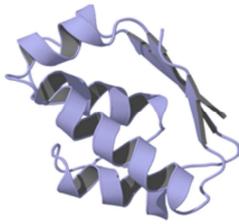
Färbung gemäß
elektrostatischem Potential
auf Proteinoberfläche.

Blau: positiv
Rot: negativ

25

Als Beispiel haben wir den Komplex von Barnase und Barstar betrachtet. In der Abbildung sind die beiden Protein-Oberflächen entsprechend dem elektrostatischen Potential auf ihrer Oberfläche eingefärbt. Barnase bindet außerhalb von Zellen an negativ geladene RNA. Daher ist Barnase stark positiv geladen und dessen intrazellulärer Inhibitor Barstar dann wieder stark negativ (wie RNA).

Experiment: untersuche Protein-Protein-Assoziation mit atomistischen MD-Simulationen

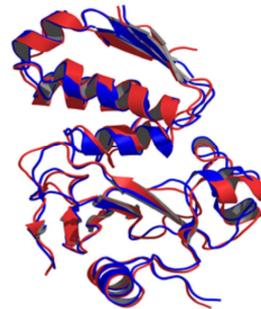


Schiebe Barnase:Barstar-Kristallkomplex um 1.3 – 2.0 nm auseinander + rotiere Barstar um bis zu 45°

9 MD-Simulationen über mehrere 100 ns Dauer.



5 Simulationen näherten sich Kristallkomplex bis auf 1.2 – 3.5 Å innerhalb von 100 ns



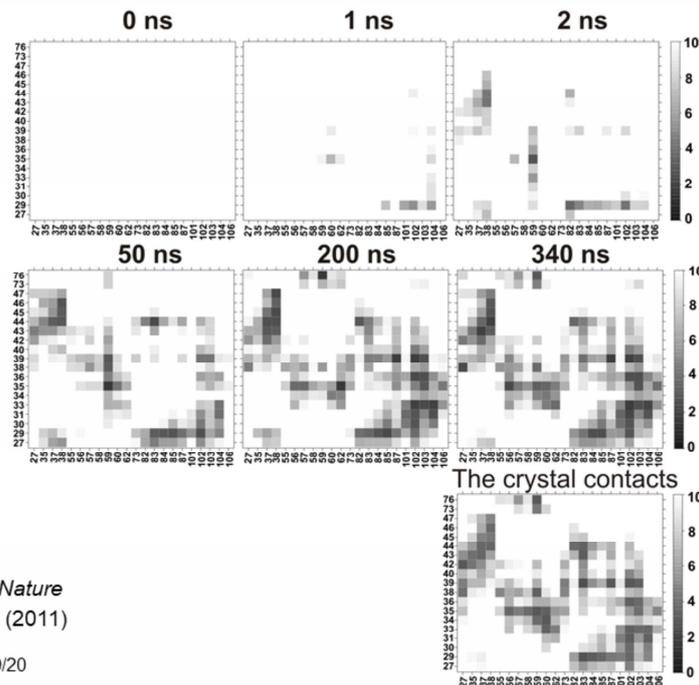
Rot: Schnappschuss aus Simulation (340 ns)

**Blau: Kristallstruktur
RMSD= 1.2 Å**

Ahmad *et al.*, *Nature Comm.* 2, 261 (2011)

Mit Moleküldynamiksimulationen kann man die Assoziation der beiden Proteine simulieren. Wenn man die beiden Proteine aus ihrer Konformation in der gebundenen Kristallstruktur etwas auseinanderzieht (1.3 bis 2.0 Nanometer Entfernung) und nur leicht gegeneinander verdreht, assoziieren sie wieder innerhalb von weniger als einer Mikrosekunde.

Schnelle Ausbildung von Kontakten zwischen Proteinen



Ahmad *et al.*, *Nature*
Comm. 2, 261 (2011)

7. Vorlesung WS 19/20

27

Hier sieht man, wie sich während der MD-Simulation die Kontakte zwischen den Kontaktresiduen ausbilden. Unten ist das Kontaktmuster (Distanzmatrix, siehe V5) in der Kristallstruktur gezeigt. Die ersten Kontakte bilden sich innerhalb weniger Nanosekunden. Es herrscht also eine starke Anziehung zwischen den entgegengesetzt geladenen Proteinen. Dann rütteln sich beide Proteine am Interface noch zurecht, bis die optimale Konformation eingenommen wird.

Energetik der Assoziation hydrophiler Proteine

JCTC

Journal of Chemical Theory and Computation

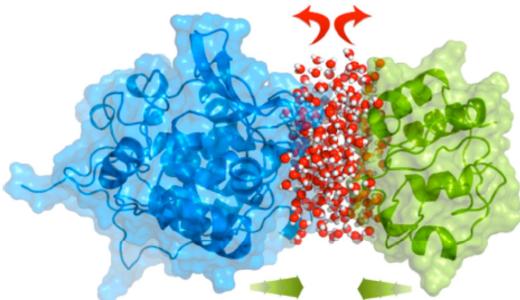
Article

pubs.acs.org/JCTC

Energetics of Hydrophilic Protein–Protein Association and the Role of Water

Özlem Ulucan, Tanushree Jaitley, and Volkhard Helms*

* Center for Bioinformatics, Saarland University, Saarbrücken, Germany



7. Vorlesung WS 19/20

Softwarewerkzeuge

28

Das ist ein link zu diesem Paper: <https://pubs.acs.org/doi/10.1021/ct5001796>

Im gerade gezeigten Beispiel wurden freie Simulationen verwendet, bei denen sich beide Proteine frei bewegen können. Nun möchten wir die Energieoberfläche der beiden Proteine als Funktion des Abstandes berechnen. Dabei verwenden wir Umbrella Sampling. Ein Protein wird dabei in seinem Massenschwerpunkt festgehalten. Auf den Massenschwerpunkt des zweiten Proteins wirkt ein harmonisches Federpotential, das das Protein bei einem bestimmten Abstand halten möchte. Man misst, wie stark die mittlere Auslenkung in diesem Potential ist. Daraus ergibt sich die Anziehung der Proteine bei diesem Abstand. Man führt viele getrennte Simulationen bei unterschiedlich großen Abständen durch und integriert dann das Ergebnis auf.

Energetik der Assoziation hydrophiler Proteine

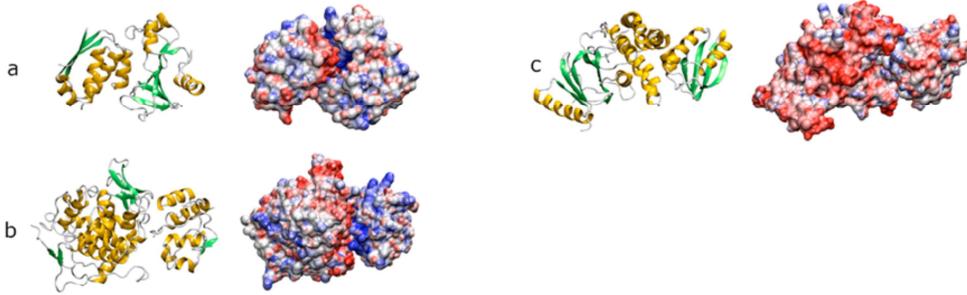


Table 1. Some Global and Interface Properties of the Three Protein–Protein Complexes^a

	BN-BS	CC-CYP	EIN-HPr
no. of amino acids in protein I	110	108	249
no. of amino acids in protein II	89	296	85
area of binding interface (Å ²)	778	570	1002
no. of interface residues in protein I	16	13	33
no. of interface residues in protein II	14	10	24
no. of H-bonds across interface	14	4	6
no. of salt-bridges across interface	12	2	5
total charge of protein I [e]	+2	-7	-19
total charge of protein II [e]	-6	+6	-2
total charge of interface I [e]	+3	+5	-5
total charge of interface II [e]	-4	-2	+4
binding constant	$1.3 \times 10^{-14} \text{ M}^{-1b}$	$6 \times 10^{-7} \text{ M}^{-1c}$	$3.1 \times 10^{-6} \text{ M}^{-1d}$

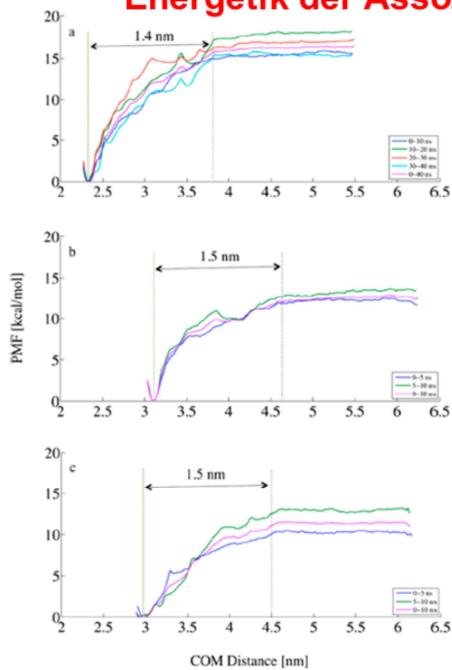
7. Vorlesung WS 19/20

Softwarewerkzeuge

29

Wir haben diesmal drei verschiedene Proteinkomplexe simuliert (Barnase–Barstar, cytochrome c–cytochrome c peroxidase, und die N-terminale Domäne von enzyme I mit histidine-containing phosphocarrier) um deren Gemeinsamkeiten und Unterschiede herauszufinden. Alle drei haben recht kleine Interfaces und die Interfaces sind entgegengesetzt geladen (rote Kreise). Die Gesamtproteine EIN und HPr jedoch nicht (pinker Kreis).

Energetik der Assoziation hydrophiler Proteine



Führe Simulationen bei festen Abständen beider Proteine durch, wobei ein Protein durch ein Federpotential festgehalten wird und das andere fest positioniert ist.

Messe die mittlere Auslenkung vom Nullpunkt des Federpotentials -> Anziehende bzw. abstoßende Wechselwirkung.

Energieprofil ist „downhill“.

Es gibt keine Energiebarriere bei nahen Abständen.

3 verschiedene Protein-Protein-Systeme verhalten sich sehr ähnlich.

7. Vorlesung WS 19/20

Softwarewerkzeuge

30

Die Energieoberfläche verläuft „downhill“ als Funktion des relativen Abstands der Massenschwerpunkte. Der gebundene Zustand liegt bei etwa 2.5 nm Abstand (oben) bzw. 3 nm (Mite und unten), da eines der beiden unteren Proteine jeweils größer ist als oben. Die Energieprofile verlaufen sehr ähnlich und haben eine Reichweite von etwa 1.5 nm. Wenn die Proteine weiter voneinander entfernt sind als 1.5 nm, „sehen“ sie sich als nicht. Die Tiefe des Potentialtopfes entspricht (in etwa, bis auf entropische Korrekturen) der freien Bindungsenthalpie ΔG_{bind} .

Können Liganden an Protein-Protein-Schnittstellen binden?

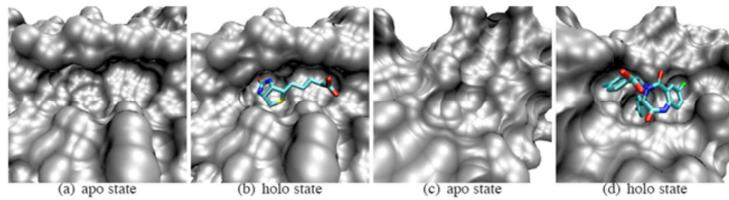


Figure 1.8: Ligand-induced conformational changes in enzyme active sites and protein-protein interaction interfaces: the binding sites of Biotin on the enzyme Streptavidin (a, b) and of the SMPPH DIZ on the protein MDM2 (c, d). (a) and (b) possess an almost identical binding site, whereas the pockets in (c) and (d) show noticeable differences.

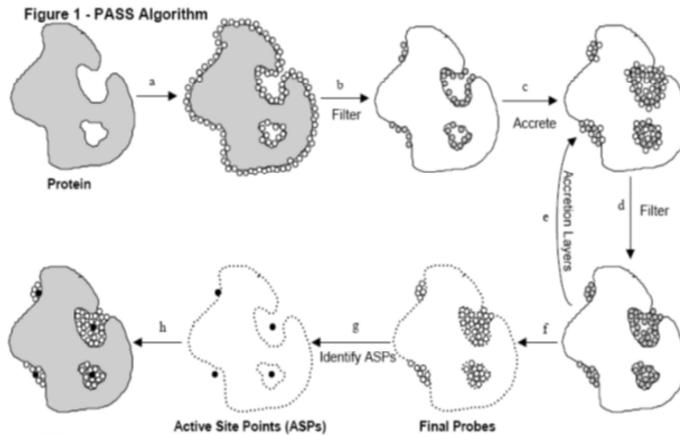
- Interfaces **zu flach**
→ Bindungstaschen?
- Interfaces **zu groß** (~1500 Å²)
→ können kleine Moleküle Interaktion inhibieren?
- Interfaces besitzen keine ausgeprägten **Features**
→ kann Spezifität erreicht werden?
- Natürliche Liganden sind Proteine; Interfaces werden von verschiedenen Abschnitten der Proteinkette gebildet
→ Liganden mimic?

H. Yin, A.D. Hamilton: Strategies for Targeting Protein-Protein Interactions with Synthetic Agents (Angew. Chem. Int. Ed. 2005, 44, 4130-4163)

Nun schauen wir uns die Bindung kleiner Ligandenmoleküle auf Proteinoberflächen an. Es gibt die “einfachen” Fälle, wie z.B. das in (a) und (b) gezeigte Streptavidin-Enzym. Bereits im ungebundenen Zustand (a) besitzt es eine wohlgeformte Bindungstasche für seinen Liganden Biotin. Dieser bindet dann genau in diese Tasche hinein (b). In anderen Fällen wie dem Protein MDM2 gibt es im ungebundenen Zustand fast keine Tasche (c). MDM2 ist übrigens ein Inhibitor von p53. Bei Bindung des Liganden DIZ bildet sich per “induced fit” eine ähnliche tiefe Tasche wie beim Biotin (d).

Dies ist eine typische Situation, wenn man small molecule inhibitors für Protein-Protein-Wechselwirkungen entwerfen möchte. Wir haben uns daher gefragt, wie man solche Taschen trotzdem identifizieren kann.

Taschen-Detektion mit PASS-Algorithmus



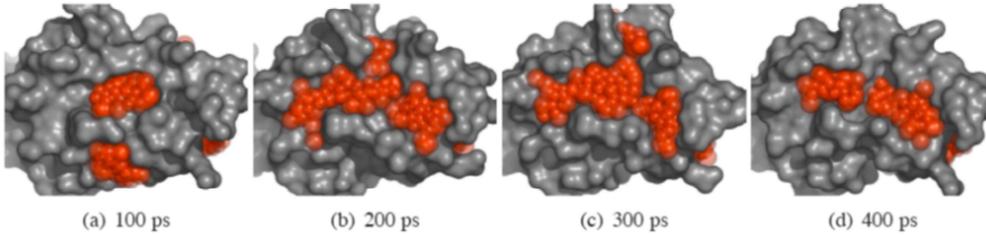
a. PASS uses three-point geometry to coat the protein with an initial layer of spherical probes. b. These probes are filtered to eliminate those that (i) clash with the protein, (ii) are not sufficiently buried, and (iii) lie within 1Å of a more buried probe. c. A new layer of spheres (white) is accreted onto a scaffold consisting of all previously-identified probes (shaded). d. The probes are filtered as described in step b. e. Accrete a new layer of spheres onto the existing probes, as in step c. f. Accretion and filtering (steps e and d) are repeated until a layer is encountered in which no newly-found probes survive the filters. This leaves the final set of probe spheres. g. Probe weights (PW) are computed for each sphere and active site points (ASPs) are identified from amongst the final probes. h. The final PASS visualization is produced. By default, the final probe spheres are first smoothed, leaving only clusters of four or more.

G.P. Brady, P.F.W. Stouten: Fast Prediction of Protein Binding Pockets with PASS, J. Comp. Aid. Mol. Des. (2000) 14, 383-401

Wir haben den PASS-Algorithmus verwendet, mit dem man Taschen/pockets auf und in Proteinstrukturen identifizieren kann. Es gibt viele ähnliche Tools.

Proteinoberflächen verhalten sich wie Flüssigkeit!

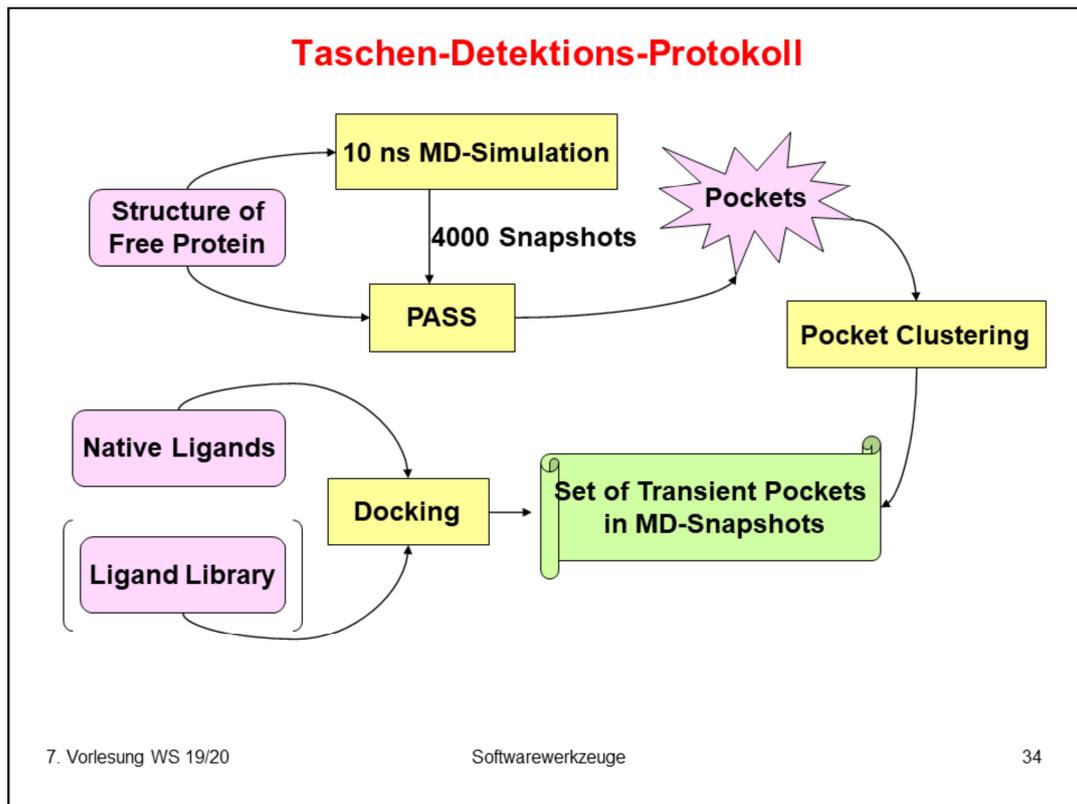
Schnappschüsse aus einer MD-Simulation des MDM2-Proteins



Rote Kugeln: mit PASS-Algorithmus bestimmte Taschen.

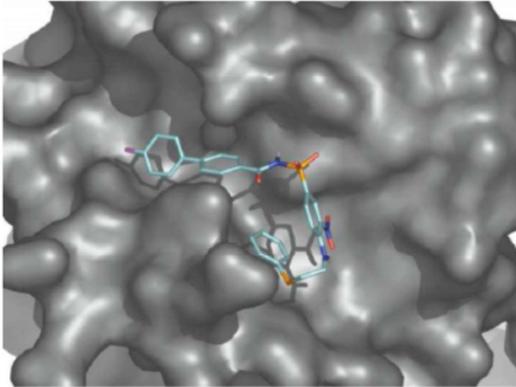
Schnappschüsse folgen in Abständen von 100 Pikosekunden.

In <https://pubs.acs.org/doi/10.1021/jm070095g> haben wir MD-Simulationen für das MDM2-Protein durchgeführt und mit dem PASS-Algorithmus in Schnappschüssen während der Simulation transienten Taschen (rot markiert) gefunden. Diese bilden sich und verschwinden wiederum in Bruchteilen von Nanosekunden. Die Oberfläche solch eines Proteins ähnelt einem Korallenriff oder einem Schwamm.

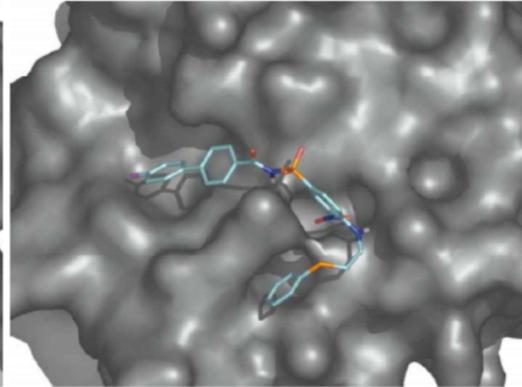


Mit einem MD / Pass-Protokoll haben wir dann zunächst eine große Anzahl an Taschen charakterisiert und dann mit Liganden-Docking versucht, die bekannten Liganden in die Taschen zu docken.

Docking an Bcl-X_L



(a) BCL-X_L (native binding mode)



(b) BCL-X_L (docking pose)

Eyrich, Helms,
J.Med.Chem. (2007)
50, 3457-3464

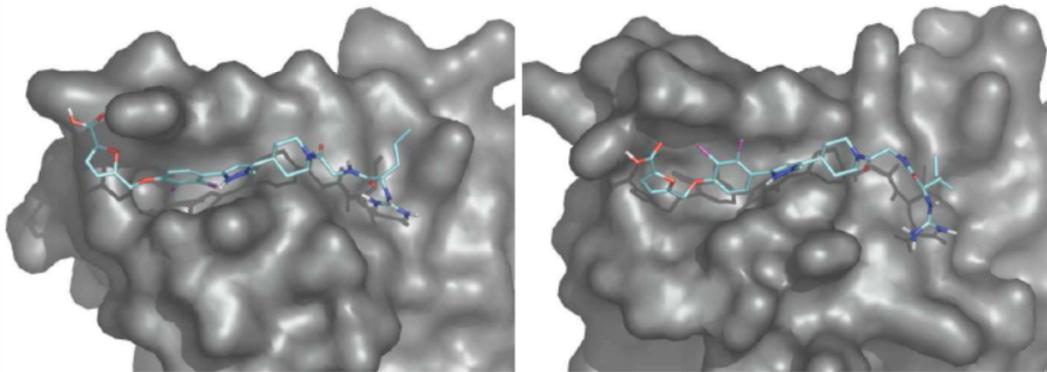
7. Vorlesung WS 19/20

Softwarewerkzeuge

35

Wie man sieht, funktioniert dies recht gut. Links ist die Kristallstruktur eines Komplexes von Bcl-XI mit einem Inhibitor gezeigt, rechts die dazu ähnlichste Docking-Konformation.

Docking an IL-2



(c) IL-2 (native binding mode)

(d) IL-2 (docking pose)

Eyrich, Helms,
J.Med.Chem. (2007)
50, 3457-3464

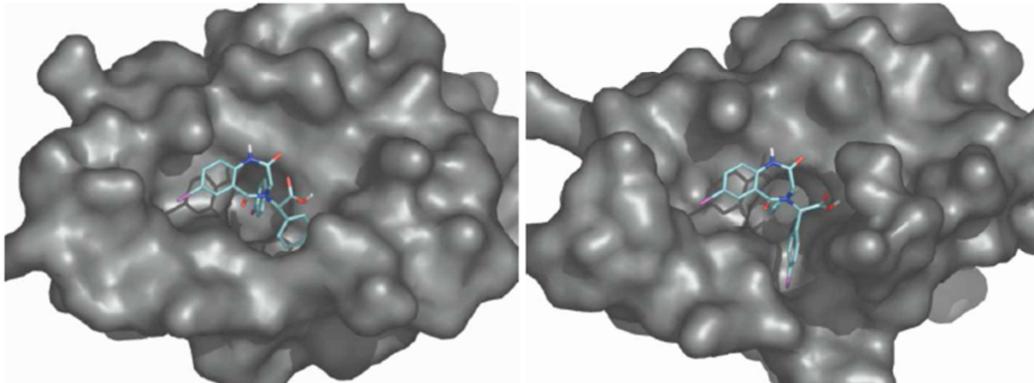
7. Vorlesung WS 19/20

Softwarewerkzeuge

36

Dasselbe für Interleukin 2.

Docking an MDM2



(e) MDM2 (native binding mode)

(f) MDM2 (docking pose)

Eyrich, Helms,
J.Med.Chem. (2007)
50, 3457-3464

7. Vorlesung WS 19/20

Softwarewerkzeuge

37

Und für MDM2. Diese Vorlesung rundet den Teil der „Softwarewerkzeuge-Vorlesung“ über Proteinstruktur ab. Im dritten Teil der Vorlesung werden wir uns dann mit der Analyse von Hochdurchsatzdaten beschäftigen.