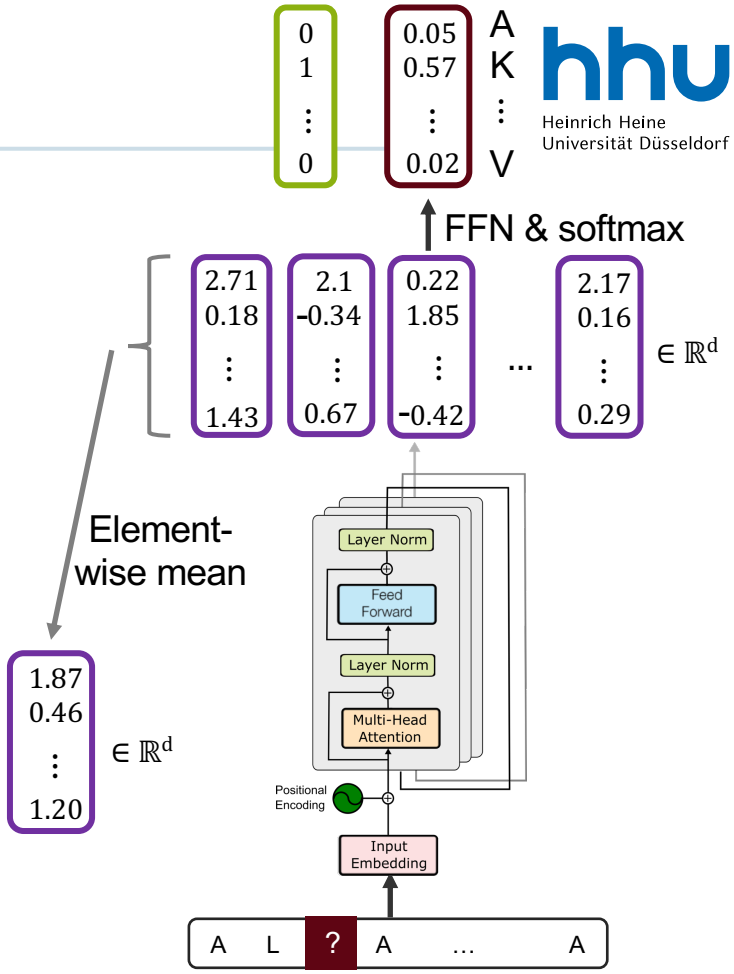
Heinrich Heine
Universität Düsseldorf

Training Transformer Networks

Supervised Fine-Tuning

Self-Supervised Training

- Advantage:
  - Meaningful protein representations without labeled training data
- Disadvantages:
  - We do not train the model for a specific task and cannot be sure that it extracts all relevant information
  - Taking the mean over many amino acid representations results in information loss
- How can we overcome the disadvantages?
  - We can train an encoder for a specific task and force it to store all task-relevant protein information in a single vector



Supervised Training

- We add a token (<cls>) at the beginning of the sentence
  - We want to use the representation of this token as a whole protein representation
- Training task: predict protein property using updated <cls>-token representation
- Training end-to-end (encoder & FFN) forces the model to store all task-relevant information in the updated <cls> representation
- After training (two options):
  - Use this model as our final prediction model
  - Extract updated <cls> representation as a task-specific protein representation and use as input for another model
- We start this training phase (fine-tuning) by using the parameters of the pre-trained Transformer Network for the mask token prediction