

Strukturvorhersage

Modellierung G-Protein-gekoppelter Rezeptoren mit ROSETTA

ANETTE SCHREIBER¹, JENS MEILER²

¹INSTITUT FÜR BIOCHEMIE, UNIVERSITÄT LEIPZIG

²CENTER FOR STRUCTURAL BIOLOGY, VANDERBILT UNIVERSITY, NASHVILLE, TN, USA

Die Verknüpfung funktioneller und struktureller Informationen über G-Protein-gekoppelte Rezeptoren (GPCRs) ist für das Verständnis dieser komplexen Systeme von Vorteil. Wir beschreiben hier die Modellierung von GPCRs mit der Software ROSETTA, wobei auch flexible Regionen abgebildet werden.

Combination of functional and structural information of G-protein coupled receptors (GPCRs) aids the understanding of these complex systems. Here we describe modeling of GPCRs using the ROSETTA software suite, which also covers variable regions like loops or termini.

■ G-Protein-gekoppelte Rezeptoren (*G-protein coupled receptors*, GPCRs) sind in eine Vielzahl biologischer Prozesse involviert und daher ein interessantes Ziel für Grundlagenforschung und Wirkstoffentwicklung. Für das Verstehen ihrer komplexen Funktion ist es vorteilhaft, funktionelle und strukturelle Informationen zu verknüpfen. Die so generierten Modelle können dazu genutzt werden, Vorhersagen über das System zu treffen und zielgerichtet Experimente zu entwerfen, mit deren Hilfe die Modelle weiter verfeinert werden können.

Dieses Verfahren setzt jedoch Strukturkenntnisse voraus. Der Durchbruch in dieser Hinsicht war die Kristallstruktur des Rinder-

Rhodopsins im Jahr 2000 durch Palczewski und Mitarbeiter [1]. Es folgten weitere Strukturen des Sehpurpurs, auch in der aktiven Konformation, sowie von weiteren Rhodopsin-ähnlichen Rezeptoren (beta1- und beta2-adrenerger Rezeptor, Adenosin A_{2A}-Rezeptor, Chemokinrezeptor CXCR4). Es sind jedoch weitere etwa 800 verschiedene GPCRs bekannt [2], deren detaillierte Struktur noch immer unklar ist. Dies gilt insbesondere für die flexiblen und für individuelle Interaktionen entscheidenden intra- und extrazellulären Loops.

Computergestützte Strukturvorhersagen ermöglichen es, auf Basis der bekannten Kristallstrukturen Modelle von GPCRs zu erstellen, die auch für die flexiblen Bereiche eine hohe Genauigkeit aufweisen. Wir beschreiben im Folgenden, wie das frei verfügbare Programm ROSETTA für das Modellieren von GPCRs genutzt werden kann (Abb. 1).

Effiziente Suche im Konformationsraum

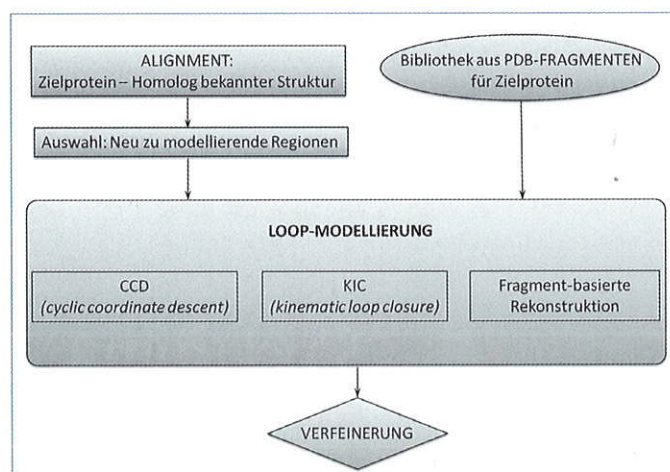
Aufgrund der Größe von GPCRs ist eine systematische Variation der Torsionswinkel des Peptidrückgrats und/oder der Seitenketten praktisch unmöglich. Stattdessen teilt ROSETTA die Suche im Konformationsraum in zwei Phasen [3]: In der ersten, größeren (*low-resolution*) Phase wird lediglich das Proteinerückgrat variiert. Dafür werden entsprechend der Sequenz und vorhergesagten Sekundärstruktur aus der Protein-Datenbank (PDB) [4] passende Rückgrat-Fragmente einer Länge von neun bzw. drei Aminosäuren gesucht. Diese werden in einem sogenannten Monte-Carlo-Metropolis-Verfahren zufällig kombiniert und nur dann akzeptiert, wenn sich die Gesamtenergie des Systems dadurch verbessert. Die dazugehörige Energiefunktion ist kein herkömmliches Kraftfeld, sondern reflektiert, wie oft diese Anordnung in der PDB vorhanden ist.

In der zweiten, verfeinernden (*high-resolution*) Phase entsteht ein *all atom model*. Es werden die Seitenkettenkonformationen variiert, wobei ebenfalls ausschließlich bekannte Konformere aus der PDB verwendet werden. Außerdem werden alle Freiheitsgrade des Modells energetisch minimiert [3].

Einbeziehung des Membranpotenzials

ROSETTA wurde ursprünglich für die Untersuchung löslicher Proteine entwickelt und optimiert. Die Umgebung eines Proteins innerhalb einer Lipidschicht unterscheidet sich jedoch erheblich von der wässrig-polaren Umgebung löslicher Proteine. Dies spiegelt sich in einer sehr spezifischen Verteilung der Aminosäuren in den verschiedenen Membranschichten sowie der Innen- bzw. Außenseite des Proteins wider.

Yarov-Yarovoy und Kollegen [5] haben auf der Basis von 28 bekannten Transmembranstrukturen einen ersten Algorithmus für ROSETTA für die *low-resolution*-Konforma-



◀ **Abb. 1:** Fließschema zum allgemeinen Vorgehen für das Modellieren eines Proteins mit dem Programm ROSETTA auf Basis einer bekannten homologen Struktur. Modifiziert nach [10].

tionsvorhersage von Membranproteinen entwickelt. Die Membran wird dabei durch eine 60 Å dicke, in sich weiter aufgegliederte (polare, Übergangs-, innere und äußere hydrophobe) Schicht repräsentiert. Nach jedem Schritt der Proteinfaltung wird die Position dieser Membran neu optimiert, sodass die Gesamtenergie des Systems möglichst gering ist. Diese energetische Bewertung ist ausreichend, um die Faltung des Proteins in die richtige Richtung zu lenken. Für die Auswahl der Modelle müssen aber alle erhaltenen Strukturen nach struktureller Ähnlichkeit gruppiert werden, wobei die Zentren der größten Cluster als beste Modelle ausgewählt werden.

In der Folge wurde auch ein Verfahren für *all atom models* entwickelt [6], sodass mit der oben beschriebenen Methode generierte Modelle verfeinert werden können. In dem verwendeten Kraftfeld werden auch schwache und verzweigte Wasserstoffbrücken inklusive ihrer Orientierung explizit bewertet. Die Interaktion mit der Membran wird über ein wissensbasiertes Kontinuum-Modell berücksichtigt, in das experimentelle Ergebnisse zum Energietransfer zwischen hydrophiler und hydrophober Phase einfließen. Diese Energiefunktion kann zwischen nativen und nicht-nativen Konformationen unterscheiden, das heißt in einer Auftragung von Energie gegen Abweichung zur nativen Struktur bildet sich ein Faltungstunnel aus, der idealerweise in einer Struktur mit der geringsten Energie und der geringsten Abweichung zur tatsächlichen Struktur mündet (**Abb. 2**). Damit sind die Voraussetzungen für die Strukturvorhersage von Membranproteinen geschaffen, die der Qualität für lösliche Proteine nicht nachsteht.

Homologie-Modelle und flexible Bereiche

Wie gelangt man nun zum Modell „seines“ G-Protein gekoppelten Rezeptors? Die derzeit gängige Methode für die Erstellung von GPCR-Modellen arbeitet auf Basis der bekannten Kristallstrukturen (**Abb. 1**). Dafür können verschiedene Kriterien hinsichtlich der Sequenz in den Transmembranbereichen genutzt werden, um das geeignete Vorbild zu finden [7]. Auf gleiche oder ähnliche Bereiche des Zielproteins werden dann die Koordinaten der Kristallstruktur übertragen. Die Loop-Regionen unterscheiden sich jedoch häufig sehr stark voneinander, sodass in diesen Bereichen

entweder das Modell komplett neu erstellt oder zumindest stark verändert werden muss. Dafür kann prinzipiell die oben beschriebene Fragment-basierte Methode verwendet werden. Die gegebenen Loop-Begrenzungen (das heißt die Koordinaten der angrenzenden Helices) erfordern jedoch eine Anpassung und Verfeinerung des Algorithmus, sodass Strangbrüche vermieden werden. Dafür werden die jeweiligen Regionen des Proteins auf Basis oben beschriebener Fragmente neu gefaltet, während der Rest des Modells unberührt bleibt. Dieses Protokoll kann in Kombination mit der *low-resolution*-Energiefunktion für Membranproteine verwendet werden.

Das ist von besonderer Bedeutung, wenn Bereiche modelliert werden sollen, die unmittelbar mit der Membran interagieren. Ein Beispiel dafür ist die strukturelle Untersuchung des C-Terminus von GPCRs. Dieser ist an vielen Rezeptorinteraktionen beteiligt und weist eine konservierte, amphiphile achte Helix auf, die direkt an der Membranoberfläche aufliegt [8]. Wir haben am Beispiel des C-Terminus von Tintenfisch-Rhodopsin (PDB: 2Z73A) erprobt, wie gut sich die Fragment-basierte Rekonstruktion in Membranumgebung für die Modellierung G-Protein-gekoppelter Rezeptoren eignet. Dafür wurden alle intrazellulären Loops sowie der in der Kristallstruktur aufgelöste Teil des C-Terminus 1.000-mal gefaltet. Danach wurden die resultierenden Strukturen nach Ähnlichkeit gruppiert und die Zentren der größten Gruppen mit der *all atom*-Energiefunktion je 1.000-mal verfeinert. Der anschließende Vergleich mit der Kristallstruktur zeigte, dass sich tatsächlich ein Faltungstunnel ausbildet. Dabei wiesen die energetisch besten Strukturen einen C_{α} -RMSD100 (*root-mean-square distance 100*, Abweichung des Proteinrückgrats normalisiert auf 100 Reste) von 3–4 Å auf, das entspricht etwa 5,5 Å auf die gesamte Sequenzlänge. **Abbildung 2** zeigt, dass die achte Helix dabei korrekt ausgebildet wurde und ihr amphiphiler Charakter deutlich ausgeprägt war. Die energetisch besten Modelle zeigten außerdem alle eine Interaktion des distalen C-Terminus mit dem dritten intrazellulären Loop. Noch bemerkenswerter war die Zuverlässigkeit, mit der der helikale Charakter dieses intrazellulären Loops vorausgesagt wurde. Alle energetisch günstigen Modelle wiesen in diesem Bereich die Konformation der Kristallstruktur auf. Diese Versuche zeigen ein-

Erfahren,
dynamisch,
erfolgsorientiert

...
Kompetenter Partner gesucht –
Allrounder im Labor, zuverlässig,
einfach zu händeln.

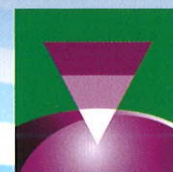
Ihre Favoriten:
SONOREX Ultraschallbäder
SONOPULS Homogenisatoren

- leistungsstark
- zuverlässig
- energiesparend
- umweltfreundlich

...
das ist ein
BANDELIN!



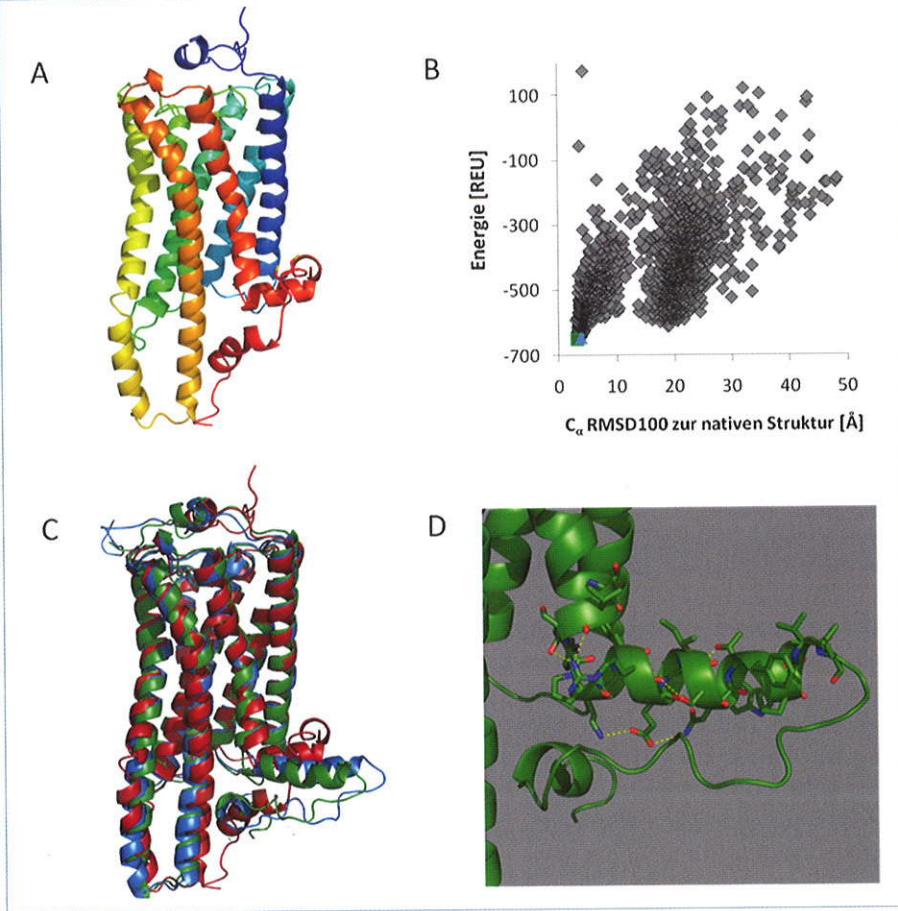
Besuchen Sie uns!



BANDELIN
The Ultrasound Company

www.bandelin.com

Analytica Vietnam
Halle A1 Stand B48



▲ **Abb. 2:** Modellierung des C-Terminus und der intrazellulären Loops des Tintenfisch-Rhodopsins mittels Fragment-basierter Rekonstruktion in Membranumgebung. **A**, native Struktur (PDB: 2Z73A), Farbverlauf vom N-Terminus (blau) zum C-Terminus (rot). **B**, Auftragung der Energien (ROSETTA Energy Units, REU) gegen die Abweichung zur Kristallstruktur. Es bildet sich ein Faltungstunnel aus, das heißt die Modelle mit der geringsten Energie haben die geringste Abweichung zur nativen Struktur. **C**, Überlagerung der beiden Modelle mit der niedrigsten Energie (grün und blau) mit der nativen Struktur (rot). **D**, Detailaufnahme der achten Helix des energetisch besten Modells. Hydrophobe Reste sind der Membranoberfläche ausgesetzt, hydrophile dem Zytosol (rot: Sauerstoff; blau: Stickstoff; grün: Kohlenstoff und Proteinrückgrat; gelb: polare Interaktionen).

drucksvoll, dass selbst mit einer verhältnismäßig geringen Zahl an gebildeten Strukturen und somit einem geringen Rechenaufwand sehr gute Modelle eines G-Protein-gekoppelten Rezeptors in der Membranumgebung erreicht werden können.

GPCR-Strukturen ganz ohne Kristall?

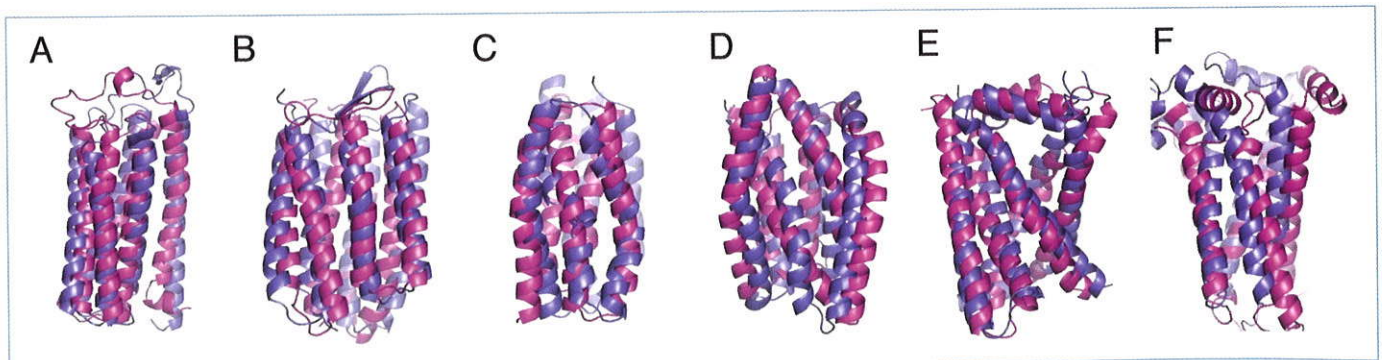
Eine weitere Herausforderung für die Strukturvorhersage ist das Erstellen von Modellen für Membranproteine ohne homologes Struktur Vorbild. Dies könnte beispielsweise von Vorteil sein für GPCRs der Klasse B oder C, für die bislang noch keine Struktur bekannt ist, oder auch für nicht-G-Protein-gekoppelte

Multitransmembranrezeptoren wie die Adiponektinrezeptoren, deren Struktur ebenfalls unbekannt ist.

Mithilfe der *low-resolution*-Energiefunktion für Membranproteine können für die *de novo*-Strukturvorhersage von Multitransmembranomänen wie die zentralen Helices von Bakteriorhodopsin bereits RMSD100-Werte von 2,8 Å (3,1 Å über 123 Reste) erreicht werden [5]. Dies entspricht in etwa der Genauigkeit von Modellen für lösliche Proteine dieser Länge. Für komplexere und längere Strukturen wie die des Rinder-Rhodopsins kann derzeit zumindest die prinzipielle Faltung und Anordnung korrekt vorhergesagt wer-

den, die besten Strukturen wiesen in unseren Tests einen RMSD100 von etwas unter 7 Å auf (11 Å über alle 348 Reste).

Eine Möglichkeit, die Qualität der *de novo*-Modelle deutlich zu verbessern, ist die Einbeziehung experimenteller Daten, beispielsweise aus Elektronenspinresonanz-Experimenten, aber auch aus Mutagenese-Studien oder Fluorophor-Interaktionen [9]. Dies gibt Auskunft über Interaktionen von in der Sequenz weit auseinanderliegenden Resten, was den möglichen Konformationsraum drastisch beschränkt. Unter Annahme nur einer solchen Interaktion konnten wirklichkeitstreuere Modelle von Proteinen mit bis zu sechs Transmembranhelices erstellt werden (**Abb. 3**, [9]). Beispielsweise wurde für die *de novo*-Strukturvorhersage der C-terminalen Domäne der Lactose-Permease (sechs Transmembranhelices; **Abb. 3D**) eine experimentell bestimmte Salzbrücke verwendet; nach Strukturverfeinerung hatte das beste Modell einen RMSD100 zur Kristallstruktur von nur 3,1 Å (4,2 Å über 185 Reste) [9]. Das Modellieren der strukturell komplexeren eukaryotischen



▲ **Abb. 3:** Beste Modelle (magenta) und Kristallstruktur (violett) von sechs helikalen Transmembranproteinen. Die Strukturen wurden mit der ROSETTA-Membran-Energiefunktion unter Annahme einer nicht lokalen Interaktion generiert. **A**, Cytochrom C, Kette A (5 Transmembranhelices/191 Reste). **B**, Bakteriorhodopsin (7/227). **C**, Lactose-Permease, N-terminaler Teil (6/190). **D**, Lactose-Permease, C-terminaler Teil (6/185). **E**, Fumarat-Reduktase, Kette H (5/216). **F**, Cytochrom bc1, Kette D (5/222). Abbildung aus [9] mit freundlicher Genehmigung der Autoren.

G-Protein-gekoppelten Rezeptoren bleibt jedoch auch mit diesem Ansatz anspruchsvoll, wahrscheinlich werden für eine genaue Strukturvorhersage mehrere Distanzen benötigt.

Danksagung

Wir danken dem DAAD für die finanzielle Unterstützung der Reisekosten nach Nashville.

Literatur

[1] Palczewski K, Kumasaka T, Hori T et al. (2000) Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* 289:739–745

- [2] Millar RP, Newton CL (2010) The year in G protein-coupled receptor research. *Mol Endocrinol* 24:261–274
- [3] Rohl CA, Strauss CE, Misura KM et al. (2004) Protein structure prediction using Rosetta. *Methods Enzymol* 383:66–93
- [4] Berman HM, Westbrook J, Feng Z et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242
- [5] Yarov-Yarovsky V, Schonbrun J, Baker D (2006) Multipass membrane protein structure prediction using Rosetta. *Proteins* 62:1010–1025
- [6] Barth P, Schonbrun J, Baker D (2007) Toward high-resolution prediction and design of transmembrane helical protein structures. *Proc Natl Acad Sci USA* 104:15682–15687
- [7] Worth CL, Kleinau G, Krause G (2009) Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. *PLoS One* 4:e7011
- [8] Bockaert J, Marin P, Dumuis A et al. (2003) The 'magic tail' of G protein-coupled receptors: an anchorage for functional protein networks. *FEBS Lett* 546:65–72

- [9] Barth P, Wallner B, Baker D (2009) Prediction of membrane protein structures with complex topologies using limited constraints. *Proc Natl Acad Sci USA* 106:1409–1414
- [10] Kaufmann KW, Lemmon GH, DeLuca SL et al. (2010) Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry* 49:2987–2998

Korrespondenzadresse:

Prof. Dr. Jens Meiler
Vanderbilt University
Center for Structural Biology
7330 Stevenson Center, Station B 351822,
Nashville, TN 37235, USA
Tel.: +1-(0)615-936-5662
Fax: +1-(0)615-936-2211
jens.meiler@vanderbilt.edu
www.meilerlab.org
www.rosettacommons.org

AUTOREN



Anette Schreiber

2005–2010 Biochemiestudium an der Universität Leipzig, Juni–August 2010 Aufenthalt bei Prof. Dr. J. Meiler an der Vanderbilt University, Nashville, TN, USA. Seit 2010 Doktorandin bei Prof. Dr. A. G. Beck-Sickingher an der Universität Leipzig.



Jens Meiler

1993–1998 Chemiestudium an der Universität Leipzig, 1998–2001 Promotion bei Prof. Dr. C. Griesinger, Universität Frankfurt a. M., 2001–2005 Postdoc bei Prof. Dr. D. Baker, University of Washington, Seattle, WA, USA. Seit 2005 Professur für Chemie, Pharmakologie, und Biomedizinische Informatik an der Vanderbilt University, Nashville, TN, USA.

Gut verpackt

...und ab in die Zellen...



Transfektionsreagenzien

- hoch effizient
- zum günstigen Preis
- schnelles und einfaches Protokoll
- schafft höchste Expressionsraten
- Serum stört die Transfektion nicht
- als ready-to-use Reagenz

AppliFect

- AppliFect – DNA/RNA
- AppliFect LowTox – DNA, für sensitive Zellen
- AppliFect Insect – DNA für Insektenzellen
- AppliFect SI – für siRNA, miRNA

AppliChem

Darmstadt hat eine weitere Topadresse: