A hierarchical transformer, consisting of multiple layers that progressively capture more complex patterns, is the core architectural component. The lower layers focus on individual motifs and simple interactions, while higher layers handle intricate grammatical structures. We augment this with **grammar-specific attention heads**, each tailored to a particular grammar type. These heads allow the model to focus on relevant parts of the sequence depending on the grammar being processed. **Motif embeddings** are enriched with biological properties, providing the model with deeper understanding of each transcription factor's behavior. A key novelty is the **dynamic positional encoding**, which adapts to represent both absolute and relative positions of motifs. This is crucial for grammars that depend on specific spacing or ordering. We also introduce a grammar-aware loss function, guiding the model to learn the unique aspects of each grammar type.

To enhance interpretability and explicit grammar learning, we incorporate a syntax tree generation module. This is complemented by an **adversarial grammar discrimination** component, pushing the model to generate realistic enhancer sequences (this may be overkill but nice to talk about).

We model complex motif interactions through a graph neural network layer, capturing pairwise and higher-order relationships between motifs. Our training strategy employs curriculum learning, gradually introducing more complex grammars as the model improves. The architecture also features attention-guided convolutions, where filter weights are dynamically adjusted based on attention patterns. For nested structures, we implement recursive transformer blocks that can be applied iteratively.

We include auxiliary tasks such as grammar transition prediction and employ motif syntax regularization to respect biological constraints. A multi-scale sequence scanning mechanism captures both fine-grained motif details and broader grammatical patterns. Finally, we also develop a continuous latent grammar space, allowing for smooth interpolation between different enhancer structures.

Grammar-specific architectural components (which part of whole architecture do I think will learn each grammar):

1. Basic Single Motif Grammar: Lower layers of the hierarchical transformer and basic attention heads.
2. Fixed-Space Two-Motif Grammar: Dynamic positional encoding and specialized attention heads.
3. Variable-Space Two-Motif Grammar: Flexible attention mechanisms and motif interaction graph.
4. Or-Logic Grammar: Multi-head attention with competing heads for different motifs.
5. Repetitive Motif Grammar: Cyclic attention patterns and recursive transformer blocks.
6. Palindromic Motif Grammar: Specialized palindromic attention heads and symmetry-aware positional encoding.
7. Ordered Multi-Motif Grammar: Hierarchical attention and grammar transition prediction.
8. Alternating Motif Grammar: Alternating attention patterns and motif syntax regularization.
9. Nested Motif Grammar: Recursive transformer blocks and syntax tree generation.
10. Complex Multi-Motif Grammar: Full utilization of the motif interaction graph and multi-scale sequence scanning.

While some elements like the basic transformer structure and attention mechanisms (ideas) are taken from existing work (LSTM+transformers paper for grammars 2021), our architecture introduces several novel components relative to those works. The grammar-specific attention heads, dynamic positional encoding, and motif interaction graph are unique to our approach. The integration of biological knowledge into the architecture, such as the motif embedding augmentation and syntax regularization, is also a key novelty.

The overarching features of our architecture include:

1. Hierarchical structure: Captures patterns at multiple levels of complexity.
2. Grammar-specificity: Tailored components for each grammar type.
3. Biological integration: Incorporates domain knowledge into the model.
4. Interpretability: Generates explicit representations of grammatical structure.
5. Adaptive learning: Employs curriculum learning and adversarial training.
6. Multi-scale analysis: Processes sequences at various resolutions.
7. Flexible positional representation: Adapts to various spatial relationships between motifs.
8. Interactive motif modeling: Captures complex relationships between transcription factors.
9. Auxiliary tasks: Enhances learning through related prediction tasks.
10. Latent space modeling: Allows for continuous representation and interpolation of grammar types.