Parallel programming with low-level interfaces has been the most viable choice in scientific computing for a long time. In such models, different parallelisms require different parallel programming interfaces, e.g., message passing for parallelism across nodes, threading for intra-node parallelism, and vector processing for SIMD and GPUs. Often applications are confronted with these multiple interfaces to fully exploit the current and future large-scale machines. We present our work toward higher-level programming models, allowing for a single program to run on different parallel platforms without much human intervention, and at the same time to achieve close to hand-tuned performance.