Strukturbasiertes Klassifizieren von Protein-Proteininteraktionen

Andreas Henschel

Technische Universität Dresden
ahenschel@gmail.com

1 Einleitung

Für das Verständnis der molekularen Wechselwirkungen in der Zelle ist die Identifikation von allgemeinen Prinzipien in Protein-Protein-Interaktionen (PPI) von größter Bedeutung. Ein integrativer Ansatz, der die Gesamtheit der bereits bekannten Protein-Protein-Interaktionen einschließt, hilft, die komplexen Aspekte des Metabolismuses und die Entstehung von Krankheiten zu verstehen. Eine wertvolle Informationsquelle von interagierenden Proteinen sind dreidimensionale Komplexe von Proteinen, die physisch miteinander in Kontakt stehen (siehe Abbildung 2). Diese Strukturen zeigen, wie Protein-Protein-Interaktionen auf atomarer Ebene funktionieren. Es ist möglich, die Atome und Residuen zu identifizieren, die in die Wechselwirkung involviert sind und damit das Interface darstellen.

Für viele PPI sind nur die Aminosäuresequenzen (kurz: Sequenzen), nicht aber die Proteinstrukturen mit dreidimensionalen Atomkoordinaten (kurz: Strukturen), bekannt. Allerdings führte der weltweit vereinte Einsatz zu stetigem Wachstum der zentralen Datenbank für Proteinstrukturen (Protein Databank, PDB), sodass die vorhandenen Datenmengen statistische Analysen zulassen.

Diese Arbeit demonstriert Lösungen für folgende drei Probleme:

1. Eine Klassifikation von Protein-Protein-Interaktionen, siehe Abschnitt 2
2. Ein systematischer Ansatz zur Identifizierung konvergent evolvierter Motive in Protein-Interaktionen, siehe Abschnitt 3
3. Die Generierung von Motiv-Deskriptoren zur Vorhersage von Proteininteraktionen auf Sequenzebene, siehe Abschnitt 4

Ein Überblick über die zueinander in Beziehung stehenden Probleme und die relevanten Ressourcen ist in Abbildung 1 dargestellt. Proteine bestehen aus einer oder mehreren Domänen, d.h. strukturell unabhängigen Einheiten, die eine modularere Sichtweise auf Proteine und Interaktionen erlauben. Diese Domänen können basierend auf sequenzieller

oder struktureller Ähnlichkeit auf einen gemeinsamen Ursprung untersucht und daraufhin in Familien eingeteilt werden.

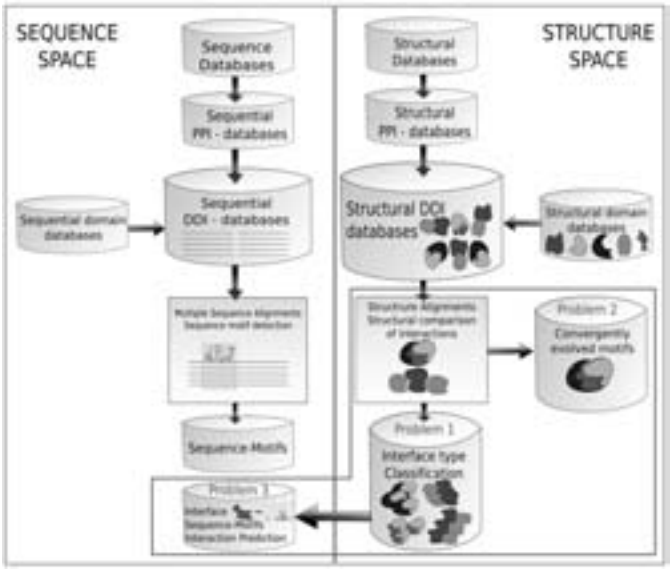


Abbildung 1: Der in dieser Arbeit dargestellte Ansatz ist struktureller Natur. Die herkömmlichen entsprechenden Schritte auf sequenzieller Ebene sind auf der linken Seite dargestellt. Der rotmarkierte Bereich zeigt den Fokus dieser Arbeit.

2 Eine strukturbasierte Protein-Protein-Interaktionsklassifikation

2.1 Einleitung und Problemstellung

Proteine wechselwirken auf sehr spezifische Weise [NT03]. Um diese Interaktionsvielfalt kategorisch zu erfassen, ist eine Klassifikation der Interfaces notwendig [Alo07]. Diese ermöglicht es, generelle Prinzipien für jede Interface-Klasse abzuleiten. Für die Proteininteraktionsanalyse existieren einige Datenbanken. Eine noch offene Herausforderung ist jedoch die Erstellung einer Interaktions-Datenbank, die Familien-Familieninteraktionen nach geometrischer Assoziation klassifiziert.

Problemstellung 1: Interface-Klassifikation
Ziel ist es, eine Klassifikation von Domän-Domäninteraktionen von allen verfügbaren Multi-Domän-Komplexen mit Strukturinformation zu erstellen. Die Interaktionen sollen nicht nur auf Basis der beteiligten Domänfamilien, sondern auch auf Basis der geometrischen Assoziation unterschieden werden. Für jede der sich ergebenden Interface-Klassen sollen die interagierenden Residuen ermittelt und visualisiert werden, sowohl in Sequenz als auch in Struktur. Der Algorithmus soll einen guten Kompromiss zwischen Rechenaufwand und Klassifikationsgenauigkeit darstellen.

Die Motivation für eine derartige Datenbank ist vielfältig. Man kann ersehen, wie viele Interfaces eine Domäne besitzt und welches für welchen Partner genutzt wird. Für eine gegebene Familien-Familien-Interaktion ist es interessant zu sehen, ob und welche alternativen Bindungsmöglichkeiten existieren, siehe z.B. Abbildung 2. Weiterhin ist es sinnvoll, beim Modellieren makromolekularer Proteinkomplexe, nicht nur eine (wie in [ABC+04]) sondern alle bekannten alternativen Bindungsmöglichkeiten in Erwägung zu ziehen.

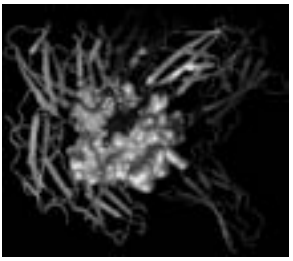
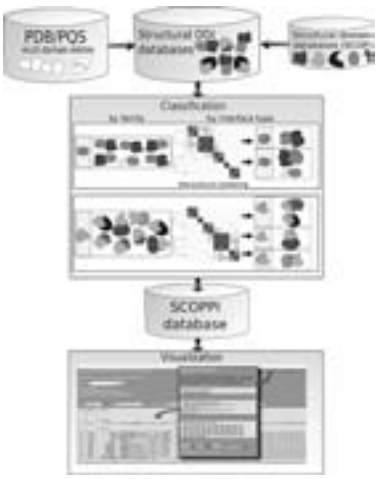


Abbildung 2: Fibronectin III (in Regenbogenfarben) bindet auf vielfältige Weise an Zytokin (zentral, grau mit entsprechend gefärbten faces).

2.2 Algorithmendesign und Definitionen

Ein *face* ist eine Menge von Interfaceresiduen (Aminosäuren) einer Domäne, die in direktem Kontakt mit einer anderen Domäne steht. Ein *interface* besteht demzufolge aus zwei in Kontakt stehenden faces. Jedes face wird einzeln klassifiziert und einem Interface wird dann die beiden jeweiligen Klassen zugewiesen.

Das Workflow-Diagramm für die Interface-Klassifizierung ist in Abbildung 3a) dargestellt. Die Daten-Akquirierung von PQS, einer PDB-Weiterentwicklung, die nicht-biologische Interfaces ausschließt, erbrachte 70.000 Interfaces. Der komplette Algorithmus ist in Abbildung 3b als Pseudocode wiedergegeben. Um die Ähnlichkeit zwischen zwei *faces* nach dem Strukturalignment zu messen, wurden zwei geometrische Maße eingeführt, die Überlappung der Bindungsstellen (*face overlap*), und der Winkel zwischen Bindungsstellen (*face angle*), siehe Zeile 10 und 11. Letzterer mißt den Winkel zwischen den Zentroiden der *faces* und dem Zentroid der gemeinsamen, alignierten Domäne. Bei dem Vergleich aller Interaktionen miteinander entstanden für jedes Ähnlichkeitsmaß die entsprechende Distanzmatrix. Hierarchisches Clustering wurde dann systematisch angewendet mit verschiedenen Schwellwerten und single, average und complete linkage (Zeilen 14-19). Ein manuell generiertes Benchmark-Set wurde dann benutzt, um zu entscheiden, welche Schwellwerte und welches Clustering (linkage) am besten ist. Die Komplexität des Algorithmuses ist quadratisch mit Bezug auf die Anzahl der Interaktionen in jeder Familie. Der Algorithmus wird durch sequenzbasierte konservative Redundanzbereinigung beschleunigt, was die Rechenzeit von ca. 3000 auf 32 CPU-Stunden reduziert.



a)

```
00:Assign SCOP families to all multi-domain PQS
   entries
01:For each SCOP family f in PQS:
02:  For each domain instance of f:
03:    identify interacting domains
04:    Store domain-domain interaction in the
       format (domain1/family1 - domain2/family2)
05:  IL_f := List of all interactions regarding f
06:  RIL_f := Representatives of IL_f after seq.
     clustering
07:  For each ddi1 in RIL_f:
08:    For each ddi2 in RIL_f:
09:      structural alignment of ddi1 and ddi2
         on the two domains belonging to f
10:      calc. angle-based sim. of ddi1 and ddi2
11:      calc. atom-overlap sim. of ddi1 and ddi2
12:      generate entry in distance-matrix
13:  save distance matrix
14:  perform hier. clustering on distance matrix
15:  produce interaction-clusters wrt some cutoff
16:  For each cluster:
17:    For each interaction ddi in cluster:
18:      For each interaction represented by ddi:
19:        Assign face type to f-domain
```

b)

Abbildung 3: Überblick Interface-Klassifikationsalgorithmus: a) Workflow-Diagramm b) Pseudocode.

2.3 Datenvisualisierung und Web-Service

Die resultierende relationale Datenbank SCOPPI (Structural Classification Of Protein-Protein Interactions) dient als Backend für ein vielfältiges webbasiertes Abfragesystem, welches die Analyse aller Interface-Klassen erlaubt (Abbildung 3a). Man kann mit der Superfamilie, der Familie, dem PDB-Eintrag, Schlüsselwörtern oder Gene-Ontology-Begriffen [ABB+00] SCOPPI-Anfragen starten. Anzeigeoptionen sind Aminosäuresequenz mit hervorgehobenen interagierenden Residuen, Konservierungsniveaus und Interface-Abbildungen. Die Interaktionen können nach Größe und Redundanzniveau gefiltert werden.

2.4 Resultate

Alle gefundenen Domäneninterfaces in Strukturen wurden in 6000 verschiedene Klassen eingeteilt. Fast 40% der Familienpaare können sich auf verschiedene Weise binden. Die Wachstumsrate für Interface-Klassen - eine Verdopplung innerhalb von ca. 4 Jahren - konnte mithilfe von Datensätzen für jedes Jahr bestimmt werden.

Die Klassifikationsgenauigkeit wurde mithilfe von 416 manuell klassifizierten Interfaces ermittelt. Im Vergleich zu einer vorangegangen Arbeit [KI05] zeigt die *face overlap*-Methode eine um fast 10% verbesserte Trefferquote (recall) bei gleicher Genauigkeit (Abbildung 4). Die hybride Methode (mit vorgeschaltetem sequenzbasiertem Clustering, Zeile 06 in Abbildung 3b) verbesserte Kim&Isons Ansatz in Trefferquote und Genauigkeit um

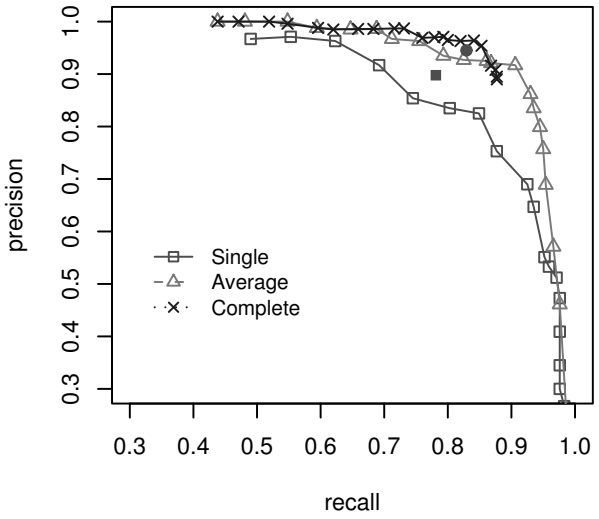


Abbildung 4: ROC-Diagramm der Interface-Klassifikation mit hierarchischem Clustering und verschiedenen Linkage-Methoden und Schwellwerten, basierend auf dem *face overlap*-Ähnlichkeitsmaß. Die Trefferquote und die Genauigkeit (recall/precision) einer rein sequenzbasierten Methode [KI05] und der hier vorgestellten hybriden Methode sind als Vergleich mit einem roten Viereck bzw. einem roten Kreis dargestellt.

jeweils 5%. Eine Reihe von Schlussfolgerungen können aus der Interface-Klassifikation gezogen werden: Hub-Proteine verwenden viele verschiedene Interfaces. Allgemein gilt, die Anzahl der *faces* und der Interaktionspartner korreliert. Weiterhin kann die scheinbar schlechte Interface-Konservierung von Hub-Proteinen wie der Ran-Familie der Diversität an Partnern und Interaktionen zugeschrieben werden. Die Sammlung aller bekannter *faces* hilft bei der Analyse der Konservierung der nichtbindenden Proteinoberfläche im Vergleich zu den Interfaces. Die Resultate sind online verfügbar unter `www.scoppi.org`.

3 Konvergent evolvierte Motive

3.1 Einleitung und Problemstellung

Ein zum oben dargelegten Ansatz verwandtes Problem ist, multifunktionale Interfaces zu finden, die in der Lage sind, verschiedene (aber analoge) Partner zu binden, was auf konvergente Evolution hinweist. Fälle von konvergenter Evolution wurden bereits einzeln dokumentiert, ein systematischer Ansatz wurde noch nicht unternommen.

Problemstellung 2: Konvergent evolvierte Motive
Ziel ist es, *faces* zu entdecken, die in nicht-homologen Proteinen existieren und an die gleiche oder an die entsprechende Stelle in homologen Proteinen binden. Alle Fälle, die mithilfe der PDB-Strukturen gefunden werden können, sollen identifiziert werden. Ähnlichkeiten zwischen den nicht-homologen Proteinen (vor allem bei viralen Proteinen) sollen dann auf Aminosäure- und Atomniveau entdeckt werden.

Neu bei dem hier gezeigten Ansatz ist, dass Proteinanalogien indirekt entdeckt werden. Da Ähnlichkeiten in nicht-homologen Proteinen schwer zu entdecken sind, basiert unser Ansatz auf dem Alignieren des gemeinsamen Bindungspartners. Die Inspiration dazu liefert ein bekanntes Beispiel für konvergente Evolution: Subtillisin und Trypsin, zwei nicht-homologe Proteine, haben die gleiche katalytische Triade, die aber mit herkömmlichen Alignment-Methoden nicht auffindbar ist. Für beide Proteine existieren jedoch Strukturen mit dem gleichen Inhibitor, der für das Alignment genutzt werden kann.

3.2 Algorithmusbeschreibung

Die PDB wird nach Paaren von Interaktionen $A - B$ und $A' - C$ durchsucht, ("ABAC" genannt), wobei B und C von verschiedenen Superfamilien, aber A und A' von der gleichen Familie stammen. Wenn B und C an die gleichen Stellen in A bzw A' binden, dann werden B und C als analoge Proteine mit konvergent evolvierten Motiven markiert. Um Äquivalenzen in ihnen zu entdecken, werden alle Interfaceresiduen paarweise miteinander verglichen und der *Motif Match Score*, MMS berechnet.

Der Algorithmus als Pseudocode:

```
 1 Compute all domain-domain interaction pairs A-B
 2 For all SCOP families Do:
 3  Sequence Alignment of all family members
 4 For all pairs A-B and A'-C Do
 5  If MSA with A,A' has ISO overlap > 30% Then
 6   Structurally align A, A'
 7   Compute centers of mass of interfaces
 8   Compute interface angle IA
 9   Compute interface atom overlap IAO
10   Compute motif match score MMS
11   Add ISO, IA, IAO, MMS to database
12 Sort database by MMS
```

3.3 Resultate

3.3.1 Statistische Betrachtungen

58.000 ABAC-Fälle wurden entdeckt, davon 3000 ohne Immunoglobulinbeteiligung. 4.2% aller nicht-homologen Proteine mit gemeinsamem Partner binden diesen an der gleichen Stelle. 15% aller Familien (270) sind in diesen Interaktionen involviert.

4 Interfacemotive für die Vorhersage von Proteininteraktionen auf Sequenzebene

4.1 Einleitung und Problemstellung

Die Vorhersage von Protein-Protein-Interaktionen ist ein zentrales Problem in der Bioinformatik und wurde auf verschiedene Weise Diese Arbeit beschreibt zum ersten Mal, wie eine Interface-Klassifikation benutzt wird, um sequenzielle Interfacemotive automatisch zu erstellen. Diese Motive können dann mit Sequenz-Suchmethoden benutzt werden, um interagierende Regionen in Sequenzen zu identifizieren und um Homology-Modeling der Interface-Regionen zu initialisieren.

Problemstellung 3: Proteininteraktionsvorhersage
Mithilfe der Interface-Klassifizierung aus Problemstellung 1 sollen Sequenzmotive für konservierte Interface-Regionen konstruiert werden, um Interaktionen vorhersagen zu können. Eine Repräsentation soll gefunden werden, die die Zerstreuung der Inteface-Teile auf Sequenzebene berücksichtigt. Weiterhin soll die Repräsentation einen wohldefinierten Wahrscheinlichkeitswert für einen Treffer in einer Sequenzdatenbank liefern, um die statistische Signifikanz dieses Treffers zu eruieren. Die Motivdeskriptoren sollen einem geeigneten Validierungsprozess unterzogen werden.

Der Hauptgrund für solch eine Analyse ist, dass viele Proteine mit Sequenzen bezüglich ihrer Funktion noch sehr unvollständig beschrieben (annotiert) sind. Funktionale Charakterisierung eines Proteins wird aber oft durch die Identifikation von Interaktionspartnern verbessert. Dafür werden alle äquivalenten sequenziellen Segmente einer Interface-Klasse, die den strukturellen Merkmalen eines Interfaces entsprechen, in ein einzelnes Hidden Markov Model (HMM) kompiliert.

4.2 Konstruktion der Protein-Protein-Interaktionsdeskriptoren

Eine Bibliothek von Deskriptoren für 740 Protein-Protein-Interaktionen und 3000 Protein-Liganden-Interaktionen wurde erstellt. Der Pseudocode für den entsprechenden Algorithmus ist:

```
1. For all interface types:
  2. Extract all non-redundant instances of an interface type from SCOPPI
  3. Obtain a multiple sequence alignment (MSA) and identify interacting residues
  4. Identify segments in the MSA predominantly comprising interacting residues
  5. Include residues that are between selected sequence neighbors
  6. Optionally add sequences identified as structural homologous by HSSP
  7. Generate HMMs for each segment
  8. Combine HMMs into one descriptor connected by insert states that reflect the
     linker regions between segments
  9. Add descriptors (one for each face) to library
10. The library can now be used to predict faces and interfaces.
```

HSSP ist eine Datenbank die zu einer Strukur "strukturell homologe" Proteinsequenzen assoziiert. Für PPI-Deskriptoren wurde SCOPPI benutzt, während für Liganden, wo keine entsprechende Klassifikation von Bindungspartnern wie SCOPPI existiert, PLI-Deskriptoren wurden aus einem einzelnen Strukturbeispiel und HSSP konstruiert. Um Interface-Regionen in einem Sequenzalignment aller Instanzen einer Interface-Klasse zu identifizieren, wurde ein Konsens von Interface-Spalten berechnet. Aus jeder zusammenhängenden Menge von Interface-Spalten wurde ein HMM generiert. Diese werden dann zu einem zusammhängenden HMM gefügt, welches dem allgemeinen probabilistischen System von HMMs genügt. Weiterhin wird die Länge der Zwischenregionen berücksichtigt. Somit wird ein Wahrscheinlichkeitswert für Treffer von multiplen Segmenten berechnet. Deskriptoren wurden auf mehrere Weise validiert: mit Cross-Validation, mit Trefferquotenanalyse in zufälligen Datenbanken und mit Vergleichen zu etablierten Interaktionsdatenbanken. Für ATP-Bindungsdeskriptoren wurde exemplarisch ein Benchmark erstellt und mit entsprechenden PROSITE-Motiven [Bai92] verglichen, wobei unsere HMM-Deskriptoren eine deutlich höhere Genauigkeit erreichten.

4.3 Resultate

Der neue Aspekt der hier präsentierten Vorhersagemethode ist, dass sie auf einer Interaktionsklassifikation aufbaut. Abgesehen von der besonderen Datenquelle (SCOPPI) unterscheidet sich diese Arbeit auch von anderen in der Kombination von automatisch generierten, von Strukturen abgeleiteten, Deskriptoren, die sich für die Sequenzsuche eignen und mehrteilige Motive repräsentieren können. Die Deskriptoren wurden benutzt, um 230 Interaktionen, die aus der Literatur bekannt sind, zu validieren. Der hinzugefügte Beitrag liegt dabei in der Identifikation der Interface-Residuen.

5 Fazit

Die Lösungen zu drei zueinander in Bezug stehenden Problemen wurden präsentiert:

1. Die Datenbank SCOPPI, eine strukturbasierte Klassifikation von Domän-Domän-Interaktionen wurde erstellt. Der hybride Clustering-Algorithmus dafür ist schnell und effizient. Die Datenbank wurde mithilfe eines manuell erstellten Benchmark-Sets validiert, was 83% Genauigkeit und 95% Trefferquote ergab, was die sequenz-

basierte Methode von Kim und Ison in beiden Fällen um jeweils 5% übertrifft. Ein Webservice steht der Allgemeinheit unter `http://www.scoppi.org` zur Verfügung.

2. Ein Suchalgorithmus für konvergent evolvierte Motive in nicht-homologen Proteinen wurde erstellt. Er basiert auf der Idee, dass ein Interface in einer Familie, dass an nicht-homologe Proteine binden kann, auf Konvergenzen in diesen Proteinen aufmerksam machen kann. Eine relationale Datenbank mit allen aus der PDB ableitbaren Fällen wurde erstellt.

3. Eine neue Methode zur Vorhersage von Protein-Proteininteraktionen wurde entwickelt. Sie erstellt eine Sequenzrepräsentation (Hidden Markov Models) für alle Instanzen einer Interface-Klasse in SCOPPI. Die Repräsentation berücksichtigt multiple Segmente, ein Merkmal, das zuvor noch nicht in strukturbasierten Vorhersagemethoden implementiert wurde.

Eine ganze Reihe von biologisch relevanten Resultaten demonstriert den Nutzen der aufgezeigten Lösungen.

1. Die SCOPPI Datenbank erlaubt den gesamtheitlichen Einblick in das Interaktionsverhalten einer Proteinfamilie. Für Hub-Proteine im besonderen ließ sich klären, dass sie vor allem durch distinkte *faces* an verschieden Partner binden. Alternative strukturelle Domänassoziationen wurden für 40% der Familien-Familien-Interaktionen festgestellt. SCOPPI bildet die Grundlage für eine akkuratere Berechnung der Konservierung der Interfaces im Vergleich zur restlichen Oberfläche [CSH+04].

2. Die Datenbank für konvergent evolvierte Motive gibt Aufschluss darüber, wie oft konvergente Evolution vorkommt: 4.2% aller nicht-homologen Proteine mit gemeinsamem Bindungsparter binden diesen an der gleichen Stelle. Fälle von viraler Mimikry wurden festgestellt.

3. PPI-Deskriptoren wurden benutzt, um 230 aus der Literatur bekannte Interaktionen zu validieren. Der hinzugefügte Beitrag liegt dabei in der Identifikation der Interface-Residuen. Proteininteraktionen für bisher nicht charakterisierte Proteine wurden vorgeschlagen. Neue Ligandenbindungsstellen können vorhergesagt werden, wie im Falle von Adenosin-Triphosphat gezeigt wurde.

Die Resultate dieser Arbeit wurden in [HKS05, HWKS07, WHKS06, KHWS06] veröffentlicht. Der Code für die Implementierung (ausschließlich der statistischen Analysen) umfasst 6700 Zeilen Python-Code.

Literatur

[ABB+00] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver,

A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin und G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–9, May 2000.

[ABC+04] P. Aloy, B. Böttcher, H. Ceulemans, C. Leutwein, C. Mellwig, S. Fischer, A. Gavin, P. Bork, G. Superti-Furga, L. Serrano und R. B. Russell. Structure-based assembly of protein complexes in yeast. *Science*, 303(5666), 2004.

[Alo07] P. Aloy. Shaping the future of interactome networks. *Genome Biol*, 8(10):316, Nov 2007.

[Bai92] A. Bairoch. PROSITE: a dictionary of sites and patterns in proteins. *Nucleic Acids Res*, 20 Suppl:2013–2018, May 1992.

[CSH+04] D. R. Caffrey, S. Somaroo, J. D. Hughes, J. Mintseris und E. S. Huang. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci*, 13(1):190–202, Jan 2004.

[HKS05] A. Henschel, W. K. Kim und M. Schroeder. Equivalent binding sites reveal convergently evolved interaction motifs. *Bioinformatics*, 2005.

[HWKS07] A. Henschel, C. Winter, W. K. Kim und M. Schroeder. Using structural motif descriptors for sequence-based binding site prediction. *BMC Bioinformatics*, 8 Suppl 4:S5, 2007.

[KHWS06] W. K. Kim, A. Henschel, C. Winter und M. Schroeder. The many faces of protein-protein interactions: A compendium of interface geometry. *PLoS Comput Biol*, 2(9):e124, Sep 2006.

[KI05] W. K. Kim und J. C. Ison. A survey of geometric association of domain-domain interfaces. *Proteins*, 2005. accepted.

[NT03] I. M. A. Nooren und J. M. Thornton. Diversity of protein-protein interactions. *EMBO Journal*, 22(14):3486–92, 2003.

[WHKS06] C. Winter, A. Henschel, W. K. Kim und M. Schroeder. SCOPPI: a structural classification of protein-protein interfaces. *Nucleic Acids Res*, 34:D310–D314, 2006.

Andreas Henschel studierte von 1995-1998 Informatik (Datalogi) an der Århus Universitet, Dänemark, 1998-1999 an der Middlesex University London und 1999-2002 an der TU Dresden, wo er den internationalen Masterstudiengang Computational Logic mit der Diplomarbeit zum Thema “Logikbasierte Roboter” mit der Note 1,1 abschloss. Seit 2003 beschäftigt sich Andreas Henschel mit Bioinformatik, zuerst im Max-Planck-Institut für Zellbiologie, Dresden, und anschliessend (ab 2004) bei Prof. Michael Schroeder am BioInnovationszentrum/TU Dresden im Rahmen seiner Promotion. Diese schloss er 2008 zum vorliegenden Thema mit “summa cum laude” ab. Seit April arbeitet Andreas als Postdoc an Textmining und Technologievorhersagemethoden mit Bezug auf erneuerbare Energien am Masdar Institute of Science and Technology, Abu Dhabi, Vereinigte Arabische Emirate.