Computer Physics Communications 199 (2016) 8–11

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

Exact enumeration of conformations for two and three dimensional lattice proteins

K. Silpaja Chandrasekar, M.V. Sangaranarayanan *

Department of Chemistry, Indian Institute of Technology – Madras, Chennai-600036, India

ARTICLE INFO

Article history:
Received 12 June 2015
Received in revised form
16 September 2015
Accepted 21 September 2015
Available online 13 October 2015

Keywords:
Lattice proteins
Hydrophobic-Polar model
Self-avoiding walks
Conformations
Polymers

ABSTRACT

We report an efficient methodology for exactly enumerating conformations of lattice proteins, taking into account the self-avoiding nature. These self-avoiding walks in square and simple cubic lattices take into account, the detailed paths by which a destination site can be reached from a starting site. The strategy employing JAVA programming enables the exact enumeration for very large lattice sizes. The estimation of these conformations is helpful in designing the protein sequences pertaining to Hydrophobic-Polar models.

Program summary

Program title: Exact enumeration of conformations in lattice proteins

No of lines in distributed program: 1027

No of bytes in distributed program: 51698

Programming language: JAVA

Computer: Tested on Intel® Core(TM) i5-4570. Will function on any computer with JAVA compilers.

Operating system: Tested on Microsoft Windows 7 Professional. Should run on any platform with JAVA

Has the code been vectorized or parallelized? : No

RAM: Depends on the system size.

Nature of the problem: The problem involves the enumeration of the conformation of lattice proteins for various amino acid chain lengths by considering them as self-avoiding walks. The methodology is illustrated for square and simple cubic lattices.

Solution method: The code was implemented using the nodes as in a tree which was arranged in the stack. Each element is visited once and is moved to the next node. The process is repeated till all nodes are visited and the process terminates when a site is revisited.

Restrictions: Runs only in JAVA compilers

Running time: Depends on the system size. Using 4 processors, the output leading to entries in Tables 1 and 2 was generated within one minute.

The listing of the codes and the output of the program are provided in the Supporting Information.

1. Introduction

The exact enumeration of conformations in two and three dimensional lattices is of paramount importance in diverse contexts; among them, mention may be made of the following: (i) Hydrophobic-Polar (HP) model for proteins [1]; (ii) analysis of secondary structures in polymers [2] and (iii) adsorption of heteropolymer on surfaces [3]. In the HP model of lattice proteins, the analysis of conformations provides an insight into the fraction of designing sequences of various amino acid chain lengths. In the field of polymer physics, the secondary structures of polymers are influenced by the number of monomer units. The kinetics and thermodynamics of adsorption of polymers is dictated by the nature of the surface while the conformational analysis provides a clue to the mechanism.

* Corresponding author.
E-mail address: sangara@iitm.ac.in (M.V. Sangaranarayanan).

http://dx.doi.org/10.1016/j.cpc.2015.09.017
0010-4655/

The calculation of conformations *vis a vis* self-avoiding random walks in any lattice is an NP hard problem [4]. While enumerations for small lattice sizes is feasible in two and three dimensions, either scaling hypothesis using critical exponents [5] or extrapolation procedures using ratio methods [6] is required for a reliable estimate of the thermodynamic properties of the system. In order to obviate this limitation, it is imperative that the computational frontier be pushed further so as to enumerate the conformations exactly. It is customarily thought that such computations for large lattice sizes are impossible with the present day computers.

The objective of this Communication is to provide a simple algorithm for exact enumerations of conformations pertaining to square and simple cubic lattices of amino acid chain lengths N. The maximum value of N is 72 in two dimensions and 41 for three dimensions. Our *tour de force* consists in identifying the destination sites in square and cubic lattices as well as employing JAVA programming whereby the computational limitations are partially overcome.

2. Methodology

(a) Square lattice

This algorithm implies the construction of the matrix which is placed in a stack. An array value is assigned to each row of a stack. The array value that corresponds to each row in the stack refers to the element in the matrix of n rows and m columns, denoted as $n * m$. For a walker typically starting at the site marked as '*' in Fig. 1, the destination side can be reached via the path shown by arrows, without visiting any site more than once. In the case of the square lattice, the destination sites can be considered as 'left', 'right' and 'adjacent', as shown in Fig. 1. The root node is defined to be the starting point and it is queued in the stack with all other nodes. The path starts from the first site and covers all the sites of the lattice. The nodes are dequeued and examined repeatedly with all the root nodes.

The algorithm calculates the number of ways by which these three paths can be reached for each site and the total number yields the total number of self-avoiding walks.

Algorithm 1 -Algorithm for computing the number of conformations square lattices

```
Function con2d(a ,int visited ,int source)
visited(source) = 1;
If visited(source) = 1 then
Stack <-push(i)
If visited[i] = 1 then
Element <- i
If not visited(i) then
     element <- false
Function Cl_con3d(int number_of_nodes, source)
For i1,j1 to 3,4
q_i <- nodes_to_visit.first()
for all i1,j1 do
for number_of_nodes = 4 to 72 do
for all edges e in G.incidentEdges(t) do
o ← G.opposite(t,e)
scanner <- close
if( [tempRow tempCol] != 0 )
          then   row = (row+1)%n;
           i1<-j1<- j1<-i1 = visited [1]
if not visited(i) then
   visited (i) <- false
```

(b) Simple cubic lattice

The methodology is essentially similar to the algorithm for two dimensions but with six different destination sites for the six faces (denoted as p1, p2, p3, p4, p5 and p6) of a simple cubic lattice as shown in Fig. 2. Each site is characterized by three indices and the destination site can be present in any of the six faces. The algorithm enumerates the total number of ways by which the destination site can be reached.

Algorithm 2 -Algorithm for computing the number of conformations for simple cubic lattices

```
Function con3d(a ,int visited ,int source)
visited(source) = 1;
If visited(source) = 1 then
Stack <-push(i)
If visited[i] = 1 then
Element <- i
If not visited(i) then
          element <- false
Function Cl_con3d(int number_of_nodes, source)
int a = (log 1.6)^(number_of_nodes) , o;
int b = (number_of_nodes)^(a) ;
for i1,j1,k1 to 3,4,5
q_i <- nodes_to_visit.first()
for all i1,j1,k1 do
for number_of_nodes = 3 to 41 do
for all edges e in G.incidentEdges(t) do
o ← G.opposite(t,e)
scanner ← close
i1<-j1 = visited[1]
j1<-k1 = visited[1]
k1<-i1 = visited[1]
   if not visited(i) then
 visited (i) <- false
```

3. Results and discussion

The counting of self-avoiding walks (SAWs) in two and three dimensions has been a fascinating exercise in statistical physics, for more than six decades. In the case of square lattices, the enumerations till $N = 71$ have been accomplished using 128 processors on HP alpha server [7]. For simple cubic lattices, the computational sophistication has led to an increase in the chain length from $N = 9$ in 1947 [8] to $N = 36$, recently [9]. The extension to large lattice sizes is required to deduce the fraction of designing sequences as well as obtain accurate thermodynamic properties in the $N \to \infty$ limits.

3.1. Enumeration of conformations

(A) Square lattice

The methodology for estimating the number of conformations of square lattices (C_N^{2d}) is shown in Algorithm 1. While these have hitherto been enumerated *exactly* till $N = 28$ [9], we have extended these calculations to $N = 72$ (Table 1, Supporting Information) here. However, the enumerations could not be carried out for $N > 72$ on account of the limitations associated with JAVA programming.

(B) Simple cubic lattice

The number of conformations for simple cubic lattices (C_N^{3d}) is estimated using the Algorithm 2. The exact values exist till $N = 36$ at present [10] while the conformations till $N = 41$ have been provided here for the first time (Table 2, Supporting Information). The conformations for $N > 41$ could not be enumerated.

3.2. Influence of the lattice size on conformations

(A) Square lattice

Fig. 3(a) depicts the variation of the conformations with lattice size in two-dimensions wherefrom it is seen that log $\{C_N^{2d}\}$ varies linearly with the chain length. It is customary to postulate a scaling relation using the universal exponent γ_{2d} and effective coordination number μ_{2d} [11] for the *generating function* of SAWs.

$$C_N^{2d} \sim \mu_{2d}^N N^{(\gamma^{2d-1})}$$

where the exponent $\gamma_{2d} \approx 1.34$ and the value of $\mu_{2d} \approx 2.638$.

(B) Simple cubic lattice

Fig. 3(b) provides the dependence of $\log\{C_N^{3d}\}$ on the number of lattice sites (N). It is seen that a linear dependence arises for all