ETH zürich

ETH Library

Development of a novel protein structure prediction force field

Doctoral Thesis

Author(s):
Ulrich, Patrick Roland

Publication date:
1996

Permanent link:
https://doi.org/https://doi.org/10.3929/ethz-a-001698339

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Diss. ETH Nr. 11884

Development of a novel protein structure prediction force field

Dissertation

for the degree of
Doctor of Natural Sciences
submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZÜRICH

ETH-BIBLIOTHEK ZÜRICH

presented by
PATRICK ROLAND ULRICH
Dipl.Chem.ETH
born 26.Dezember 1966
citizen of St.Antoni FR

accepted on the recommendation of
Prof. Dr. W.F. van Gunsteren, examiner
Prof. Dr. E. Pretsch, co-examiner

Zürich 1996

Kurzfassung

Die hier vorgelegte Arbeit befasst sich mit der Ausarbeitung vereinfachter Kraftfelder zur Simulation und Untersuchungen des Verhaltens von Proteinmolekülen.

Proteine sind grösstenteils ausserordentlich komplexe Systeme, obschon sie aus nur 20 Grundbausteinen, den Aminosäuren, und einigen wenigen anderen Atomen (z.B. Ionen) oder Molekülen (prosthetische Gruppen) aufgebaut sind. Somit kann mit den heute verfügbaren Methoden und einem sogenannten atomaren Kraftfeld, mit dem die Kräfte zwischen mehreren tausend Atomen berechnet werden müssen, nur über sehr kurze Zeiten (< 100 ns) simuliert werden. Dies im Gegensatz zu Faltungsprozessen, die in einem Zeitbereich von Millisekunden bis Minuten stattfinden.

Mit dem Entwurf eines vereinfachten Kraftfeldes erhofft man sich einerseits Rückschlüsse auf übergeordnete, d.h. nicht auf lokal atomaren Details beruhenden Mechanismen, die die weitgehend noch unverstandene Proteinfaltung beeinflussen, andererseits einen Schritt zu machen in Richtung Vorhersage der tertiären aus der primären Struktur, der Aminosäuresequenz.

Das vorliegende Kraftfeld ist halb statistisch, halb empirisch. Es beruht auf der Annahme einfacher van der Waals-artiger Wechselwirkungen zwischen den einzelnen Aminosäuren, die als Massenpunkte, lokalisiert in der C^{α}-Position, dargestellt werden. Die notwendigen Parameter dieser Paar-Wechselwirkungen, nämlich die Tiefe (ε-Parameter) und der Ort (σ-Parameter) des lokalen Minimums, wurden auf Grund statistischer Berechnungen aus einem Satz Proteinen bekannter Struktur geschätzt.

Den Kern dieser Arbeit bildet die sogenannte Parametrisierung des Kraftfeldes. Dabei wird eine Funktion beschrieben, die die Energiedifferenz zwischen der Kristallstruktur und inkorrekt gefalteten Alternativ-Strukturen aufsummiert. Da diese Funktion nach obigen Parametern abgeleitet werden kann, besteht die Aufgabe darin, diese Funktion zu optimieren, was durch eine Molekül-Dynamik Simulation im Parameterraum bei niedriger Temperatur geschieht.

Da die Zahl der Parameter viel kleiner ist als die Zahl der Alternativ-Strukturen und da bei dieser Optimierung das globale Minimum nicht erreicht wird, bleibt ein gewisser Prozentsatz an inkorrekt gefalteten Strukturen, die eine niedrigere Energie aufweisen als der Wildtyp. Für einen Satz Proteine ausserhalb des für die Parametrisierung verwendeten Satzes wurde das Kraftfeld auf seine Tauglichkeit getestet. Ausserdem wurde untersucht, wie gut die bekannten Tests von Eisenberg und Novotny sich nachvollziehen lassen.

Im zweiten Teil befasst sich die Arbeit mit der Anwendung des berechneten Kraftfeldes auf Suchmethoden. Ein ungelöstes Problem ist die Frage nach dem globalen Minimum einer Hyperfläche mit mehreren hundert Variabeln, wie sie auch nur schon ein mittelgrosses Protein in diesem einfachen Kraftfeld darstellt. Mehrere bekannte einfache Methoden wurden angewendet und getestet.

Der dritte Teil schliesslich ist Anwendungen des Kraftfelds gewidmet. Dabei werden Proteine verschiedener Sets auf ihre Eigenschaft geprüft, gegenüber falsch gefalteter Alternativ-Strukturen besser abzuschneiden. Im weiteren werden vereinzelte Fälle aus der Literatur übernommen und auf ähnliche Vorhersagekraft getestet.

Summary

The studies performed within the framework of this thesis aim at the development of simplified force fields for simulation and examination of protein molecules.

Proteins are extremely complex molecular systems, although their framework is limited to only 20 units, the amino acids, and a few other atoms (metal ions) or molecules (prosthetic groups). This is the reason why today's computer power does not allow simulation times of more than 100 nanoseconds with a so-called „all atom force field", where thousands of interactions between atoms have to be calculated. In contrast, protein folding processes have a time scale of milliseconds up to minutes.

The aim of a simplified force field is to understand global mechanisms, i.e. not depending on local atomic arrangement, that influence protein folding on the one hand, and on the other hand form a basis to prediction of tertiary structure from primary structure, that is, the amino acid sequence.

The force field presented is empirical and statistical as well. It is based on the assumption of simple interactions of van der Waals type between the amino acid sites, represented as a mass point in the C^{α}-position. The parameters of these interactions, the depth (ε-parameter) and the location (σ-parameter) of the local minimum, were initially estimated from a set of proteins of known structure.

The core of this work is the parameterization of the force field. To this end, a function is implemented, that sums the potential energy difference between the crystal structure and incorrectly folded alternative structures. This function can be derived with respect to the parameters. This function is optimized using a molecular dynamics simulation in parameter space at low temperature.

Since the number of parameters is much lower than the number of alternative structures and since the global minimum of this function remains unreached, a small percentage of incorrectly folded structures have a lower energy than the wild type. A set of proteins not included in the calibration set was used to test the force field. Finally, the well-known tests of Eisenberg and Novotny were performed for the force field.

The second part concerns the use of the force field in combination with searching methods. The question of finding the global minimum of a high-dimensional function with several hundred variables, as for a medium-sized protein in a simplified force field, is still unsolved. Several known simple search methods were applied and tested.

The third part is dedicated to applications of the force field. In this, protein of different sets are tested to score better than wrongly folded alternative structures in terms of potential energy. Furthermore, some work found in the literature is repeated applying the force field on the considered proteins.