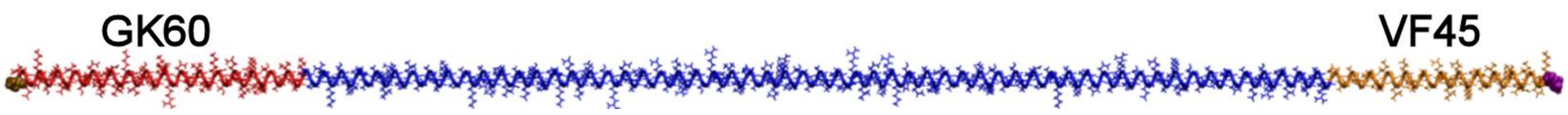
Computational Study of Honeybee Silk Protein Terminal Domain Secondary Structure and Dynamics

Andrew T. Church, Philip C. Church and Jeffrey S. Church

Introduction

Recombinant versions of one of the honeybee silk proteins, Apis mellifera Fibrion 3, can be fabricated into a variety of material forms[1,2] making it of interest for the design of functionally active protein-based materials. Unlike spider silks which typically supports a β-sheet secondary structure, honeybee silk is primarily observed as having a tetrameric coiled-coil structure.[3] The implications of this difference is significant, leading to questions regarding how the terminal domains (GK60 and VF45) must differ in structure and function, to facilitate aggregation of these proteins into silk fibers. This work presents a bioinformatic and molecular dynamic study of the folding of theses regions in order to better understand how honey bee silk tetramers self-assemble. Our findings include the first detailed atomic level tertiary structure predictions in honeybee silk terminal domains, as well as a detailed comparison of bioinformatic and molecular dynamic prediction capabilities for these terminal peptides.



Method

MD simulations[4-6]

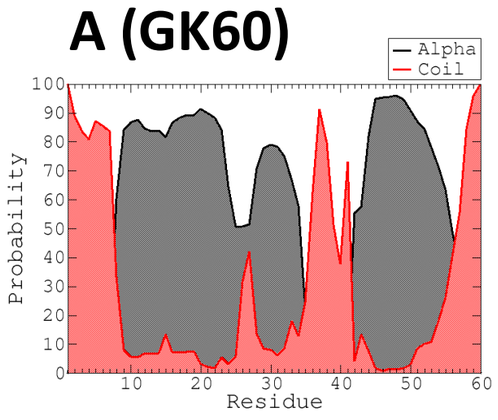
- MD Code - Gromacs 2020
- Force Field - CHARMM22*
- Water Model - tips3p
- Clustering – 0.4 nm

Bioinformatics[7-10]

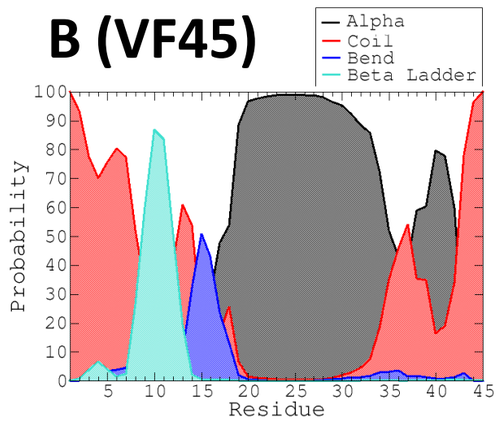
- Raptor X
- Porter 5
- Paircoil 2
- Marcoil – MTK matrix

Results

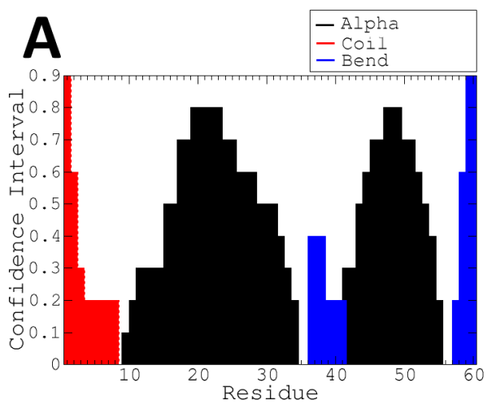
A (GK60)

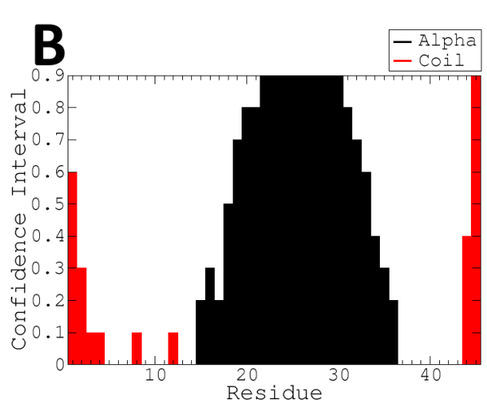


B (VF45)

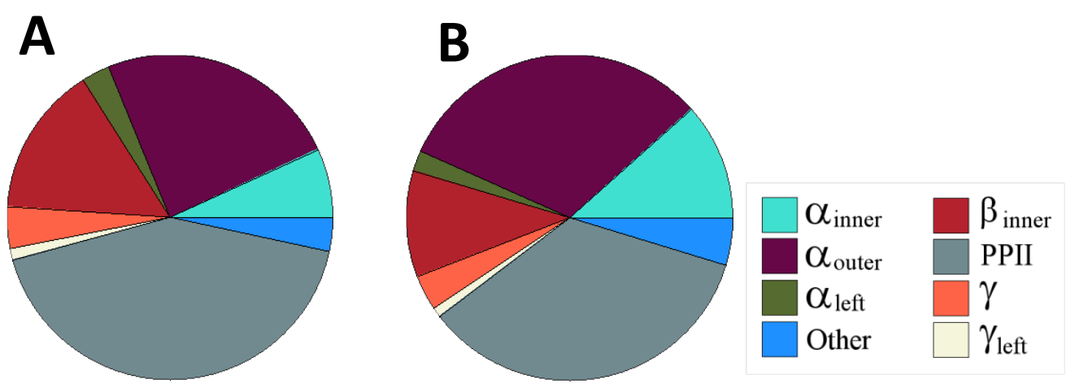


Raptor X – conditional neural network





Porter 5 - bidirectional recurrent neural network

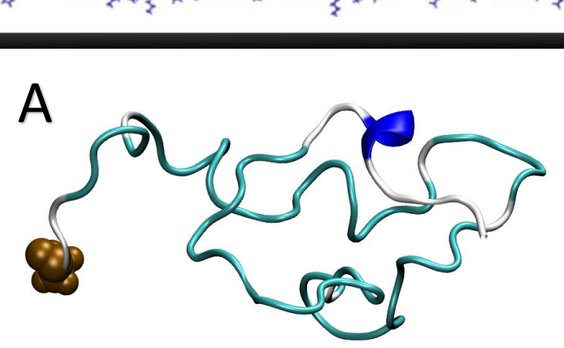


- α-helix – 24.3%
- β-sheet – 15.0%
- PPII – 42.3%

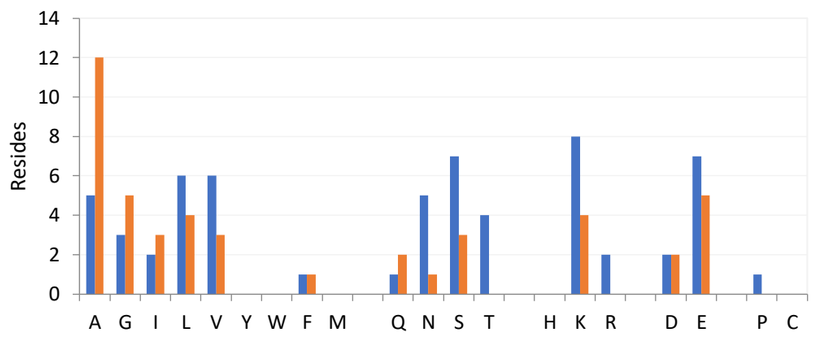
- α-helix –31.7%
- β-sheet -10.5%
- PPII – 34.9%

Molecular Dynamics - Ramachandran

- MD VF45 – α-helix and PPII/unstructured ~even
- MD GK60 – dominated by PPII/unstructured
- Bioinformatics - similar secondary structure prediction
- Agreement in α-helix quantity in VF45
- Disagreement in α-helix quantity in GK60



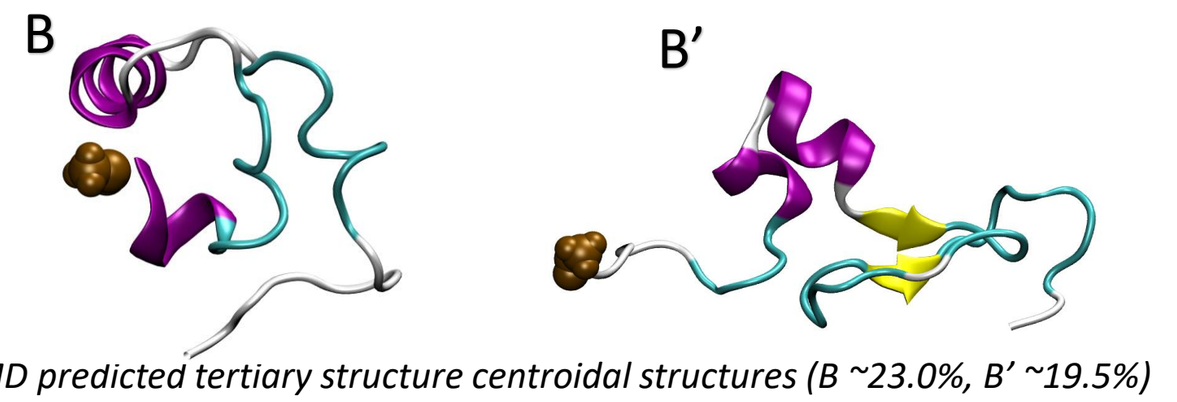
MD (left) predicted (73.6%), Bioinformatics (right) predicted tertiary structure (0.086p)



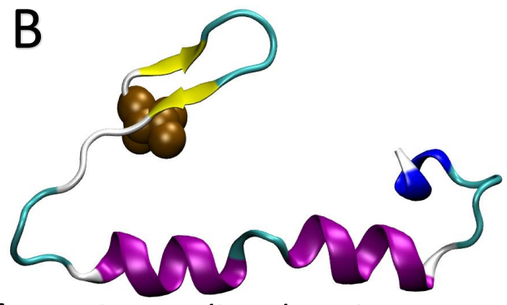
Amino acid content

GK60 — 43% MALEK, 8% A

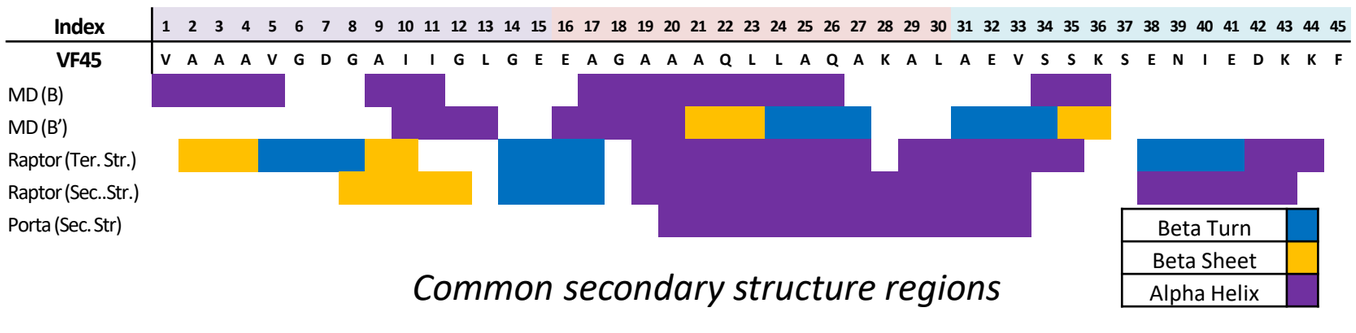
VF45 — 55% MALEK, 26% A



MD predicted tertiary structure centroidal structures (B ~23.0%, B' ~19.5%)



Bioinformatics predicted tertiary structure (0.061p)



Common secondary structure regions

- MD GK60 – Globular, lacking secondary structure
- MD VF45 – Rapid folding, transient helices
- Paircoil : Coiled-coil - 0% within confidence range across sequence, both terminals
- Marcoil : Coiled-coil - ~0.3% GK60, ~24.7% VF45

Conclusions

- ❑ VF45 good correlation between MD and bioinformatics; secondary and tertiary
- ❑ GK60 significant disagreement,
  - ❑ α-helix promoting AA content comparably low
  - ❑ Non-neighboring intra-protein interactions are significant contributors

www.JPAscientific.com

Contact : ATC@jpascientific.com

JPA Scientific

Software Development & Computational Modeling

References

1. Poole J, Church JS, Woodhead AL, et al. Macromol Biosci. 2013;13(10):1321-1326.
2. Rapson TD, Church JS, Trueman HE, Dacres H, Sutherland TD, Trowell SC. Biosens Bioelectron. 2014;62:214-220.
3. Woodhead AL, Church AT, Rapson TD, Trueman HE, Church JS, Sutherland TD. Polymers. 2018;10(7):776.
4. Hess B, Kutzner C, Van Der Spoel D, Lindahl E. J Chem Theory Comput. 2008; 4(3):435-447.
5. MacKerell AD, Bashford D, Bellott M, et al. J Phys Chem B. 1998;102(18):3586-3616.
6. Piana S, Lindorff-Larsen K, Shaw DE. Biophys J. 2011;100(9):L47-L49.
7. Peng J, Xu J. Proteins. 2011;79 Suppl 1(Suppl 10):161-171.
8. Torrisi M, Kaleel M, Pollastri G. bioRxiv. Published online January 1, 2018:289033.
9. Delorenzi M, Speed T. Bioinformatics. 2002;18(4):617-625.
10. McDonnell AV, Jiang T, Keating AE, Berger B. Bioinformatics, vol. 22, 3, 1 2006