Automated reaction prediction has the potential to elucidate complex reaction networks for many applications in chemical engineering, including materials degradation, drug design, combustion chemistry and biomass conversion. Unlike traditional reaction mechanism elucidation methods, which rely on manual setup of quantum chemistry calculations, automated reaction prediction avoids tedious trail-and-error learning processes and greatly reduces the risk of leaving out important reactions. Despite these promising advantages, the potential of automated reaction prediction as a general-purpose tool is still largely unrealized, due to high computational cost and inconsistent reaction coverage. Therefore, this thesis develops methods to simultaneously reduce the computational cost and increase the reaction coverage. Specifically, the computational cost is reduced by the development of more efficient transition state (TS) localization workflows and fast molecular and reaction property prediction packages, while the reaction coverage is increased by a comprehensive reaction space exploration based on mathematically defined elementary reaction steps. These components are implemented in two open-source packages, one is TAFFI (Topology Automated Force-Field Interactions) component increment theory (TCIT), and the other is Yet Another Reaction Program (YARP).

The power of TCIT and YARP has been demonstrated by a broad range of applications. In the first application, YARP was used to explore the reactivity of unimolecular and bimolecular reactants, comprising a total of 581 reactions involving 51 distinct reactants. The algorithm discovered all established reaction pathways, where such comparisons are possible, while also revealing a much richer reactivity landscape. Secondly, YARP was applied to the

search for prebiotic chemical pathways, which is a long-standing puzzle that has generated a menagerie of competing hypotheses with limited experimental prospects for falsification. With YARP, the space of organic molecules that can be formed within four polar or pericyclic reactions from water and hydrogen cyanide (HCN) was comprehensively explored. In the third application, predicting the reaction network of glucose pyrolysis, YARP generated by far the largest and most complex reaction network in the domain of biomass pyrolysis and discovered many unexpected reaction mechanisms. Further, motivated by the fact that existing reaction transition state databases are comparatively small and lack chemical diversity, YARP, together with the concept of a graphically defined model reaction, were utilized to address the data gap by comprehensively characterizing a reaction space associated with C, H, O and N containing molecules with up to 10 heavy (non-hydrogen) atoms. The resulting dataset, namely Reaction Graph Depth 1 (RGD1) dataset, represents the largest and most chemically diverse TS dataset published to date. In addition to exploring the molecular reaction space, YARP was also extended to explore reaction networks in heterogeneous catalysis systems. With ethylene oligomerization on silica-supported single site Ga^{3+} catalysts as a model system, YARP illustrates how a comprehensive reaction network can be generated by using only graph-based rules for exploring the network and elementary constraints based on activation energy and system size for identifying network terminations. The diverse scope of these applications and milestone quality of many of the reaction networks produced by YARP, illustrating that automated reaction prediction is approaching a general-purpose capability.