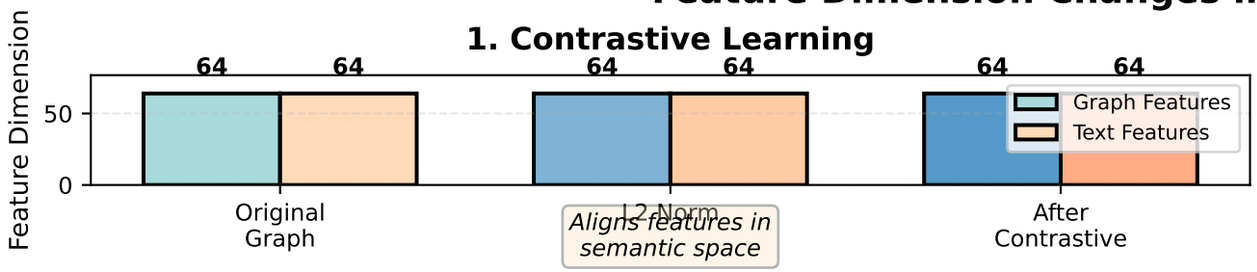
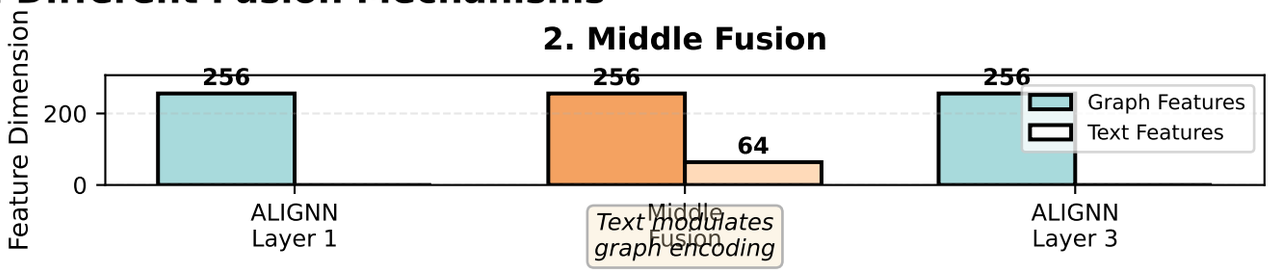
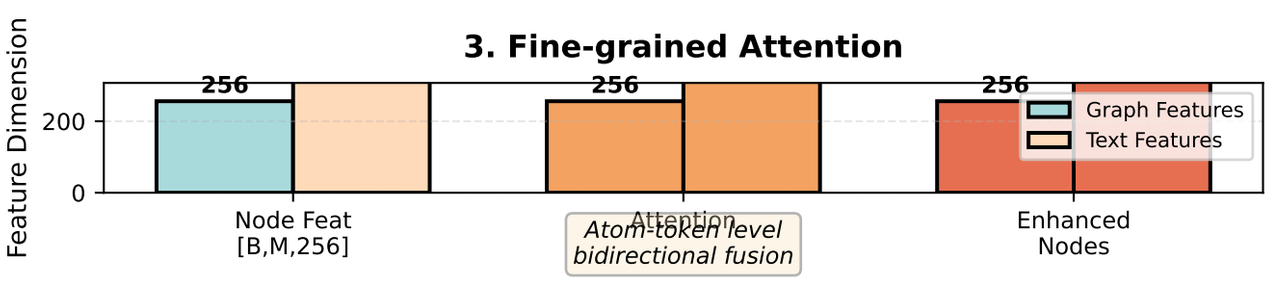
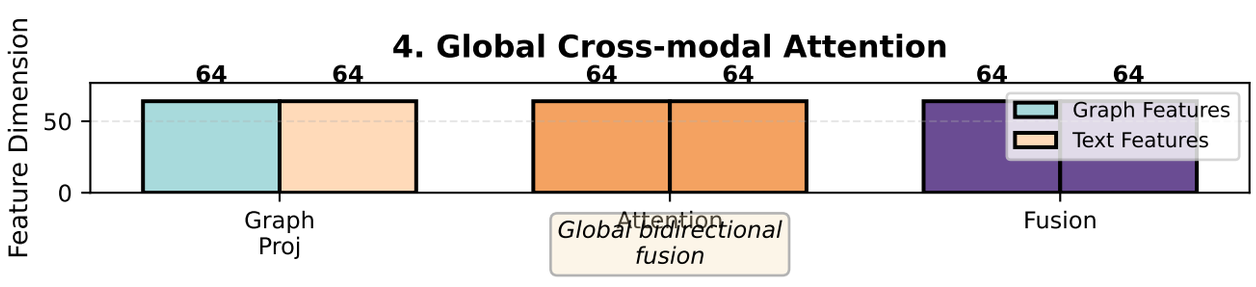

1. Contrastive Learning
Feature Dimension
64
64
64
64
64
64
50
0
Graph Features
Text Features
Original Graph
L2 Norm
After Contrastive
Aligns features in semantic space


2. Middle Fusion
Feature Dimension
256
256
64
256
200
0
Graph Features
Text Features
ALIGNN Layer 1
Middle Fusion
ALIGNN Layer 3
Text modulates graph encoding


3. Fine-grained Attention
Feature Dimension
768
768
768
256
256
256
200
0
Graph Features
Text Features
Node Feat [B,M,256]
Attention
Enhanced Nodes
Atom-token level bidirectional fusion


4. Global Cross-modal Attention
Feature Dimension
64
64
64
64
64
64
50
0
Graph Features
Text Features
Graph Proj
Attention
Fusion
Global bidirectional fusion