



Shantheya Balasupramaniam (Autor)  
**Validierung der Leistungsfähigkeit von  
Bewertungsfunktionen zur Vorhersage von  
Bindungsaffinitätsdifferenzen auf Basis von Ligand-  
Paar-Datensätzen**



<https://cuvillier.de/de/shop/publications/8006>

Copyright:  
Cuvillier Verlag, Inhaberin Annette Jentsch-Cuvillier, Nonnenstieg 8, 37075 Göttingen,  
Germany  
Telefon: +49 (0)551 54724-0, E-Mail: [info@cuvillier.de](mailto:info@cuvillier.de), Website: <https://cuvillier.de>

# 1. Einleitung

Die Bindungsaffinität eines Moleküls zu einem Zielprotein lässt sich auf experimentellem Wege bestimmen oder mittels theoretischer Ansätze vorhersagen (Abbildung 1). Bei den experimentellen Verfahren zur Affinitätsbestimmung wird zwischen Methoden, die thermodynamische Eigenschaften der Protein-Ligand-Komplexbildung direkt vermessen (z. B. isotherme Titrationskalorimetrie [1]) und denen, die anhand von Bindungs- oder Verdrängungsassays indirekt auf die Bindungsaffinität schließen lassen, unterschieden.



**Abbildung 1. Methoden zur Bestimmung der Bindungsaffinität.**

Zur Bestimmung der Bindungsaffinität stehen experimentelle sowie theoretische Verfahren zur Verfügung. Die experimentellen Methoden gliedern sich in direkte und indirekte Methoden. Die Wahl des theoretischen Verfahrens richtet sich danach, ob die Struktur des Zielproteins bekannt ist oder nicht. Bewertungsfunktionen finden bei Methoden, die auf bekannte Zielstrukturen zurückgreifen, Anwendung.

Bei den theoretischen Ansätzen, durch die eine Vorhersage der zu erwartenden Bindungsaffinität gelingen soll, richtet sich das verwendete Verfahren nach dem Aufklärungsgrad der dreidimensionalen Struktur des Zielproteins (Rezeptor). Liegt die Rezeptorstruktur nicht vor, so kommen unter der Annahme, dass strukturell verwandte Verbindungen sich auch in ihrer biologischen Aktivität ähneln sollten, ligandbasierte Ansätze (z. B. QSAR Verfahren) zum Einsatz. Ist die Rezeptorstruktur aufgeklärt, bieten strukturbasierte Ansätze (z. B. De-novo-



Design) die Möglichkeit, das Verhalten des betrachteten Proteins gegenüber verschiedenen Liganden zu charakterisieren [2].

Zunächst soll auf die Anfänge der theoretischen Ansätze, die auf Strukturinformationen eines Proteins zurückgriffen, eingegangen werden. Als eine der Pioniere entwickelten Eisenberg und McLachlan von den Atomkoordinaten eines Proteins ausgehend ein einfaches Additionsschema, um den Beitrag eines jeden Proteinatoms zur freien Solvationsenergie des gesamten Proteins zu berechnen und somit Struktur und Energie in einen empirischen Zusammenhang zu bringen [3, 4].

Anfang der 90er Jahre wurden erstmals Protein-Ligand-Interaktionen additiv verknüpft, um die freie Bindungsenergie (und somit die Bindungsaffinität) eines Protein-Ligand-Komplexes (Abk. PLK) bekannter Raumstruktur quantitativ abzuschätzen [5]. Zu den heute weitverbreiteten Methoden des strukturbasierten Wirkstoffdesigns, die über die Beschreibung von Protein-Ligand-Interaktionen eine Vorhersage der Bindungsaffinität ermöglichen, zählen die Bewertungsfunktionen (engl.: *scoring functions*) [6, 7]. In den folgenden Abschnitten soll auf diese computergestützten Methoden näher eingegangen werden.



## 1.1 Bewertungsfunktionen

Bewertungsfunktionen sind einfache mathematische Modelle zur Vorhersage der Stärke von nicht-kovalenten Interaktionen zwischen zwei Bindungspartnern [8]. Sie berechnen die freie Bindungsenthalpie einer gegebenen Wechselwirkungsgeometrie zwischen einem Liganden und einem Protein und beurteilen, ob diese als energetisch günstig einzuordnen ist [9, 10]. Anhand dieser Bewertung kann anschließend die zu erwartende Bindungsaffinität näherungsweise abgeschätzt werden [10]. Diesem Ansatz liegt die Annahme zu Grunde, dass die Bindungsaffinität eines Liganden zu seiner Zielstruktur von enthalpischen und entropischen Faktoren abhängt und dabei die elektronischen Eigenschaften beider Bindungspartner sowie die strukturelle Komplementarität eine entscheidende Rolle spielen [11–13].

Weitere Möglichkeiten zur Bestimmung der freien Bindungsenergie stellen unter anderem die Freie-Energie-Störungsrechnung (engl.: *free energy perturbation* (Abk. FEP)) sowie Methoden der linearen Interaktionsenergie (Abk. LIE) und thermodynamischen Integration (Abk. TI) dar [2, 8, 9, 14, 15]. Diese besitzen zwar eine höhere Vorhersagegenauigkeit, sind aber auch rechenintensiver als Bewertungsfunktionen, die wiederum dem Anwender einen Kompromiss zwischen Geschwindigkeit und Genauigkeit anbieten [6, 16].

### 1.1.1 Einsatzgebiete

Die Einsatzgebiete von Bewertungsfunktionen sind heutzutage so vielfältig wie unterschiedlich und sollen im Folgenden kurz erläutert werden:

Zu den Haupteinsatzgebieten gehört das **Docking**, bei dem unter Beachtung der physikochemischen Eigenschaften sowie der räumlichen Beschaffenheit beider Bindungspartner versucht wird, ein kleines Molekül in eine Proteinbindetasche optimal einzupassen. Klassischerweise soll hierbei das Verhalten eines Liganden gegenüber einem strukturaufgeklärten Protein untersucht werden, wenn dafür keine Bindungsstudien vorliegen. Da es sich bei dem betrachteten Liganden häufig um ein flexibles Molekül handelt, werden dreidimensionale Konformere benötigt bzw. während des Dockingprozesses erzeugt und in der (vermuteten)



Bindetasche des Proteins platziert. Die Molekülflexibilität kann auf verschiedenste Weise gehandhabt werden wie etwa durch die Verwendung von Liganden aus einer Multikonformdatenbank (z. B. FRED [17]), durch den inkrementellen Aufbau des Liganden in der Bindetasche (z. B. FlexX [18]) oder durch stochastische Methoden, die den Konformationsraum des Liganden abtasten (z. B. GOLD [19]).

Während des Dockingprozesses werden Lösungsansätze, die aufgrund von Abstoßungskräften und räumlicher Diskrepanz als problematisch eingestuft werden, verworfen und akzeptable Bindungsgeometrien (Posen) beibehalten. Die energetische Beurteilung sowie Optimierung der generierten Posen des Liganden erfolgen unter Anwendung von Bewertungsfunktionen [8, 9, 20]. Ziel hierbei ist es, neben der tatsächlich zu erwartenden Bindungspose eines Liganden (engl.: *pose prediction*) auch seine entsprechende Bindungsaffinität möglichst genau vorherzusagen [21] und dadurch die Konformation und Ausrichtung zu finden, die bei der nativen Bindung an das Protein vorliegen würde [9].

Während die dreidimensionale Struktur eines Liganden durch diverse Programme wie z. B. MOE Builder [22] erzeugt werden kann, wird vom Protein eine experimentell bestimmte Raumstruktur benötigt. Derartige Raumstrukturen von biologischen Makromolekülen werden meist durch Methoden wie der NMR-Spektroskopie oder der Röntgenstrukturanalyse gewonnen, wobei beide Methoden ihre eigenen Vorteile haben. Der Einsatz der NMR-Spektroskopie zur Strukturaufklärung ist auf kleinere Proteine mit geringerem Molekulargewicht beschränkt, bietet aber den Vorteil, dass durch die Vermessung in Lösung auch die dynamischen Eigenschaften des Proteins miterfasst werden können. Daher liefert eine NMR-spektroskopische Untersuchung gleich eine Vielzahl von Strukturvorschlägen [10]. Im Gegensatz hierzu muss für eine Röntgenstrukturanalyse zunächst das Protein im kristallinen Zustand vorliegen, wobei diese Methode auch bei der Strukturaufklärung großer Proteine sowie Proteinkomplexe Anwendung findet. Auch im Hinblick auf das Ergebnis zeigen beide Methoden Unterschiede. Während die NMR-Spektroskopie also viele „Lösungen“ anbietet, wird bei der Röntgenstrukturanalyse ein einzelner Strukturvorschlag (Kristallstruktur) erhalten.

Eine weitere, wenn auch weniger gängige, Alternative zur experimentellen Strukturbestimmung stellt die Elektronenmikroskopie dar, welche aber aufgrund der aufwändigen Probenvorbereitung sowie der Notwendigkeit vieler Proteinkristalle der Röntgenstrukturanalyse als



führende Methode klar unterlegen ist [10]. Weiterhin bietet auf der virtuellen Ebene die Homologiemodellierung eine theoretische Möglichkeit zur Strukturvorhersage.

Raumstrukturen von Proteinen zum Zweck von Dockingstudien können aus der *Protein Data Bank* (Abk. PDB) extrahiert werden [23]. Die PDB stellt eine freierfügbare Sammlung experimentell bestimmter dreidimensionaler Strukturen von biologischen Makromolekülen wie Proteinen oder Nukleinsäuren dar [24–26]. Sie enthält derzeit 135.519 Raumstrukturen, von denen 125.799 auf Proteine entfallen (Stand: Dezember 2017). Über 90% dieser Proteinstrukturen gehen auf die Strukturaufklärung durch die Röntgenstrukturanalyse zurück. Jeder Eintrag in der Datenbank wird durch eine vierstellige Bezeichnung (Abk. PDB-Code) eindeutig gekennzeichnet.

Eine Übersicht über die verschiedenen Docking Programme sowie Docking Webserver wird vom Swiss Institute of Bioinformatics (Abk. SIB) zur Verfügung gestellt [27].

Ein weiteres Haupteinsatzgebiet von Bewertungsfunktionen auf computergestützter Ebene ist das **virtuelle Screening**, welches ligandbasiert (z. B. mittels Ähnlichkeitssuche), strukturbasiert (z. B. mittels Docking) oder als Kombination aus beidem durchgeführt werden kann. Dabei werden große Moleküldatenbanken (Screening-Datenbank) meist sukzessiv mit zunehmend komplexeren Filtermethoden nach vermeintlich bioaktiven Verbindungen durchforstet [28], die im späteren Verlauf biologischen Testungen unterzogen werden können. Im Fokus stehen hierbei zum einen die Entdeckung arzneistoffähnlicher (engl.: *druglike*) Moleküle, die im Hinblick auf ihre Eigenschaften und Wirkung zu einem arzneilich bedeutsamen Wirkstoff optimiert werden können und zum anderen die Identifizierung neuartiger chemischer Strukturen, die sich als geeignete Leitstrukturen erweisen [29]. Moleküle für derartige Screening-Datenbanken werden aus einem enorm großen chemischen Raum gewonnen, der sowohl existente Verbindungen als auch hypothetische Molekülstrukturen umfasst [30]. Dabei stellen die Datenbanken nur einen Bruchteil des chemischen Raumes dar. Um eine vage Vorstellung vom Umfang des chemischen Raumes zu bekommen, sei auf die von Ruddigkeit et al. entwickelte Datenbank GDB-17 verwiesen, die sich lediglich auf Moleküle mit bis zu 17



Kohlenstoff-, Stickstoff-, Sauerstoff-, Schwefel- sowie Halogenatomen „beschränkt“, aber bereits mehr als 166 Milliarden Einträge umfasst [31]!

Ligandbasierte Ansätze kommen im virtuellen Screening zum Einsatz, wenn für die untersuchte Zielstruktur mindestens eine aktive Verbindung (Referenzmolekül) bekannt ist und diese genutzt werden kann, um Datenbanken nach ähnlichen Molekülen zu durchforsten. Die gängigsten Methoden übersetzen aus Gründen der Vergleichbarkeit die Informationen der untersuchten Moleküle in Deskriptoren (z. B. 2-D- oder 3-D-Deskriptoren) und nutzen anschließend Ähnlichkeitsmaße (z. B. den Tanimoto-Koeffizienten), um die Verbindungen nach abnehmender Ähnlichkeit zum Referenzmolekül zu sortieren. Sind für die Zielstruktur neben aktiven Verbindungen auch inaktive bekannt, so kann unter Verwendung von Methoden des Maschinellen Lernens (z. B. QSAR) eine entsprechende Klassifizierung dieser Verbindungen erfolgen [32, 33].

Auf der Seite der strukturbasierten Ansätze zum virtuellen Screening können Docking Programme eingesetzt werden, um für eine vorliegende dreidimensionale Proteinstruktur die Screening-Datenbank nach aktiven Liganden zu durchsuchen und anschließend Lösungen hinsichtlich des zu erwartenden Bindungsmodus zu erzeugen. Für die Verarbeitung der dabei entstehenden Vielzahl an Lösungsansätzen werden schnelle Methoden wie die Bewertungsfunktionen benötigt. Neben der energetischen Beurteilung und Optimierung der generierten Bindungsweisen, um hierdurch bindende und nicht bindende Moleküle voneinander unterscheiden zu können, dienen sie weiterhin zur Rangbildung der Moleküle entsprechend ihrer vorhergesagten freien Bindungsenergie bzw. Bindungsaffinität [9]. Dabei wird für die Moleküle auf den obersten Rängen angenommen, dass sie mit hoher Wahrscheinlichkeit biologisch aktiv sind.

Neben diesen Haupteinsatzgebieten finden Bewertungsfunktionen weiterhin Anwendung bei der Vorhersage von Bindetaschen [34], bei der Leitstrukturoptimierung [35] sowie bei der Beurteilung der therapeutischen Tauglichkeit von Targets [36, 37].



### 1.1.2 Anforderungen und Validierungsgrundlagen

Die soeben beschriebenen Einsatzgebiete lassen erkennen, dass Bewertungsfunktionen für eine Vielzahl von Anwendungen geeignet sein sollten. Idealerweise sollten sie gegenüber kleinen strukturellen Unstimmigkeiten unempfindlich sein und ein breites Affinitätsspektrum abdecken. Darüber hinaus ist gerade beim virtuellen Screening die benötigte Rechenzeit der gewählten Methode von großer Bedeutung.

Um die Eignung der Bewertungsfunktionen für die diversen Zwecke zu beurteilen, können verschiedene Ansätze der Validierung verfolgt werden. Die gewählte Validierungsstrategie richtet sich dabei nach dem jeweiligen Ziel.

Wird die Fähigkeit einer Bewertungsfunktion zur Identifizierung der nativen bzw. der dieser am nächsten kommenden Bindungspose inmitten von „Köder“-Posen, die mittels Docking generiert wurden, bewertet, so wird dies als *docking power* jener Funktion bezeichnet. Hierbei werden alle Bindungsposen betrachtet, wobei der native bzw. nahezu native Bindungsmodus im Idealfall die beste Bewertung (Score) erhalten sollte [38, 39].

Die *scoring power* beschreibt die Fähigkeit einer Bewertungsfunktion, Scores für Komplexe mit bekannter Raumstruktur zu generieren, die mit ihren experimentellen Bindungsaffinitätsdaten (linear) korrelieren. Dabei wird für eine Vielzahl von diversen Zielstrukturen (Targets) überprüft, ob die Methode sich zur Vorhersage von Bindungsaffinitäten eignet und wie präzise sie arbeitet [38, 39].

Die *ranking power* repräsentiert die Fähigkeit einer Bewertungsfunktion, verschiedene Liganden eines Proteins gemäß ihrer Bindungsaffinität richtig zu reihen, wobei die native Bindungspose dieser Liganden bekannt sein sollte. Hierbei steht die Leistung der betrachteten Methode auf den einzelnen Targets im Vordergrund [38, 39].

Unter der *screening power* wird die Fähigkeit einer Bewertungsfunktion zur Identifizierung von aktiven Verbindungen inmitten von Zufallsmolekülen verstanden [16, 39].





### 1.1.3 Klassen von Bewertungsfunktionen

Bewertungsfunktionen können grob in drei Hauptklassen gegliedert werden. Dabei wird zwischen *kraftfeldbasierten*, *empirischen* und *wissensbasierten* Methoden unterschieden. Alle Ansätze beruhen auf der Annahme, dass sich die freie Bindungsenergie einer Protein-Ligand-Bindung als Summe einzelner Beiträge darstellen lässt [7]. Im Folgenden soll kurz auf die verschiedenen Klassen eingegangen werden. Detaillierte Beschreibungen der verschiedenen Klassen mit entsprechenden Vertretern finden sich in diversen Publikationen und Lehrbüchern [2, 7, 28, 40–42].

#### 1.1.3.1 Kraftfeldbasierte Bewertungsfunktionen

Kraftfeldbasierte Bewertungsfunktionen nutzen klassische Molekülmechanik-Kraftfelder (z. B. AMBER [43–45], CHARMM [46]), um direkte nicht-kovalente Protein-Ligand-Interaktionen zu berechnen. Die gesamte freie Bindungsenergie wird über die Summe aller Terme für die intermolekularen van-der-Waals-Wechselwirkungen sowie elektrostatischen Interaktionen (Abk. Energieterme) abgeschätzt [2, 6, 8, 38]. Hierbei werden auch gelegentlich intramolekulare Energiebeiträge (Spannungsenergien) miteinbezogen. Die Solvatation bzw. Desolvatation kann durch das verwendete Molekülmechanik-Kraftfeld implizit mitbetrachtet werden. Alternativ hierzu lässt sich der Beitrag der Solvatation sowie einiger entropischer Faktoren durch die Verwendung der Poisson-Boltzmann-Gleichung (Abk. PB) [47] bzw. des „Generalisierten Born-Modells“ (Abk. GB) [48, 49] in Kombination mit einem Term für die „Solvens zugängliche Oberfläche“ (engl.: *solvent-accessible surface area* (Abk. SA)) bestimmen [8].

Nachteile kraftfeldbasierter Ansätze ergeben sich aus ihrer unzureichenden Beschreibung entropischer Effekte des Protein-Ligand-Bindungsprozesses und ihrer damit verbundenen Neigung, großen stark polaren Molekülen sehr hohe Scores zuzuweisen. Hinzu kommen die intrinsischen Fehler, die individuelle Energieterme der verschiedenen Bewertungsfunktionen in sich bergen. Gegenüber den anderen Hauptklassen besitzen kraftfeldbasierte Ansätze den Vorteil, dass sie keinen Trainingsdatensatz benötigen. Da sie auf verfügbare Kraftfelder und Solvatationsmodelle zurückgreifen, profitieren sie auch von deren fortschrittlichen Entwick-



lung und besitzen daher das größte Potenzial, künftig die Bindungsenergien zwischen einem Protein und einem Liganden zuverlässig vorherzusagen [6, 9, 50].

Beispiele für kraftfeldbasierte Bewertungsfunktionen sind *GoldScore* [19], *AutoDock* [51] und frühere Versionen von *DOCK* [52, 53].

### 1.1.3.2 Empirische Bewertungsfunktionen

Empirische Bewertungsfunktionen zerlegen die gesamte freie Bindungsenergie in Beiträge, die wichtige energetische Eigenschaften einer nicht-kovalenten Protein-Ligand-Bindung charakterisieren. Jeder Term setzt sich dabei aus einem Deskriptor und einem Gewichtungsfaktor zusammen. Der Deskriptor beschreibt strukturelle Merkmale eines PLK und seine geometrischen Charakteristika. Der Gewichtungsfaktor wird durch eine Regressionsanalyse (z. B. multiple lineare Regression (Abk. MLR)) auf Basis eines Trainingsdatensatzes bestehend aus PLKs bekannter Raumstruktur und experimentell bestimmter Bindungsaffinität bestimmt. Anschließend werden die Gewichtungsfaktoren mit ihren zugehörigen Deskriptoren multiplikativ verknüpft [6, 8].

Der Umfang sowie die Proteinvierfalt des Trainingsdatensatzes haben einen ebenso großen Einfluss auf die Qualität der daraus resultierenden Bewertungsfunktion wie die Auflösung der enthaltenen Kristallstrukturen, die Genauigkeit der experimentellen Bindungsdaten und der Affinitätsbereich, der durch den Trainingsdatensatz abgedeckt wird. Diese Aspekte entscheiden schließlich auch über die Anwendbarkeit einer empirischen Bewertungsfunktion auf Proteine, die außerhalb des „Chemo- bzw. Genotyps“ [6] des Trainingsdatensatzes liegen. Des Weiteren bleiben weniger gängige Interaktionen (wenn nicht sehr stark und spezifisch) bei der Regressionsanalyse für gewöhnlich außen vor. Auch die Quantifizierung des entropischen Beitrages sowie die Erfassung sämtlicher am Bindungsprozess beteiligten Faktoren auf Deskriptorebene, erweisen sich als schwierig wenn nicht sogar als unmöglich [6].

Zu den empirischen Bewertungsfunktionen zählen populäre Methoden wie *ChemScore* [54, 55] sowie *Glide-Score* [56, 57] und *X-Score* [58]. Arbeiten von Fornabaio et al., Kerzmann et al. und Catana et al. enthalten Beispiele für benutzerdefinierte Formen dieser Klasse von



Bewertungsfunktionen, die durch das Hinzufügen bzw. Entfernen gerichteter Terme an spezielle Fragestellungen angepasst wurden [59–61].

### 1.1.3.3 Wissensbasierte Bewertungsfunktionen

Wissensbasierte Bewertungsfunktionen unterliegen der Annahme, dass Informationen, die aus einem Trainingsdatensatz von strukturell aufgeklärten PLKs extrahiert werden können, nützliche Erkenntnisse zur Beurteilung von intermolekularen Interaktionen liefern. Die Häufigkeit mit der verschiedene Proteinatom-Ligandatom-Kontakte im Trainingsdatensatz auftreten, wird dabei als Maß für den Beitrag dieser Kontakte zur jeweiligen Protein-Ligand-Bindung angesehen [6]. So wird angenommen, dass Atompaaungen, die in einem bestimmten Abstand häufig auftreten (z. B. im Vergleich zum Zufall als Referenzzustand), auf eine energetisch günstige Interaktion hinweisen, während selten auftretende Kontakte als Indiz für den umgekehrten Fall gelten [8]. So können den verschiedenen Atompaaungen über eine statistische Auswertung und unter Verwendung des inversen Boltzmann Gesetzes distanzabhängige Potenziale zugewiesen werden [6]. Die Gesamtheit aller distanzabhängiger Potenziale dient schließlich der Bewertung eines PLKs [8].

Wie bereits im Abschnitt 1.1.3.2 zu den empirischen Bewertungsfunktionen beschrieben, spielt auch bei den wissensbasierten Bewertungsfunktionen die Qualität des Trainingsdatensatzes eine entscheidende Rolle im Hinblick auf seine Anwendbarkeit. Im Gegensatz zu den empirischen Ansätzen, bei denen neben der Raumstruktur der PLKs auch deren experimentell bestimmte Bindungsaffinität benötigt wird, bedarf es bei den wissensbasierten Methoden nur deren dreidimensionalen Struktur [6–9, 38]. Um zuverlässige Aussagen hinsichtlich der Häufigkeit des Auftretens verschiedener Proteinatom-Ligandatom-Paarungen treffen zu können, sollte sich der Trainingsdatensatz idealerweise aus einer Vielzahl von diversen Komplexen zusammensetzen, deren Auflösung zur Bestimmung distanzabhängiger Potenziale geeignet ist. Für Atomtypen, die in den PLKs selten vertreten sind, ist eine solide statistische Auswertung jedoch kaum möglich [8].

Zu den wissensbasierten Bewertungsfunktionen gehören die Methoden *S<sub>MOG</sub>* [62], *PMF* [63, 64] sowie *DrugScore* [65]. Die verschiedenen Ansätze dieser Klasse können sich unter ande-