Investigations into the Secondary Structure of Human Fibronectin FN1 with the Replica Exchange Algorithm TIGER2h

***M. Kulke, N. Geist, W. Langel**, Institut für Biochemie, Universität Greifswald*

In Short

- Determination of the complete structure of the extracellular protein fibronectin
- Step 1 and 2: Replica exchange simulation of several secondary segments
- Step 3: Constraint simulations for correcting the folding of several modules
- Step 4: Molecular dynamics equillibration for validation of the complete structure

Fibronectin FN1 (P02751) is a glycoprotein which acts as a mediator for different biomolecules in the extracellular matrix (ECM)[1]. The protein plays an important role in the adhesion complexes of cells and connects the transmembrane signal protein integrin with other ECM molecules like collagen. We aim to build a model of the cell adhesion to titanium dioxide and this connector protein is of crucial importance. In the frame of this work, we want to establish a stable atomistic structure of the whole molecule from numerous elements known so far.

Natural fibronectin is a dimer, and its two subunits are linked by disulfide bonds. The monomers contain subunits of 31 modules of type fibronectin I-III (Figure 2). The subunits form a chain, which consists of 2354 amino acids with a flexible structure. These subunits provide various interaction sites for fibrin, collagen, heparin, integrin and self-interactions. Several experiments[2–9] report structures of fragments of fibronectin, but no study on the full molecule is known to us. For some parts of the sequence, no 3D structures are available. We built up the molecule by merging the structural information for fragments from these studies and applied homology modeling on the SWISS Server[10] to the missing parts. Molecular dynamics simulations of the resulting complete fibronectin molecule have shown that the homology model of some modules containing mostly sheet structures is inaccurate and results in a partially unstable conformation. We want to cure the defects by molecular dynamics simulations of selected segments. The standard method for protein folding is replica exchange molecular dynamics (REMD). Previously we have shown that folding of peptides with 12-14 residues in explicit solvent with REMD is at the limit of a simulation project at HLRN (to be submitted). Simulations in implicit solvent were much faster but did not reproduce the folding with reasonable precision. We introduce a newly developed very efficient replica exchange method TIGER2h (to be submitted) that affords a smaller number of replicas and calculates the potential energies for the Metropolis sampling criterion in implicit water. With this algorithm the folding of a peptide with 41 amino acids was simulated at HLRN.

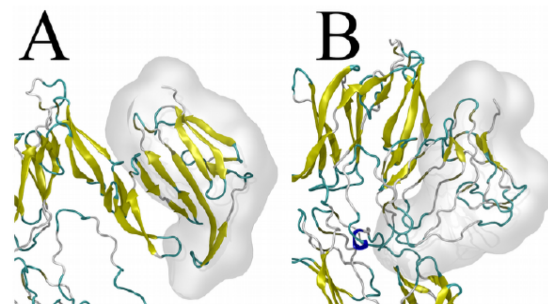

***Figure 1:** Illustration of the interconnecting disulfide region I_{12} (white shaded) A before and B after equilibration. The secondary structure elements are indicated with yellow sheet, blue 3-10-helix, cyan turn and white random coil.*

Already after short equilibrations of fibronectin in vacuum, the secondary structure of module I_{12} diverges from the homology model (Figure 1). Other sheet structures are stable in vacuum and it is unlikely that this instability is only due to the omission of the solvent. Therefore, at this stage we search for a stable folded structure of the amino acid sequence around module I_{12}, which is unknown so far, but is presumably similar to other modules of type I. Extended to this barrel is a sequence of unknown structure that includes the interchain disulfide bonds. We thus split the sequence of 94 amino acids with poorly known secondary structure into two segments with 44 and 50 amino acids respectively, and proceed in two steps: First the secondary structure of the barrel I_{12} containing 44 amino acids (2293 to 2336) is calculated. In a second step the sequence 2337 to 2386 is folded with constrained I_{12} in the simulation cell. The TIGER2h method shall be applied, and we want to demonstrate the usefulness of this new algorithm for these larger peptides. The result is further validated by checking the condition that the cysteine residues 2367 and 2371 should be easily accessible to interchain connections in the final structure.

In stage 2 and 3 the secondary structure of several other misfolded modules (I_4-I_8, I_{10}, III_2, III_3, III_5,

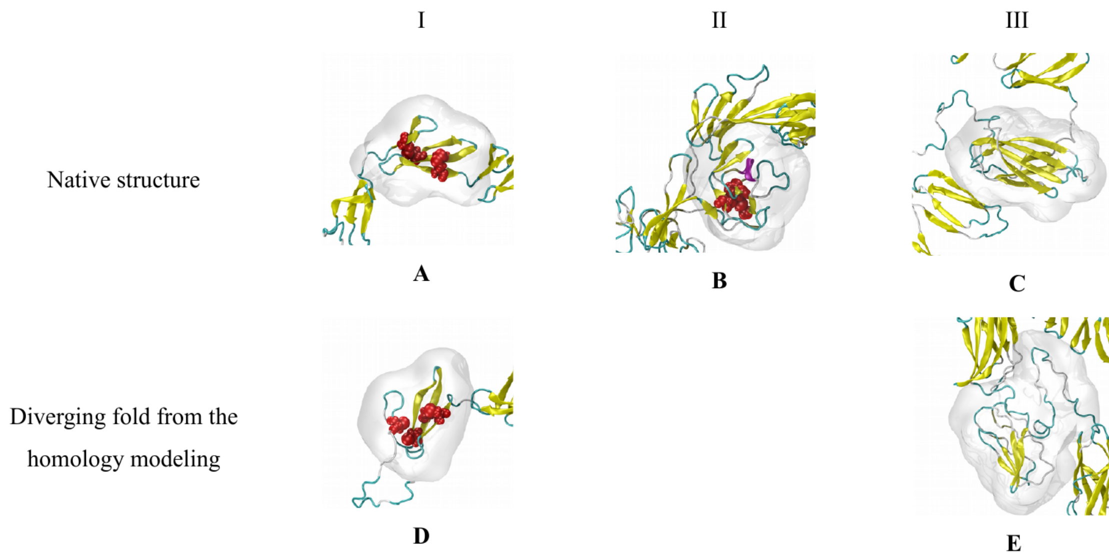

Figure 2: *Each of the structure motifs (type I-III), which form the fibronectin chain, consists of 40-90 amino acids and have a distinct β-sheet barrel conformation. The type I module has five stacked β-sheets with an enclosed, highly conserved region containing two disulfide bonds[2]. In type II modules, X-like oriented sheets are linked by two disulfide bonds. In contrast the type III modules have seven antiparallel sheets without disulfide bridges. The secondary structure elements are indicated as yellow sheet, purple helix, cyan turn and white random coil. Cysteines in bridges are red. The native conformations of modules I-III have structures A-C, respectively. D and E result from the homology model and contain conformations with mostly the same secondary structures elements as A and C, but strongly diverging tertiary structures. The type II module B only occurs two times in the fibronectin sequence, which both come from X-ray studies[3] and are correctly folded.*

III_7, III_8) shall be established with a combination of secondary structure constraints (rubber bands) and replica exchange.

In the final stage 4, the results of stage 1-3 are integrated in the existing model of fibronectin and simulations on the nanosecond scale of the full fibronectin mono- and dimer have to be performed for verifying that a stable secondary and tertiary structure was attained.

WWW

https://biochemie.uni-greifswald.de/institut-und-struktur/forschung/forschung-in-den-arbeitskreisen/ordner-aks-lehrstuehle/biophysikalische-chemie/forschung/

More Information

[1] Mao, Y.; Schwarzbauer, J. E., *Matrix Biology* **2005**, *24* (6) 389-399.

[2] Potts, J. R.; Bright, J. R.; Bolton, D.; Pickford, A. R.; Campbell, I. D., *Biochemistry* **1999**, *38* (26), 8304-8312.

[3] Graille, M.; Pagano, M.; Rose, T.; Ravaux, M. R.; van Tilbeurgh, H., *Structure* **2010**, *18* (6), 710-718.

[4] Rudino-Pinera, E.; Ravelli, R. B. G.; Sheldrick, G. M.; Nanao, M. H.; Korostelev, V. V.; Werner, J. M.; Schwarz-Linek, U.; Potts, J. R.; Garman, E. F., *J. Mol. Biol.* **2007**, *368* (3), 833-844.

[5] Erat, M. C.; Slatter, D. a; Lowe, E. D.; Millard, C. J.; Farndale, R. W.; Campbell, I. D.; Vakonakis, I., *Proc. Natl. Acad. Sci. USA* **2009**, *106* (11), 4195-4200.

[6] Vakonakis, I.; Staunton, D.; Rooney, L. M.; Campbell, I. D., *EMBO J.* **2007**, *26* (10), 2575-2583.

[7] Leahy, D. J.; Aukhil, I.; Erickson, H. P., *Cell* **1996**, *84* (1), 155-164.

[8] Niimi, T.; Osawa, M.; Yamaji, N.; Yasunaga, K.; Sakashita, H.; Mase, T.; Tanaka, a; Fujita, S., *J. Biomol. NMR* **2001**, *21* (3), 281-284.

[9] Sharma, A.; Askari, J. A.; Humphries, M. J.; Jones, E. Y.; Stuart, D. I., *EMBO J.* **1999**, *18* (6), 1468-1479.

[10] Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T., *Bioinformatics* **2006**, *22* (2), 195-201.