Journal of Advanced Research in Applied Chemistry & Chemical Engineering
Volume 6, Issue 2 - 2023, Pg. No. 13-18
Peer Reviewed Journal

Research Article

# Perceptive On Artificial Intelligence In Chemic Molecular Design

S. Brindha[1], Akrati Shrivastava[2], S.Ravichandran[3]

[1]Assistant Professor, Department of Computer Applications, SRM Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Potheri, Chengalpattu, Tamilnadu.
[2]B.Sc. Student in Agriculture, School of Agriculture, Lovely Professional University, Jalandhar, Punjab.
[3]Professor in Chemistry, School of Mechanical Engineering, Lovely Professional University, Jalandhar, Punjab.

## I N F O

**Corresponding Author:**
S.Ravichandran, Professor in Chemistry, School of Mechanical Engineering, Lovely Professional University, Jalandhar, Punjab.
**E-mail Id:**
ravichandran.23324@lpu.co.in
**Orcid Id:**
https://orcid.org/0000-0001-7281-2778
**How to cite this article:**
Brindha S, Shrivastava A, Ravichandran S. Perceptive On Artificial Intelligence In Chemic Molecular Design. *J Adv Res Appl Chem Chem Eng* 2023; 6(2): 13-18.

Date of Submission: 2023-09-16
Date of Acceptance: 2023-01-15

## A B S T R A C T

The need for custom compounds is growing, which makes molecular design work challenging. To be computationally manageable, earlier approaches to molecular design relied on simplified thermodynamic models. The most thorough molecular picture is provided by quantum mechanics in comparison, but it is difficult to directly incorporate it into computer-aided molecular design (CAMD). The current use of artificial intelligence to create an automated molecular design in chemistry has been met with skepticism on a number fo fronts, but deep learning and machine learning approaches also increase conceptual, technical, scalability and end to end error quantification issues. This article seeks to identify the most current and innovative technological advancements made by each of the parts of such an autonomous artificial intelligence and machine learning process. Furthermore, it can be integrated to significantly speed up the protein target.

**Keywords:** Artificial Intelligence, Molecular Design, Revocable Tree Representation



## Introduction

Artificial Intelligence is normally utilized more as well as more chemists to perform various type of chemical related working activities of drugs Figure 1. Purely AI in research is applied to chemistry has highly been fueled by the need to accelerate the drug discovery and reduce the large amount of costs to find and the time to invest in customize for new type of drugs. AI to chemistry has been grown tremendously in current years growth and distribution of AI related to past two decades in research areas considered that analytical chemists as well as bio related chemistry[1-4] are integrating AI to the greatest extent of the growth rates. The data driven models to make predicted the various tasks in the field of chemistry and artificial intelligence systems are helps to compare relationships are often present in datasets. Artificial Intelligence and chemistry both are great bond helps to identify the design the diagrams automatically. It mainly focuses on drug discovery and development in the healthcare industry. The technology mingling with the field of medicine, finding in formulating and creating drugs has become far and more advanced Figure 1.

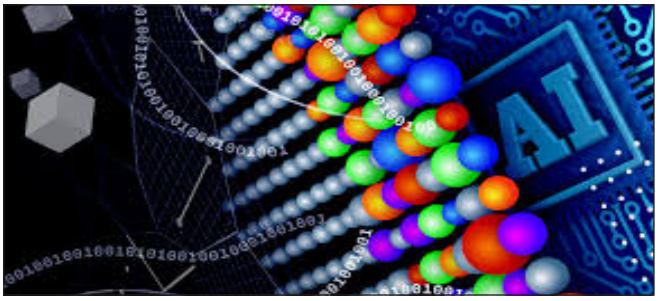


**Figure 1.Artificial Intelligence in materials**

Artificial intelligence should not just be used to find new drugs in the realm of chemistry. The fundamental components of chemical bonds are what make up science and are well suited to problems for which the physical laws that determine the molecular properties that must be predicted are not exactly well known with empirical relationships because it would be too difficult to establish sufficiently strong non linearities between parameters. Examine fresh developments in interdisciplinary research and spot recurring combinations of topics or subfields of study in literature. Chemical bonding, which form the foundation of science and technology, are the subject of an AI analysis. Artificial intelligence-based algorithms are frequently well suited to challenges where it is difficult to establish empirical relationships and it is difficult to uncover the molecular attributes that need to be anticipated based on physical laws. Strong parameter nonlinearities are used in conjunction with other prediction techniques, such as physical equations or empirical relationships, to control even more precise forecasts. The prevalence of several substance classes and their AI-related chemistry studies are assessed, highlighting how widely AI is being used in the sciences and analytical chemistry. The area of chemistry is prepared to develop molecular design technologies to levels much beyond what present synthetic offerings can provide.[5-7]

## Drawbacks of Manual Process

To make large advancements instead of incremental improvement is not a result of a lack of effort. It is one of the biggest limitations of the human mind. We are unable to design molecular structures containing hundreds of thousands of atoms. Each atom with specific purpose. Properties regarding the electronic structure, reactivity, toxicity, receptor binding capabilities, protein folding structure from its amino acid backbone, are now routinely predicted by ML programs.

## Novel Field of Artificial Intelligence

Although there are a huge number of applications for artificial intelligence in the realm of technology.

## Detection in Molecular Properties

The first and foremost applications of AI of chemistry is the detection of molecular properties, scientists have been performing the process of detecting chemical properties of molecules are explained in Figure 2. Artificial Intelligence has facilitated this procedure as a well as enabled scientist to detect molecular properties detecting properties has also enabled scientists to evaluate the potential or hypothetical molecule. The prediction of molecular properties can also be done using machine learning algorithms. Numerous other types of attributes, including bioactivity, toxicity, solubility, melting temperatures, atomization energies, HOMO/LUMO molecular orbital energies, and many others, have also been predicted using these algorithms. They are wholly data-driven and not dependent on physical principles or carefully crafted empirical relationships. In essence, supervised learning is the Predicted process by which these AI systems are taught by being fed numerous instances of molecules and the associated attributes. Different techniques for regression or classification can be utilized, including neural networks, support vector machines, random forests, linear regression.

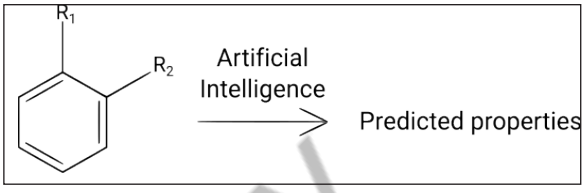


**Figure 2.Artificial Intelligence in Predicted Properties**

AI-based algorithms are particularly well-suited to problems for which the physical laws that determine the molecular properties to be predicted are not exactly known, or when empirical relationships would be too complicated to establish, e.g. for of strong non- linarites or correlations between parameters. It's interesting to note that using AI in conjunction with other prediction techniques, such as mathematical formulas or empirical relationships, can produce forecasts that are even more precise.

## Designing Molecules

Chemistry benefits greatly from the detection of molecular properties, and the application of artificial intelligence to molecular design has led to ground breaking discoveries. Scientists may now collect historical data as well as create chemical bonds for building molecules to molecular design. The design of molecules also results in additional practical applications that have aided in the advancement of sustainability in the chemistry community. One of the channel's most value added responsibilities is creating new compounds. Both their subject knowledge and their knowledge of chemistry are regularly used. The user utilizes their own imagination to suggest unique molecular structures are subsequently tested in the right applications either practically. The drawbacks are that human creativity is typically constrained a scientist may choose molecular patterns or functional groups that are certain or confident as well as remove particles that they find weird. The human mind is not capable of manipulating such large sets of molecules in a reasonable amount of time and large sets of molecules are needed for virtual screening.[13-16] Human creativity is less subjective it is also possible to create large numbers of molecules quickly. Calculation stable structural conformations and creating grained models for molecular dynamics simulations are examples for machine

learning. Numerous deep learning methods, including variation auto encoders, adversarial auto encoders, recurrent neural networks and graph networks, have been developed as well as utilized to create novel molecules. These Algorithms generate molecular structures as well as strings or another modern method, straight graphs. To create a statistical distribution of the molecules, these deep learning algorithms needed to be trained on a huge number of molecules, generally millions. Millions of molecules are present in a large number of datasets, including the ZINC database and the QM9 dataset from the ChEMBL database, which can also be used as training datasets. When creating SMILES strings, it's important to consider the authority of the results in order to ignore invalid SMILES.

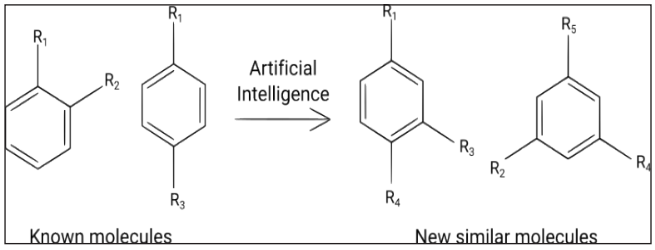


**Figure 3.Structure of molecules using Artificial Intelligence**

The deep learning algorithm has been trained; by choosing the educated statistical distribution, which is frequently close to molecules of interest because those molecules are known to contain desired attributes, it can generate large numbers of novel compounds. The designed molecules can be produced artificially or used for virtual screening to determine their desirable features. Small amounts of molecules are preferred over obtaining large numbers, and this is computed using Bayesian optimization Figure 3.

## Discovering Drugs

The process of finding novel, potent antibiotic compounds that can destroy many of the world's troublesome disease-causing bacteria is one of the top applications of artificial intelligence in chemistry. Scientists are attempting to combat the surface-level emergence of new illnesses.

## Retro Synthesis Reaction

To find the building blocks of their molecules, a procedure is broken down. To find a molecule, the synthesis is carried out backward. It takes a lot of time and resources and is laborious and challenging because it calls for vast experience. Computer assistance in the Artificial Intelligence process may result in greater accuracy and time-saving efficiency.[17]

## Automated and Technology

Another subset of artificial intelligence, called AI planning, deals with automated generation of action sequences that can be carried out by unmanned artificial intelligence vehicles, such as automatic robots used in factories and other types of laboratories. AI planning also includes automated planning and scheduling. Though it cannot be guaranteed, prediction is more precise and effective in terms of probability. With the use of machine learning or deep learning algorithms, computers are fed data over time that improves their comprehension abilities and equips them to collaborate with human intellect.

## Chemical Reaction Optimization

The search for the ideal reaction conditions might take a very long time and involve many distinct parameters that are difficult to regulate between experiments. This process is entirely automated, just like a blog article. The optimal reaction conditions for chemical processes are currently being found iteratively by neural networks. For existing optimization algorithms or for optimizing one variable at a time, this deep learning model was more efficient. AI-based approaches offer the advantage of allowing multiple parameters to be changed simultaneously to obtain the best outcome in chemical reactions.

## Proposed Revocable Tree Representation of Molecules

A cyclic structure is produced if the bond nodes meet at many nodes. Call the nodes that represent the bond and ring, respectively, "bond nodes" and "ring nodes." The dataset's distinctive atom clusters can be used to create a language. Based on the atom cluster within the node, a word ID (w i) is given to each node (N i). By using this coarse-grained representation, we may create molecules from chemically sound fragments rather than by building them atom by atom from chemically flawed intermediates.

The JT representation, however, cannot be instantly changed back into the original molecule. By expanding the JT decomposition of, we introduce the revocable JT (RJT) representation () of molecules in this study. The ID of the atoms shared between two adjacent nodes (Ni and Nj) should be recorded as the "site information" i,j in order to remove arbitrariness in the node connection (i.e., to determine the atoms shared between two nodes connected by an edge). The original JT implementation shown in figure 3. entails sharing one or two atoms between nearby nodes connected by the JT edge. In addition, type-1 edges can be divided into three groups: edges linking singleton nodes, two bond nodes, and ring nodes. In these three instances, the site information (i,j) is represented by the indices of the shared atom in both nodes. The indices of the two atoms in both ring nodes are maintained for type-2 edges, which always connect two ring nodes. The first atom's index and the second atom's direction ID (+1 or 1) are saved as site information i,j since the shared atoms between the

nodes are always situated next to one another in the ring. Despite the fact that just one atom is shared across the nodes, a spiro connection between two rings is handled as a special case of the type-2 edge and given the direction ID of 0. The most appropriate location for storing this site information is edge $E_{i,j}$, which represents the existence of shared atoms in $N_i$ and $N_j$ in . Thus, in this study, the site information is encoded as edge features Figure 4.

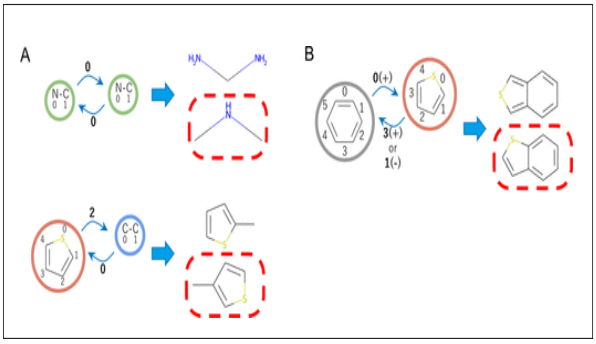


**Figure 4.Definition of the site information that enables the reversible conversion to the original molecular structure**

Details about the site for type-1 edges. Circles and arrows on the left panels, respectively, denote the nodes and edges of the tree representation. The arrows' proximity to numbers denotes location information. The potential molecular structures formed from the tree representation are displayed in the left panels. The original building denoted by the dashed line can be chosen without hesitation using site information. Site data for type-2 edges (B). Near the arrows, numerals and (+)/ marks denote site information. The following is a description of the RJT representation to molecular graph conversion algorithm (Algorithm 1): This approach is comparable to the Jin et al. assembly procedure, with the exception that it does not need listing all possible combinations of node-to-node attachments.

The assembly techniqueis similar to this method's algorithm for converting the RJT representation to molecular graph (technique 1), except that it does not need listing every possible combination of node-to- node attachments. By adding a graph to the created graph that corresponds to the child node, our technique traverses the predicted in depth-first order. The atom(s) shared by two nodes can be identified uniquely using the site information, and the nodes could be built deterministically to convert to. The projected site information might, however, not match the types of nodes (for instance, if more than four atoms are linked to a single carbon atom).

As an alternative, a specific RJT could have its invalid information flagged by raising an exception. The code was made simpler by using the latter implementation. The latter implementation was used to simplify the code.

**Table 1.Algorithm 1 Depth-first Assembly**

| Algorithm 1 Depth-first Assembly of Ť | |
|---|---|
| Require: | Reversible Tree Representation of mol Ť |
| Step 1: | Initialize: i ← One node selected from Ť |
| Step 2: | Initilialize: j← Φ |
| Step 3: | Ğ← mol graph of node i |
| Step 4: | Function DepthFirstAssembler(I,j,G) |
| Step 5: | for k€ N(i)\j do |
| Step 6: | Set $\sigma_{i,k}$ ←site information of i against k |
| Step 7: | Set $\sigma_{k,i}$ ←site information of k against i |
| Step 8: | Set Ç← mol graph of node k |
| Step 9: | Attach mol graph Ğ to Ğ using $\sigma_{i,k \text{ and }} \sigma_{k,i}$ |
| Step 10: | DepthFirstAssemble (k,i Ğ) |
| Step 11: | end for |
| Step 12: | end function |

By adding a graph to the created graph that corresponds to the child node, the suggested method traverses the predicted in depth-first order. The atom(s) shared by two nodes can be identified uniquely using the site information, and the nodes could be built deterministically to convert to. The projected site information may, however, not match the types of nodes (for instance, if more than four atoms are coupled to a single carbon). Based on the results of the soft max logits produced by the neural network the second (or third) probable site information can be used in such circumstances. As an alternative, a specific RJT could have its invalid information flagged by raising an exception. The code was made simpler by using the latter implementation.

Importantly, in Table 1. the algorithm describes the reversible tree plantation, function depth first assembler of artificial intelligence and machine learning has become common place in the pharmaceutical sector. A brief explanation of how machine learning has developed and how it is being used to model the various features of molecules that cut across industries and their daily utilities. Different molecular architectures will be covered by the potential for tightly integrating AI with machinery as well as automated workflows to construct generative models. To illustrate how machine learning has impacted and inspired our job, list a few of the companies that are at the forefront of employing AI. Although they are still in the early stages of development, AI algorithms for quantum computational simulations are showing great promise. When various machine learning techniques are combined, the applications

of these simulations become highly appealing. It is possible to implement ML to find the coarse-grain models that simplify the molecule in molecular dynamics simulations. The potential energy surface and hence the molecular configuration are then determined by a second ML.

The time-steps that machine learning is best suited to anticipate can be found by using this technique. In theory, this might lead to the processing power no longer being the limiting factor in molecular dynamics simulation due to the employment of several ML algorithms. Large molecules can be simulated on longer time scales to an ML algorithm than was previously conceivable. Genetic Engineering Fast structure determination and improved molecular dynamic simulations are made possible by computationally simple and precise quantum-chemical simulations utilizing ML. The ability to create millions of compounds for a chosen attribute has been made possible by advancements in molecular property structuring.

By combining these applications, it is possible to analyze and count randomized molecules in relation to brand-new types of medications, materials, proteins, and catalysts. Randomized compounds are ineffective and do not offer the benefits of having a molecule that is specially created. Even while mechanical engineers can experiment with various combinations of metal, electronics, and other materials, the best outcome will still result in a poorly designed vehicle. The majority of chemistry-related AI combines genetic programming, another AI technique, and molecular property design with quantum chemical molecular property design. New types of molecules can be created via genetic programming for a certain set of work modules.

Artificial intelligence, also known as machine intelligence, would recreate the evolutionary process to enable creative computer imitation. Genetic programming is closely related to feeding the computer a random assortment of chemicals. Every single compound is simulated in order to assess it for the desired attribute. Depending on the score or fitness, it will be ranked. The molecules with lower scores are eliminated, mimicking the evolutionary method of fitness-based survival.

The survival of the fittest evolutionary principle is replicated by discarding molecules that receive a lower score for this attribute. are chosen to pass on their chemical structure to the following iteration in order to convey genes to progeny in order to get a high score. There are three ways to create the next iteration. The parent can be modified, cloned, or coupled with an additional evolving design. This stage implements the genetic variation of the program and is governed by the GP's architecture. The elements are then assessed, and the cycle is repeated thousands of times. The artificial intelligence program can produce innovative, useful, and creative designs.

## Discussions

### Machine learning is particularly effective at solving particular chemical issues

**Table 2.Molecular Design Trails with various properties**

| Molecules | Trials | Trials |
|---|---|---|
| Mass of Psycnometer | 420 gm | 420 gm |
| Mass of Psycnometer with Solid Samples | 830 gm | 756 gm |
| Mass of Psycnometer with Solid Samples and Water | 1726 gm | 1680 gm |
| Mass of Psycnometer with Water | 1466 gm | 1466 gm |

Significant applications for an AI for molecular design are provided by the consolidation of many of these systems. The concurrent research in these individual fields of property prediction, quantum simulation, and genetic programming is also explained in Figure. The main benefit of combining these technologies into an artificial intelligence for molecular design is that it would open up practically limitless applications and possibilities. Applying evolutionary principles would enable a molecular design process that was supported by biological observations. The primary tool for molecular design, which is fully unrestricted by human limitations and allows for the creation of sophisticated and efficient molecular designs, is arguably more interesting Table2.

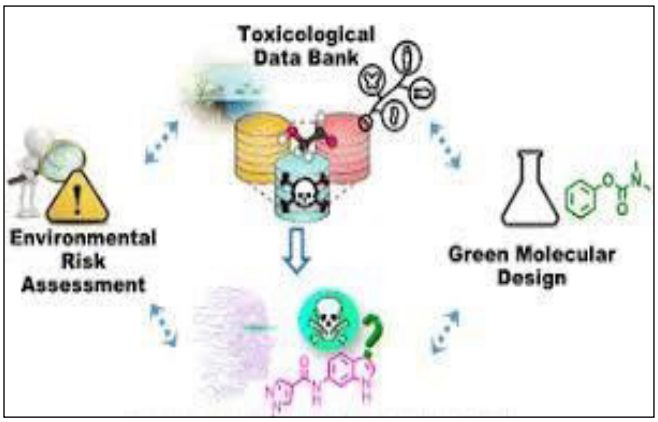


**Figure 5.AI Based Chemical Toxicity Prediction**

Each component of the AI has been recognized, and it has already been used to build catalysts and successful medication candidates. Finding a molecule with high polarity and selecting other molecules with the best efficiency based on size are two very straightforward jobs that an AI could complete. Candidates could very well outperform comparable human-made compounds in terms of functionality because GP

has the capacity to test and develop intelligently designed molecules. As many fewer molecules would undergo and fail experimental testing, the time spent on manufacturing and screening would be significantly reduced, testing and health screenings.

## Conclusion

In the healthcare sector, the application of artificial intelligence and chemistry is mostly focused on medication development. The discipline of medicine has embraced technology, which has greatly increased drug research and production. Due to the highly developed technology and equipment employed by scientists, this technique has also resulted from sophisticated research and development in the pharmaceutical business. The conclusion is that nano chemistry, which is present in biology created through evolution, would need aid beyond what humans are capable of providing in order to challenge human- made molecular designs. As a result, a different tool is required that can better take into consideration those constraints and build molecules using a process other than human inventiveness. Numerous tools are becoming available as a result of developments in computer science and machine learning (ML) technology. A breakthrough in Revocable Tree Representation of Molecules was made by each of the parts of this artificial intelligence-based machine learning approach. Quantum chemistry simulations using ML approaches show promise for much less computationally intensive simulations with sufficient accuracy for chemical applications.

## References

1. Mehrazma B,Rauk, A. Exploring amyloid-β dimer structure using molecular dynamics simulations. The Journal of Physical Chemistry A, 2019 ;123(22):, 4658-4670.
2. Dr. S Brindha, Dr.S.Ravichandran, “Artificial Intelligence in Chemic molecular Design”, International Conference on Chemistry and Allied Sciences”, (ICCAS-2022), DST-SERB, CSIR,INSA& TSCHE Sponsored to Pingle Government College for Women (Autonomous), Hanumkonda, Warangal, India, during 25-27 August 2022.
3. Chu Y, & He, X. (2019). MoleGear: a java-based platform for evolutionary de novo molecular design. Molecules, 24(7), 1444.
4. Copeland B J. Alan Turing and the Beginning of AI. URl: https://www. britannica. com/technology/artificial-intelligence/Alan-Turing-andthe-beginning-of-AI 2020; (visited on 07/04/2020).
5. Cova T F, Pais A A. Deep learning for deep chemistry: optimizing the prediction of chemical patterns. Frontiers in chemistry 2019; 7, 809.
6. Allison Duettmann and James B. Lewis. Artificial Intelligence for Nanoscale Design. 2018. url:https:// foresight.org/wp- content/uploads/2018/11/AI-for-Nanoscale- Designerport-2.pdf.
7. Freeze J G, Kelly, H R, & Batista V S. Search for catalysts by inverse design: artificial intelligence, mountain climbers, and alchemists. Chemical reviews, 2019; 119(11):, 6595-6612.
8. Granda J M, Donina L, Dragone V, et al. Controlling an organic synthesis robot with machine learning to search for new reactivity. Nature, 2018; 559(7714), 377-381.
9. Wang J, Olsson S, Wehmeyer C, et al . Machine learning of coarse-grained molecular dynamics force fields. ACS central science, 2019; 5(5):, 755-767.
10. Jin, W, Barzilay R, & Jaakkola T. (2018, July). Junction tree variational autoencoder for molecular graph generation. In International conference on machine learning (pp. 2323-2332). PMLR.
11. Hauptfleisch K. This Is How Artificial Intelligence Will Shape eLearning For Good. Publisher: Elearning Trends, https://elearningindustry. com/artificial-intelligence-will-shape-elearning 2016;
12. Rupp M, Tkatchenko A, Müller K R, et al.Fast and accurate modeling of molecular atomization energies with machine learning. Physical review letters, 2012; 108(5):, 058301.
13. Senior A W, Evans R, Jumper J, et al. Improved protein structure prediction using potentials from deep learning. Nature, 2020; 577(7792), 706-710.
14. Shindo H, & Matsumoto Y. Gated graph recursive neural networks for molecular property prediction. arXiv preprint arXiv: 2019; 1909.00259.
15. Segler M H, Waller M P. Neural-symbolic machine learning for retrosynthesis and reaction prediction. Chemistry–A European Journal, 2017; 23(25):, 5966-5971.
16. Morawietz T, Singraber A, Dellago C, et al. How van der Waals interactions determine the unique properties of water. Proceedings of the National Academy of Sciences, 2016 ;113(30):, 8368-8373.
17. Torrisi M, Pollastri G, Le Q. Deep learning methods in protein structure prediction. Computational and Structural Biotechnology Journal 2020; 18, 1301-1310.