Proteiny jsou nezbytné pro život, protože hrají zásadní roli v mnoha biologických procesech. Navrhování nových proteinů s požadovanou funkcí je důležitým problémem ve vývoji léků a biologickém výzkumu. Velké databáze proteinových sekvencí lze použít k trénování velkých jazykových modelů převzatých ze zpracování přirozeného jazyka na řeči proteinů zapsané v abecedě aminokyselin. V této práci demonstrujeme, jak lze velké jazykové modely založené na předtrénovaných hlubokých neuronových sítích efektivně vyladit pro kontrolovatelné generování proteinových sekvencí z několika odlišných proteinových rodin. Pomocí bioinformatických metod a metod založených na hlubokém učení ukazujeme, že model je schopen generovat vysoce kvalitní proteinové sekvence, které vykazují nízkou podobnost s existujícími proteiny.