

## 6. Zusammenfassung

Der Malaria verursachende Organismus *Plasmodium falciparum* (*P. falciparum*) besitzt in seinem Kerngenom für die Mitochondrien bestimmte Proteine, die als Transportsignal ein mitochondriales Transitpeptid enthalten. Durch die kürzlich erfolgte Sequenzierung des Genoms von *P. falciparum* ist es wünschenswert, Vorhersagealgorithmen für verschiedene Proteinlokalisationen zur Verfügung zu haben. Für andere Organismen etablierte Programme zur Vorhersage von mitochondrialen Transitpeptiden, MitoProtII und TargetP, lieferten bei Anwendung auf Sequenzen aus *P. falciparum* nur unbefriedigende Ergebnisse. MitoProtII erzielte in einer 20-fachen Kreuzvalidierung einen Matthews-Koeffizienten von  $cc = 0,49$ , TargetP erzielte in diesem Fall einen Matthews-Koeffizienten von  $cc = 0,60$ . TargetP erzielte für die Sequenzen aus *P. falciparum* nur eine Selektivität von 47%, MitoProtII nur eine Sensitivität von 35%. Dieser Ergebnisse haben die Entwicklung eines speziell auf *P. falciparum* trainierten Vorhersagemodells wünschenswert gemacht.

Kerncodierte mitochondriale Precursorproteine aus *P. falciparum* wurden mit statistischen Methoden, Hauptkomponentenanalyse, selbstorganisierenden Karten und überwachten neuronalen Netzen analysiert und mit solchen aus anderen Organismen verglichen. Zwei Repräsentationen der Datensätze wurden gewählt, Aminosäurehäufigkeiten und 19 physikochemische Eigenschaften. Ein grundsätzlich unterschiedlicher Aminosäuregebrauch konnte festgestellt werden. Glycin, Alanin, Prolin und Arginin werden in *P. falciparum* mit weniger als 60% der Häufigkeit in der Swiss-Prot-Datenbank, Version 36, verwendet. Isoleucin, Tyrosin, Asparagin und Lysin werden hingegen mit mehr als 150% der Häufigkeit in der Referenzdatenbank verwendet. Diese Häufigkeitsmuster wurden, mit Variationen, auch in allen Targetingsequenzen beobachtet.

In der Datenanalyse mittels Hauptkomponentenanalyse und selbstorganisierenden Karten ließen sich cytoplasmatische Proteine in beiden Repräsentationen klar von der Gruppe mitochondrialer, extrazellulärer und apicoplastischer Proteine trennen. Die Trennung innerhalb der zweiten Gruppe war weniger deutlich.

Ein neuronales Netz (PlasMit) zur Vorhersage mitochondrialer Transitpeptide in *P. falciparum* wurde entwickelt. Basierend auf der relativen Aminosäurehäufigkeitsverteilung innerhalb der ersten 24 N-terminalen Aminosäuren lieferte es einen Matthews-Korrelationskoeffizienten von 0,74 (86% korrekt vorhergesagte Sequenzen) in einer 20fachen Kreuzvalidierung. Dieses Netz sagte 2449 (24%) der 10276 vorhergesagten Open Reading

Frames aus dem Genom von *P. falciparum* als mögliche mitochondrial lokalisierte Proteine voraus. Ein Netz mit identischer Topologie wurde auf eine geringere Anzahl falsch-positiver Vorhersagen trainiert und erzielte einen Mathews-Koeffizienten von 0,51 (84% korrekte Vorhersagen) in einer 10fachen Kreuzvalidierung. Dieses Netz sagte 903 (8,8%) potentielle mitochondriale Precursorproteine unter den 10276 vorhergesagten Open Reading Frames voraus.

Sämtliche Trainingsdatensätze, die Open Reading Frames des Genoms von *P. falciparum*, sowie das Netz, das den höchsten Mathews-Koeffizienten erzielt hat, sind per Web unter <http://www.modlab.de>, Menüpunkt PlasMit, erreichbar.