

BIOWEPRO

Biomolekulare Wechselwirkungen von Proteinen¹

Teilprojekt Generierung und Bewertung von diskreten Dockingpositionen & Proteindocking mit Korrelationsmethoden

Dietmar Schomburg, Michael Meyer

Gesellschaft für Biotechnologische Forschung, Molekulare Strukturforchung

Mascheroder Weg 1, 38124 Braunschweig

Kurzfassung

Das Projekt BIOWEPRO als Kooperationsprojekt der Arbeitsgruppe um Prof. Schomburg mit den Gruppen Prof. Sagerer, Universität Bielefeld, Technische Fakultät, Dr. Soumpasis, Max-Planck Institut für Biophysikalische Chemie Göttingen und Prof. Kriegel, Universität München, Institut für Informatik diente der Entwicklung von geeigneten Algorithmen und eines Datenbanksystems, das die effiziente Beantwortung der Frage nach den möglichen Wechselwirkungen eines gegebenen Moleküls bekannter Raumstruktur mit den in der Datenbank erfaßten Proteinen mit bekannter Faltung erlauben soll. Das hier beschriebene Projekt befaßt sich vornehmlich mit Protein-Protein Wechselwirkungen, wobei sich während der Projektbearbeitung gezeigt, hat, daß mit den erarbeiteten Methoden auch Nichtprotein-Anfragemoleküle von mindestens ca. 20 Nichtwasserstoffatomen erfolgreich bearbeitet werden können. Die Projektteilbereiche der Datenbankkonstruktion, der Grobidentifizierung von Docking-Bereichen und der Feinbestimmung der Komplexstruktur von Proteinkomplexen konnten im Gesamtprojekt erfolgreich abgeschlossen werden, wobei deutlich effizientere und zuverlässige Methoden als bisher in der Literatur beschrieben, entwickelt werden konnten. Hierbei wird im Normalfall ein beliebiger experimentell bestimmter Komplex eines Proteins mit ca. 0.5-1.5 Å Genauigkeit reproduziert.

¹ *Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Forschung und Technologie unter dem Förderkennzeichen 01 IB 307 A7 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.*

1. Einleitung

Neben den in den Organismen als Informationsspeicher und -überträger benutzten Nucleinsäuren stellen die Proteine die größte und bei weitem vielfältigste Klasse biologischer Makromoleküle dar. Sie weisen in Struktur und Funktion eine enorme Variabilität auf. Ob sie aber nun ihre biologische Funktion als Enzyme (Biokatalysatoren), Hormone, Strukturproteine, Antikörper, Muskelproteine, Toxine, Sauerstoffträger oder anderes ausüben, tun sie dies, indem sie Wechselwirkungen mit anderen Molekülen eingehen und molekulare Komplexe bilden, die entweder nur kurzfristig oder auch länger stabil sind.

Die experimentelle Bestimmung solcher Wechselwirkungen ist heute mit mehreren Methoden möglich, die genaue Bestimmung der wechselwirkenden Oberflächenbereiche gelingt aber nur mit sehr aufwendigen analytischen Verfahren beim Vorliegen relativ großer Proteinmengen. Für das Verständnis von molekularen Wechselwirkungen, aber auch für eine Vielzahl von Anwendungen im Bereich vor allem der medizinischen Diagnose und Therapie wäre es von hoher Bedeutung, wenn die Wechselwirkungen von Proteinen mit anderen Molekülen akkurat vorhergesagt werden könnten. Als Beispiele seien hier nur die *in vitro* Reaktion von Antikörpern mit Proteinen genannt, die heute zum schnellen Nachweis von z.B. Infektionen oder allgemein in der biochemischen Analytik eingesetzt wird, wo die Vorhersage von zu erwarteten Wechselwirkungen der eingesetzten Antikörper mit anderen Proteinen und somit die Vermeidung der Gefahr von Fehldiagnosen eminent wichtig ist. Auch die Frage, ob ein in der Entwicklung befindliches Medikament, das als Antagonist einen bestimmten Rezeptor blockieren oder ein bestimmtes Enzym inhibieren soll, Nebenwirkungen über Wechselwirkungen mit anderen Proteinen erwarten läßt, ist von hoher Bedeutung.

In den letzten Jahren ist die Anzahl der bekannten Proteinraumstrukturen stark angestiegen. Eine weitere Beschleunigung dieses Anstiegs sowie eine Identifizierung der meisten Humanproteine im Rahmen der weltweiten Sequenzierungsprojekte wird für die nächsten Jahre erwartet. Somit erscheint es notwendig, Methoden zu entwickeln, mit deren Hilfe die oben skizzierten Fragestellungen bearbeitet werden können. Hierzu wurde im hier beschriebenen hochambitösen Projekt ein integriertes Konzept verfolgt, das es erlaubt, eine Reihe von nach dem heutigen Stand der Technik ungelösten Problemen anzugehen.

- So mußten effiziente Algorithmen entwickelt werden, die eine möglichst genaue Identifizierung der Kontaktstellen zweier miteinander wechselwirkenden Proteine sowie die Vorhersage der Struktur ihres Komplexes erlauben.
- Die vorhergesagten Komplexe mußten mit Hilfe einer biophysikalisch sinnvollen und allgemeinverwendbaren Energiefunktion bewertet werden. Hierzu muß diese Energiefunktion entwickelt werden.
- Da das Projektziel über die effiziente Beschreibung eines 1:1 Komplexes weit hinausgeht, mußte ein Datenbanksystem entworfen und entwickelt werden, das effiziente Anfragen mit dem Ziel erlaubt, Proteine zu identifizieren, die mit einem gegebenen Molekül eine Wechselwirkung eingehen, die energetisch über einer vorgegebenen Schranke liegt. Hierzu mußten Verfahren zur Grobidentifizierung von Kontaktstellen entwickelt werden.
- Die entwickelten Methoden mußten anhand eines gegebenen möglichst umfangreichen Satzes von experimentell bestimmten Proteinkomplexen überprüft und optimiert werden.

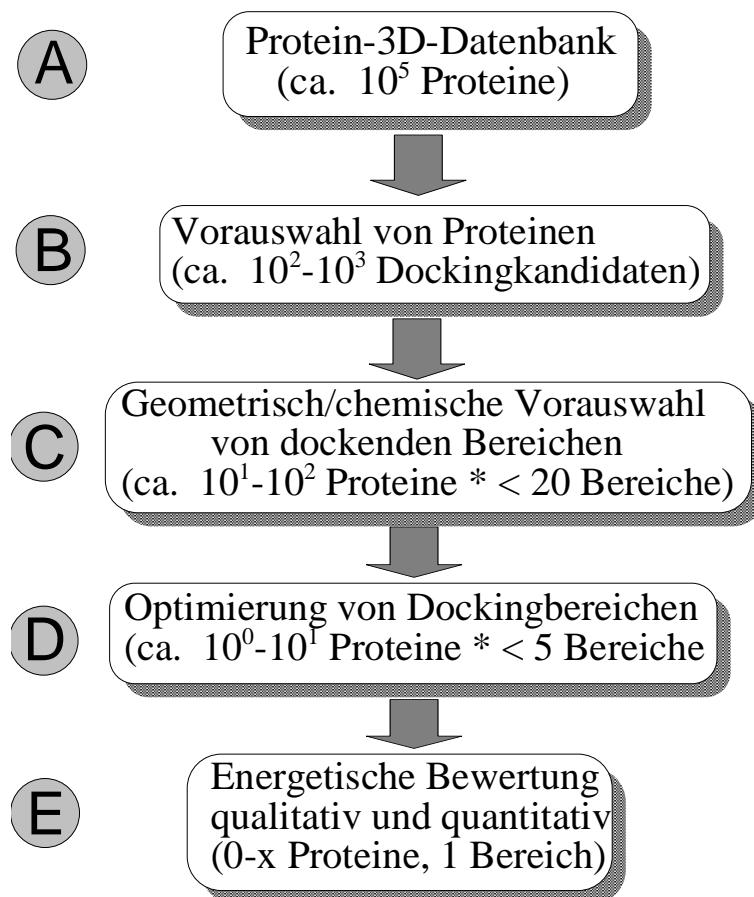


Abbildung 1, BIOWEPRO Gesamtkonzept

Die skizzierten Aufgabenstellungen wurden im Projekt durch eine enge Kooperation von vier Gruppen bearbeitet, die von ihren Vorerfahrungen her ideal geeignet sind. So liegen in München Erfahrungen in numerischer Repräsentation, Speicherung und Suche komplexer räumlicher Objekte in Datenbanksystemen vor, in Bielefeld Erfahrungen mit numerisch subsymbolischen und wissensbasierten Techniken zur Musteranalyse, in Braunschweig Erfahrungen in der molekularen Strukturchemie, der Protein-Strukturvorhersage in der interaktiven Molekülgraphik sowie im Protein-Design und in Göttingen Erfahrungen in den Methoden zur effektiven Modellierung von Umgebungseffekten auf biologische Makromoleküle (PMF-Verfahren).

Im folgenden werden hier die Ergebnisse des Braunschweiger Teilprojektes beschrieben.

2. Generierung und Bewertung von diskreten Dockingpositionen

Vorselektion von Verwendung von geometrischen und H-Brücken Deskriptoren

Bei der schnellen Identifizierung von geeigneten Dockingbereichen kann man sich die Informationen über typische Dockingstellen zunutze machen, die aus experimentell bestimmten Proteinkomplexen zu Beginn des Projektes abgeleitet wurden. So weisen typische Kontaktstellen von Proteinen folgende Eigenschaften auf (Schomburg und Jakob, in Vorbereitung):

- Sie umspannen eine Fläche von typischerweise 700-1000 Å². Kontaktstellen mit einer Fläche von mehr als ca. 1200 Å² sind typisch für Protein-Quartärstruktur.
- Sie weisen einen signifikant höheren hydrophoben Anteil im Verhältnis zur nichtbindenden Oberfläche auf (Aminosäure an Kontaktflächen überrepräsentiert: Phe, Cys, Trp, Leu, Met, Ile, Tyr, unterrepräsentiert: Lys, Glu, Asp, Ser)
- Kontaktbereiche sind meist signifikant stärker gekrümmt

Hydrophobe Wechselwirkungen, die einen großen Anteil zu der freien Enthalpieänderung bei der Komplexbildung beitragen, erfordern durch ihre kurze Reichweite hohe Komplementarität der Oberflächen. Somit ist es mit einem schnellen Verfahren der Erkennung von Oberflächenkomplementaritäten möglich, mögliche Kontaktstellen zu identifizieren, die dann in den Folgeschritten optimiert und energetisch bewertet werden.

Das für diesen Schritt im Projekt erarbeitete Verfahren hat sich als äußerst erfolgreich und schnell erwiesen (Meyer et al. 1997). Nach der Oberflächenberechnung nach einem modifizierten Pearl & Honneger Algorithmus, wird diese trianguliert und es werden flächengewichtete Oberflächennormalen berechnet. Hierbei hat sich die Optimierung und Wahl eines geeigneten Radius der und die Dichte des Gitterrasters als besonders wichtig erwiesen. Diese werden mit Hilfe von Oberflächenparametern aus der Mustererkennung (Zweites Winkelmoment, Kontrast, Entropie, Autokorrelation) zu Parametern umgewandelt, die typische Kontaktzonen (s.o.) beschreiben. Durch eine Klassifizierung dieser Parameter und Sortierung in die gleichen Klassen werden Kontaktflächen durch hohe lokale Punktdichte gut korrelierender Punkte beschrieben. Dieses Verfahren vermeidet die kombinatorische Explosion

beim Vergleich von jedem Punkt auf der Oberfläche des einen Moleküls mit jedem Punkt auf der des zweiten. Hierbei wird nur bei ca. 10% zufällige Übereinstimmung gefunden.

Zweites Winkelmoment $f_1 = \sum_{i=0}^k \sum_{j=0}^k (g_{ij})^2$

Kontrast $f_2 = \sum_{i=0}^k \sum_{j=0}^k (i - j)^2 \cdot g_{ij}$

Entropie $f_3 = - \sum_{i=0}^k \sum_{j=0}^k g_{ij} \cdot \log g_{ij}$

Autokorrelation $f_4 = (\sum_{i=0}^k \sum_{j=0}^k (ij \cdot g_{ij}) - m_I m_J) / (\sigma_I \cdot \sigma_J)$

wobei:

$$\sigma_I = \left[\sum_{i=0}^k \sum_{j=0}^k i^2 \cdot g_{ij} - \left(\sum_{i=0}^k \sum_{j=0}^k i \cdot g_{ij} \right)^2 \right]^{1/2}$$

$$\sigma_J = \left[\sum_{i=0}^k \sum_{j=0}^k j^2 \cdot g_{ij} - \left(\sum_{i=0}^k \sum_{j=0}^k j \cdot g_{ij} \right)^2 \right]^{1/2}$$

$$m_I = \sum_{i=0}^k \sum_{j=0}^k i \cdot g_{ij} \quad m_J = \sum_{i=0}^k \sum_{j=0}^k j \cdot g_{ij}$$

Die nach Ausschluß zufälliger, lokal eng begrenzter Parameterübereinstimmungen berechnete Reorientierungsmatrix ergab bei 63 von 65 untersuchten Komplexen eine eindeutige Identifizierung der Kontaktregion (mit einem durchschnittlichen Rechenzeitbedarf von ca. 12-18 Minuten auf einer Workstation). Das Verfahren wird momentan im Hinblick auf eine Reduzierung der Punktdichte und die Berücksichtigung von potentiellen H-Brücken optimiert. Ein Zwischenstand ergibt hier bei einem stark reduzierten Rechenzeitbedarf von 2-8 Minuten die korrekte Identifizierung von 50 der getesteten 65 Proteinkomplexe.

3. Proteindocking mit Korrelationsmethoden

Ziel dieses Teilprojektes war die Vorhersage der Struktur des Komplexes, den zwei Proteine bilden können. Das Dockingverfahren zielt speziell auf eine Analyse von Protein-Protein-Komplexen ab. Es ist jedoch so flexibel, daß auch ein Docking mit kleinen organischen Molekülen oder mit Nucleotiden simuliert werden kann. Unter der Annahme einer starren Struktur der beiden Ausgangsmoleküle werden drei Rotations- und drei Translationsfreiheitsgrade optimiert, die die Orientierung der Moleküle relativ zueinander beschreiben. Das Verfahren kann wahlweise als eigenständige Methode eingesetzt werden oder auf Informationen von anderen Methoden aufbauen, die potentielle Kontaktstellen der Proteinoberflächen grob lokalisieren können. In diesen Fall ist eine Berechnung der genauen Komplexstruktur besonders schnell möglich, da nur ein begrenzter Bereich der Eulerwinkel, die zur Beschreibung der Rotation verwendet werden, durchsucht werden muß.

Die zu dockenden Moleküle werden aus Dateien im PDB-Format eingelesen und je nach Art der Moleküldarstellung unterschiedlich behandelt. Liegen vollständige Molekülstrukturen mit Wasserstoffatomen vor, wie Sie aus NMR-Untersuchungen oder theoretischen Arbeiten bestimmt werden können, dann wird jedem Atom der entsprechende van der Waals Radius zugewiesen ("all atom" Modell). Bei Strukturen aus der Proteinkristallographie ohne Wasserstoff wird ein "united atom" Modell verwendet, bei dem den Atomen, an die Wasserstoff gebunden ist, ein vergrößerter van der Waals Radius zugeordnet wird. Für Standardaminosäuren, Nucleotide, Cofaktoren und einige häufige Moleküle sind die Radien in das Programm integriert worden, für andere Moleküle werden die Radien einer Datei entnommen, die von dem Anwender auf beliebige Moleküle erweitert werden kann.

Der Einsatz von Korrelationsmethoden zum Docking hat sich als besonders geeignet erwiesen, da das Verfahren sehr effizient ist [1]. Dazu stellen wir die Moleküle auf einem Gitter dar, wobei der Abstand der Gitterpunkte maximal 1.6 \AA für eine grobe Vororientierung der beiden Komponenten beträgt und 1.0 \AA für eine genaue Berechnung der Komplexstruktur nicht übersteigen sollte. Die optimale Translation wird durch die Berechnung der Korrelation bestimmt, die über eine vektorisierte FFT berechnet wird. Anschließend wird ein Protein gedreht, erneut auf dem Gitter dargestellt und erneut die Korrelation berechnet. Die Ergebnisse werden

nach der Größe der Korrelation geordnet.

Da die bisher übliche schrittweise Variation [1] der Eulerwinkel θ_1, θ_2 und θ_3 nicht zu einer gleichmäßigen Abtastung führt, wird mit den Linearkombinationen $\theta_+ = \theta_1 + \theta_2$ und $\theta_- = \theta_1 - \theta_2$ der Eulerwinkel gearbeitet. Dieses verringert den erforderlichen Rechenaufwand um ein Drittel und ermöglicht eine äquidistante Suche mit einem Abstand d , indem jeder Summand der Gl. (1) für ein festes θ_2 gleich groß gewählt wird.

$$d^2 = \Delta\theta_+^2 \cos^2(\theta_2/2) + \Delta\theta_2^2 + \Delta\theta_-^2 \sin^2(\theta_2/2)$$

Eine Suche über den gesamten Bereich der Eulerwinkel ist nur notwendig, wenn keine Vorinformation über den mögliche Kontaktstellen vorliegt. Anderenfalls wird nur eine Korrelation für einen engen Bereich von Eulerwinkeln berechnet.

In einem letzten Schritt wird eine Verfeinerung der besten Orientierung der beiden Proteine durchgeführt. Dazu werden die Eulerwinkel über eine Kombination von Simplexsuche und Simulated Annealing verbessert und die optimale Translation wiederum über die Korrelation berechnet. Diese Kombination von Simplexsuche und Simulated Annealing ist effizienter als ein konventionelles Simulated Annealing Verfahren, da es grundsätzlich die Tendenz hat, die Eulerwinkel so zu variieren, daß ein Maximum der Korrelation gesucht wird. In Abhängigkeit von der Temperatur werden jedoch auch Schritte akzeptiert, die zu einer Verringerung der Korrelation führen und lokale Minima können überwunden werden.

| Komplex | Molekül 1 Atomzahl | Molekül 2 Atomzahl | Rang Fischer et al. 1995 | Rang Biowepro | RMS Biowepro |
|---------|-----------------------|-----------------------|--------------------------------|------------------|-----------------|
| 2MHB | 1178 | 1113 | 1 | 1 | 0.82 |
| 4PHV2 | 760 | 760 | 4 | 1 | 1.52 |
| 2SEC | 1923 | 522 | 20 | 1 | 0.67 |
| 1TEC | 2007 | 522 | 134 | 1 | 0.62 |
| 4TPI | 1644 | 456 | 20 | 1 | 0.66 |
| 2PTC | 1629 | 454 | 27 | 1 | 0.38 |
| 1TGS | 1652 | 416 | 2 | 1 | 0.79 |
| 1CHO | 1750 | 400 | 60 | 1 | 0.64 |
| 4SGB | 1310 | 380 | 10 | 1 | 0.38 |
| 4CPA | 2438 | 290 | 281 | 1 | 2.56 |
| 4PHV | 1520 | 92 | 3082 | 1 | 0.90 |
| 2CPK | 2665 | 75 | 7 | 1 | 0.46 |
| 4MBN | 1217 | 44 | 2 | 1 | 0.97 |
| 4HVP | 1516 | 54 | 1 | 1 | 1.52 |
| 3DFR | 1342 | 33 | 22 | 1 | 0.61 |
| 2IGF | 3378 | 58 | 19 | 1 | 0.61 |
| 1FDL | 3308 | 1001 | 28 | 33/3* | |
| 2HFL | 3227 | 1001 | 105 | 28/3* | |
| 3HFM | 3295 | 1001 | 1146 | >75/1 | |

Tabelle 1: Ergebnisse von ausgewählten Dockingsimulationen

In der Tabelle wird ersichtlich, daß im Vergleich zu einem kürzlich publizierten Literaturverfahren die korrekte Kontaktregion mit hoher Genauigkeit mit Ausnahme der letzten drei speziellen Fälle immer korrekt identifiziert wird. In den in der Literatur beschriebenen Verfahren wird die korrekte Kontaktregion teilweise auf einen Rang über 1000 gesetzt. Eine Kombination der beiden unter 3.4.2 und 4. beschriebenen Verfahren ist erstmalig auch in der Lage, die Kontaktstellen der letzten drei Komplexe (Antikörper/Lysozym) automatisch auf eine der ersten Positionen zu setzen.

Für 40 Proteinkomplexe wurden Dockingsimulationen durchgeführt. Die Komplexstrukturen wurden dazu in ihre Komponenten zerlegt, einer Koordinatentransformation mit zufälligen

Eulerwinkeln unterworfen und mit dem Dockingverfahren wieder rekonstruiert. Eine Auswahl der Ergebnisse ist in der Tabelle 1 zusammengestellt.

Statt einer Verfeinerung in einem engen Bereich von Eulerwinkeln kann das Simulated Annealing Verfahren auch für eine globale Suche nach dem absoluten Maximum der Korrelation bezüglich der Eulerwinkel eingesetzt werden, da die Periodizität der Eulerwinkel in das Suchverfahren integriert wurde.

Die untersuchten Komplexe zeigen, daß sich das Verfahren sehr gut für Enzym-Inhibitor-Komplexe und zur Berechnung der Struktur von Dimeren verwenden läßt. Es konnten jedoch auch Komplexe aus Proteinen mit kleinen organischen Molekülen berechnet werden, sofern die kleinen Moleküle aus mindestens 20 Atomen bestehen, und es wurde auch ein Protein-DNA-Komplex (PDB-Eintrag 1gat) erfolgreich gedockt. Beim Docking eines Liganden in das bekannte aktive Zentrum eines Proteins kann auch ein kleines Segment der Proteinstruktur verwendet werden, da es sich um ein rein geometrisches Dockingverfahren handelt. Diese Vorgehensweise führt zu einer weiteren Beschleunigung der Rechnungen. Die berechneten Strukturen zeigen nur sehr geringe Abweichung von den experimentellen Daten. Typische rms-Abweichungen liegen zwischen 0.5 und 1.5 Å. Bei größerem Abstand der Gitterpunkte kann in kürzerer Zeit eine rms-Abweichung im Bereich von 1.5 bis 3.0 Å erzielt werden. Im Gegensatz zu anderen aktuellen Verfahren [2] liegen die richtigen Strukturvorschläge fast immer auf dem ersten Rang. Eine Ausnahme bilden Antigen-Antikörper-Komplexe. Die Besonderheit dieser Komplexe besteht darin, daß die in der Natur realisierte Orientierung nicht diejenige ist, bei der die Kontaktfläche zwischen den Proteinen am größten ist. Statt dessen ist die Anzahl der Wasserstoffbrücken zwischen den beiden Proteinen besonders hoch. Dieses ermöglicht eine Vorhersage der Dockingregion auf der Basis von zu einander komplementären Donor- und Akzeptoratom an den Oberflächen der beiden Proteine. Zunächst wird eine Analyse der intramolekularen Abstände der Donor- und Akzeptoratome auf den Oberflächen der beiden Komponenten durchgeführt und die Komplementarität der Oberflächen über Normalenvektoren bewertet, um potentielle Kontaktregionen zu berechnen. Über eine Hauptachsentransformation der entsprechenden Dockingregionen können dann die Eulerwinkel abgeschätzt werden und durch eine nachfolgende Korrelationsrechnung kann die richtige Komplexstruktur ermittelt werden. Diese Vorgehensweise reduziert die Anzahl der Rotationen, für die die Korrelationsuche durchgeführt wird, um etwa 80 bis 90%.

Zur Untersuchung der Toleranz des Verfahrens bezüglich geringer Strukturänderungen von Seitenketten an der Proteinoberfläche haben wir an 15 Beispielen von Komplexen die Inhibitoren gegen die Strukturen der isolierten Inhibitoren bzw. der isolierten Enzyme ausgetauscht und den Einfluß einer Low-Pass-Filterung untersucht. Die Strukturen dieser Komplexe konnten ebenfalls erfolgreich gedockt werden. Die Methode ist damit in der Lage, begrenzte Konformationsänderungen von Seitenketten an der Proteinoberfläche bei der Komplexbildung zu tolerieren.

4. Zusammenfassung und Ausblick

Im Projekt BIOWEPRO im Rahmen einer sehr fruchtbaren Kooperation eine Vielzahl von Algorithmen und methodischen Beiträgen für die Bearbeitung des Protein-Dockingproblems erarbeitet. Das Gesamtprojekt beinhaltet ein so bisher noch nicht verfügbares Paket zur Bearbeitung von Docking-Anfragen und die bisher erreichten Teilschritte übertreffen bisher in der Literatur bekannte Verfahren deutlich. So wurde(n):

- ein geeignetes Datenbanksystem für Dockinganfragen entwickelt,
- schnelle Grobdocking-Verfahren zur Vorselektierung und Filterung von Molekülen und Dockingbereichen entwickelt,
- ein Fourier-basiertes Verfahren zur Berechnung und Optimierung von Strukturen von Proteinkomplexen geschaffen,
- ein Konzept und Verfahren zur energetischen Bewertung von Proteinkomplexen erarbeitet.

Prinzipiell ist das Verfahren zur Vorhersage von Dockingpartnern für ein gegebenes Molekül aus der Datenbank aller gegebenen Protein-3D Strukturen geeignet. Allerdings ist die Berücksichtigung von "Induced Fit", d.h. größeren Strukturanpassungen bei der Komplexbildung noch ein weitgehend ungelöstes Problem.

5. Literatur

Ackermann, F., Herrmann, G., Posch, S., Sagerer, G. 1995: '*3D-Segmentierungstechniken und vektorwertige Bewertungsfunktionen für symbolisches Protein-Protein-Docking*'. In Schomburg, D. and Lessel, U., editors. *Bioinformatics: From Nucleic Acids and Proteins to Cell Metabolism*, Weinheim: VCH, 105-124

Ackermann, F., Herrmann, G., Kummert, F., Posch, S., Sagerer, G., Schomburg, D., 1995: '*Protein docking combining symbolic descriptions of molecular surfaces and grid-based scoring functions*', in Rawlings, C., Clark, D., Altmann, R., Hunter, L., Lengauer, Th., and Wodak, S., editors, *Proceedings Third International Conference on Intelligent Systems for Molecular Biology*, Menlo Park: AAAI Press, 3-11

Ackermann, F., Porczynski, G., Posch, S., Sagerer, G., 1994: '*Hierarchische Modellierung und Bewertung von Proteindockingpositionen*'. In Haubensak, F. and Sühnel, J., editors, *Bioinformatik - Computereinsatz in den Biowissenschaften*, Jena, IMB, 19-29

Aldinger, K., Ester, M., Förstner, G., Kriegel, H.-P., Seidl, T., 1994: '*Datenbankunterstützung für das Protein-Protein-Docking: ein effizienter und robuster Feature-Index*', Proc. "Bioinformatik - Computereinsatz in den Biowissenschaften", 2. GI-Fachtagung "Informatik in den Biowissenschaften", Jena, 41-52

Ester, M., Kriegel, H.-P., Seidl, T., Xu, X., 1995: '*Formbasierte Suche nach komplementären 3D-Oberflächen in einer Protein-Datenbank*'. Proc. GI-Fachtagung "Datenbanken in Büro, Technik und Wissenschaft" (BTW '95), Reihe Informatik aktuell, Springer, 373-382

Fischer, D. Lin, S. L., Wolfson H. L. und Nussinov, R., 1995: '*A Geometry-based Suite of Molecular Docking Processes*'. *J. Mol. Biol.* 248, 459

García, A., Hummer, G., Soumpasis, D.M., (1994): '*Theoretical Description of Protein Hydration. A PMF Calculation based on Two and Three Particle Correlation Functions*'. *Biophys. J.* 66, 130

García, A., Hummer, G., Soumpasis, D.M., (1995): '*Theoretical Description of Biomolecular Hydration. Application to A-DNA*'. In: *Neutrons in Biology*, B. Ed: Schoenborn. Elsevier, Amsterdam (im Druck)

Holm, L. Ouzounis, C. Sander, C. Tuparev, G. Vriend, G., 1992: '*A database of protein structure families with common folding motifs*'. *Protein Science*, 1, 1691

Hummer, G., Soumpasis, D.M., García, A., (1995): '*Potentials of Mean Force Description of Ionic Interactions and Structural Hydration in Biomolecular Systems*'. In: *Nonlinear Excitations*

in Biomolecules. Ed: M. Peynard. Springer Verlag, Berlin, 82 - 99

Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C., Vakser, I. A., 1992: '*Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation technique*'. Proc. Natl. Acad. Sci. USA, Bd. 89, 2195-2199

Kriegel, H.-P., Schmidt, T., Seidl, T., 1995: '*Moleküloberflächen in einer objektorientierten Proteindatenbank*', GI-Fachtagung Bioinformatik, Braunschweig, erweiterte Zusammenfassung

Meier, R., Ackermann, F., Herrmann, G., Posch, S., Sagerer, G., 1995: '*Segmentation of molecular surfaces based on their convex hull*'. In Proceedings International Conference on Image Processing, Washington, D.C., IEEE Computer Society Press, 552-555

Meyer, M., Schomburg, D., 1995: '*Proteindocking mit Korrelationsmethoden*'. GI-Fachtagung Bioinformatik, Braunschweig, erweiterte Zusammenfassung

M. Meyer; P. Wilson; D. Schomburg 1996: Hydrogen Bonding and Molecular Surface Shape Complimentarity as a Basis for Protein Docking, *J. Mol. Biol.* **264**, 199-210

Richards, F. M., 1977: '*Areas, volumes, packing, and protein structure*'. Ann. Rev. Biophys. Bioeng., Bd. 6, 151-176

Schillo, C., Herrmann, G., Ackermann, F., Posch, S., Sagerer, G., 1995: '*Statistical classification and segmentation of biomolecular surfaces*'. In Proceedings International Conference on Image Processing, Washington, D.C., IEEE Computer Society Press, 560-563

Seidl, T., Kriegel, H.-P., 1995: '*Solvent Accessible Surface Representation in a Database System for Protein Docking*', Proc. 3rd Int. Conference on Intelligent Systems for Molecular Biology (ISMB-95), Cambridge, UK, AAAI Press, 350-358, Literatur Bielefeld

Seidl, T., Kriegel, H.-P., 1995: '*A 3D Molecular Surface Representation Supporting Neighborhood Queries*', Proc. 4th Int. Symposium on Large Spatial Databases (SSD '95), Portland, Maine, USA, Lecture Notes in Computer Science, Vol. 951, Springer, 240-258

Soumpasis, D.M., García, A., Klement, R., Jovin, T.M., (1990): '*The Potentials of Mean Force (PMF) Approach for treating Ionic Effects in Biomolecular Structures in Solution*'. In: Theoretical Biochemistry & Molecular Biophysics. Adenine Press, Guilderland NY, 343 - 360

Soumpasis, D.M., (1993): '*Formal Aspects of the Potential of Mean Force Approach*'.

In: Computation of Biomolecular Structures Achievements, Problems and Perspectives'. Ed: Soumpasis, D.M. & Jovin, T.M. Springer Verlag, Heidelberg, 223 - 239

Wilson, P., Meyer, M., Schomburg, D., 1995: *'The Geometric of Protein-Protein Docking Candidates by Molecular Surface Shape Comparison on the Basis of Normal Vector. A Derived Shape Descriptors'*. GI-Fachtagung Bioinformatik, Braunschweig

ANHANG: Liste der bisher aus dem Teilprojekt entstandenen Veröffentlichungen

Ackermann, F., Herrmann, G., Kummert, F., Posch, S., Sagerer, G., Schomburg, D., 1995: *'Protein docking combining symbolic descriptions of molecular surfaces and grid-based scoring functions'*, in Rawlings, C., Clark, D., Altmann, R., Hunter, L., Lengauer, Th., and Wodak, S., editors, Proceedings Third International Conference on Intelligent Systems for Molecular Biology, Menlo Park: AAAI Press, 3-11

Meyer, M., Schomburg, D., 1995: *'Proteindocking mit Korrelationsmethoden'*. GI-Fachtagung Bioinformatik, Braunschweig, erweiterte Zusammenfassung

M. Meyer; P. Wilson; D. Schomburg 1996: Hydrogen Bonding and Molecular Surface Shape Complimentarity as a Basis for Protein Docking, *J.. Mol. Biol.* **264**, 199-210

Wilson, P., Meyer, M., Schomburg, D., 1995: *'The Geometric of Protein-Protein Docking Candidates by Molecular Surface Shape Comparison on the Basis of Normal Vector. A Derived Shape Descriptors'*. GI-Fachtagung Bioinformatik, Braunschweig