

## Zusammenfassung

Im Zentrum der vorliegenden Arbeit steht die Strukturaufklärung von Domänen aus humanen Proteinen mittels NMR-Spektroskopie im Kontext von „*structural and functional genomics*“. Ein Weg zur strukturellen Aufklärung von neuen Proteindomänen ist es, multiple Domänenkonstrukte auf der Basis von bioinformatischen Vorhersagen zu erstellen, um nach Expression im analytischen/präparativen Maßstab die korrekten Domänengrenzen und die Struktur zu bestimmen. In diesem Rahmen wurden die verschiedenen Arbeitsschritte unter Anwendung modernster Expressionsverfahren parallelisiert. Diese Methoden wurden am Beispiel der humanen RUN-Domäne etabliert. Es wurden 16 RUN-Konstrukten aus unterschiedlichen Zielproteinen hergestellt. Dies ergab ein lösliches und gefaltetes Proteindomänenkonstrukt des humanen Proteins RUFY1, welches in weiteren Untersuchungen verwendet wurde. Die RUN-Domäne findet man gehäuft in Proteinen, die mit GTPasen der Rap- und Rab-Familie, sowie mit Tyrosinkinasen assoziiert sind, wie zum Beispiel das Effektorprotein RUFY1 aus der ETK-Tyrosinkinase. Weiter wurde in Zusammenarbeit mit der Proteinstrukturfabrik (PSF) Berlin die bis dahin uncharakterisierte, humane p47 SEP-Domäne mittels NMR-Spektroskopie strukturell bestimmt. p47 ist ein aus den drei Domänen UBA, SEP und UBX modular aufgebautes Adaptorprotein der AAA-ATPase p97. Am Beispiel der SEP-Domäne wurden neue Aminosäure-selektive NMR-Experimente getestet und in ein modifiziertes Konzept der automatisierten Resonanzzuordnung implementiert. Anschließend wurde die NMR-Struktur bestimmt und als „*novel fold*“ anhand einer DALI-Datenbanksuche identifiziert. Darüberhinaus wurde für die humane p47 SEP-Domäne erstmalig eine strukturbasierte Hypothese zur Funktion anhand eines Topologievergleiches mit Cystatininhibitoren erarbeitet und mittels enzymkinetischer und NMR-spektroskopischer Methoden bestätigt. Dabei zeigte sich, dass die humane p47 SEP-Domäne ein reversibler kompetitiver Inhibitor der humanen Cysteinprotease Cathepsin L ist. Die Inhibitionskonstante  $K_i$  für Cathepsin L beträgt  $1,5 \pm 0,2 \mu\text{M}$  und wurde mit fluorimetrischen Methoden ermittelt. An der Wechselwirkung mit Cathepsin L beteiligte Reste der Domäne konnten mittels NMR-Titration bestimmt werden. Sie befinden sich in der aufgrund des Topologievergleiches mit Cystatininhibitoren postulierten Interaktionsfläche. Bislang wurde keine Funktion für SEP-Domänen vorgeschlagen. Die hier erarbeitete Hypothese stellt einen Anfang dar und erweitert erheblich den bisher bekannten biologischen Kontext des p47 Proteins um die Möglichkeit der Wechselwirkung mit Cysteinproteasen. Sie macht weitere Experimente zur genauen Einordnung dieser neuen Interaktion denkbar und plausibel.

Das dritte Kapitel der vorliegenden Arbeit behandelt die Bestimmung der NMR-Strukturen des Komplexes der humanen Gas7 (*Growth-arrest-specific-protein 7*) WW-Domäne mit dem prolinreichen Liganden LIPPPPPL. Das *hGas7* Protein enthält eine WW-Domäne und ist ein Fusionspartner des Proteins MLL (*myeloid leukemia linkeage*), das bei der Krebserkrankung *acute myeloid leukemia* eine Rolle spielt. WW-Domänen werden aufgrund ihrer unterschiedlichen Bindungsspezifität in 4-6 Klassen eingeteilt. Die *hGas7* WW-Domäne mit dem Bindungsmotiv –IPPPPx- ist ein Vertreter der Klasse II. Im Rahmen dieser Arbeit wurde die Bindungsspezifität der *hGas7* WW-Domäne erstmalig strukturell charakterisiert. Damit wurde die Abgrenzung zu Vertretern der Klassen I und IV sowie innerhalb der Klasse II zu dem zwischen zeitlich charakterisiertem Vertreter FBP11 strukturell unterlegt. Die hier erarbeiteten strukturellen Grundlagen zur Sequenz- und Bindungsspezifität der *hGas7* WW-Domäne und dem prolinreichen Liganden dient als Ausgangspunkt für die weitere Entwicklung von peptidischen Inhibitoren.

## Summary

The main topic of this thesis is the structure determination of domains from human proteins by NMR-spectroscopy in the context of structural and functional genomics. One way to determine the structure of new protein domains is to generate multiple domain constructs derived from bioinformatic predictions and subsequent expression on an analytical/preparative scale. After analysing the domain constructs by biophysical methods the fold and the correct domain boundaries can be determined. In this context, all necessary working steps were carried out in parallel to produce protein domain constructs with modern expression methods. This is shown and established by generating 16 RUN domain constructs from different mammalian target proteins. One of these constructs from the human protein RUFY1 was soluble and characterized as a folded protein domain construct. The RUN domain is frequently found in proteins associated with GTPases of the Rap and Rab family and with tyrosine-kinases. Among these, RUFY1 is an effector protein of the tyrosine-kinase ETK. Moreover, in collaboration with the Protein Structure Factory (PSF) in Berlin the NMR solution structure of the uncharacterized human p47 SEP domain was determined. The protein p47 is a modular multidomain adaptorprotein of the AAA-ATPase p97 comprising an UBA, SEP and UBX domain. The resonance assignment of the SEP domain was used as a case study to test new amino acid selective NMR-experiments and to implement them into a modernised concept for the automated resonance assignment process. The NMR structure of the human p47 SEP domain was determined and identified by a DALI-database search as a novel fold. Furthermore in this thesis a structure-based functional hypothesis was formulated for the human p47 SEP domain to inhibit cysteine proteases based on a topological comparison with cystatin inhibitors. The hypothesis was confirmed by enzymekinetik fluorescence measurements and NMR-titration experiments. These results clarified that the human p47 SEP domain is a reversible competitive inhibitor of the human cysteine protease cathepsin L with a inhibition constant  $K_i$  of  $1.5 \pm 0.2 \mu\text{M}$ . The SEP domain interaction surface was investigated by NMR-titration experiments and specific residues were identified in line with the suggested hypothesis. Taken together, the hypothesized thesis confirmed by the obtained results considerably expand the biological role of the p47 protein interacting with cysteine proteases. This allows new ideas to be explored and experiments to be feasible.

The third chapter presented in this thesis deals with the complex structure determination of the human *Gas7* (*growth-arrest-specific-protein 7*) WW domain with its prolinerich ligand LIPPPPPL. The *hGas7* protein comprises one WW domain and is seen to play an important role in the disease of acute myeloid leukaemia via interacting with the oncoprotein MLL

(*myeloid lineage linkage*). WW domains are classified in 4-6 binding classes according to their specific molecular binding recognition. The *hGas7* WW domain is a representative of class II and recognizes a specific poly-proline-binding motif. This work presents for the first time the binding specificity of the *hGas7* WW domain was determined. This result thus enabled us to distinguish the latter class II from class I and IV representatives and within the class II representative FBP11 whose structure was solved recently in our group. The described structural clarification between the sequence and binding specificity of the WW domain *hGas7* and its prolinerich ligand LIPPPPPL serves as a starting point for the further development of specific peptidic inhibitors.