

10 Zusammenfassung

Diese Arbeit setzt sich aus sieben Studien zusammen, die sich mit der Identifizierung von Proteindomänen oder -motiven in Proteinsequenzen befassen. Ziel der ersten fünf Studien ist es, durch detaillierte Proteinsequenzanalysen neue Proteindomänen zu entdecken (a-e). Zwei weitere Manuskripte haben Anwendungen der genomweiten Identifizierung von Proteindomänen zum Thema (f-g).

(a) Die „Domain in Apoptosis and Interferon Response“ (DAPIN) wurde als gemeinsames Proteinmodul von Proteinen identifiziert, die in Krankheitsprozessen von Vertebraten auffällig geworden sind, wie dem Pyrin Protein (hereditäres familiäres mediterranes Fieber), dem ASC Protein (Apoptose, Brustkrebs), einigen Interferon-induzierbaren Proteinen (Entzündung und Virusantwort), oder dem AIM2 Protein (Melanome). Aufgrund der Kombination der DAPIN mit bekannten Domänen aus Apoptoseproteinen in einigen Proteinen und den Ergebnissen der Strukturvorhersage folgerte ich, dass die DAPIN-Familie eine vierte Subfamilie von Adapterdomänen der Death-Domain-Superfamilie darstellt.

(b) Die Spin/Ssty-Proteinfamilie spielt eine Rolle in der Ausbildung des Spindelapparats während der Gametogenese von Vertebraten. Im Zuge dieser Arbeit wurden vier menschliche Proteine der Spin/Ssty-Familie beschrieben, sowie drei der Maus, zwei des Huhns, eins des Rinds und eins des Medaka-Fischs. Alle Spin/Ssty-Proteine bestehen aus einer sich dreifach wiederholenden Einheit, dem Spin/Ssty-Repeat, der wahrscheinlich eine β -Faltblattstruktur bildet. Die phylogenetische Analyse der Repeats und die Analyse der Genstrukturen ergaben, dass die repetitive Architektur von Spin/Ssty-Proteinen durch zwei aufeinanderfolgende Duplikationen von Exons in einem gemeinsamen Vorfahren der heutigen Vertebraten entstanden sein muss.

(c) Das humane MCSP Protein und das orthologe NG2 Protein der Ratte spielen sowohl in der Wundheilung als auch in der Entwicklung von Tumoren eine Rolle. Die Entdeckung einer neuen repetitiven Proteindomäne, dem CSPG-Repeat, ermöglichte eine Feineinteilung der Domänenstruktur der NG2/MCSP-Proteine. Der zentrale flexible Teil der NG2-Ektodomäne besteht aus 15 Kopien des CSPG-Repeats. Jede bildet höchstwahrscheinlich eine β -Faltblattstruktur aus. Der CSPG-Repeat ist entfernt verwandt mit dem Cadherin-Repeat. Eine Kopie des CSPG-Repeats in einem cyanobakteriellen Protein entstammt wahrscheinlich einem horizontalem Gentransfer von einem marinen Vielzeller zu einem Cyanobakterium.

(d) Mutationen im humanen Gen LGI1 führen zu einer veränderten Expression des C-Terminus des LGI1 Proteins. Ich entdeckte eine repetitive Sequenzeinheit, den

EPTP-Repeat, im C-Terminus von LGI1. Der EPTP-Repeat ist das gemeinsame Sequenzmodul der Proteine LGI1, LGI2, LGI3 und LGI4, des G-Protein gekoppelten Rezeptors VLGR1 und des TNEP1 Proteins. In einem Mausmodell für Epilepsie ist das VLGR1 Gen mutiert. Auch das humane Gen LGI4 liegt einer chromosomalen Region, die mit einem Epilepsie-Syndrom assoziiert wird. Es ist somit wahrscheinlich, dass der EPTP-Repeat eine essentielle Funktion in der Aufrechterhaltung der Gehirnfunktion hat.

(e) Die Suche nach den molekularen Ursachen der Histidin-Phosphorylierung in Eukaryonten war die Triebfeder zur Sequenzanalyse der Proteine HIG und NtrY. Durch verschiedene Methoden der Sequenzanalyse konnte gezeigt werden, dass die sensorische Region von bakteriellen Histidinkinase-Rezeptoren der NtrY-Familie und der membran-proximale Bereich von eukaryotischen HIG-ähnlichen Proteinen eine signifikante Sequenzähnlichkeit besitzen. Da die Sequenzanalyse auch Regionen von Transmembranhelices umfasst, sollte die Homologie der beiden Proteinfamilien zusätzlich auf experimenteller Ebene belegt werden. HIG-Proteine wären die ersten Proteine von Metazoen, die homolog zu Histidinkinase-Rezeptoren von Bakterien sind.

(f) Immunorezeptor Tyrosin-basierte inhibitorische Motive (ITIMs) haben eine wichtige Funktion in der Kontrolle der Aktivierung von Immunzellen. Die von mir verwendete Strategie zur Suche nach ITIMs in großen Sequenzdatenbanken nutzt den Sequenzkontext, also Informationen über eventuelle Signalpeptide, Transmembranhelices oder Domänen in einem Protein, um die hohe Rate von falsch-positiven ITIM-Vorhersagen herkömmlicher Verfahren der Proteinmotivsuche zu reduzieren. So konnten 109 humane ITIM-Rezeptoren identifiziert werden. Von diesen wurden 36 bereits in der Literatur als ITIM-Rezeptoren beschrieben. Nur zwei bekannte Typ-I ITIM-Rezeptoren wurden nicht gefunden. Eine Analyse öffentlicher Datenquellen über die Gewebeexpression humaner Gene ergab, dass die Expression von ITIM-Rezeptoren nicht auf Blutzellen beschränkt ist.

(g) In 271 Proteinen, die kürzlich in einer massenspektrometrischen Studie des Nukleolus entdeckt wurden, konnten im Zuge dieser Arbeit 115 bekannte und 91 neue Proteindomänen identifiziert werden. Die phylogenetischen Profile dieser Domänen in Archaeobakterien, Eubakterien und Eukaryonten deuten auf einen archaeobakteriellen Ursprung derjenigen Proteine des Nukleolus hin, die urtümliche Funktionen in der Ribosomenreifung besitzen, zeigen aber deutlich seinen insgesamt chimären Ursprung aus eubakteriellen, archaeobakteriellen und eukaryotischen Proteindomänen. Die Ergebnisse sprechen für eine langsame, kontinuierliche Entwicklung des Nukleolus in den ersten Eukaryonten und damit gegen die umstrittene Hypothese eines endosymbiotischen Ursprungs des Nukleus.