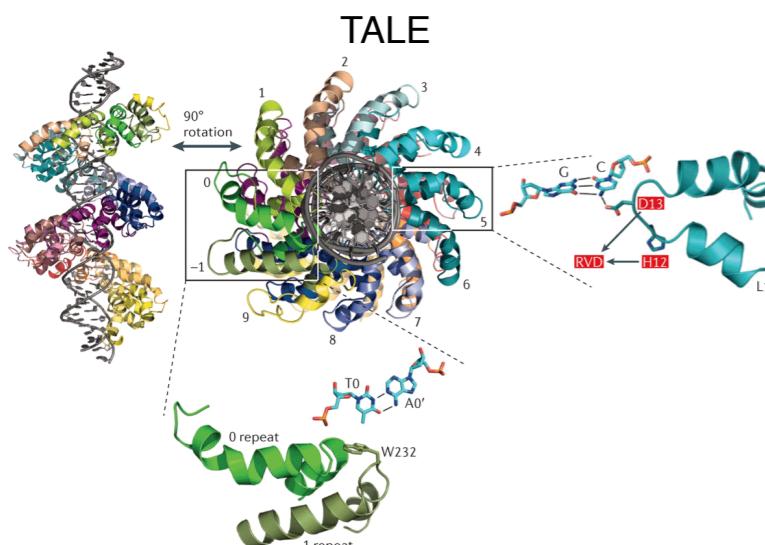
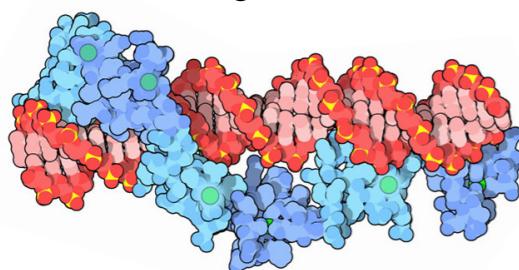
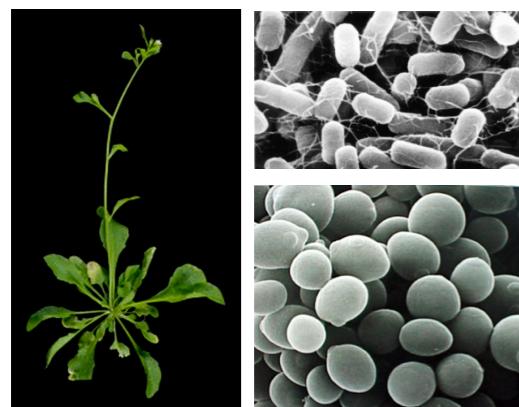


①構造化領域自動探査ソフトウェアSPADE (Search for PAttered DNA Elements) の開発 (1/4)

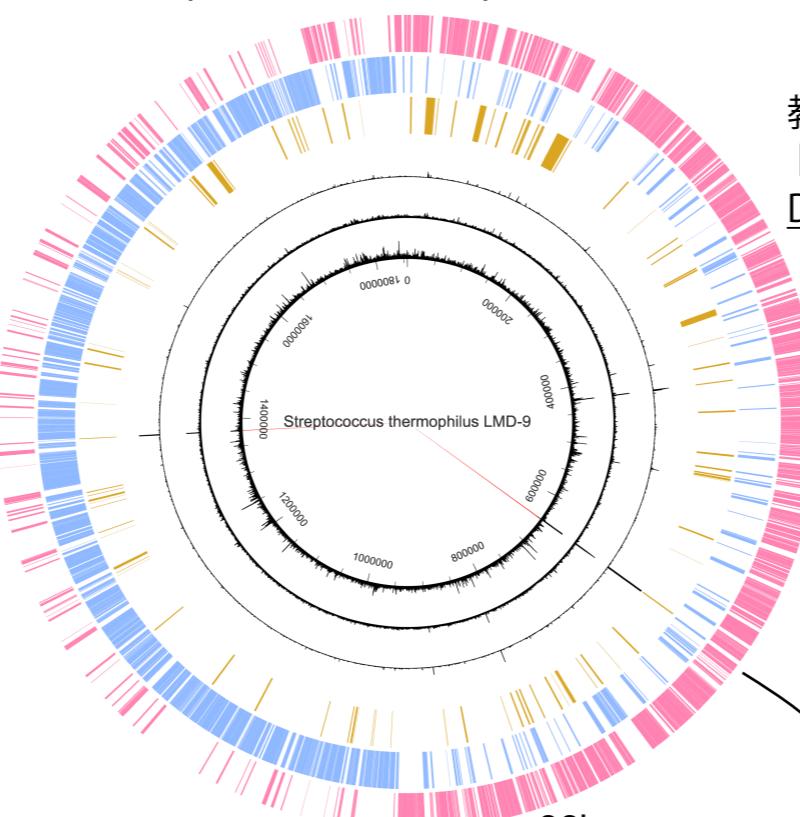
Zinc finger domain



これまでに発見・応用された全てのゲノム編集関連遺伝子は周期的リピート配列を持つ



Streptococcus thermophilus LMD-9



東京大学先端科学技術研究センター
谷内江 望 (yachie@synbiol.rcast.u-tokyo.ac.jp)

教師なしでメタゲノム情報から高速に周期的構造化リピートを発見するソフトウェアSPADE (Search for PAttered DNA Elements) の開発

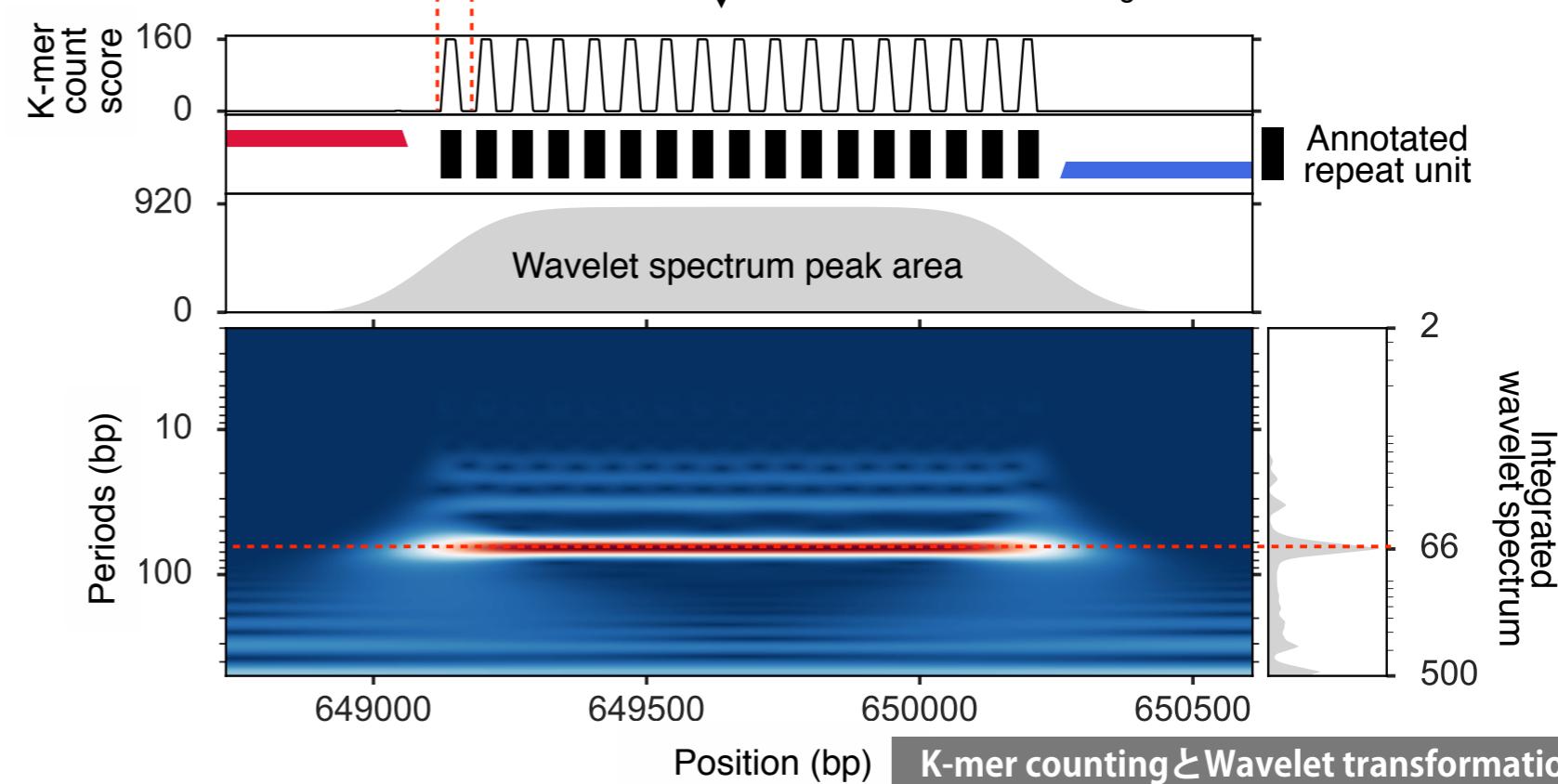
これまでにリピート配列をゲノムスケールでアノテーションするソフトウェアはある

周期的な既知モチーフリピートをタンパク質構造についてアノテーションするソフトウェアはある

周期的リピートをゲノムスケールで高速探査するソフトウェアはない

- Genes on forward strand
- Genes on reverse strand
- Annotated repeat region

1st circle: 5-mer counting score
2nd circle: 7-mer counting score
3rd circle: 10-mer counting score



Strategy for periodic repeat sequence

(1) Generation of possible k-mer sequence in the search space

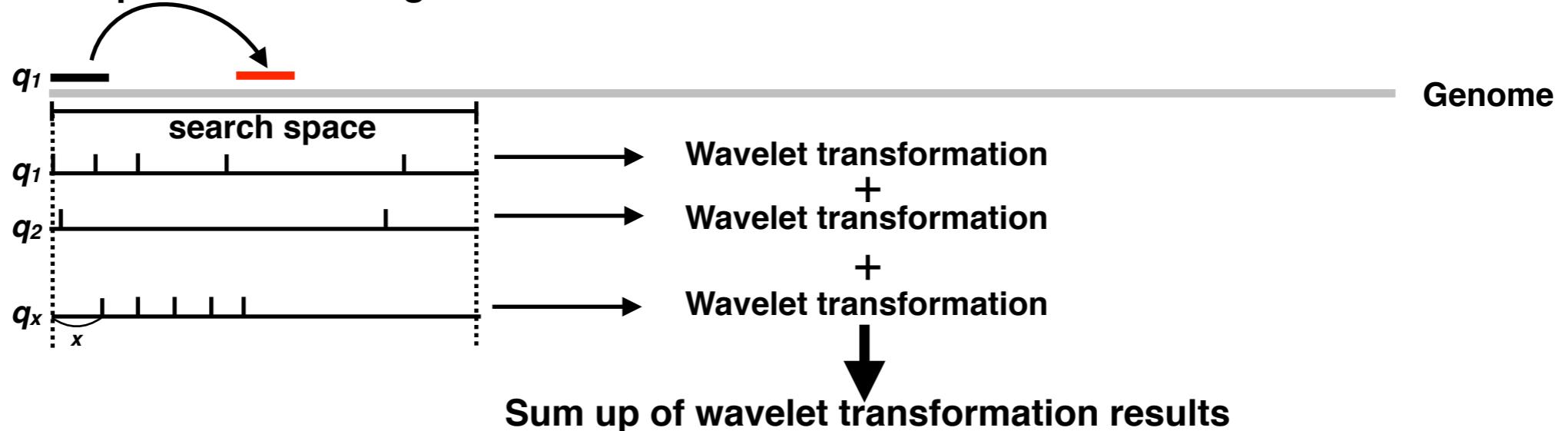


(2) Detection of similar k-mer in search space



(3) Scoring

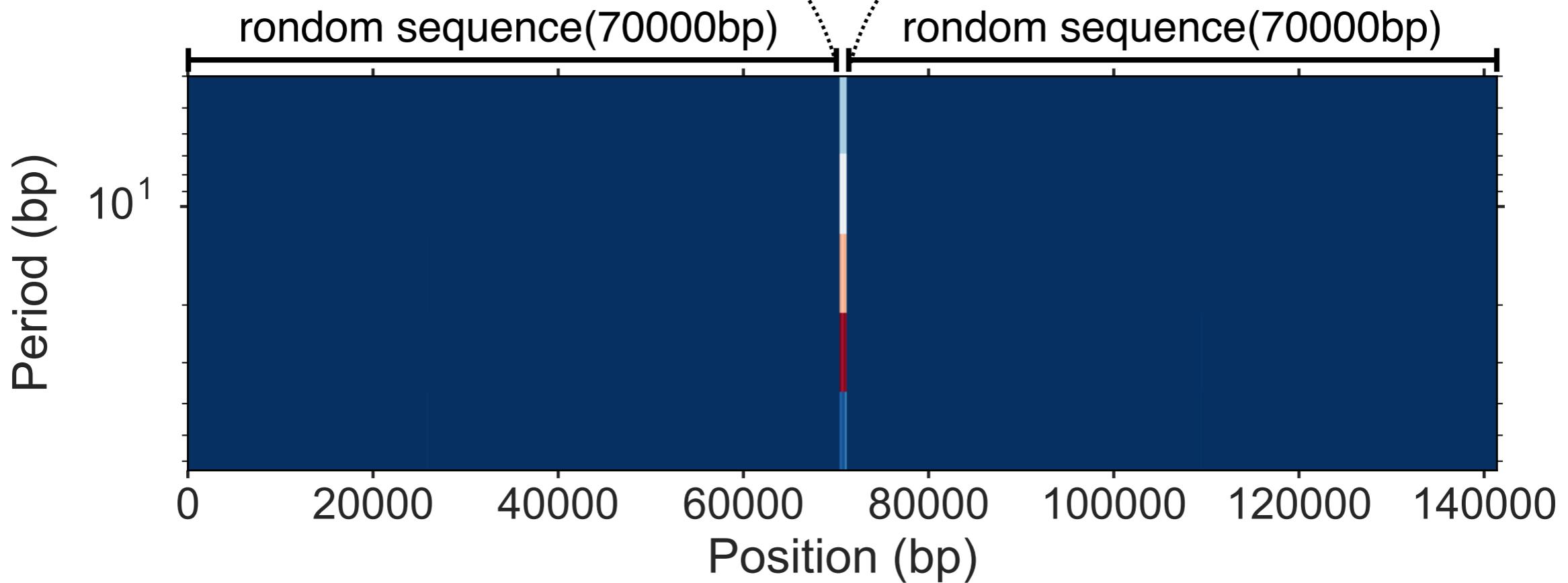
+1 at first position in hit region

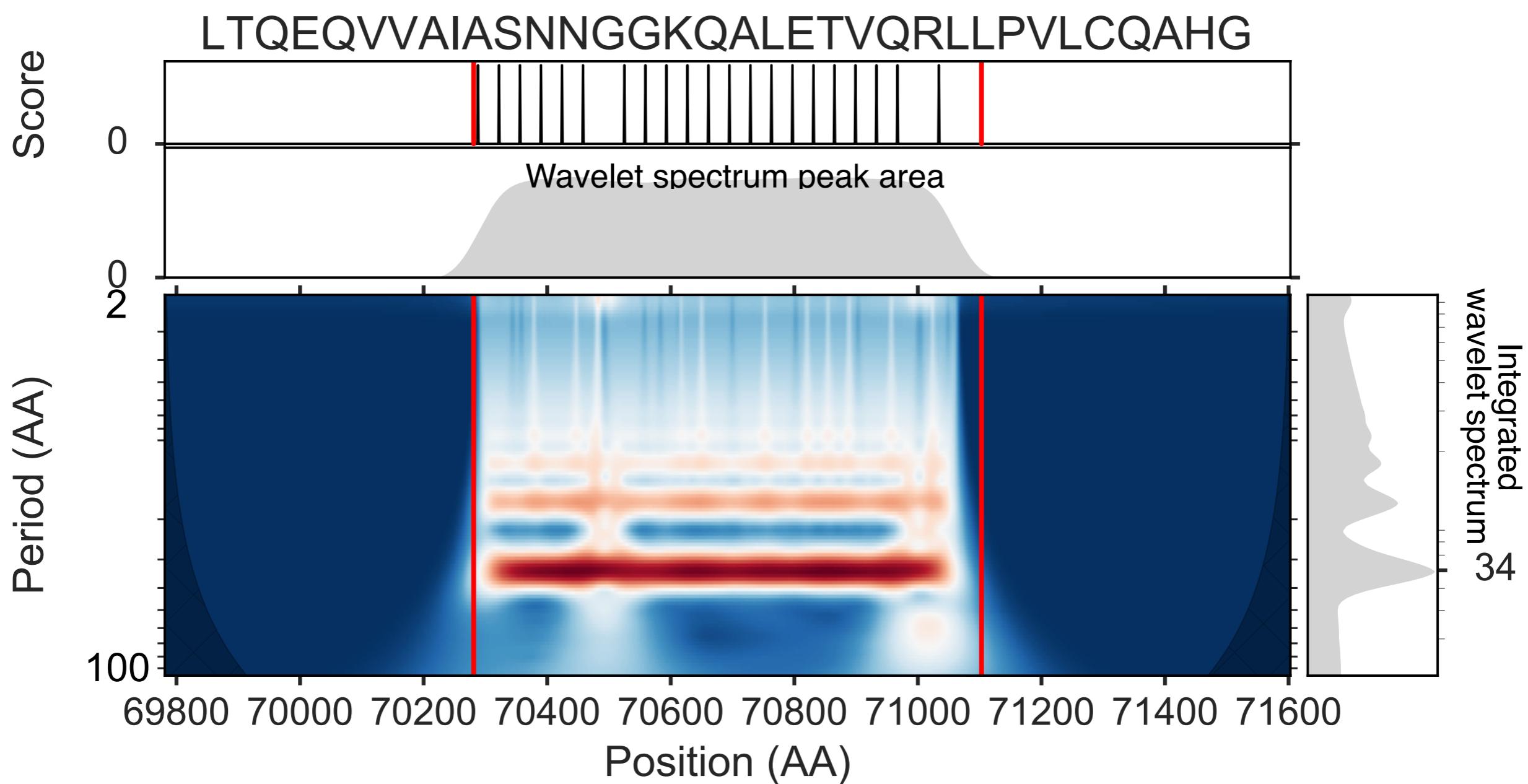


(4) Sliding search space

Test result 1

TAL effector protein PthXo1 of *Xanthomonas oryzae* pv. *oryzae* PXO99A





Test result 2

Streptococcus thermophilus LMD-9 (64000bp ~ 66000bp)

