

BIOINFORMATIKA – ZADÁNÍ

3. SAMOSTATNÉ PRÁCE

Skládání sekvencí

Podmínky

Úkoly budete odevzdávat na 13. cvičení, přičemž pro pozdní odevzdání budou platit podmínky na webu Marka Cravena ([tady](#)) - za každý den navíc se vám bude odečítat 1 bod.

Skládání sekvencí

1. Máme dány následující 3-mery:

{AGT, AAA, ACT, AAC, CTT, GTA, TTT, TAA}

Zkonstruujte graf překryvů („overlap graph“) a nalezněte hamiltonovskou cestu (obsahující 7 hran). Zapište nadsekvenci odpovídající této hamiltonovské cestě).

2. Máme zadáno následující spektrum:

$S = \{ATG, GGG, GGT, GTA, GTG, TAT, TGG\}$

Ukažte, jak lze pomocí metody založené na eulerovských tazích nalézt sekvenci s takovou, že **$Spectrum(s,3) = S$** (nepovinně: jak lze nalézt všechny takové sekvence?).

3. Vyberte si jeden z postupů uvedených výše a zautomatizujte jej (tj. napište program pro řešení úloh daného typu). V případě, že si zvolíte první možnost, tak nezapomeňte hledat optimální Hamiltonovskou cestu. K tomu můžete využít například Torsche toolbox katedry řídicí techniky (<http://rtime.felk.cvut.cz/scheduling-toolbox/>).