

Aus dem Institut für Biochemie
der Medizinischen Fakultät der Charité – Universitätsmedizin Berlin

Habilitationsschrift

Bioinformatische Verfahren bei der Entwicklung von Proteinliganden

zur Erlangung der Lehrbefähigung (venia legendi)
für das Fach
Biochemie

vorgelegt der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

von

Herrn Dr. Robert Preißner
geboren am 1.8.1962 in Berlin

Dekan: Prof. Dr. med. Martin Paul

eingereicht: 07/2005

Gutachter: 1. Prof. Dr. J. Selbig, Potsdam

2. Prof. Dr. K. Schulze-Osthoff, Düsseldorf

Widmung

Meinen Kindern

1. Inhaltsverzeichnis

WIDMUNG	2
1. INHALTSVERZEICHNIS	3
2. ZUSAMMENFASSUNG	5
3. EINLEITUNG	6
4. ERGEBNISSE	8
4.1. Proteinstruktur	8
4.1.1. Columba: eine integrative Datenbank von Proteinen, Strukturen und Annotationen. .	8
4.1.2. Inverse Sequenzähnlichkeit in Proteinen in Relation zum dreidimensionalen Faltungstyp.	9
4.1.3 Inverse Sequenzähnlichkeit zwischen Proteinen impliziert keine strukturelle Ähnlichkeit.	9
4.1.4. Loops In Proteins (LIP): eine umfassende Loop-Datenbank für den Homologie- Modellbau.....	10
4.1.5. Voronoi-Zelle – eine neue Fehler-vermeidende Methode zur Volumen- und Dichte- Berechnung in Proteinen.	11
4.1.6. Alpha-Helices terminiert durch Glycin: Identifikation gemeinsamer struktureller Motive.	12
4.2. Wechselwirkungsflächen in Proteinen	13
4.2.1. Dictionary of Interfaces in Proteins (DIP) – eine Datenbank komplementärer molekularer Oberflächen.	13
4.2.2. Homonyme und Synonyme im Wörterbuch der Wechselwirkungsflächen in Proteinen (DIP).	14
4.2.3. Ersatzteile für Helix-Helix-Packungen.	14
4.2.4. Beschleunigte 3D-Suche in Proteindaten mit graphentheoretischen Ansätzen.	15
4.2.5. Inhomogene molekulare Dichte: Referenzwerte für Packungsdichte und die Verteilung von Kavitäten in Proteinen.	15
4.2.6. Gegabelte Wasserstoff-Brücken in Proteinen.	16
4.3. Das Proteasom	17
4.3.1. Die gemeinsame RTP-Bindungsstelle von HIV-1 Tat und 11S Regulator ist essentiell für deren Effekte auf das Proteasom einschließlich Antigen-Prozessierung.	17
4.3.2. Konvergenz atomarer Substrukturen in den Wechselwirkungsflächen der Sekundärstrukturen proteosomaler Untereinheiten.	18
4.3.3. Ein umfassender Überblick über proteasomale Sequenzen: Schlussfolgerungen zur Evolution des Proteasoms.	19
4.3.4. Sequenzinformationen innerhalb der proteasomalen Prosequenzen vermitteln die effiziente Integration der beta-Untereinheiten in das 20S Proteasom.	20

4.4. Membran-Proteine	21
4.4.1. Molekulare Packung und Packungsdefekte in helikalen Membran-Proteinen.....	21
4.4.2. Strukturelle Charakteristika von Trans-Membran-Helices	21
4.4.3. Analyse und Vorhersage von Helix-Helix Interaktionen in Membran-Kanälen und Transportern	22
4.5. Lokale Sequenzmuster	23
4.5.1. Vorhersage von Prolin in cis-Konformation in Proteinstrukturen auf Basis der Aminosäure-Sequenz.	23
4.5.2. Alpha-Helices terminiert durch Glycin: Erkennung durch Sequenzmuster.	23
4.5.3. Konservierung von cis-Prolin Bindungen in Proteinen im Verlauf der Evolution. ...	24
4.6. Liganden und Medikamente	25
4.6.1. Struktureller Vergleich organischer Bibliotheken mit Protein-Substrukturen.....	25
4.6.2. SuperLigands: eine durchsuchbare Liganden-Datenbank der Proteinstrukturdatenbank.	25
4.6.3. SuperDrug: eine Konformationsdatenbank WHO-klassifizierter Medikamente.	26
4.6.4. Vergleich von 2D-Ähnlichkeit und 3D-Alignment: Anwendung zur Suche in einer Konformationsdatenbank von Medikamenten.	27
4.6.5. In silico Screening von Medikamenten-Datenbanken nach TSE-Inhibitoren.....	27
5. DISKUSSION	29
6. FAZIT UND AUSBLICK	32
7. LITERATUR	33
8. DANKSAGUNG	37
9. EIDESSTATTLICHE VERSICHERUNG	38
10. EIGENE PUBLIKATIONEN.....	39
10.1. Eigene Originalarbeiten (peer-reviewed).....	39
10.1.1. Erst-/Letztautorschaften	39
10.1.2. Co-Autorschaften	41
10.2. Patente	43
11. ANHANG	44

2. Zusammenfassung

In der postgenomischen Phase der biomedizinischen Forschung sind die Proteine, insbesondere als Zielstrukturen für viele Medikamente, zunehmend in den Brennpunkt des wissenschaftlichen Interesses gerückt. Insofern sind hochwertige und schnell verfügbare Informationen über Evolution, Stabilität, Dynamik und Wechselwirkungen der Proteine von eminenter Bedeutung. In diesem Zusammenhang wurden entsprechende Verfahren und Datenbanken erarbeitet, die international genutzt werden.

Dabei spielen 3D-Strukturen sowohl der Proteine als Zielstrukturen, als auch der Liganden als Leitstrukturen für Medikamente eine außerordentliche Rolle. Zur Generierung verlässlicher Zielmolekül-Strukturen wurden neue Methoden erarbeitet. Die Berücksichtigung der Liganden-Flexibilität über explizite Konformere ermöglichte die Implementierung schneller Verfahren. Dadurch konnten für das *in silico* Screening durchsuchbare 3D-Datenbanken aufgebaut werden, die Millionen verfügbarer Naturstoffe sowie synthetischer Substanzen enthalten. Eine von uns generierte Web-basierte Medikamentendatenbank, die über eine schnelle Ähnlichkeitssuche die direkte Verbindung zu den medizinischen Zielstrukturen herstellt, ermöglicht durch die Einbindung der WHO-definierten Indikations-Kodierung (ATC-codes) eine breite Nutzung.

Das Struktur-gestützte Design von Peptid-Bibliotheken, die auch nichtlineare Bindungsstellen abbilden, ist etabliert worden. Damit ergeben sich neue Möglichkeiten, peptidische Binder für Proteine aufzufinden. Um die ungünstigen pharmakologischen Eigenschaften der gefundenen Peptide zu verbessern, wurden automatisierte Verfahren zum Design von Peptid-Mimetika entwickelt. Der Einsatz von konformationellen Photo-Schaltern sowohl in biologisch wirksamen Peptiden als auch in kleinen, nicht-peptidischen Inhibitoren ist durch von uns entwickelte Verfahren erheblich vorangekommen und erschließt diverse neue medizinische Applikationen.

Neben der Entwicklung bioinformatischer Verfahren hat deren Anwendung auf konkrete, experimentell validierbare Projekte maßgeblich zum Erfolg beigetragen. Dabei konnten Protein-Protein-Interaktionen durch Peptid-Bibliotheken aufgeklärt werden und Liganden für verschiedene Zielmoleküle vorgeschlagen werden, die bereits in mehreren Fällen zu Patentanmeldungen geführt haben.

3. Einleitung

Die Röntgenstrukturaufklärung und insbesondere die annähernd industrielle Exprimierung, Reinigung, Kristallisierung und Synchrotron-gestützte Röntgenbeugung¹ sowie NMR haben einen riesigen Bestand an öffentlich zugänglichen, atomar aufgelösten 3D-Daten von etwa 30000 Proteinstrukturen im Komplex mit fast doppelt so vielen Liganden geschaffen. Dabei handelt es sich um etwas mehr als 5000 verschiedene, kleine Liganden, von denen ca. 750 zugelassene Medikamente oder Derivate davon sind. Damit wird der unmittelbare Bezug der Strukturforschung zur medizinischen Forschung deutlich, denn nur durch diese Komplexstrukturen kann die Beziehung zwischen Zielstruktur (Target) und Effektor (Drug) letztendlich detailliert untersucht werden.

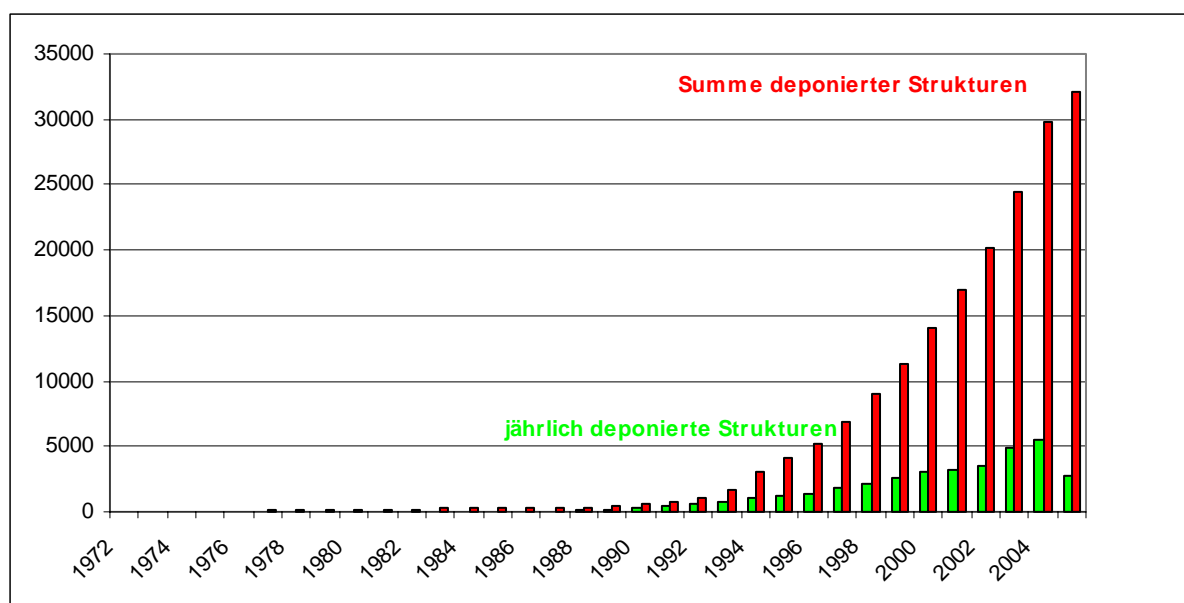


Abb. 1: Entwicklung des Datenbestandes der Proteinstrukturdatenbank (PDB)².

Methoden zur Vorhersage von Proteinstrukturen stehen weiterhin im Mittelpunkt des Interesses, da viele (humane) Proteine schwer kristallisierbar sind^{3, 4}. Dabei sind membranständige Rezeptoren von besonderem Interesse, denn diese pharmakologisch besonders relevanten Proteine sind strukturell bisher weitgehend unbekannt. Ein tiefgehendes Verständnis für die strukturellen und funktionellen Besonderheiten von Membranproteinen ist besonders wichtig, um erste Erfolge in der Strukturaufklärung z.B. durch gezielte Veränderungen der Löslichkeit⁵ ausbauen zu können.

Bisher werden nur einige Hundert der möglichen Proteine als Targets für therapeutische Ansätze genutzt⁶, obwohl NMR-Untersuchungen die meisten Proteine als medikamentös beeinflussbar („druggable“) darstellen⁷. Zurzeit werden große Anstrengungen unternommen, um nicht nur Proteine, die ohnehin kleine Moleküle binden, sondern auch Protein-Protein-Wechselwirkungen zu adressieren⁸. Die Methoden, um Liganden zu entwickeln, die in den Hot-Spot-Bereichen von Protein-Interaktionsflächen binden, befinden sich noch in frühen Entwicklungsphasen⁹.

Ebenso lässt sich feststellen, dass einige Millionen chemischer Substanzen zur Verfügung stehen, die auf Grund ihrer physiko-chemischen Eigenschaften als Medikamente in Betracht kämen (Leads), woraus sich eine experimentell nicht zu beherrschende Vielfalt von etwa 5000 Targets und jeweils 60 Millionen möglichen Leitstrukturen ergibt¹⁰. Aus diesen Überlegungen wird die Bedeutung des *in silico* Screening als Ergänzung zum *in vitro* High-Throughput-Screening (HTS) sichtbar. Dafür bedarf es schneller, angereicherter Datenbanken, um bereits frühzeitig irrelevante Substanzen mit geeigneten Filtern auszuschließen¹¹, so dass aufwändigere Verfahren, wie beispielsweise Docking¹², für die letztendliche Auswahl experimentell zu untersuchender Substanzen zum Einsatz kommen können. Dabei spielen wechselwirkende molekulare Oberflächen, die auch im Zentrum unserer Arbeiten stehen, sowohl bei Protein-Liganden- als auch bei Protein-Protein-Erkennungsprozessen eine entscheidende Rolle¹³.

Neben der Entwicklung neuer Algorithmen¹⁴ oder der Generierung integrierter Datenbanken¹⁵ besteht eine zentrale Aufgabe und Stärke der modernen Bioinformatik in der engen Interaktion mit experimentell arbeitenden Gruppen aus Molekularbiologie¹⁶, Strukturaufklärung¹⁷, Chemie¹⁸ bzw. Molekularer Medizin^{19, 20}. Im Umfeld medizinischer Forschung an der Charité ist es möglich, diesem internationalen Trend Rechnung zu tragen und die kontinuierlich ansteigenden Erfolge in Form von Publikationen und Patenten zeigen, dass die Bioinformatik ein wichtiges Bindeglied zwischen Theorie und Experiment sein kann.

4. Ergebnisse

4.1. Proteinstruktur

4.1.1. Columba: eine integrative Datenbank von Proteinen, Strukturen und Annotationen.

- Trissl S, Rother K, Mueller H, Steinke T, Koch I, **Preissner R**, Frömmel C, Leser U. Columba: an integrated database of proteins, structures, and annotations. BMC Bioinformatics. 6, 81 (2005)

Strukturelle und funktionelle Forschung erfordert oftmals die Generierung geeigneter Sätze von Proteinstrukturen auf Basis bestimmter Eigenschaften. Dazu gehören sowohl Sequenzcharakteristika als auch Klassifikationen bezüglich Faltung oder Funktion. Die Zusammenstellung solcher Datensätze ist aufwändig, da Recherchen in verschiedenen Datenbanken angestellt und zusammengeführt werden müssen. Um diesen Prozess zu erleichtern, haben wir eine integrierte Datenbank erstellt. Columba integriert zur Zeit 12 Datenquellen. Darunter sind die Proteinstrukturdatenbank (PDB), eine Datenbank biochemischer Stoffwechselwege (KEGG), eine annotierte Sequenzdatenbank (Swiss Prot) sowie Klassifikationsdatenbanken bezüglich Strukturklasse, Faltungstyp und Enzymdaten (CATH, SCOP, Gene Ontology, ENZYME). Unsere Datenbank kann per Schlüsselwort-Suche oder via Quellen-spezifischem Web-Formular durchsucht werden. Dadurch können Nutzer schnell und einfach PDB-Datensätze selektieren und herunterladen, die z.B. in einem bestimmten Stoffwechselweg wichtig sind, einen definierten Faltungstyp und eine bestimmte Funktion haben und deren Strukturbestimmung gewissen Qualitätsanforderungen genügt. Die Ergebnisse solcher Anfragen werden sowohl in Nutzer-freundlichem Webformat als auch in Maschinen-lesbarem Format zur automatisierten Weiterverarbeitung angeboten. Die COLUMBA-Datenbank erlaubt die einfache und zuverlässige Generierung von Proteindatensätzen für Struktur-basierte Studien. Ein solches System existierte bisher nicht und wird die Arbeit für kleine und große Datensätze wesentlich effizienter gestalten. Das Web-Interface ist erreichbar unter <http://www.columba-db.de>.

4.1.2. Inverse Sequenzähnlichkeit in Proteinen in Relation zum dreidimensionalen Faltungstyp.

- **Preissner R.**, Goede, A, Michalsky, E, Frömmel, C. Inverse sequence similarity in proteins and its relation to the three-dimensional fold. FEBS Lett. 414, 425-429 (1997)

Derzeit ist die erfolgreichste Strategie zur Vorhersage der Tertiärstruktur von Proteinen die Homologie-Modellierung unter Verwendung bekannter Raumstrukturen. Eine gute Chance auf ein verlässliches Modell besteht nur, wenn zwischen Ausgangs- (Template) und Zielstruktur eine signifikante Sequenzähnlichkeit besteht (z.B. 27% über 100 Aminosäuren). In dieser Analyse untersuchen wir das Phänomen inverser Sequenzähnlichkeit in Proteinen und seine strukturelle Bedeutung. In Proteinsequenzdatenbanken haben wir dieses Phänomen erstaunlich häufig und ausgeprägt aufgefunden (bis zu 34% über 204 Aminosäuren). Bei Untersuchung der invers-ähnlichen Protein Paare mit bekannter Struktur oberhalb der sonst geltenden Grenze für Strukturähnlichkeit konnten wir in der Regel keinen gleichen oder ähnlichen Faltungstyp feststellen. Aus dieser Analyse schlussfolgern wir, dass die direkte Verwendung inverser Sequenzähnlichkeit für die Strukturvorhersage – auch jenseits der Grenze für „sichere Ähnlichkeit“ – nicht möglich ist.

4.1.3 Inverse Sequenzähnlichkeit zwischen Proteinen impliziert keine strukturelle Ähnlichkeit.

- Lorenzen S., Gille C., **Preissner R.** and Frömmel C. Inverse sequence similarity of proteins does not imply structural similarity. FEBS Lett. 545, 105-109 (2003)

Es gibt eine Debatte über die Faltbarkeit inverser Proteinsequenzen. Theoretische Arbeiten und Experimente kommen zu unterschiedlichen Ergebnissen. Viele Proteine in der Protein Strukturdatenbank (PDB) zeigen inverse Sequenzähnlichkeit zueinander. In dieser Arbeit untersuchen wir, ob und inwieweit dies mit struktureller Ähnlichkeit korreliert. Zu diesem Zweck haben wir die C-alpha-Atome aller Proteine in normaler und umgekehrter Durchlaufrichtung überlagert und den Grad der Übereinstimmung bzgl. zugeordneter Sekundärstruktur untersucht. Vergleicht man Proteine mit weniger als 50% Sequenzidentität, so zeigen nur 0.5% der invers alignierten Paare einen ähnlichen Faltungstyp (99 von 19073) während 9% der normal alignierten Paare (1731 von 19248) dies zeigen. Diese Beobachtungen unterstützen die Ansicht, dass die Inversion der Sequenz fast immer zu veränderten Faltungseigenschaften führt. Insofern sind inverse Sequenzen eine interessante Kontrollgruppe mit ähnlichen lokalen Sequenzeigenschaften.

4.1.4. Loops In Proteins (LIP): eine umfassende Loop-Datenbank für den Homologie-Modellbau.

- Michalsky, E., Goede, A. and **Preissner, R.** Loops In Proteins (LIP) - a comprehensive loop database for homology modelling. Protein Engineering 16, 1-7 (2003)

Eine der wichtigsten und zugleich schwierigsten Aufgaben beim Protein-Modellbau ist die Vorhersage der Loops (Schleifenstrukturen zwischen den repetitiven Sekundärstrukturen), wie auch aus der Vielzahl existierender Ansätze zu entnehmen ist. Loops In Proteins (LIP) ist eine neue Datenbank, die alle Protein-Segmente der Protein Data Bank (PDB) bis zur Länge von 15 Aminosäuren enthält (~107). In dieser Arbeit untersuchen wir die Verwendbarkeit dieser Datenbank beim Homologie-basierten Modellbau. Das Durchsuchen die von LIP nach einem geeigneten Kandidaten für einen Loop dauert weniger als 1s und die vergleichende Bewertung wenige Minuten. Der Geschwindigkeitsvorteil gegenüber existierenden Ansätzen liegt bei einer Größenordnung. Als Qualitätskriterium wird die mittlere quadratische Abweichung (RMSD) zwischen vorhergesagter und experimentell beobachteter Loop-Struktur gewählt. Für ein international genutztes Test-Set von Loops bis zu einer Länge von 9 Aminosäuren lag dieser Wert unter 1 Angström, bei der Länge 14 unter 2 Angström. Die Ergebnisse wurden im Detail mit einem kürzlich publizierten *ab initio* Ansatz und zwei weiteren renommierten Methoden verglichen. Die LIP-Methode produzierte insgesamt sehr gute Resultate und liefert insbesondere für längere Loops die besten Ergebnisse.

4.1.5. Voronoi-Zelle – eine neue Fehler-vermeidende Methode zur Volumen- und Dichte-Berechnung in Proteinen.

- Goede, A., Preissner, R., Frömmel, C. Voronoi cell – a new error avoiding method for volume and density calculation in proteins. J. Comp. Chem. 19, 1113-1123 (1996)

Bei der Berechnung von Atom-Volumina und der Dichte in Proteinen ist man mit dem Problem schneidender Kugeln konfrontiert. Zur Berechnung lokaler Packungsdichten muss das Volumen zwischen den Atomen entsprechend der Atom-Lokalisierung zwischen den Atomen verteilt werden. Verschiedene Methoden, die auf Voronoi's Idee basieren, die Atome durch Polyeder zu approximieren, wurden für diesen Zweck vorgeschlagen. In dieser Arbeit vergleichen wir Methoden zur Zuweisung des nicht-atomaren Volumens zu verschiedenen Atomen und finden Abweichungen von bis zu 100% für bestimmte Atome. Außerdem zeigt sich, dass die vorgeschlagenen Teilungsebenen zwischen den Atomen nicht regelmäßig die Schnittkreise zwischen den kovalent gebundenen Atomen treffen. Für Atome, die sich deutlich in der Größe unterscheiden (z.B. C-H) führt das zu eindeutigen Fehlzuweisungen des Volumens um bis zu 7%. Verschiedene Algorithmen haben den großen Nachteil, nicht alle Volumina zuzuweisen. Diese kleinen, nicht zugeordneten Polyeder („vertex error“) summieren sich bereits für ein kleines Protein zu etwa 60 Kubik-Angström, was der Deletierung einer kompletten Aminosäure entspräche. Um die erwähnten Probleme bei Volumen- und Dichte-Berechnungen zu beseitigen, haben wir ein neues Verfahren, als Voronoi-Zelle bezeichnet, entwickelt. Dabei ersetzen wir die teilende Ebene zwischen Atomen durch eine gewölbte Fläche, die dadurch charakterisiert wird, dass ihr Abstand von den van der Waals Oberflächen der beiden Atome stets gleich ist. Der Vorteil dieser Definition besteht darin, dass nicht nur die Schnittkreise von Atomen, sondern auch der Schnittkreis von vergrößerten Atomen (für die Berechnung des Lösungsmittel-unzugänglichen Volumens („solvent excluded volume“) um den Wasser-Radius erweitert) getroffen wird. Die resultierenden hyperbolischen Oberflächen, die eine Voronoi-Zelle bilden, können leicht konstruiert und berechnet werden, wobei sie folgende Vorteile aufweisen: keine Fehlzuweisung von Volumen für Atome unterschiedlicher Größe, geometrisch plausible Zuweisung innerhalb der Atome und verbesserte Anwendbarkeit auf Wasser-zugängliche Protein-Oberflächen.

4.1.6. Alpha-Helices terminiert durch Glycin: Identifikation gemeinsamer struktureller Motive.

- **Preissner, R.,** Bork, P. On alpha-helices terminated by glycine. 1. Identification of common structural features. *Biochem Biophys Res Commun.* 180, 660-665 (1991)

Etwa ein Drittel aller Helices endet mit einem Rest, der einen positiven Torsionswinkel ϕ aufweist. Drei Viertel davon sind Glycin-Reste. Diese starke Überhäufigkeit kann mit typischen gegabelten Wasserstoff-Brücken-Mustern erklärt werden, die nur bei positiven ϕ -Winkeln auftreten können und zum Abbruch der Helix führen. Hier stellen wir einen von uns entwickelten Algorithmus vor, der diese Charakteristika in Proteinstrukturen identifiziert. In einem nicht-redundanten Datensatz 79 hochaufgelöster Proteinstrukturen haben wir 456 Helices untersucht, von denen 158 ein Glycin in der entsprechenden Konformation am Ende aufweisen. Eine Untersuchung der evolutionären Konservierung über strukturbasierte multiple Sequenzvergleiche ergab eine deutlich stärkere Konservierung des Glycin-Restes gegenüber der Umgebung.

4.2. Wechselwirkungsflächen in Proteinen

4.2.1. Dictionary of Interfaces in Proteins (DIP) – eine Datenbank komplementärer molekularer Oberflächen.

- **Preissner, R.,** Goede, A. and Frömmel, C. Dictionary of interfaces in proteins (DIP). Data bank of complementary molecular surface patches. J. Mol. Biol. 280, 535-550 (1998)

Molekulare Oberflächen von Proteinen sind verantwortlich für die selektive Bindung von Liganden, für Protein-Protein Erkennung, aber auch für spezifische Interaktionen zwischen Proteinteilen. Diese grundlegende Bedeutung führt dazu, die bisher bekannten inneren Wechselwirkungsflächen als Lernmenge für intermolekulare Erkennungsprozesse beispielsweise von Substraten oder Co-Enzymen aufzufassen. Darüber hinaus werden Vorhersagen von Kontakten möglich, die bei Protein-Faltung und -Assoziation auftreten. Zu diesem Zweck haben wir eine Definition für Wechselwirkungsfläche („interface“) erarbeitet. Darunter verstehen wir ein Paar passender Oberflächenteile, „patches“, die von räumlich benachbarten Sekundärstrukturen stammen. Alle solche Wechselwirkungsflächen von bekannten Proteinstrukturen haben wir in einer umfassenden Datenbank abgelegt: Data bank of Interfaces in Proteins (DIP). Gegenwärtig enthält die Datenbank einen nicht-redundanten Datensatz von 350 Proteinen, woraus sich etwa 160000 Oberflächenteile ergeben, die von etwa 12500 Sekundärstrukturen gebildet werden. Für spezielle Zwecke können leicht weitere Strukturen hinzugefügt oder Subgruppen gebildet werden. Sowohl die Atomkoordinaten der Oberflächensegmente als auch die Abstände zu den Atomen der benachbarten Sekundärstruktur sind direkt zugänglich. Überraschender Weise ist die generelle Form der Segmente weitgehend unabhängig von der jeweiligen Sekundärstruktur und lässt sich als flacher Körper mit einem Längen-Breiten-Höhen-Verhältnis von 3:2:1 beschreiben. Die relative Orientierung der Segmente zueinander schwankt nur minimal bezüglich der Abstände der Schwerpunkte und der Winkel der Normalen. Das existierende Recherche-System für DIP erlaubt die Selektion der molekularen Oberflächensegmente entsprechend verschiedener Kriterien wie geometrischer Charakteristika, atomare Zusammensetzung, Typ der Sekundärstruktur, Art und Anzahl der Kontakte etc. Ein schneller, Sequenz-unabhängiger 3D-Überlagerungsalgorithmus wurde entwickelt und implementiert, um geometrisch ähnliche Wechselwirkungsflächen zu identifizieren. Damit sind wir in der Lage, eine große Zahl ähnlicher Wechselwirkungsflächen in völlig verschiedenen Proteinen zu identifizieren.

4.2.2. Homonyme und Synonyme im Wörterbuch der Wechselwirkungsflächen in Proteinen (DIP).

- **Preissner R.**, Goede A. and Frömmel C. Homonyms and synonyms in the Dictionary of Interfaces in Proteins (DIP). *Bioinformatics*. 15, 832-836 (1999)

Sind verschiedene Berichte über molekulare Mimikry - z.B. im Zusammenhang mit Kreuzreaktivität - als Zufallsbefunde anzusehen oder steht dahinter ein generelles Prinzip der Protein-Evolution? Um diese Frage zu beantworten, muss man zwei Arten von molekularen Ähnlichkeiten betrachten: Ähnlichkeiten zwischen Homologen (Synonyme) und zwischen Flächen (patches) von nicht-verwandten Proteinen (Homonyme). Alle Wechselwirkungsflächen aus bekannten Proteinen wurden in einer Datenbank (DIP) erfasst. Ein schneller, Sequenz-unabhängiger 3D-Überlagerungsalgorithmus wurde entwickelt, um diese automatisiert nach ähnlichen geometrischen Oberflächen suchen zu können. Überraschenderweise findet man eine große Zahl ähnlicher Flächen auf der Oberfläche nicht-verwandter Proteine. Sogar Oberflächensegmente von unterschiedlichen Sekundärstrukturtypen ähneln einander. Die mögliche Bedeutung dieser Homonyme wird an Hand von überzeugenden Beispielen diskutiert. Die Daten zu den Wechselwirkungsflächen sind - zusammen mit einem Programm zum Betrachten und Manipulieren - verfügbar unter <http://www.charite.de/ch/biochem>. Dadurch sind einzelne Segmente anwählbar und können in verschiedenen Formaten exportiert werden. Solche Daten für weitere Proteine sind auf Anfrage von den Autoren erhältlich, Datenbank-Suchen und -Analysen können als Kooperationsprojekt durchgeführt werden.

4.2.3. Ersatzteile für Helix-Helix-Packungen.

- **Preissner R.**, Goede A. and Frömmel C. Spare parts for helix-helix interaction. *Protein Eng.* 12, 825-832 (1999)

Etwa 6000 Kontakt-Regionen (patches) von Helix-Helix-Packungen aus 300 hochaufgelösten, nicht-homologen Proteinen werden in dieser Arbeit untersucht. Diese Segmente werden durch räumliche Nachbarschaften zwischen den Helices definiert und unter Verwendung eines variablen Abstandskriteriums auf atomarer Basis berechnet. Folgende Fragen werden beantwortet: (1) Sind die Aminosäure-Präferenzen und Atomzusammensetzungen bestimmter Typen helikaler Segmente zur Vorhersage des Typs der Nachbar-Sekundärstruktur geeignet?

(2) Kann man helikale Segmente entsprechend ihrer strukturellen Ähnlichkeit klassifizieren? Zu diesem Zweck wurden die Segmente mit einem automatischen Überlagerungsalgorithmus klassifiziert, woraus ein deutlich reduziertes Set repräsentativer Wechselwirkungsflächen gebildet werden kann. Auf dieser Basis wird die Detektion austauschbarer Segmente in verschiedenen Proteinen dargestellt.

4.2.4. Beschleunigte 3D-Suche in Proteindaten mit graphentheoretischen Ansätzen.

- Frömmel, C., Gille, Ch., Goede, A., Gröpl, C., Hougardy, S., Nierhoff, T., **Preissner, R.** and Thimm, M. Accelerating screening of 3D protein data with a graph theoretical approach. *Bioinformatics* 19, 2442-2447 (2003)

Das Dictionary of Interfaces in Proteins (DIP) ist eine Datenbank, in der 3D-Strukturen von wechselwirkenden Proteinteilen gespeichert sind – so genannten Segmenten (patches). Das Auffinden ähnlicher Segmente ist eine wichtige Suchmöglichkeit. Allerdings dauert das Durchsuchen der gesamten Datenbank Stunden. In dieser Arbeit untersuchen wir, ob Ähnlichkeiten bereits sicher identifiziert werden können, indem nur ein Teil von DIP durchsucht wird. Zu diesem Zweck haben wir die Verteilung der Ähnlichkeit zwischen den Segmenten untersucht. Die Bewertung der Ähnlichkeit erfolgt über einen „Score“, der den Prozentsatz überlagerter Atome (sowie deren Typ) und den Wert der mittleren quadratischen Abstände der überlagerten Atome (RMSD) berücksichtigt. Die Verteilung der Score-Werte in DIP ist trimodal entsprechend der verschiedenen Niveaus der strukturellen Ähnlichkeit. Interessanterweise lassen sich die Isolinien, die die verschiedenen Niveaus trennen mit unabhängigen Methoden reproduzieren. Die Anwendung der Theorie der Zufallsgraphen erlaubt die Ableitung verschiedener struktureller Charakteristika der Daten in DIP. Diese können genutzt werden, um die Suche nach ähnlichen Segmenten zu beschleunigen. Je nach verlangtem Ähnlichkeitsniveau ergeben sich Beschleunigungen um den Faktor 10 bis mindestens 25.

4.2.5. Inhomogene molekulare Dichte: Referenzwerte für Packungsdichte und die Verteilung von Kavitäten in Proteinen.

- Rother, K., **Preissner, R.**, Goede, A. and Frömmel, C. Inhomogeneous molecular density: reference packing densities and distribution of cavities within proteins. *Bioinformatics* 19, 2112-2121 (2003)

Es gibt in der Literatur widersprüchliche Angaben zur Packungsdichte im Innern von Proteinen. Unter Verwendung einer verbesserten Voronoi-Prozedur berechnen wir in dieser Arbeit Referenzpackungswerte für verschiedene Regionen des Proteininnern. Außerdem soll geklärt werden, ob und wo Kavitäten auftreten. Hier stellen wir Sätze von Referenzpackungsdichten in Abhängigkeit vom Abstand zur Oberfläche und zu internen „Löchern“ zur Verfügung. Das Proteininnere ist dicht gepackt, aber nicht homogen. Es gibt im Durchschnitt 4.4 Löcher pro 100 Aminosäuren. Am häufigsten treten diese in einer Tiefe von 2.5-3.6 Angström unterhalb der Connolly-Oberfläche auf. Nichtsdestotrotz ist die tiefste Region der Proteine weniger dicht gepackt als die Umgebung des „Kerns“.

4.2.6. Gegabelte Wasserstoff-Brücken in Proteinen.

- **Preissner, R.,** Egner, U, Saenger, W. Occurrence of bifurcated three-center hydrogen bonds in proteins. FEBS Lett. 288, 192-196 (1991)

Hier unternehmen wir eine detaillierte Analyse der Wasserstoff-Brücken-Geometrie in hochaufgelösten Proteinstrukturen. Dabei haben wir in einem Datensatz von 13 Proteinen festgestellt, dass ein Viertel (1204) aller Wasserstoff-Brücken (4974) gegabelt ist. Das heißt, dass ein Wasserstoff-Brücken-Donor X-H zwei Akzeptoren A1 und A2 gegenübersteht. Systematisch treten diese Wasserstoff-Brücken in alpha-Helices auf, wo etwa 90% von diesem Typ sind. Die Hauptkomponente ist $(n+4)\text{N-H}\cdots\text{O}=\text{C}(n)$, wie für 3.6₁₃ alpha-Helices zu erwarten ist. Die vorherrschende Nebenkomponekte, eigentlich aus 3₁₀-Helices bekannt, ist $(n+4)\text{N-H}\cdots\text{O}=\text{C}(n+1)$. Verzerrungen an den Helix-Enden werden durch typische Wasserstoff-Brücken-Muster stabilisiert. In beta-Faltblättern sind 40% aller Wasserstoff-Brücken gegabelt. Das häufige Auftreten der gegabelten Wasserstoff-Brücken verdeutlicht deren Bedeutung für Struktur, Stabilität und Funktion von Proteinen.

4.3. Das Proteasom

4.3.1. Die gemeinsame RTP-Bindungsstelle von HIV-1 Tat und 11S Regulator ist essentiell für deren Effekte auf das Proteasom einschließlich Antigen-Prozessierung.

- Huang X, Seifert U, Salzmann U, **Preissner R**, Henke W, Sijts AJ, Kloetzel PM, Dubiel W. The RTP site shared by the HIV-1 Tat protein and the 11S regulator subunit alpha is crucial for their effects on proteasome function including antigen processing. *J Mol Biol.* 323, 771-82 (2002)

Das menschliche HI-Virus Protein Tat hemmt die Peptidase-Aktivität des 20S Proteasoms konkurriert mit dem 11S Regulator PA28 um die Bindung an das Proteasom. Strukturelle Vergleiche detektierten eine Ähnlichkeit zwischen dem Tat-Protein und der 11S Regulator-Untereinheit (REG-alpha) als die REG/Tat-Proteasom (RTP) Bindungsstelle. Kinetische Untersuchungen konnten die Aminosäuren Lys51, Arg52 und Asp67 von Tat als RTP-Site bestätigen. Das Pendant im REG-alpha wird durch die Aminosäuren Glu235, Lys236 und Lys239 gebildet. Mutanten nach Ala an diesen Positionen in REG-alpha ergeben Proteine, die REG-Komplexe bilden und an das Proteasom binden, aber keine aktivierende Wirkung mehr auf das Proteasom ausüben. Die REG-alpha RTP-site ist für eine verstärkte Präsentation eines Cytomegalo-Virus-Protein-Epitops aus pp89 durch MHC-Moleküle in Maus-Fibroblasten verantwortlich. Zell-Experimente zeigen, dass die Aminosäuren 37-72 des Tat-Proteins für die Interaktion des viralen Proteins mit dem Proteasom in vivo notwendig sind. Das vollständige Tat-Protein und das Fragment 37-72 unterdrückt die 11S regulierte Präsentation des pp89 Epitops. Im Gegensatz dazu sind Varianten des Tat-Peptids 37-72, bei denen die Aminosäuren Lys51, Arg52 und Asp67 zu Alanin mutiert sind, nicht in der Lage, die Antigen-Präsentation zu reduzieren.

4.3.2. Konvervierung atomarer Substrukturen in den Wechselwirkungsflächen der Sekundärstrukturen proteosomaler Untereinheiten.

- Gille, C., Goede, A., **Preissner, R.**, Rother, K. and Frömmel, C. Conservation of atomic substructures interfaces of secondary structural elements in proteasomal subunits. *J. Mol. Biol.* 299, 1147-1154 (2000)

Es lässt sich beobachten, dass die divergente Evolution zweier Proteine die strukturelle Ähnlichkeit der Hauptketten konserviert, während die Sequenzähnlichkeit auf ein verschwindendes Maß absinkt. In dieser Arbeit untersuchen wir, inwieweit lokale Substrukturen zwischen Sekundärstrukturelementen von dieser evolutionären Konservierung betroffen sind. Als Fallstudie benutzen wir die Strukturfamilie der Proteasomen-Untereinheiten, für die 17 verschiedene Kristall-Strukturen vorliegen. Davon stammen 14 aus *Saccharomyces cerevisiae*, 2 Untereinheiten aus *Thermoplasma acidophilum* und eine Untereinheit aus *E. coli*. Der strukturelle Kern der Untereinheiten besteht aus 23 Sekundärstrukturen. Jeweils zwei davon bilden eine Wechselwirkungsfläche („interface“), bestehend aus zwei komplementären Segmenten („patches“). Wir finden 61 solche Wechselwirkungsflächen, die in allen 17 Proteasomen-Untereinheiten vorkommen und die strukturell deutlich konservierter sind als zu erwarten wäre. Es ist bekannt, dass die Veränderung der Sequenz mit der mittleren quadratischen Abweichung (RMSD) der Hauptketten korreliert. Bildet man einen mittleren patch-RMSD so ist dieser stärker mit der Sequenzänderung korreliert als der RMSD-Wert der Protein-Hauptkette.

4.3.3. Ein umfassender Überblick über proteasomale Sequenzen: Schlussfolgerungen zur Evolution des Proteasoms.

- Gille C, Goede A, Schloetelburg C, **Preissner R**, Kloetzel PM, Gobel UB, Frömmel C. A comprehensive view on proteasomal sequences: implications for the evolution of the proteasome. *J Mol Biol.* 326, 1437-48 (2003)

Das Proteasom ist eine große, multimere, selbst-kompartimentierende Protease, die eine essentielle Rolle bei der Beseitigung falsch gefalteter Proteine, dem Abbau regulatorischer Proteine, der Prozessierung von Proteinen bei der spezifischen partiellen Proteolyse, der Zell-Zyklus-Kontrolle und der Vorbereitung der Peptide zur Immun-Präsentation spielt. Man kann zwei grundlegende Strukturtypen unterscheiden: das 20S Proteasom und das Proteasom-ähnliche Hitze-Schock Protein, das von Hitze-Schock-Lokus V kodiert wird (hslV). Im Allgemeinen ist jedes biologische Reich durch einen spezifischen Typ des Proteasoms gekennzeichnet. Das 20S Proteasom tritt bei Eukaryota und Archebakterien auf, während hslV vornehmlich in Bakterien vorkommt. Um diese Regel zu validieren, haben wir in dieser Arbeit eine umfassende Analyse aller zurzeit vorliegenden Genome vorgenommen. Dabei haben wir eine Reihe von Ausnahmen von dieser Regel gefunden. (1) Protista: zusätzlich zum 20S Proteasom haben *Leishmania*, *Trypanosoma* und *Plasmodium* auch hslV, welches von einem Mitochondrien-Vorläufer stammen könnte. (2) Bakterien: bei *Magnetospirillum magnetotacticum* und *Enterococcus faecium* fanden wir zwei verschiedene Varianten des hslV, was auf horizontalen Gentransfer oder Genduplikation zurückzuführen sein könnte. Durch Einbeziehung von Rohdaten der Sequenzierung konnten wir bestätigen, dass einige bakterielle Genome - offenbar auf Grund von Genverlust - kein Proteasom haben. (3) G+C Gram-positive besitzen eher ein 20S Proteasom als hslV. Der strukturelle Kern des Proteasoms besteht aus zwei verschiedenen, homologen Ketten alpha und beta, die durch Genduplikation in weitere sieben Sub-Typen differenzierten. Durch detaillierte Analyse des Genoms des intrazellulären Keims von *Encephalitozoon cuniculi* waren wir in der Lage, zu zeigen, dass die Differenzierung der beta-Untereinheiten früher erfolgte als die der alpha-Kette. Ein wichtiges methodisches Ergebnis unserer Analyse ist die Notwendigkeit zur Einbeziehung der Rohdaten der Sequenzierung, da sonst wichtige Proteine „übersehen“ werden können. Das Struktur-basierte Alignment der 433 Sequenzen aus 143 Organismen steht auf unserer Website zur Verfügung und wird regelmäßig aktualisiert.

4.3.4. Sequenzinformationen innerhalb der proteasomalen Prosequenzen vermitteln die effiziente Integration der beta-Untereinheiten in das 20S Proteasom.

- Schmidt M, Zantopf D, Kraft R, Kostka S, **Preissner R.** and Kloetzel PM. Sequence information within proteasomal prosequences mediates efficient integration of beta-subunits into the 20 S proteasome complex. J Mol Biol. 288, 117-128 (1999)

Die Reifung von Proteasen wird über Pro-Sequenzen geregelt. Während der Biogenese des hochgradig oligomeren 20S Proteasoms müssen fünf Untereinheiten integriert werden, die Pro-Sequenzen enthalten und entweder durch Autokatalyse oder durch benachbarte Untereinheiten prozessiert werden müssen. In dieser Arbeit untersuchen wir die Bedeutung der proteasomalen Pro-Sequenzen für die Komplex-Bildung, indem wir die Pro-Sequenz der fakultativen Untereinheit beta1i/LMP2 auf neun Aminosäuren kürzen oder ganz deletieren. Zusätzlich werden die geladenen Reste innerhalb der Pro-Sequenz durch neutrale ersetzt. Während die komplette Deletierung der Pro-Sequenz die korrekte Inkorporation nicht behinderte, stören die geladenen Reste in der gekürzten Variante den Einbau, was sich durch Ersatz dieser mit neutralen Aminosäuren kompensieren lässt. Bei der Bildung des Immuno-Proteasoms sind Bildung, Inkorporation und Prozessierung der induzierbaren Untereinheiten ein kooperativer Prozess. Wir finden eine lineare Korrelation zwischen dem Niveau von beta1i/LMP2 und dem von beta2i/MECL1, was eine direkte Interaktion dieser Untereinheiten als Basis für den konzertierten Einbau nahe legt. In der Abwesenheit von beta5i/LMP7 häufen sich Vorläufer-Komplexe mit unprozessiertem beta1i/LMP2 an. Der Beitrag von beta5/LMP7 zur kooperativen Bildung einer homogenen Population des Immuno-Proteasoms ist daher höchstwahrscheinlich auf eine beschleunigte Kinetik der Prozessierung von beta1i/LMP2 und wohl auch beta2i/MECL1 zurückzuführen.

4.4. Membran-Proteine

4.4.1. Molekulare Packung und Packungsdefekte in helikalen Membran-Proteinen.

- Hildebrand, P.W., Rother, K., **Preissner, R.**, Goede, A., Frömmel, C. Molecular packing and packing defects in helical membrane proteins. *Biophys J.* 88, 1970-7 (2005)

Die Packung von Transmembran-Helices ist essentiell für die Stabilität und Funktion alpha-helikaler Membran-Proteine. Durch Verwendung einer modifizierten Voronoi-Prozedur wird es möglich, die Packungsdichte von Transmembran-Domänen zu berechnen. Unsere Ergebnisse zeigen, dass die Helices von Membran-Kanälen und -Transportern deutlich lockerer gepackt sind als diejenigen von globulären Proteinen. Diese gefundenen Packungsdefizite finden auch ihren Ausdruck in einer höheren Anzahl von Kavitäten an funktionell bedeutsamen Punkten. Die Umgebung dieser Hohlräume, die die Poren und Kanäle säumen, wird durch polare Aminosäuren geprägt, die in "offenem Zustand" der Transporter für das wässrige Medium zugänglich sind. Hohlräume, die im Gegensatz dazu, weiter im Inneren liegen und funktionell bedeutsame Umlagerungen (an Scharnieren) erleichtert, sind durch hydrophobe Umgebungen geprägt. Diese Korrelationen stützen die Annahme, dass die geringeren Packungsdichten in Membran-Transportern und -Kanälen für die benötigte Helix-Beweglichkeit beim Transportvorgang essentiell sind.

4.4.2. Strukturelle Charakteristika von Trans-Membran-Helices

- Hildebrand, P.W., **Preissner, R.**, Frömmel, C. Structural features of trans-membrane helices. *FEBS Lett.* 559, 145-51 (2004)

In dieser Arbeit wird der zur Zeit verfügbare Satz von experimentell hochaufgelösten 3D-Strukturen nicht homologer, helikaler Membran-Proteine analysiert. Das sind derzeit 15 Proteinstrukturen mit 160 Transmembran-Helices. Im Unterschied zu löslichen, globulären Proteinen befinden sich diese Bereiche in der hydrophoben Umgebung der Lipid-Doppelschicht und hier sollen daraus resultierende strukturelle Besonderheiten der Helices hinsichtlich der Wasserstoff-Brücken und der Dihedralwinkel untersucht werden. Der Ramachandran-Plot von Protein-Kanälen und -Transportern zeigt eine signifikante Verschiebung bezüglich der phi- und psi-Winkel gegenüber vergleichbaren Helices aus globulären Proteinen von $+4.5^\circ$ und -5.4° . Am C-Terminus von Transmembran-Helices findet

man Motive, die den Gly-Helix-Caps von Helices globulärer Proteine entsprechen. Allerdings finden sich aufgrund der hydrophoben Umgebung andere Sequenzmuster. Die dargestellten Besonderheiten sind für die Erstellung räumlicher Modelle von Membran-Proteinen von Bedeutung.

4.4.3. Analyse und Vorhersage von Helix-Helix Interaktionen in Membran-Kanälen und Transportern

- Hildebrand PW, Lorenzen S, Goede A, **Preissner R**. Analysis and prediction of helix-helix interactions in membrane channels and transporters. *Proteins*. 64, 253-62 (2006)

Membranproteine erfüllen eine Vielzahl verschiedener Funktionen als Zelloberflächen-Rezeptoren, Redoxproteine, Ionenkanäle und Transporter. Membranproteine mit funktionellen Poren zeigen andere Charakteristika der Helix-Helix-Packung als andere helikale Membranproteine. Wir fanden, dass die Helix-Helix-Kontakte in 13 nicht-homologen, hochaufgelösten Membranproteinen mit Transportfunktion überwiegend durch schwach polare Aminosäuren (G>S>T>F) vermittelt werden. Durch die überwiegend rechtshändige Helixkreuzung entstehen typische Kontakte bei jeder vierten Aminosäure.

Darüber hinaus findet sich bei diesen Membranproteinen eine starke Korrelation zwischen der Hydrophobizität und dem Bestreben zum Kontakt mit der Lipidmembran. Die polaren Reste hingegen sind in der Regel zum polaren Lösungsmittel (Kanal) orientiert.

Im Gegensatz dazu fanden wir für 14 nicht-homologe, hochaufgelöste und helikale Membranproteine ohne Transportfunktion auch große und/oder polare Aminosäuren als dominante Kontaktdeterminanten (D>S>M>Q). Dadurch entstehen eher linkshändige Helix-Kreuzungen, die durch Kontakte nach 3.5 Resten charakterisiert sind.

Dementsprechend scheinen in Abhängigkeit von der Funktion unterschiedliche Konzepte der Faltung und Stabilisierung realisiert zu sein. Durch Abbildung dieser Unterschiede über eine Matrix-basierte Methode gelang uns eine Verbesserung der Vorhersage der Helix-Helix-Kontakte um 16% auf 76%.

4.5. Lokale Sequenzmuster

4.5.1. Vorhersage von Prolin in cis-Konformation in Proteinstrukturen auf Basis der Aminosäure-Sequenz.

- Frömmel, C, **Preissner, R.** Prediction of prolyl residues in cis-conformation in protein structures on the basis of the amino acid sequence. FEBS Lett. 277, 159-163 (1990)

In Proteinen befinden sich die meisten Peptidbindungen in trans-Konformation, bei einem Torsionswinkel $\omega=180^\circ$. Nur einige treten in cis-Konformation auf ($\omega=0^\circ$), die meisten davon sind Proline. Etwa 6% der 4000 Prolin-Reste in der Protein Data Bank zeigen cis-Konformation. Die Sequenzumgebungen unterscheiden sich signifikant zwischen den cis- und trans-Prolinen. Beispielsweise treten aromatische Reste gehäuft N-terminal von cis-Prolinen auf, während diese bei trans-Prolinen häufiger C-terminal vorkommen. Aber in allen Fällen handelt es sich um komplexe Muster, die für cis- oder trans-Konformation typisch sind. Aus der Untersuchung der lokalen Umgebung (± 6 Reste) haben wir 6 Muster abgeleitet, mit denen wir 75% aller cis-Proline vorhersagen können, ohne dass eine falsch-positive Vorhersage auftritt.

4.5.2. Alpha-Helices terminiert durch Glycin: Erkennung durch Sequenzmuster.

- Bork, P, **Preissner, R.** On alpha-helices terminated by glycine. 2. Recognition by sequence patterns. Biochem Biophys Res Commun. 180, 666-672 (1991)

In dieser Arbeit haben wir Konsensus-Sequenz-Muster entwickelt, um Helix-Enden mit typischem Wasserstoff-Brücken-Mustern zu beschreiben. Dieses spezielle Muster tritt bei einem Drittel aller Helices auf und ist in der Regel mit Glycin und einem positiven Torsionswinkel phi assoziiert. In der Proteinstrukturdatenbank (PDB) haben wir 575 Helices identifiziert, die diesen Kriterien entsprechen. Ein einfacher Algorithmus zur Clusterung findet sechs Gruppen mit definiertem Eigenschaftsmuster. Mit Hilfe dieser Muster ist es möglich, 87% (501) der Helix-Enden korrekt vorherzusagen, ohne falsch-positive Vorhersagen.

4.5.3. Konservierung von cis-Prolin Bindungen in Proteinen im Verlauf der Evolution.

- Lorenzen, S., Peters, B., Goede, A., **Preissner, R.**, Frömmel, C. Conservation of cis proline bonds in proteins during evolution. *Proteins*. 58, 589-95 (2005)

In Proteinen und Peptiden befindet sich die überwiegende Mehrheit der Peptidbindungen in trans-Konformation. Nur etwa 5% der X-Prolin Bindungen nimmt die cis-Konformation ein. In dieser Arbeit untersuchen wir den Konservierungsgrad von Prolin-Resten evolutionär verwandter Proteinen. Insgesamt finden wir einen hohen Konservierungsgrad für cis-Proline. Dabei gibt die Ähnlichkeit der Proteine in der gesamten Sequenz ein deutlich stärkeres Signal als entsprechende lokale Sequenzmuster. Der Konservierungsgrad von trans-Prolin Resten liegt zwischen dem bei cis-Prolin gefundenen Wert und dem Wert umgebender Aminosäuren. Auch die Austauschmuster unterscheiden sich für cis- und trans-Proline, so dass für Proteinfamilien mit einer Sequenzidentität oberhalb 20% klare Hinweise auf die Konformation der X-Prolin Bindungen gefunden werden können. Dieser Befund unterstreicht die Bedeutung von cis-Peptid Bindungen für die Struktur, Faltung und Funktion von Proteinen.

4.6. Liganden und Medikamente

4.6.1. Struktureller Vergleich organischer Bibliotheken mit Protein-Substrukturen.

- **Preissner, R.**, Goede, A., Rother, K., Osterkamp, F., Koert, U. and Frömmel, C.
Matching organic libraries with protein-substructures. *J. Comp.-Aided Mol. Design.*
15, 811-817 (2001)

Hier präsentieren wir einen allgemeinen Ansatz zur automatischen Identifikation von Substrukturen in Proteinen, die gegebenen 3D-Templates ähneln. Diese Arbeit dokumentiert die erfolgreiche Verwendung nicht-peptidischer Grundkörper („Templates“) wie beispielsweise beta-Turn-Mimetika. Wir vergleichen gut untersuchte Turn-Mimetika wie das bicyclische turned Dipeptid (BTD) und Spiro-Laktam (Spiro) mit dem 2,5-disubstituierte Tetrahydrofuran (THF), einem neuen kürzlich entwickelten und charakterisierten Furan-Derivat. Die detektierte geometrische Ähnlichkeit zwischen den Templates und den Protein-Patches entspricht einer mittleren quadratischen Abweichung (RMSD) von 0.3 Angström, was der Ähnlichkeit aktiver Zentren homologer Proteine entspricht. Dieses automatische Verfahren ist von Wert für biomedizinische Fragestellungen wie das Auffinden spezieller Mimetika für bestimmte Protein-Substrukturen, aber auch für den Entwurf semiartifizieller Peptide oder Proteine („Template-Assembled Synthetic Proteins“, TASP).

4.6.2. SuperLigands: eine durchsuchbare Liganden-Datenbank der Proteinstrukturdatenbank.

- Michalsky, E., Dunkel, M., **Preissner, R.** SuperLigands: a searchable database of ligands from the Protein Data Bank. *BMC Bioinformatics* 6, 122 (2005)

Gegenwärtig enthält die Strukturdatenbank (PDB) etwa 30000 Proteinstrukturen, die ihrerseits etwa 70000 experimentell bestimmte Kleinstrukturen enthalten. Dabei handelt es sich um etwa 5000 verschiedene - überwiegend organische - Liganden mit geringem Molekulargewicht. Informationen über diese Liganden, davon einige Hundert zugelassene Medikamente, sind wichtig für die Modellierung, insbesondere für die Vorhersage von Bindung und Funktion. Hier stellen wir eine neue Datenbank vor, SuperLigands, die per Web-Zugang Informationen über diese Liganden liefert und das Abspeichern der Strukturen unter Anreicherung topologischer und chemischer Informationen erlaubt. Das Recherche-System

erlaubt Ähnlichkeitssuchen (Tanimoto-Koeffizient) und 3D-Überlagerungen - auch mit bekannten Medikamenten - aus der SuperDrug Datenbank. SuperLigands ergänzt die existierenden Informationsquellen der biomedizinischen Forschung um Informationen über PDB-gebundene Kleinstrukturen und eröffnet durch Such-Optionen in Medikamenten und Kombination von 2D- und 3D-Ähnlichkeit wertvolle neue Möglichkeiten.

4.6.3. SuperDrug: eine Konformationsdatenbank WHO-klassifizierter Medikamente.

- Goede, A., Dunkel, M., Mester, N., Frömmel, C., **Preissner, R.** SuperDrug: a conformational drug database *Bioinformatics*. 21, 1751-3 (2005)

Es gibt eine Reihe von Quellen für experimentell bestimmte und berechnete 3D-Strukturen von kleinen Molekülen, aber freie, öffentlich zugängliche Datenbanken für zugelassene Medikamente und deren 3D-Strukturen existierten bisher nicht. Außerdem fehlte eine klare Zuordnung von Indikations-bezogenen Klassifikationen zu Strukturklassen. Zu diesem Zweck haben wir die Anatomisch-Therapeutisch-Chemische Klassifikation (ATC-Code) der Weltgesundheitsorganisation (WHO) benutzt, um eine Datenbank von etwa 2500 3D-Strukturen der wirksamen Komponenten zugelassener Medikamente aufzubauen. Um deren strukturelle Flexibilität zu berücksichtigen, haben wir 100000 repräsentative Konformere berechnet. In dieser Arbeit präsentieren wir das Web-Recherche-System, das Suchen z.B. nach Medikamenten-Namen, Synonymen, Handelsnamen, Trivialnamen, Summenformel, CAS-Nummer und ATC-Code ermöglicht. Darüber hinaus sind Ähnlichkeitssuchen (Tanimoto-Koeffizient) und 3D-Überlagerungen implementiert. Medikamente (z. B. oberhalb einer gewissen Ähnlichkeit) können im MOL-Format exportiert werden.

Die Datenbank ist zugänglich unter <http://bioinformatics.charite.de/superdrug>.

4.6.4. Vergleich von 2D-Ähnlichkeit und 3D-Alignment: Anwendung zur Suche in einer Konformationsdatenbank von Medikamenten.

- Thimm, M., Goede, A. Hougardy, S. **Preissner, R.** Comparison of 2D Similarity and 3D Superposition. Application to Searching a Conformational Drug Database. J. Chem. Inf. Comput. Sci. 44, 1816-1822 (2004)

In dieser Arbeit vergleichen wir die Ergebnisse von 2D-Ähnlichkeitssuchen in einer Datenbank von mehr als 2000 Medikamenten mit denen von 3D-Überlagerungen der entsprechenden 100000 Konformere. Für eine Klasse von Medikamenten unternehmen wir hier eine detaillierte Analyse der Korrelation zwischen struktureller Ähnlichkeit und ähnlicher Wirkung. Insgesamt ergeben die 2D-Suche mittels Tanimoto-Koeffizient und die 3D-Score basierte Suche ähnliche Resultate, allerdings werden wichtige strukturelle und funktionelle Charakteristika durch 2D-Ähnlichkeitsmaße außer Acht gelassen. Beispiele für das Über- und Unterschätzen der Ähnlichkeit durch 2D-Repräsentationen werden im Detail diskutiert. Der erforderliche Mehraufwand für 3D-Überlagerungen wird für unterschiedliche Algorithmen (fast, 0.01s; best, 0.5s) mit dem Qualitätsgewinn verglichen. Aufgrund dessen kommen wir zu dem Schluss, dass – in Anbetracht der Geschwindigkeit moderner PCs – 3D Überlagerungen die kommende Methode der Wahl für Bibliotheksgenerierung und Ähnlichkeitssuchen sein werden.

4.6.5. In silico Screening von Medikamenten-Datenbanken nach TSE-Inhibitoren.

- Lorenzen, S., Dunkel, M., **Preissner, R.** In silico screening of drug databases for TSE inhibitors. Biosystems. 80, 117-22 (2005)

Tausende von Wirkstoffen werden für verschiedene Indikationen eingesetzt. In dieser Arbeit demonstrieren wir die Verwendung einer neuen Konformationsdatenbank zur Suche nach Substanzen mit positivem Einfluss auf Transmissible Spongiforme Enzephalopathien (TSE). Durch Kombination von 2D- und 3D-Ähnlichkeitssuchen mit wirksamen Substanzen haben wir 16 Kandidaten abgeleitet. Die Identifizierung einer kleinen Zahl von bereits zugelassenen Medikamenten für eine mögliche neue Indikation erlaubt die experimentelle Validierung mit vergleichsweise aufwändigen Methoden.

4.6.6. Darstellung der bindungsrelevanten Konformationsdynamik von Medikamenten in Konformationsbibliotheken.

- Gunther, S., Senger, C., Michalsky, E., Goede, A., **Preissner, R.** Representation of target-bound drugs by computed conformers: implications for conformational libraries. BMC Bioinformatics. 7, 293 (2006)

Die wachsende Zahl bekannter Proteinstrukturen stellt wertvolle Informationen über pharmazeutische Zielstrukturen zur Verfügung. Bindungsstellen können im atomaren Detail untersucht und zur Entwicklung neuer Leitstrukturen genutzt werden. Die Flexibilität von Liganden ist ein kritischer Punkt bei der Selektion möglicher Medikamente. Da berechnete 3D-Konformere für Millionen von Substanzen vorliegen, wäre die Kenntnis der tatsächlich an die Zielstruktur bindenden Konformere von großem Vorteil für ein effizientes *in silico* Screening.

In der Proteindatenbank konnten wir annähernd 200 Medikamente identifizieren, die im Komplex mit den Zielstrukturen kristallisiert wurden. So konnten wir einen Vergleich von etwa 5.000 Kristallstrukturen mit 10.000 berechneten Konformeren dieser Medikamente durchführen. Die kraftfeldbasierte Konformerberechnung (<100 Konformere pro Medikament) erfolgte unter dem Gesichtspunkt der optimalen Abdeckung des Konformationsraumes. Für drei Viertel der Medikamente konnte die bioaktive Konformation mit einer Genauigkeit < 1 Angström dargestellt werden. Diese Quote steigt für Medikamente mit höchstens durchschnittlichem Molekulargewicht auf über 90%.

Wir konnten verschiedene Medikamente identifizieren, deren kristallographisch bestimmte Raumstruktur an unterschiedliche Zielmoleküle gebunden vorliegt. Dabei stellte sich heraus, dass in solchen Fällen durchaus verschiedene bioaktive Konformationen eingenommen werden, die einen breiten Bereich des Konformationsraumes und der Energie einnehmen. Trotzdem kann die Mehrzahl der Medikamente durch bis zu 100 Konformere gut abgebildet werden, was für 3D-Ähnlichkeits- oder Pharmakophorsuchen von großer Bedeutung ist.

Die 10.000 berechneten Medikamenten-Konformere und die zugehörigen Überlagerungen mit den Liganden der Proteindatenbank stehen frei zur Verfügung:
<http://bioinformatics.charite.de/superdrug>.

5. Diskussion

Der enorme Datenfluss, der durch Genom-^{21, 22}, Metabolom-^{23, 24} und Proteom-Projekte²² sowie Array- und Hochdurchsatz-Experimente²⁵ entsteht, hat der Bioinformatik einen festen Platz in den Lebenswissenschaften gesichert²⁶. Dabei spielen sowohl physikalisch-chemische Konzepte zur Behandlung von (Makro-)Molekülen als auch Techniken aus Mathematik, Statistik und Informatik eine wichtige Rolle, um Biomoleküle im großen Maßstab behandeln, verstehen und zugehörige Informationen organisieren zu können^{27, 28}.

Die exakte Berechnung der Packungsdichte in Molekülkomplexen ist ein wichtiges Instrument zur Beurteilung der Bindung²⁹. Für eine detaillierte lokale Bewertung ist die Voronoi-Zerlegung die Methode der Wahl³⁰. Eine Reihe von Tools steht dafür zur Verfügung^{31, 32}, wobei verschiedene Anwendungen im Vordergrund stehen:

- die Bestimmung von Sekundärstrukturelementen³³
- die Berechnung von effektiven Bindungspotentialen^{34, 35}
- Anwendungen zur Strukturvorhersage über Faltungserkennung³⁶, Faltungskompatibilität (Threading)³⁷ oder Faltungssimulationen

Leider muss man konstatieren, dass die Verfügbarkeit und einfache Bedienung entsprechender Tools und nicht die Qualität des Verfahrens die Verwendungshäufigkeit dominieren, weshalb wir unsere Bemühungen zur Nutzerfreundlichkeit und Kompatibilität verstärkt haben – die Veröffentlichung eines entsprechenden Servers ist in Vorbereitung.

Der Bereich der Proteinstrukturvorhersage bleibt von herausragender Bedeutung, da die experimentelle Strukturbestimmung beispielsweise trotz der enormen Anstrengungen in 11 Konsortia zur strukturellen Genomik nur einige Hundert Zielmolekül-Strukturen beitragen konnte³⁸. Auch wenn die *ab initio* Vorhersage bisher nur grobe Modelle liefern kann, ist die Qualität der Strukturmodelle, die durch Homologie-gestützten Modellbau entstehen, nahe an experimentell-bestimmten Strukturen³⁹. Dabei sind die Lösungsmittel-zugänglichen Bereiche der Proteinkette, die sogenannten Loops, der schwierigste Teil der Vorhersage. Analytische *ab initio* Vorhersagen sind auf kurze Loops beschränkt und machen Fragmentierungen nötig⁴⁰. Seit einigen Jahren ist die Abdeckung der Loops aus Protein Data Bank als so gut anzusehen, dass Vorhersagen mit geringer Abweichung auf dieser Basis für bis zu 10 Aminosäuren lange Loops möglich sind⁴¹. Dementsprechend schneiden PDB-Fragment-basierte Verfahren für verschiedene Test-Sets besonders gut ab^{42, 43}.

Protein-Protein-Interaktionen sind zentraler Regulationsbestandteil zellulärer Netzwerke. Deren Beeinflussung ist in den Mittelpunkt des Interesses der pharmazeutischen Forschung gerückt⁴⁴. Verschiedene Datenbanken widmen sich diesem Thema⁴⁵⁻⁴⁸. Unser Ansatz, Sekundärstruktur-Kontakte in solche Analysen einzubeziehen, vergrößert den zur Verfügung stehenden Datensatz um mindestens eine Größenordnung⁴⁹.

Das Ubiquitin-Proteasom-System gehört zu den wichtigsten funktionellen Einheiten der Zelle. Die Beeinflussung dieses Systems zur Auslösung von Apoptose durch neue kleine, organische Moleküle ist uns bereits gelungen⁵⁰. Die direkte Interaktion von oligomeren, oligofunktionellen Proteinen wie dem COP9 Signalosom, dem Proteasom und den proteasomalen Regulatoren⁵¹ stellt eine Herausforderung für die Interaktionsvorhersage und – im nächsten Schritt – für die Entwicklung von Interaktionsinhibitoren dar. Bei dieser noch am Anfang stehenden Entwicklung⁵² werden sowohl Verfahren des Struktur-gestützten Peptid-Bibliothek-Entwurfs als auch Verfahren zum Design von Peptidomimetika zum Einsatz kommen.

Auf den Gebieten der effizienten Suche in Protein- und Liganden-Datenbanken, der Bewertung von 3D-Strukturen und dem Protein-Modellbau haben wir methodische Arbeit geleistet⁵³⁻⁶⁰. Außerdem konnten wir international genutzte Protein-, Liganden- und Medikamenten-Datenbanken etablieren^{61, 62}. Parallel, in Zusammenarbeit mit experimentell arbeitenden Gruppen, haben wir konkrete Anwendungen, beispielsweise für die TSE-, HIV-, oder Krebs-Therapie, verfolgt^{52, 63-66, 77}. Diesen Zyklus von eigener Algorithmen-Entwicklung, Testung an medizinisch relevanten Objekten und daraus resultierender Rückkopplung auf methodische Weiterentwicklung halte ich für unverzichtbar und möchte ihn weiter ausbauen.

Einerseits stellen Membran-Proteine etwa ein Drittel aller humanen Proteine und auf Grund ihrer Rezeptor-Funktion zwei Drittel der aktuellen medizinischen Zielstrukturen⁶⁷. Andererseits sind die 3D Strukturen von Membran-Proteinen besonders schwierig zu bestimmen, weshalb Vorhersage-Verfahren, die die Spezifika des Membranamilieus, der besonderen Aminosäure-Komposition usw. berücksichtigen, von besonderer Bedeutung sind⁶⁸.

Synergien zwischen experimentellen und theoretischen Arbeiten in der Biomedizin werden an verschiedenen Stellen sichtbar⁶⁹. Dabei müssen möglichst viele Informationen aus Zell-(Teilungs-)Experimenten, Bio-Chip-Analysen sowie molekulare als auch systembiologische Modelle in einander greifen, damit geeignete Theorie-basierte Vorhersagen gelingen⁷⁰. Da der öffentlich zugängliche Datenbestand zur systematischen Testung an 60 verschiedenen Zelltypen mit assoziierten DNA-array Daten etwa 100.000 Substanzen umfasst⁷¹, sehe ich hier einen ungehobenen Schatz, dessen fundierte theoretische Auswertung und Vernetzung mit bereits vorliegenden Strukturdaten (Substanzen / Zielmoleküle) enorme Erfolge, insbesondere in der Apoptose-Forschung verspricht. Wir haben dazu das Konzept des zellulären Fingerabdrucks entwickelt. Als Beispiel seien hier Methotrexat und Desoxyuridinphosphat, zwei chemisch sehr unterschiedliche Substanzen genannt, die durch die Ähnlichkeit ihres zellulären Fingerabdrucks auffielen. Tatsächlich konnten wir in diesem Fall eine Bindung beider Moleküle an das gleiche Zielmolekül, die Thymidilat-Synthetase, durch Kristallstrukturen belegen. In anderen Fällen lassen sich ähnliche zelluläre Fingerabdrücke durch Adressierung unterschiedlicher Zielmoleküle im gleichen Signalweg erklären. Erstaunlicherweise werden solche Präparate teilweise bereits auf Grund empirischer Daten als Kombinationspräparate eingesetzt. Es scheint also von Vorteil zu sein, eine Signalkaskade nachhaltig an mehreren Stellen zu blockieren, um so Nebenwege und Gegenregulationen zu blockieren. Sowohl die Mechanismus-Aufklärung als auch die Validierung der Eignung als medizinisches Zielmolekül durch Simulation der Wirkung im gesamten Regulationsnetzwerk wird die Entwicklung neuer wirksamer Substanzen und optimierter Kombinationstherapien ermöglichen.

In letzter Zeit gibt es einen Trend von kuratorischen zu integrativen Ansätzen⁷². Etablierte Methoden der automatisierten Text-Analyse und -Verarbeitung müssen zur Anwendung auf die Beziehungen zwischen Medikamenten und ihren Zielstrukturen zum so genannten „Relation Mining“ weiterentwickelt werden⁷³. Nur dann wird Informationsextraktion aus wissenschaftlicher (und Patent-) Literatur bzgl. der Rolle in Stoffwechselwegen oder Protein-Protein-Interaktions-Netzwerken⁷⁴ effizient möglich. Zu Lösungen in diesem Problemkreis haben wir bereits beigetragen^{75,76} und bereiten gerade eine vernetzte Datenbank vor.

6. Fazit und Ausblick

Proteine als Rezeptoren und Enzyme stehen – auch als Drug-Targets – im Mittelpunkt des medizinischen Interesses. Obwohl das Wissen über Sequenz-Struktur-Beziehungen enorm gewachsen ist, gilt der Faltungscodex der Proteine im Allgemeinen weiterhin als Herausforderung. Wir konnten eine Reihe methodischer Beiträge leisten, die verbesserte Strukturmodelle für interessante Targets ermöglichen.

Peptide als wirksame, aktive Teile von Proteinen spielen eine wichtige Rolle als Ausgangspunkt bei der Entwicklung von Medikamenten. Es ist uns gelungen, ein struktur-basiertes Verfahren zu entwickeln, das die Ermittlung – auch nicht-linearer – Epitope und Bindungsstellen über Peptidbibliotheken ermöglicht. Für den folgenden Schritt der Entwicklung vom Peptid zum oral applizierbaren Wirkstoff konnten wir ein schnelles automatisches Verfahren entwickeln, das Synthesen geeigneter Peptidomimetika vorschlägt.

Zurzeit werden weltweit etwa 3000 Wirkstoffe als Medikamente eingesetzt. Dem gegenüber stehen einige Millionen Substanzen zur Verfügung, die potentiell als Medikamente einsetzbar wären. Durch die Entwicklung von 3D-Konformationsdatenbanken und schnellen *in silico* Screening Protokollen konnten wir erfolgreich neue Substanzen für mehrere experimentelle Validierungen vorschlagen, die bereits zu einigen Patentanmeldungen geführt haben.

Nachdem sich der Fokus der biomedizinischen Forschung von Sequenz-orientierten Verfahren auf Proteine in Struktur, Funktion, Dynamik und Wechselwirkung verlagert haben, setzen sich in jüngster Zeit ganzheitliche, systembiologische Ansätze verstärkt durch. Dabei spielt sowohl die Modellierung hochvernetzter Stoffwechselwege als auch die Berücksichtigung der Protein-Metabolit-Interaktion eine wichtige Rolle bei der Vorhersage der Wirkungen und Nebenwirkungen von (potentiellen) Medikamenten. Hierbei werden die integrativen Verfahren der Bioinformatik eine Schlüsselrolle spielen. Gegenstand unserer Arbeit ist derzeit die Integration von Daten zu Medikamenten, ihren Indikationen, Nebenwirkungen, molekularen Zielstrukturen und den Signalwegen, in denen sie eine Rolle spielen. Die erste dazu in Vorbereitung befindliche Publikation zeigt bereits auf, dass die einfache Vorstellung vom Triggern einer bestimmten Schlüsselstellung einer Signalkette durch die Simulation der graduellen Beeinflussung diverser Regulationspunkte ersetzt werden muss.

7. Literatur

1. Heinemann U, Bussow K, Mueller U, Umbach P (2003) Facilities and methods for the high-throughput crystal structural analysis of human proteins. *Acc Chem Res* **36**: 157-163.
2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* **28**: 235-242.
3. Chayen NE (2004) Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol* **14**: 577-583.
4. Chayen NE (2003) Protein crystallization for genomics: throughput versus output. *J Struct Funct Genomics* **4**: 115-120.
5. Slovic AM, Stayrook SE, North B, Degrado WF (2005) X-ray structure of a water-soluble analog of the membrane protein phospholamban: sequence determinants defining the topology of tetrameric and pentameric coiled coils. *J Mol Biol* **348**: 777-787.
6. Betz UA, Farquhar R, Ziegelbauer K (2005) Genomics: success or failure to deliver drug targets? *Curr Opin Chem Biol*.
7. Hajduk PJ, Huth JR, Fesik SW (2005) Druggability indices for protein targets derived from NMR-based screening data. *J Med Chem* **48**: 2518-2525.
8. Chen L, Yin H, Farooqi B, Sebti S, Hamilton AD, Chen J (2005) p53 alpha-Helix mimetics antagonize p53/MDM2 interaction and activate p53. *Mol Cancer Ther* **4**: 1019-1025.
9. Keskin O, Ma B, Nussinov R (2005) Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol* **345**: 1281-1294.
10. Hann MM, Oprea TI (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol* **8**: 255-263.
11. Teckentrup A, Briem H, Gasteiger J (2004) Mining high-throughput screening data of combinatorial libraries: development of a filter to distinguish hits from nonhits. *J Chem Inf Comput Sci* **44**: 626-634.
12. Cavasotto CN, Kovacs JA, Abagyan RA (2005) Representing Receptor Flexibility in Ligand Docking through Relevant Normal Modes. *J Am Chem Soc* **127**: 9632-9640.
13. Cai W, Shao X, Maigret B (2002) Protein-ligand recognition using spherical harmonic molecular surfaces: towards a fast and efficient filter for large virtual throughput screening. *J Mol Graph Model* **20**: 313-328.
14. Della Vedova G, Dondi R (2003) A library of efficient bioinformatics algorithms. *Appl Bioinformatics* **2**: 117-121.
15. Fox JA, Butland SL, McMillan S, Campbell G, Ouellette BF (2005) The Bioinformatics Links Directory: a compilation of molecular biology web servers. *Nucleic Acids Res* **33**: W3-24.
16. Cohen MS, Zhang C, Shokat KM, Taunton J (2005) Structural bioinformatics-based design of selective, irreversible kinase inhibitors. *Science* **308**: 1318-1321.
17. Altman RB, Dugan JM (2003) Defining bioinformatics and structural bioinformatics. *Methods Biochem Anal* **44**: 3-14.
18. Lahana R (2002) Cheminformatics--decision making in drug discovery. *Drug Discov Today* **7**: 898-900.
19. Luscombe NM, Greenbaum D, Gerstein M (2001) What is bioinformatics? A proposed definition and overview of the field. *Methods Inf Med* **40**: 346-358.
20. Altman RB (1998) Bioinformatics in support of molecular medicine. *Proc AMIA Symp* 53-61.

21. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2005) GenBank. *Nucleic Acids Res* **33**: D34-38.
22. Rohlff C (2004) New approaches towards integrated proteomic databases and depositories. *Expert Rev Proteomics* **1**: 267-274.
23. Nobeli I, Ponstingl H, Krissinel EB, Thornton JM (2003) A structure-based anatomy of the E.coli metabolome. *J Mol Biol* **334**: 697-719.
24. Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmuller E, Dormann P, Weckwerth W, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* **21**: 1635-1638.
25. Schuffenhauer A, Popov M, Schopfer U, Acklin P, Stanek J, Jacoby E (2004) Molecular diversity management strategies for building and enhancement of diverse and focused lead discovery compound screening collections. *Comb Chem High Throughput Screen* **7**: 771-781.
26. Galperin MY (2005) The Molecular Biology Database Collection: 2005 update. *Nucleic Acids Res* **33**: D5-24.
27. Stahura FL, Bajorath J (2002) Bio- and chemo-informatics beyond data management: crucial challenges and future opportunities. *Drug Discov Today* **7**: S41-47.
28. Stahura FL, Bajorath J (2005) New methodologies for ligand-based virtual screening. *Curr Pharm Des* **11**: 1189-1202.
29. Gore SP, Burke DF, Blundell TL (2005) PROVAT: a tool for Voronoi tessellation analysis of protein structures and complexes. *Bioinformatics* **21**: 3316-3317.
30. Poupon A (2004) Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. *Curr Opin Struct Biol* **14**: 233-241.
31. Dupuis F, Sadoc JF, Jullien R, Angelov B, Mornon JP (2005) Voro3D: 3D Voronoi tessellations applied to protein structures. *Bioinformatics* **21**: 1715-1716.
32. Angelov B, Sadoc JF, Jullien R, Soyer A, Mornon JP, Chomilier J (2002) Nonatomic solvent-driven Voronoi tessellation of proteins: an open tool to analyze protein folds. *Proteins* **49**: 446-456.
33. Dupuis F, Sadoc JF, Mornon JP (2004) Protein secondary structure assignment through Voronoi tessellation. *Proteins* **55**: 519-528.
34. Verdonk ML, Cole JC, Watson P, Gillet V, Willett P (2001) SuperStar: improved knowledge-based interaction fields for protein binding sites. *J Mol Biol* **307**: 841-859.
35. Jernigan RL, Bahar I (1996) Structure-derived potentials and protein simulations. *Curr Opin Struct Biol* **6**: 195-209.
36. Zheng W, Cho SJ, Vaisman, II, Tropsha A (1997) A new approach to protein fold recognition based on Delaunay tessellation of protein structure. *Pac Symp Biocomput* 486-497.
37. Tropsha A, Singh RK, Vaisman, II, Zheng W (1996) Statistical geometry analysis of proteins: implications for inverted structure prediction. *Pac Symp Biocomput* 614-623.
38. Todd AE, Marsden RL, Thornton JM, Orengo CA (2005) Progress of structural genomics initiatives: an analysis of solved target structures. *J Mol Biol* **348**: 1235-1260.
39. Moulton J (2005) A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. *Curr Opin Struct Biol* **15**: 285-289.
40. Tosatto SC, Bindewald E, Hesser J, Manner R (2002) A divide and conquer approach to fast loop modeling. *Protein Eng* **15**: 279-286.
41. Du P, Andrec M, Levy RM (2003) Have we seen all structures corresponding to short protein fragments in the Protein Data Bank? An update. *Protein Eng* **16**: 407-414.
42. Heuser P, Wohlfahrt G, Schomburg D (2004) Efficient methods for filtering and ranking fragments for the prediction of structurally variable regions in proteins. *Proteins* **54**: 583-595.

43. Michalsky E, Goede A, Preissner R (2003) Loops In Proteins (LIP)--a comprehensive loop database for homology modelling. *Protein Eng* **16**: 979-985.
44. Pommier Y, Cherfils J (2005) Interfacial inhibition of macromolecular interactions: nature's paradigm for drug discovery. *Trends Pharmacol Sci* **26**: 138-145.
45. Thorn KS, Bogan AA (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* **17**: 284-285.
46. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol* **280**: 1-9.
47. Ogmen U, Keskin O, Aytuna AS, Nussinov R, GURSOY A (2005) PRISM: protein interactions by structural matching. *Nucleic Acids Res* **33**: W331-336.
48. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D (2004) The Database of Interacting Proteins: 2004 update. *Nucleic Acids Res* **32**: D449-451.
49. Preissner R, Goede A, Frommel C (1998) Dictionary of interfaces in proteins (DIP). Data bank of complementary molecular surface patches. *J Mol Biol* **280**: 535-550.
50. Fuellbeck M, Huang, X., Dumdey, R., Frömmel, C., Dubiel, W., Preissner, R. (2005) Novel curcumin- and emodin-related compounds identified by in silico 2D/3D conformer screening induce apoptosis in tumor cells. *BMC Cancer* **accepted**.
51. Huang X, Hetfeld BK, Seifert U, Kahne T, Kloetzel PM, Naumann M, Bech-Otschir D, Dubiel W (2005) Consequences of COP9 signalosome and 26S proteasome interaction. *FEBS J* **272**: 3909-3917.
52. Huang X, Seifert U, Salzmann U, Henklein P, Preissner R, Henke W, Sijts AJ, Kloetzel PM, Dubiel W (2002) The RTP site shared by the HIV-1 Tat protein and the 11S regulator subunit alpha is crucial for their effects on proteasome function including antigen processing. *J Mol Biol* **323**: 771-782.
53. Frommel C, Gille C, Goede A, Gropl C, Hougardy S, Nierhoff T, Preissner R, Thimm M (2003) Accelerating screening of 3D protein data with a graph theoretical approach. *Bioinformatics* **19**: 2442-2447.
54. Lorenzen S, Gille C, Preissner R, Frommel C (2003) Inverse sequence similarity of proteins does not imply structural similarity. *FEBS Lett* **545**: 105-109.
55. Lorenzen S, Peters B, Goede A, Preissner R, Frommel C (2005) Conservation of cis prolyl bonds in proteins during evolution. *Proteins* **58**: 589-595.
56. Thimm M, Goede A, Hougardy S, Preissner R (2004) Comparison of 2D similarity and 3D superposition. Application to searching a conformational drug database. *J Chem Inf Comput Sci* **44**: 1816-1822.
57. Rother K, Preissner R, Goede A, Frommel C (2003) Inhomogeneous molecular density: reference packing densities and distribution of cavities within proteins. *Bioinformatics* **19**: 2112-2121.
58. Preissner R, Goede A, Rother K, Osterkamp F, Koert U, Froemmel C (2001) Matching organic libraries with protein-substructures. *J Comput Aided Mol Des* **15**: 811-817.
59. Hildebrand PW, Preissner R, Frommel C (2004) Structural features of transmembrane helices. *FEBS Lett* **559**: 145-151.
60. Hildebrand PW, Rother K, Goede A, Preissner R, Frommel C (2005) Molecular packing and packing defects in helical membrane proteins. *Biophys J* **88**: 1970-1977.
61. Goede A, Dunkel M, Mester N, Frommel C, Preissner R (2005) SuperDrug: a conformational drug database. *Bioinformatics* **21**: 1751-1753.
62. Michalsky E, Dunkel M, Goede A, Preissner R (2005) SuperLigands - a database of ligand structures derived from the Protein Data Bank. *BMC Bioinformatics* **6**: 122.
63. Lorenzen S, Dunkel M, Preissner R (2005) In silico screening of drug databases for TSE inhibitors. *Biosystems* **80**: 117-122.
64. Fuellbeck M, Preissner, R., Dubiel, W., Frömmel, C. DE10200463269.3; 2005).

65. Jundt F, Acikgoez, O., Preissner, R., Dorken, B. (2004) Manipulation of the notch pathway by gamma-secretase inhibitors as a novel therapeutic approach in multiple myeloma. *Blood* **104**: 186-186.
66. Goede A, Dubiel, W., Frommel, C., Preissner, R. (Charité, DE10290563D; 2004).
67. Lundstrom K (2004) Structural genomics on membrane proteins: mini review. *Comb Chem High Throughput Screen* **7**: 431-439.
68. Zhou C, Zheng Y, Zhou Y (2004) Structure prediction of membrane proteins. *Genomics Proteomics Bioinformatics* **2**: 1-5.
69. Wallqvist A, Rabow AA, Shoemaker RH, Sausville EA, Covell DG (2003) Linking the growth inhibition response from the National Cancer Institute's anticancer screen to gene expression levels and other molecular target data. *Bioinformatics* **19**: 2212-2224.
70. Covell DG, Wallqvist A, Huang R, Thanki N, Rabow AA, Lu XJ (2005) Linking tumor cell cytotoxicity to mechanism of drug action: an integrated analysis of gene expression, small-molecule screening and structural databases. *Proteins* **59**: 403-433.
71. Holbeck SL (2004) Update on NCI in vitro drug screen utilities. *Eur J Cancer* **40**: 785-793.
72. Butcher EC (2005) Can cell systems biology rescue drug discovery? *Nat Rev Drug Discov* **4**: 461-467.
73. Koehler J, Rawlings C, Verrier P, Mitchell R, Skusa A, Ruegg A, Philippi S (2005) Linking experimental results, biological networks and sequence analysis methods using Ontologies and Generalised Data Structures. *In Silico Biol* **5**: 33-44.
74. Moore JH, Boczko EM, Summar ML (2005) Connecting the dots between genes, biochemistry, and disease susceptibility: systems biology modeling in human genetics. *Mol Genet Metab* **84**: 104-111.
75. Trissl S, Rother K, Muller H, Steinke T, Koch I, Preissner R, Frommel C, Leser U (2005) Columba: an integrated database of proteins, structures, and annotations. *BMC Bioinformatics* **6**: 81.
76. Gunther S, Senger C, Michalsky E, Goede A, Preissner R. (2006) Representation of target-bound drugs by computed conformers: implications for conformational libraries. *BMC Bioinformatics*. **7**: 293.
77. Herre S, Schadendorf T, Ivanov I, Herrberger C, Steinle W, Ruck-Braun K, Preissner R, Kuhn H (2006) Photoactivation of an Inhibitor of the 12/15-Lipoxygenase Pathway. *Chembiochem*. **7**: 1089-95.

8. Danksagung

Mein besonderer Dank gilt Herrn Prof. Frömmel, der nicht nur Lehrer im besten Sinne, sondern Leit- und Identifikationsfigur ist. Herrn Prof. Wittig danke ich für die freundliche Aufnahme an seinem Institut. Herrn Dr. Goede danke ich für 15 Jahre guter Zusammenarbeit. In unserer Abteilung herrscht eine hervorragende Arbeits-Atmosphäre, wofür ich allen Kollegen herzlich danke. Vielen Kollegen aus der Charité und insbesondere dem Institut für Biochemie schulde ich Dank für viele Jahre guter Zusammenarbeit.

9. EIDESSTATTLICHE VERSICHERUNG

Gemäß Habilitationsordnung der Medizinischen Fakultät Charité

Hiermit erkläre ich, dass

- keine staatsanwaltschaftlichen Ermittlungsverfahren gegen mich anhängig sind,
- weder früher noch gleichzeitig ein Habilitationsverfahren durchgeführt oder angemeldet wurde,
- die vorgelegte Habilitationsschrift ohne fremde Hilfe verfasst, die beschriebenen Ergebnisse selbst gewonnen wurden, sowie die verwendeten Hilfsmittel, die Zusammenarbeit mit anderen Wissenschaftlerinnen und Wissenschaftlern und technischen Hilfskräften und die Literatur vollständig angegeben sind,
- mir die geltende Habilitationsordnung bekannt ist.

.....
Datum

.....
Unterschrift

10. Eigene Publikationen

10.1. Eigene Originalarbeiten (peer-reviewed)

10.1.1. Erst-/Letztautorschaften

Impact-Faktor

Seite im Anhang dieser Arbeit

1. Goede, A., Dunkel, M., Mester, N., Frömmel, C., **Preissner, R.** SuperDrug: a conformational drug database *Bioinformatics*. 21, 1751-3 (2005)
IF*: **6.701**
11-25
2. Michalsky, E., Dunkel, M., **Preissner, R.** SuperLigands: a searchable database of ligands from the Protein Data Bank. *BMC Bioinformatics* 6, 122 (2005)
IF 04: **5.423**
11-2
3. Lorenzen, S., Dunkel, M., **Preissner, R.** In silico screening of drug databases for TSE inhibitors. *Biosystems*. 80, 117-22 (2005)
IF 04: **1.016**
11-8
4. Thimm, M., Goede, A. Hougardy, S., **Preissner, R.** Comparison of 2D Similarity and 3D Superposition. Application to Searching a Conformational Drug Database. *J. Chem. Inf. Comput. Sci.* 44, 1816-1822 (2004)
IF: **3.078**
11-43
5. Michalsky, E., Goede, A., **Preissner, R.** Loops In Proteins (LIP) - a comprehensive loop database for homology modelling. *Protein Engineering* 16, 1-7 (2003)
IF: **2.065**
11-50
6. **Preissner, R.**, Goede, A., Rother, K., Osterkamp, F., Koert, U., Frömmel, C. Matching organic libraries with protein-substructures. *J. Comp.-Aided Mol. Design*. 15, 811-817 (2001)
IF 01: **3.194**
11-130
7. **Preissner R.**, Goede A., Frömmel C. Homonyms and synonyms in the Dictionary of Interfaces in Proteins (DIP). *Bioinformatics*. 15, 832-836 (1999)
IF: **6.701**
11-145
8. **Preissner R.**, Goede A., Frömmel C. Spare parts for helix-helix interaction. *Protein Eng.* 12, 825-832 (1999)
IF 99: **3.209**
11-150
9. **Preissner, R.**, Goede, A., Frömmel, C. Dictionary of interfaces in proteins (DIP). Data bank of complementary molecular surface patches. *J. Mol. Biol.* 280, 535-550 (1998)
IF: **5.239**
11-85

* Impact-Faktoren von 2003 (wenn nicht anders angegeben)

10. **Preissner R.**, Goede, A, Michalsky, E, Frömmel, C. Inverse sequence similarity in proteins and its relation to the three-dimensional fold. FEBS Lett. 414, 425-429 (1997)
IF: **3.609**
11-101
11. **Preissner, R.**, Bork, P. On alpha-helices terminated by glycine. 1. Identification of common structural features. Biochem Biophys Res Commun. 180, 660-665 (1991)
IF: **2.836**
11-180
12. Bork, P, **Preissner, R.** On alpha-helices terminated by glycine. 2. Recognition by sequence patterns. Biochem Biophys Res Commun. 180, 666-672 (1991)
IF: **2.836**
11-186
13. **Preissner, R.**, Egner, U, Saenger, W. Occurrence of bifurcated three-center hydrogen bonds in proteins. FEBS Lett. 288, 192-196 (1991)
IF: **3.609**
11-198
14. Frömmel, C, **Preissner, R.** Prediction of prolyl residues in cis-conformation in protein structures on the basis of the amino acid sequence. FEBS Lett. 277, 159-163 (1990)
IF: **3.609**
11-193
15. Füllbeck, M., Huang, X., Dumdey, R., Frömmel, C., Dubiel, W., **Preissner, R.** Novel curcumin- and emodin-related compounds identified by in silico 2D/3D conformer screening induce apoptosis in tumor cells. BMC Cancer 5, 97 (2005)
IF 04: **2.293**
16. Gunther S, Senger C, Michalsky E, Goede A, **Preissner R.** Representation of target-bound drugs by computed conformers: implications for conformational libraries. BMC Bioinformatics. 7: 293 (2006)
IF 04: **5.423**
17. Fullbeck M, Michalsky E, Dunkel M, **Preissner R.** Natural products: sources and databases. Nat Prod Rep. 23: 347-356 (2006)
IF 04: **7.890**
18. Hildebrand PW, Lorenzen S, Goede A, **Preissner R.** Analysis and prediction of helix-helix interactions in membrane channels and transporters. Proteins. 64: 253-62 (2006)
IF: **4.313**
19. Goede A, Michalsky E, Schmidt U, **Preissner R.** SuperMimic--fitting peptide mimetics into protein structures. BMC Bioinformatics. 7: 11 (2006)
IF 04: **5.423**
20. Dunkel M, Fullbeck M, Neumann S, **Preissner R.** SuperNatural: a searchable database of available natural compounds. Nucleic Acids Res. 34: D678-83 (2006)
IF 04: **7.260**
21. Goede A, Jaeger IS, **Preissner R.** SUPERFICIAL--surface mapping of proteins via structure-based peptide library design. BMC Bioinformatics. 6: 223 (2005)
IF 04: **5.423**

10.1.2. Co-Autorschaften

1. Trissl S, Rother K, Mueller H, Steinke T, Koch I, **Preissner R**, Frömmel C, Leser U. Columba: an integrated database of proteins, structures, and annotations. *BMC Bioinformatics*. 6, 81 (2005)
IF: **5.423**
11-14
2. Hildebrand, P.W., Rother, K., **Preissner, R.**, Goede, A., Frömmel, C. Molecular packing and packing defects in helical membrane proteins. *Biophys J*. 88, 1970-7 (2005)
IF: **4.463**
11-35
3. Lorenzen, S., Peters, B., Goede, A., **Preissner, R.**, Frömmel, C. Conservation of cis proline bonds in proteins during evolution. *Proteins*. 58, 589-95 (2005)
IF: **4.313**
11-28
4. Hildebrand, P.W., **Preissner, R.**, Frömmel, C. Structural features of transmembrane helices. *FEBS Lett*. 559, 145-51 (2004)
IF: **3.609**
11-57
5. Frömmel, C., Gille, Ch., Goede, A., Gröpl, C., Hougardy, S., Nierhoff, T., **Preissner, R.**, Thimm, M. Accelerating screening of 3D protein data with a graph theoretical approach. *Bioinformatics* 19, 2442-2447 (2003)
IF: **6.701**
11-64
6. Rother, K., **Preissner, R.**, Goede, A., Frömmel, C. Inhomogeneous molecular density: reference packing densities and distribution of cavities within proteins. *Bioinformatics* 19, 2112-2121 (2003)
IF: **6.701**
11-70
7. Lorenzen S., Gille C., **Preissner R.**, Frömmel C. Inverse sequence similarity of proteins does not imply structural similarity. *FEBS Lett*. 545, 105-109 (2003)
IF: **3.609**
11-80
8. Gille C, Goede A, Schloetelburg C, **Preissner R**, Kloetzel PM, Gobel UB, Frömmel C. A comprehensive view on proteasomal sequences: implications for the evolution of the proteasome. *J Mol Biol*. 326, 1437-48 (2003)
IF: **5.239**
11-106
9. Huang X, Seifert U, Salzmann U, **Preissner R**, Henke W, Sijts AJ, Kloetzel PM, Dubiel W. The RTP site shared by the HIV-1 Tat protein and the 11S regulator subunit alpha is crucial for their effects on proteasome function including antigen processing. *J Mol Biol*. 323, 771-82 (2002)
IF: **5.239**
11-118
10. Gille, C., Goede, A., **Preissner, R.**, Rother, K., Frömmel, C. Conservation of atomic substructures interfaces of secondary structural elements in proteasomal subunits. *J. Mol. Biol*. 299, 1147-1154 (2000)
IF: **5.239**
11-137

11. Schmidt M, Zantopf D, Kraft R, Kostka S, **Preissner R.**, Kloetzel PM. Sequence information within proteasomal prosequences mediates efficient integration of beta-subunits into the 20 S proteasome complex. J Mol Biol. 288, 117-128 (1999)

IF: **5.239**
11-157

12. Goede, A., **Preissner, R.**, Frömmel, C. Voronoi cell – a new error avoiding method for volume and density calculation in proteins. J. Comp. Chem. 19, 1113-1123 (1996)

IF: **3.186**
11-169

13. Herre S, Schadendorf T, Ivanov I, Herrberger C, Steinle W, Ruck-Braun K, **Preissner R.**, Kuhn H. Photoactivation of an Inhibitor of the 12/15-Lipoxygenase Pathway. Chembiochem. 7: 6-14 (2006)

IF 05: **3.940**

Summe: 150

10.2. Patente

- **Determination of ligands for proteins**

Inventor: FROMMEL CORNELIUS (DE); GOEDE ANDREAN (DE); **PREISSNER ROBERT** (DE)

Patent number: US2002048776 Publication date: 2002-04-25

Classification: - international: G01N33/53; G01N33/542; G01N21/00; C12Q1/00 - european: C07K1/22, G01N33/53 Application number: US20010772538 20010129

Priority number(s): DE19981031758 19980715

- **Mittel zur Hemmung von Tumorzellen**

Inventor: GOEDE ANDREAN (DE); DUBIEL WOLFGANG (DE); FROEMMEL CORNELIUS (DE); **PREISSNER ROBERT** (DE)

Patent number: DE10290563D Publication date: 2004-04-29

Applicant: UNIVERSITAETSKLINIKUM CHARITE (DE)

Classification: - international: A61K31/165; A61P35/00 - european: Application number: DE20021090563 20020221

Priority number(s): DE20021090563 20020221; DE20011009883 20010222; DE20011017266 20010330; WO2002DE00653 20020221

- **Agent for inhibiting tumor cells**

Inventor: GOEDE ANDREAN (DE); DUBIEL WOLFGANG (DE); FROEMMEL CORNELIUS (DE); **PREISSNER ROBERT** (DE)

Patent number: WO02066027 Publication date: 2002-08-29

Applicant: UNIVERSITAETSKLINIKUM CHARITE (DE)

Classification: - international: A61K31/165; A61P35/00 - european: A61K31/05, A61K31/121, A61K31/122, A61K31/167 Application number: WO2002DE00653 20020221

Priority number(s): DE20011009883 20010222; DE20011017266 20010330

- **Signalosom als Zielkomplex zur Hemmung von Tumorzellen und Angiogenese**

Inventor: GOEDE ANDREAN (DE); DUBIEL WOLFGANG (DE); FROEMMEL CORNELIUS (DE); **PREISSNER ROBERT** (DE)

Patent number: DE10290562D Publication date: 2004-04-29

Applicant: UNIVERSITAETSKLINIKUM CHARITE (DE)

Classification: - international: A61K31/05; A61K31/12 - european: Application number: DE20021090562 20020221

Priority number(s): DE20021090562 20020221; DE20011009883 20010222; DE20011017266 20010330; WO2002DE00652 20020221

- **Neues Mittel zur Tumorsuppression und Hemmung der Tumorangiogenese**

Inventor: FUELLBECK MELANIE (DE), **PREISSNER ROBERT** (DE), DUBIEL WOLFGANG (DE), FRÖMMEL CORNELIUS (DE)

Patent number: DE102004063269.3 Publication Date: 2005-06-29

Applicant: Charite - Universitaetsmedizin Berlin (DE)

Classification: -international: A61K31/122; A61P35/00; C07C50/34

11. Anhang

Eigene Originalarbeiten (peer-reviewed)

Titel	Seite
SuperLigands: a searchable database of ligands from the Protein Data Bank	11-2
In silico screening of drug databases for TSE inhibitors	11-8
Columba: an integrated database of proteins, structures, and annotations	11-14
SuperDrug: a conformational drug database	11-25
Conservation of cis proline bonds in proteins during evolution	11-28
Molecular packing and packing defects in helical membrane proteins	11-35
Comparison of 2D Similarity and 3D Superposition. Application to Searching a Conformational Drug Database	11-43
Loops In Proteins: a comprehensive loop database for homology modelling	11-50
Structural features of transmembrane helices	11-57
Accelerating screening of 3D protein data with a graph theoretical approach	11-64
Inhomogeneous molecular density: reference packing densities and distribution of cavities within proteins	11-70
Inverse sequence similarity of proteins does not imply structural similarity	11-80
Dictionary of interfaces in proteins (DIP): Data bank of complementary molecular surface patches	11-85
Inverse sequence similarity in proteins and relation to the three-dimensional fold	11-101
A comprehensive view on proteasomal sequences: implications for the evolution of the proteasome	11-106
The RTP site shared by the HIV-1 Tat protein and the 11S regulator subunit alpha is crucial for their effects on proteasome function including antigen processing	11-118
Matching organic libraries with protein-substructures	11-130
Conservation of atomic substructures interfaces of secondary structural elements in proteasomal subunits	11-137
Homonyms and synonyms in the Dictionary of Interfaces in Proteins (DIP)	11-145
Spare parts for helix-helix interaction	11-150
Sequence information within proteasomal prosequences mediates efficient integration of beta-subunits into the 20 S proteasome complex	11-157
Voronoi cell – a new error avoiding method for volume and density calculation in proteins	11-169
On alpha-helices terminated by glycine: Identification of common structural features	11-180
On alpha-helices terminated by glycine: Recognition by sequence patterns	11-186
Prediction of prolyl residues in cis-conformation in protein structures on the basis of the amino acid sequence	11-193
Occurrence of bifurcated three-center hydrogen bonds in proteins	11-198