STUDIA INFORMATICA 2002
Volume 23 Number 2A (48)

Lech ZNAMIROWSKI
Silesian University of Technology, Institute of Informatics

NETWORK SIMULATION OF THE POST-TRANSLATIONAL MODIFICATIONS IN ORGANIC NANOSTRUCTURES*)

Summary. In this paper the paradigm for simulation of post-translational conformations in the polymer chain of amino acids, in a dynamic of the protein torsion angles aspect is discussed. The polypeptide synthesized in the ribosome is modeled using torsion angles as a degrees of freedom. The model construction paradigm for modeling of the post-translational modifications (conformations) by dynamic programming is presented. The computer network structure contains a workstation fulfilling the needs of Graphic User Interface, and a supercomputer working under control of the workstation to fulfill the numerical tasks of the model implementation.

SYMULACJA SIECIOWA MODYFIKACJI POTRANSLACYJNYCH NANOSTRUKTUR ORGANICZNYCH

Streszczenie. W pracy przedstawiono metodę symulacji potranslacyjnych konformacji łańcucha polimeru aminokwasów w aspekcie kątów torsyjnych przyjętych jako stopnie swobody układu. Przedstawiono konstrukcję modelu dla określenia konformacji potranslacyjnych metodą programowania dynamicznego. Model sieciowy, bazujący na wymianie komunikatów, zbudowany został w konfiguracji stacji roboczej spełniającej zadania graficznego interfejsu użytkownika i sterowania oraz superkomputera realizującego czasochłonne procedury numeryczne.

1. Introduction

The mechanism of post-translational modifications in nanostructures takes place after translation process performed in the reactor (ribosome) generating the polypeptide chain.

*) Work supported by State Committee for Scientific Research, Poland, Grant No. KBN 7T11C01721".

The polypeptide synthesis is a process called translation because the string of letters of the four-letter alphabet of nucleic acids accordingly with genetic code, is translated into string of amino acids of the 20-amino acid alphabet, forming proteins. The translation process is performed in a ribosome moving along the mRNA chain (powered by the hydrolysis of GTP), with the activated precursor driven by ATP, and in presence of at least one kind of tRNA and activating enzyme for each amino acid. Protein synthesis takes place in initiation, elongation, and termination stages. The *initiation stage* results in the connecting of the initiator tRNA to the start signal in mRNA. The *termination stage* takes place when a stop signal in the mRNA is read by the protein release factor. Each nucleotide triplet, or codon, in mRNA chain encodes a specific amino acid. In the *elongation stage*, each molecule of tRNA binds only the amino acid proper to a particular codon, and tRNA recognize a codon by means of a complementary nucleotide sequence named anticodon. When the termination stage occurs, the completed polypeptide is released from the ribosome [3, 10]. Next, the pro??cesses of modifications after translation take place. In the paper the spatial modifications called conformations are discussed.

2. Computer Model for the Structure Prediction

The goal of model construction is investigation of conformation in the post-translational processes parallely, when the massive numerical procedures have to be performed [2]. The hierarchical model may have a form of shell containing the objects representing material elements (molecules or groups of molecules), and the objects representing messages, timely interdependent. The objects of the model have a form of the elements of specialized library. In the paper, a network structure of the model for modeling of the post-translation process are described. The integrated network model performs two basic tasks in the two-computer system: the Graphic User Interface (GUI) and computations control is realized in the PC workstation (Win98/2000), and the time-consuming, numerical computing for application are performed in the multiprocessor based parallel computer (supercomputer) - the Sun Enterprise 6500 server (12 high performance, 64-bit SPARC™ [Scalable Processor ARChitecture] V9 RISC, 6 GFLOPS) working under the Solaris™ 7.

3. Polypeptide Structure

The three-dimensional structure of a polypeptide can be completely described by placing it in a Cartesian coordinate system and listing *x, y, z* coordinates for each atom in the chain.

However, the synthesis programs generating the conformation of the main chain of polypeptide use the amino acids library and need the information on the angles between α-carbons bonded with their side chains and a contiguous peptide groups. The backbone of a chain of amino-acid residues determined by set of torsion angles, is presented in Fig. 1.

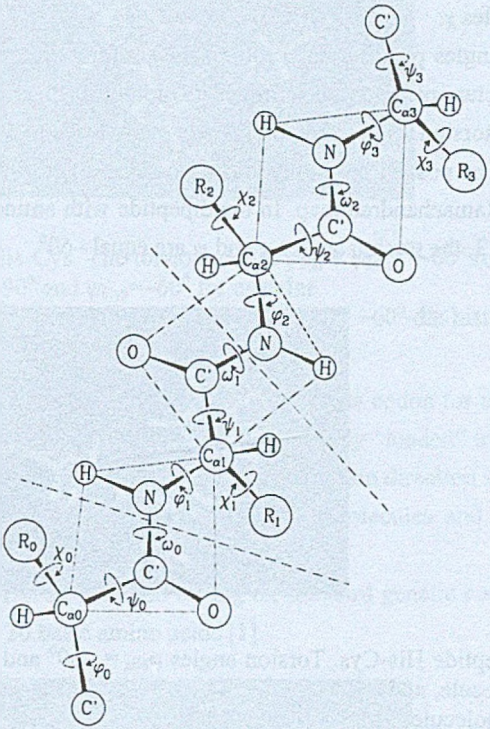

Fig. 1. Definition of protein torsion angles φ, ψ, ω and χ in the polypeptide sequence. The limits of a residue #1 are indicated by dashed lines, the chain is shown in a fully extended conformation ($\varphi_1=\psi_1=\omega_1=180^0$)

Rys. 1. Definicja kątów torsyjnych φ, ψ, ω and χ w sekwencji polipeptydu. Linie przerywane oznaczają granice reszty aminokwasowej (#1), a łańcuch przedstawiono w formie rozciągniętej ($\varphi_1=\psi_1=\omega_1=180^0$)

The backbone conformation of an amino-acid residues can be specified by listing the torsion angle φ (rotation around the nitrogen-α-carbon bond in the main chain), and ψ (rotation around the α-carbon-carbon bond in the main chain of the polypeptide). The relationship between the peptide groups, α-carbons and torsion angles can be expressed in a following form:

$$\rightarrow \mathrm{PB} \rightarrow \varphi_1 \rightarrow \mathrm{C}_{\alpha 1} \rightarrow \psi_1 \rightarrow \mathrm{PB} \rightarrow \varphi_2 \rightarrow \mathrm{C}_{\alpha 2} \rightarrow \psi_2 \rightarrow \mathrm{PB} \rightarrow \quad (1)$$

where PB denotes the peptide bond and $C_{\alpha i}$ is the i-th α-carbon atom.

The zero for torsion angle φ is defined with the N—H bond trans to the C_α—C' bond, and the zero position for ψ is defined with the C_α—N trans to the C=O bond. The peptide-bond torsion angle ω is generally 180°, this is with C=O bond trans to the N—H bond [7]. A complete description of the spatial structure of a protein also require a knowledge of the side-chain torsion angles χ.

The torsion angles play a crucial role in the conformation of proteins because the three-dimensional structure of protein determines its biological functions. On the other hand not all combinations of torsion angles are possible, as many leads to collision between atoms in adjacent residues. The possible combinations of φ and ψ angles that do not lead to collision can be plotted on a Ramachandran map. In the dipeptide with amino acid [14] sequence His-Cys presented in Fig. 2, the torsion angles φ and ψ are equal –60°.

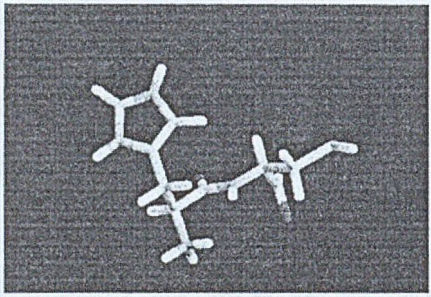

Fig. 2. Dipeptide His-Cys. Torsion angles $\varphi_{His} = -60°$ and $\psi_{His} = -60°$ for histidine molecule, and $\varphi_{Cys} = -60°$ and $\psi_{Cys} = -60°$ for cysteine. Stick representation of molecule [9]

Rys. 2. Dipeptyd His-Cys. Kąty torsyjne dla histydyny wynoszą $\varphi_{His} = -60°$ oraz $\psi_{His} = -60°$, natomiast dla molekuły cysteiny $\varphi_{Cys} = -60°$ i $\psi_{Cys} = -60°$. Reprezentacja belkowa molekuł [9]

It can be simply observed in Fig. 3 and 4 that the small changes of the torsion angles cause fundamental changes in conformation of the polypeptide in a case, when the chain of amino acids is very long.

4. Translation Process

The flow of genetic information from DNA (data bank and replication) through the transcription of genetic code triplets (codons) into the mRNA chain (mobile data transmitter) finally is expressed in the material form as a synthesized proteins. Synthesis of proteins is

performed on ribosomes in the environment, where the all aminoacylated tRNAs, activating enzymes and protein factors are accessible.

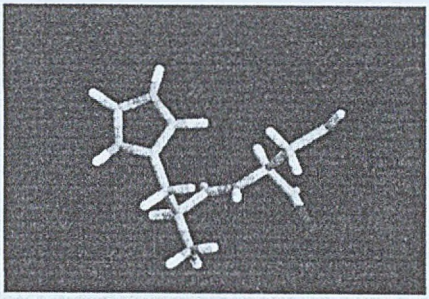

Fig. 3. Dipeptide His-Cys. The torsion angles $\varphi_{His} = \psi_{His} = -60^{o}$ for histidine, and $\varphi_{Cys} = -90^{o}$ and $\psi_{Cys} = -60^{o}$ for cysteine

Rys. 3. Dipeptyd His-Cys. Kąty torsyjne $\varphi_{His} = \psi_{His} = -60^{o}$ dla histydyny oraz $\varphi_{Cys} = -90^{o}$ i $\psi_{Cys} = -60^{o}$ dla cysteiny

The initial interaction occurs in such a way as to allow the codon for the initiating amino acid (methionine) to interact with corresponding appropriately "loaded" transfer RNA to begin polypeptide synthesis. Next, consecutive codons (translation direction is from 5' to 3' end of mRNA) interact with proper transferred amino acid molecules and the peptide chain shoves up forming the protein.

The translation process can be modeled basing on standard genetic code [10] and one or two letter description of the 20 basic amino acids [1].

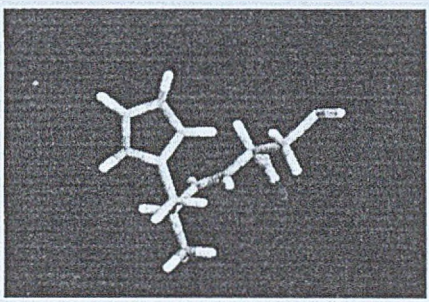

Fig. 4. Dipeptide His-Cys. The torsion angles $\varphi_{His} = -60^{o}$ and $\psi_{His} = -85^{o}$ for histidine and $\varphi_{Cys} = \psi_{Cys} = -60^{o}$ for cysteine

Rys. 4. Dipeptyd His-Cys. Kąty torsyjne $\varphi_{His} = -60^{o}$ i $\psi_{His} = -85^{o}$ dla histydyny oraz $\varphi_{Cys} = \psi_{Cys} = -60^{o}$ dla cysteiny

The long chains of nucleotide sequence, can be effectively processed with sufficient software with many parameters, and next processed [8] into PDB files for consecutive computations [12, 13].

Visualization can be effectively performed with the RasMol [9] software.

5. Post-translational Conformations

When the translation process reaches the last, termination stage, the chain of amino-acid residues forming the polypeptide is inserted into an environment which is usually the fluid. This fluid, cytosol, can be treated in first approximation as a solution. The polypetide forms a chain of the molecules of residues connected together through the rotational bond realized by peptide bonds.

In this case (Fig. 5) possible conformation depends on mutual interaction between residues of amino acids and molecules of solution like hydrophobic or hydrophilic interactions. The force powered changes, is a tendency of the system chain of polypeptide and solution molecule in some neighbourhood to get minimal free energy [5]. This can be called free conformation of the polypeptide chain.

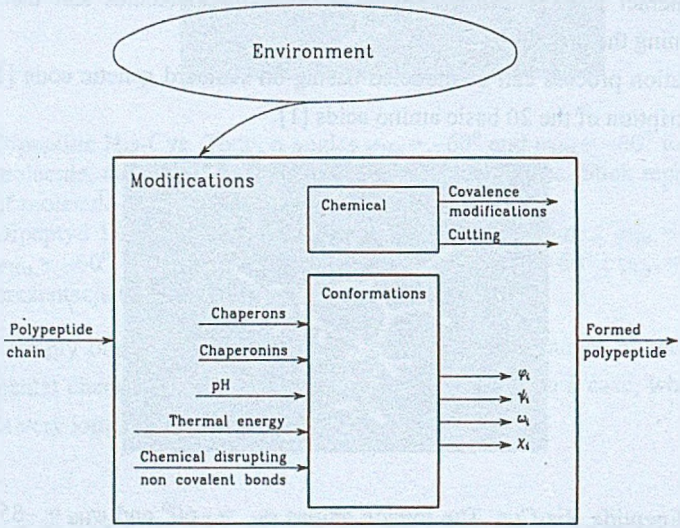

Fig. 5. Torsion angles determination in the interaction of polypeptide chain and the environment

Rys. 5. Wyznaczanie kątów torsyjnych w wyniku interakcji łańcucha polipeptydu i środowiska

In a case when an external force interacts on the polypeptide chain (like chaperons or enzymes) that extended system (chain, solution and enzyme) tends also to reach minimal internal energy. This can be called forced conformation. It is important that internal bonds e.g. disufide bonds or hydrogen bonds can interrupt the conformation transient state.

In Fig. 5 the two main groups of post-translational modifications are presented: firstly the chemical modifications based on exchange of side-chains in a polypeptide chain and cutting the selected amino acids, and secondly, conformation modifications leading to forcing the set of protein torsion angles φ_i, ψ_i, ω_i and χ_i in an amino-acid sequence. The computer model presented in the paper is capable to simulate those processes.

6. Determining the Free Conformation by Dynamic Programming

Basing on Fig. 1 we assume, that the first amino-acid residue in the peptide chain appearing from the ribosome has a number 0, the next is 1 and so on.

Let we denote the potential energy function [6] of the amino-acid residue #0 by E_0.

We assume in first approximation, that the potential energy is a function of torsion angles $\varphi_0, \psi_0, \omega_0$ and χ_0. The shape of the backbone of the polypeptide chain depends only from the pairs of the φ and ψ angles, because ω is usually equals 180° and χ is void.

In consequence, the energy of the first residue of the peptide chain can be expressed as:

$$E_0 = E_0(\varphi_0, \psi_0). \tag{2}$$

We will quantize the E_0 for $\varphi_0 = \{\varphi_0^0, \varphi_0^1, \ldots \varphi_0^K\}$ and $\psi_0 = \{\psi_0^0, \psi_0^1, \ldots \psi_0^L\}$. Thus we have:

$$E_0^{k,l} = E_0^{k,l}\left(\varphi_0^k, \psi_0^l\right).$$

The grid is not fully filled because of Ramachandran restrictions.

When the residue of next amino acid (#1) appears, the minimal energy for bonded residues of amino acids #0 and #1 has a form:

$$E_1^{k,l}\left(\varphi_1^k, \psi_1^l\right) := \min_{k,l}\left(E_0^{k,l} \Rightarrow E_1^{k,l}\right). \tag{3}$$

where symbol $\Rightarrow$ means to compute total energy in point $E_1^{k,l}$ reached from points $E_0^{k,l}$ $(k = 1, 2, \ldots K, l = 1, 2, \ldots L)$, and symbol := means to compute right side expression (3) and assign a result to the left.

When the third residue appears (#2), the minimal energy for three bonded residues of amino acids (in all accessible points on a grid) has a form:

$$E_2^{k,l}\left(\varphi_2^k,\psi_2^l\right):=\min_{k,l}\left[\min_{k,l}\left(E_0^{k,l}\Rightarrow E_1^{k,l}\right)\Rightarrow E_2^{k,l}\left(\varphi_2^k,\psi_2^l\right)\right]. \tag{4}$$

When we will reach the last residue (#n), we have:

$$E_n^{k,l}\left(\varphi_n^k,\psi_n^l\right):=\min_{k,l}\left[\min_{k,l}\left(E_{n-2}^{k,l}\Rightarrow E_{n-1}^{k,l}\right)\Rightarrow E_n^{k,l}\left(\varphi_n^k,\psi_n^l\right)\right]. \tag{5}$$

This is the key point of procedure, for $\min_{k,l} E_n^{k,l}\left(\varphi_n^k,\psi_n^l\right)$ in (5), when we recall to the E_0, we can find the optimal "trajectory" (minimal energy) for set of pairs (φ_i,ψ_i) in a form:

$$(\varphi_n,\psi_n)_{\text{opt}}\rightarrow(\varphi_{n-1},\psi_{n-1})_{\text{opt}}\rightarrow\ \ldots\ldots\ \rightarrow(\varphi_0,\psi_0)_{\text{opt}}, \tag{6}$$

and the conformation (1) of the backbone of the peptide chain is determined.

The procedure is illustrated in Fig. 6.

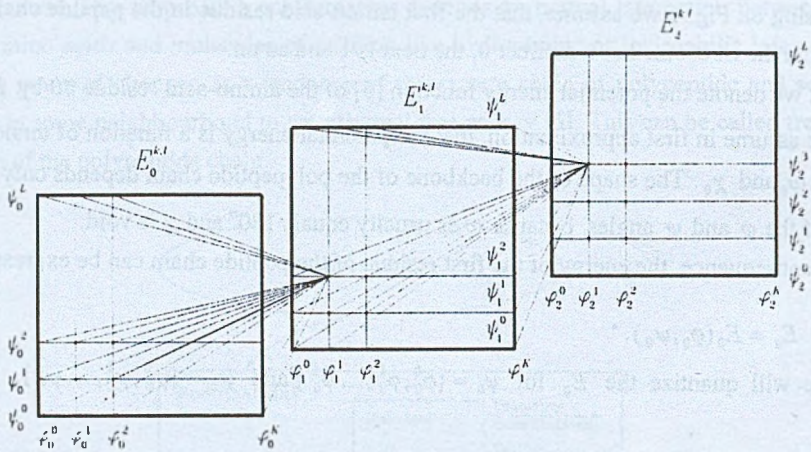

Fig. 6. Conformation determined by dynamic programming (e.g. steps #0 to #2)
Rys. 6. Wyznaczanie konformacji metodą programowania dynamicznego (przykładowo kroki #0 do #2)

6.1. Hydrogen and Disulfide Bonds

During the procedure of determining of the trajectory (6), the condition of existing hydrogen or disulfide bond have to be checked. When the condition is fulfilled, the sub-backbone is fixed, and the minimization is continuing, but the optimal part of the sub-backbone stays unchanged.

7. Message-Passing Interface in the Model Construction

Completion of the system for simulation of dynamics of the translation and conformation polypeptide chain processes, requires adoption of existing stand alone numerical programs and writing a few new modules for system integration [11, 15].

7.1. Software Tools

In the implemented system we have six main modules:

Control_PC - Control program in the integrated network maintaining the main menu in the workstation, organizing operation on the input/output data, and responsible for communication between operating systems (Visual C++).

Programs_C - Programs for numerical computing. Programs_C modules reside on PC workstation (Visual C++).

Control_Comp - Program working on a Unix platform organizing computing process by division the data set into subsets, and ordering to the subsets separate processes. This function is fulfilled in the MPI (Message-Passing Interface) environment [4].

Server_Comp - Program for listening the demands of client in a client-server communication model (Sun).

Programs_CC and **Programs_f77** - Batch programs for numerical computing (Sun supercomputer).

In the computer network applications, the standard model of data exchange is a client-server communication model. Server is a process waiting for requirement to connect from process called client to fulfill the determined tasks. In integrated network model the selected communication protocol is TCP/IP. Login to the supercomputer can be performed from arbitrary computer working in the Internet. The client workstation and computational server working in the network, communicate with each other using the package of subroutines that provide access to TCP/IP i.e. sockets interface.

7.2. Parallel Computations

In the multiprocessor system, the queues of messages resulting from processes' activity are created and supervisor process gathers the waiting results together. The Solaris 7 Operating Environment is very well adapted to the tasks of this kind. In our case, the massive computations are divided between eleven processors and indicate last processor (twelve) as a master gathering results. Each parallel process is ordered to the physical processor and the processes are mutually independent. Also, all variables in the processes are unique in each module. Data are gathered on a disk, and can be simultaneously read by processes. The par-

allel process of computing in the integrated network model is coded using MPI by a program module Control_Comp [4].

8. Conclusions

The paper presents a selected paradigms for modeling composite processes in protein synthesis: translation and post-translational conformations.

Presented integrated network model is framework for computer simulations fully using network computer system's features, where depending on the task division between resources of a system, it is possible to build friendly user interfaces, and independently improve system performance. In implemented system, through modules integration and parallel operation, the time of the system interaction was considerably shortened. Integrity of the system is satisfied through the compatibility of the data structures, the control of the system from main graphic application, and automatic transfer of data between PC workstation and the supercomputer through the network. The user can get results which are basic for design, but it also is possible to access to intermediate results of computation. This is necessary for checking the correctness of actual algorithms as well as in future development of the system's capabilities.

REFERENCES

1. Berman H. M. et al.: The Protein Data Bank. Nucleic Acid Res., Vol. 28, No. 1, 2001, pp. 235-242.
2. Clote P. and Backofen R.: Computational Molecular Biology. John Wiley & Sons, Chichester 2000.
3. Green G. and Noler H. F.: Ribosomes and Translation. Annual Rev. Biochem., Vol. 66, 1997, pp. 679-716.
4. Gropp W., Lusk E., and Skjellum A.: Using MPI. The MIT Press, Cambridge/Massachusetts, London 1994.
5. Güntert P. and Wüthrich K.: Sampling of Conformation Space in Torsion Angle Dynamics Calculations. Comp. Phys. Commun., Vol. 138, 2001, pp. 155-169.
6. Levitt M., Hirshberg M., Sharon R., and Daggett V.: Potential Energy Function and Parameters for Simulations of the Molecular Dynamics of Proteins and Nucleic Acids in Solution. Comp. Physics Comm., Vol. 91, 1995, pp. 215-231.
7. Liebecq C. (Ed.): The White Book. Biochemical Nomenclature and Related Documents, IUPAC-IUBMB Joint Commission, 2nd Edition, Portland Press, London 1992.

8. Ponder J.: TINKER – Software Tools for Molecular Design, Dept. of Biochemistry & Molecular Biophysics. Washington University, School of Medicine, St. Louis, June 2001, http://dasher.wustl.edu/tinker
9. Sayle R.: RasMol, Molecular Graphics Visualization Tool. Biomolecular Structures Group, Glaxo Wellcome Research & Development, H. J. Bernstein v.2.7.1.1, Stevenage, Hartfordshire 1998. rasmol@bernstein-plus-sons.com
10. Stryer L.: Biochemistry. W. H. Freeman and Company, New York 1994.
11. Stevens W. R.: Programming of Network Application in a Unix System. WNT, Warszawa 1995 (in Polish).
12. Znamirowski L.: Modeling of the Informatic Process of Translation in Ribosomes. Archiwum Informatyki Teoretycznej i Stosowanej, Vol. 13, No. 2, 2001, pp. 175-193.
13. Znamirowski L. and Zukowska E.: Parallel Nano-modeling of Ribosomal Translation Processes. Proceedings of the ICSEE '2002, 2002 Western MultiConference, The Society for Computer Simulation Intern., San Antonio, January 27-31, 2002, pp. 111-116.
14. Klotho: Biochemical Compounds Declarative Database – Alphabetical Compound List. January 2001, http://www.ibc.wustl.edu/moirai/klotho/compound_list.html
15. MSDN, the Microsoft Developer Network. Microsoft Corp., 3xCD-ROM, Technical Documentation for Microsoft Visual Studio 6.0, Redmond 1998.

Recenzent: Prof. dr hab. inż. Tadeusz Czachórski

Wpłynęło do Redakcji 24 kwietnia 2002 r.

Streszczenie

W pracy przedstawiono strukturę modelu sieciowego pozwalającego na badania symulacyjne konformacji liniowych polimerów aminokwasów. Konformacja ciągu generowanych w reaktorze (rybosom) reszt aminokwasów tworzących polipeptyd wyznaczana jest metodą programowania dynamicznego [relacje (5) i (6)]. Jako stopnie swobody szkieletu powstającego łańcucha polipeptydowego przyjęto kąty torsyjne φ oraz ψ [7, 10, 12, 13]. Przedyskutowano zagadnienie modyfikacji potranslacyjnych w ogólności. W szczególności przedstawiono metodę wyznaczania struktury konformacyjnej stanowiącej formę natywną pierwszorzędowej struktury polipeptydu.

Implementacja modelu w sieci komputerowej zawiera dwa podstawowe elementy: stację roboczą spełniającą funkcję graficznego interfejsu użytkownika (GUI) i sterowania procesem

symulacji oraz superkomputer (Sun Enterprise 6500), który pełni rolę jednostki wykonującej masywne obliczenia numeryczne pod kontrolą stacji roboczej.