A Predictive Method for the Evaluation of Peptide Binding in Pocket 1 of HLA-DRB1 via Global Minimization of Energy Interactions

I.P. Androulakis[1], N. N. Nayak, M. G. Ierapetritou,
D.S. Monos[2] and C.A. Floudas [3]

Department of Chemical Engineering
Princeton University
Princeton, N.J. 08544-5263

Proteins (in press)

Keywords: Peptide docking, Global optimization, HLA-DRB1 class II protein

[1]Current Address: Corporate Research Science Laboratories, Exxon Research & Engineering Co., Annandale, NJ 08801.

[2]Department of Pediatrics, University of Pennsylvania, The Children's Hospital of Philadelphia, Philadelphia, PA 19104.

[3]Author to whom all correspondence should be addressed; Tel: (609) 258-4595; Fax:(609) 258-0211.

Abstract

Human leucocyte antigens (HLA) or histocompatibility molecules are glycoproteins that play a pivotal role in the development of an effective immune response. An important function of the HLA molecules is the ability to bind and present antigen peptides to T cells. Presently there is no comprehensive way of predicting and energetically evaluating peptide binding on HLA molecules. To quantitatively determine the binding specificity of a class II HLA molecule interacting with peptides, a novel decomposition approach based on deterministic global optimization is proposed that takes advantage of the topography of HLA binding grove, and examines the interactions of the bound peptide with the five different pockets. In particular, the main focus of this paper is the study of pocket 1 of HLA DRB1*0101 allele. The determination of the minimum energy conformation is based on the ECEPP/3 potential energy model that describes the energetics of the atomic interactions. The minimization of the total potential energy is formulated on the set of peptide dihedral angles, euler angles, and translation variables to describe the relative position. The deterministic global optimization algorithm, α**BB**, which has been shown to be ϵ-convergent to the global minimum potential energy through the solution of a series of nonlinear convex optimization problems, is utilized. The PACK conformational energy model that utilizes the ECEPP/3 model but also allows the consideration of protein chain interactions is interfaced with α**BB**. MSEED, a program used to calculate the solvation contribution via the area accessible to the solvent, is also interfaced with α**BB**. Results are presented for the entire array of naturally occurring amino acids binding to Pocket 1 of the HLA DRB1*0101 allele and very good agreement with experimental binding assays is obtained.

1 Introduction

The docking problem has received a lot of attention in the open literature. The presented methods can be classified as shape-based methods that are based on molecular surface representation and energy-based methods that optimize interaction energy in order to determine good dockings. Shape-based methods have the advantage of being less computationally intensive since the number of possible different binding modes can be greatly reduced by using a simplified model for the shapes of the receptor and binder. Based on this idea, are the works of Lee and Richard [22], Connolly [8], Bacon and Moult [5], Jiang and Kim [19], Kuntz and co-workers [17].

Energy-based methods on the other hand, represent a more precise way of determing good dockings but they are more computationally demanding. Due to this fact most of the proposed approaches are based on Monte Carlo simulation and Simulated Annealing such as the works of Goodsell and Olson [12], Hart and Read [16] and Calfisch [7]. Rosenfeld et al, [35], present a peptide binding study based on random selection and minimization among potential peptide structures. More recently, dynamic programming optimization, [21], is used for optimizing the overall free energy based on a fragment assembly algorithm and molecular dynamics simulation is also utilized for studying the binding afinity of the HLA-B*2705 protein,[34]. All the proposed approaches identify the importance of accurate prediction which leads to the need of establishing efficient and systematic ways of predicting the global energetically most favorable docking mode.

Histocompatibility molecules or human leucocyte antigens (HLA) are cell surface molecules that form complexes with self and non-self peptides. The HLA-peptide complex is recognized by the T-cell receptor and initiates antigen specific immune responses. HLA molecules are very polymorphic and each of them may interact with large number of peptides. Both characteristics, their polymorphism and their binding promiscuity serve the basic function of presenting a wide range of antigens to the immune system. Appropriately presented antigens induce either tolerance or an active immune response. HLA molecules are classified as class I and class II. This distinction relates to the mode of interaction with peptides as well as to their function and distribution in the tissues. The study of the differences between the two classes as well as a trial to predict the structure of class II MHC based on the structure of class I is the subject of a recent publication of [20].

In this study the presented results involve the class II molecule HLA-DRB1*0101. A deterministic global optimization approach is proposed for determining the conformation of the binding complex with the global minimum of interaction energy. This approach is applied on evaluating the total potential energy of the entire array of amino acids interacting with pocket 1 of HLA-DRB1*0101. A detailed description of the HLA-DRB1 allele is presented in the next section. Following this introduction the problem definition and formulation are presented in sections

2 and 3, respectively, whereas section 4 presents the proposed global optimization approach. The results for all naturally occurred amino acids binding in pocket 1 of HLA-DR1 are given in section 5 and discussed in comparison with provided experimental data.

2 Problem Definition and Proposed Approach

2.1 HLA-DR1 molecule

The binding of an influenza virus peptide to the MHC protein HLA-DR1 [38] is illustrated in Figure 1. The HLA-DR1 molecule is in white, while the bound peptide is in grey with the different locations of the major protein pockets defined by the residues shown in different shades. Notice that the peptide obtains an extended conformation in the binding groove on this comlex. (All pictures were created in the molecular graphics program GRASP, [31]).

Histocompatibility proteins are organized into two major classes. Polymorphic residues in both class I and II proteins are clustered in the peptide-binding region and are responsible for the different peptide specificities. The major distinctive features of the class I and class II locci are: (i) allograft rejection properties, (ii) relative tissue distributions, and (iii) differing chemical compositions, [32]. Class I proteins generally bind fragments that range from eight to ten residues in length. Additionally, the protein pockets in this class show allele defined tendencies to bind particular amino acid side chains, or to bind unspecifically. In contrast, class II molecules bind much longer fragments and it has proven difficult to define the binding tendencies of the various pockets, [38]. The different binding properties of these two classes are conjectured to be a result of the more open structure of class II peptides. This allows longer peptides to be situated in the MHC binding groove, [32].

In humans, the major histocompatibility proteins are referred to as Human Leukocyte Antigens (HLA). These proteins are then further distinguished by serological studies into HLA-x. Class II locci fall under the classification of HLA-D. The first alleles discovered at this locus were termed HLA-DR (D-related). Subsequently, different class II locci were discovered and now define the categories of HLA-DQ and HLA-DP, [32]. The HLA-DR1 protein will be the major concern of this paper.

The peptide binding site of HLA-DR1 molecule consists of an α chain (33-35 kD), and a β chain (26-28 kD) consisting of 366 amino acid residues. The $\beta 1$ chain of the HLA-DR1 locus is highly variable, while all other regions tend to be relatively invariant.

Crystallographic studies [38] have shown that peptide binding is accommodated by five polymorphic pockets on the surface of the HLA-DR1 molecule. Each

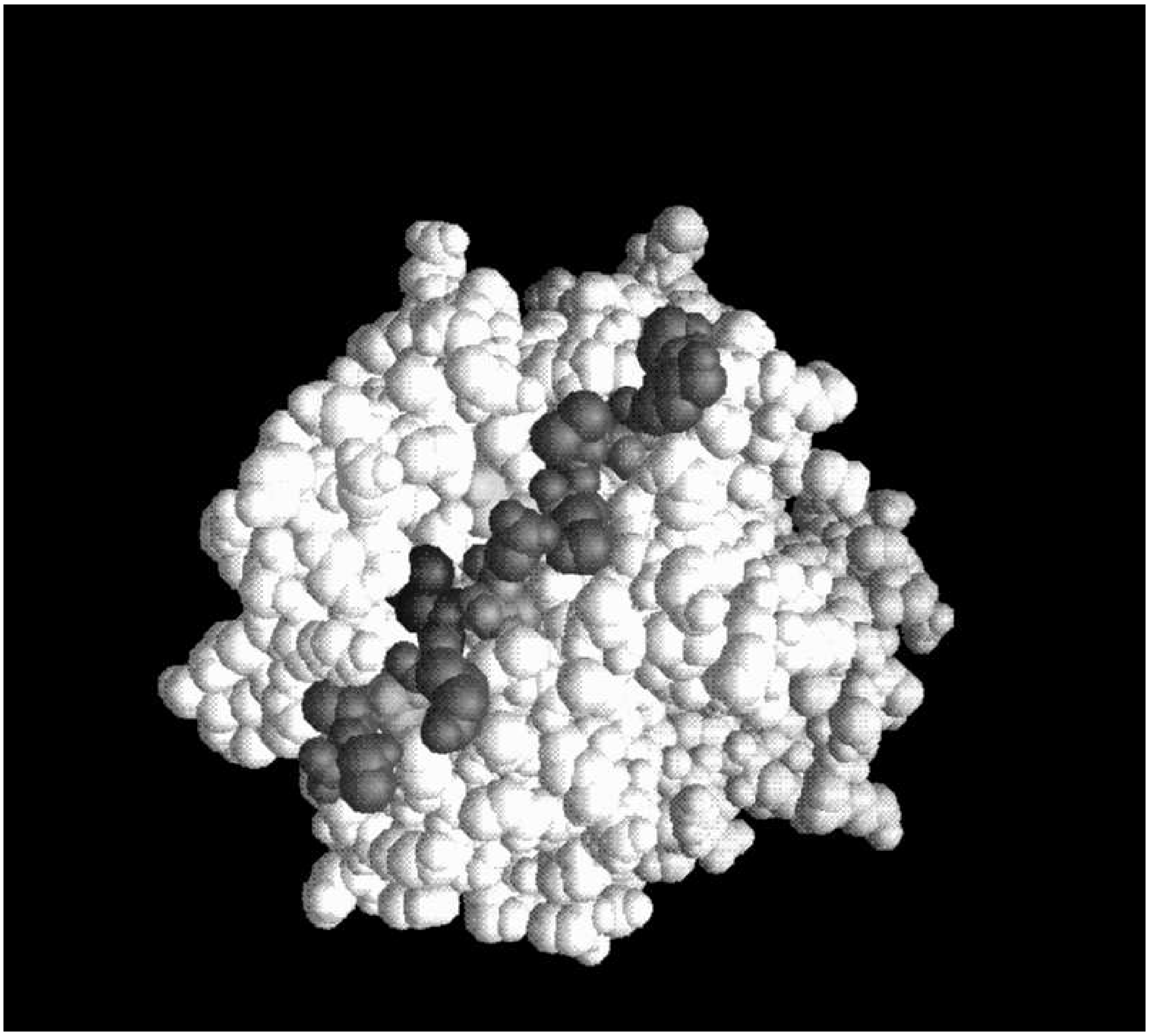

Figure 1: **HLA-DR1 bound to an influenza virus peptide**

of these pockets can accommodate a single amino acid residue when a particular peptide is bound, [38]. Accordingly, these pockets play a major role in determining the peptide specificity of class II molecules. Both pocket 1 and pocket 4 have been implicated as playing vital roles in peptide binding and subsequent recognition by T cells, [14, 10].

Pocket 1 is the largest and deepest pocket of the HLA-DR1 molecule. The area of contact for potential binders has been estimated at 200 $\mathring{A}^2$, [38]. The pocket has been implicated as being an "anchor" peptide. It has been postulated that the residues that bind to the other four pockets are mainly determined by which residue in the binding peptide attaches to this pocket, [14]. Pocket 1 consists of hydrophobic residues including several phenylalanine groups. This accounts for the preference of this pocket to accommodate hydrophobic residues, such as tyrosine and phenylalanine, [38]. The large size of this pocket makes it the most solvent accessible of the five pockets.

Pocket 4 is a relatively large pocket that is much shallower than pocket 1. The pocket consists of predominately hydrophobic amino acids, except for a positively charged arginine group. This accounts for this pocket's tendency to bind residues that have large, aliphatic side chains, or negatively charged side chains. Thus, residues such as glutamate and aspartate bind favorably to this pocket while positively charged groups such as lysine, or arginine are repelled, [38]. Pocket 4 has been shown to play an important role in the recognition of the bound peptide by T cells, [10]. Pocket 4, in addition to pockets 6, 7, and 9, has been shown to be 90% inaccessible to solvent, [38].

The other three pockets (6, 7, and 9) are considered to affect to a lesser degree the determination of peptide binding. Pocket 6 is a shallow pocket that prefers smaller residues. The similarly shallow pocket 7 is nondiscerning in its binding activities, and only partially accommodates side chains. Pocket 9 binds aliphatic side chains due to its small, hydrophobic nature, [38].

A large array of diseases have been linked to HLA mutations by statistical association. One study has established a definite link between the HLA-DR1 locus and rheumatoid arthritis. It has been shown that a particular peptidic sequence between residues 67 and 71 on the β strand of the HLA molecule plays a significant role the determination of an individual's susceptibility to this disease, [32].

Mechanisms for increased disease susceptibility can be posed, but are much more difficult to verify. For example, it is not difficult to postulate that if a mutation weakens antigen binding, then the process of disease identification and eradication will be retarded. But quantitatively establishing binding potential in the laboratory is a difficult task. Thus, diseases have been associated with mutations to various locii within the MHC molecule, but precise mechanisms for disease manifestation have yet to be established. A sampling of the diseases associated with the HLA-DR1 locus is present in Table I. The HLA-DR locus

Table I: Mutations to the HLA-DR1 locus and Associated Diseases

Changes in Pocket 4 of HLA-DR1	**Associated Disease**	**HLA Locus**
β13: Phe $\rightarrow$ Arg β71: Arg $\rightarrow$ Ala	Multiple Sclerosis	DR1(1501)
β13: Phe $\rightarrow$ Arg β71: Arg $\rightarrow$ Ala	Leprosy	DR1(1501)
β13: Phe $\rightarrow$ His β71: Arg $\rightarrow$ Lys	Rheumatoid Arthritis	DR4(0401)
β13: Phe $\rightarrow$ His β70: Gln $\rightarrow$ Asp β71: Arg $\rightarrow$ Glu	Rheumatoid Arthritis	DR4(0402)
β13: Phe $\rightarrow$ His β74: Ala $\rightarrow$ Glu	Rheumatoid Arthritis	DR4(0403)
β13: Phe $\rightarrow$ His	Rheumatoid Arthritis	DR4(0404)

that coincides with each of these mutations is also presented. Note that the locus presented in this paper corresponds to DR1(0101). Specifics of the research performed to solidify these associations are given in [44, 15] for rheumatoid arthritis, in [43, 44] for multiple sclerosis, and in [45] for leprosy.

2.2 Proposed approach

The modeling and optimization studies of the interactions between the HLA-DR1 protein and a virus peptide are based on a novel decomposition scheme. As it has been described in the previous section, the binding specificity of the HLA-DR1 molecule is mainly determined by the binding characteristics of its five pockets which enables the investigation of each one separately. This paper will concentrate on the study of pocket 1. The key ideas in the proposed decomposition approach are: (i) to consider the binding at each pocket separately, (ii) to study the binding of each amino acid to each pocket by considering one at a time, and (iii) to create a rank ordered list of the binding amino acids for each pocket, based on an energetic criterion that reflects the binding specificity.

The proposed decomposition approach consists of the following stages:

Stage 1 In this stage the pocket of the HLA-DR1 protein is represented by a number of residues as described in detail in section 3.2. The work of Stern et al. [38] provides information on the constituent amino acids of each pocket of HLA-DR1 protein. Furthermore, this work provides the cartesian coordinates of the atoms that participate in each amino acid of the HLA-DR1 protein.

Stage 2 For the specific pocket interacting with each naturally occurring amino acid a mathematical model is formulated that represents all the energetic atom-to-atom interactions. These interactions are classified as (i) inter-interactions between the atoms of the residues that define the pocket of HLA-DR1 protein and the atoms of the considered naturally occurring amino acid, and (ii) intra-interactions between the atoms of the considered naturally occurring amino acid. These interactions consist of electrostatic, nonbonded, hydrogen-bonding, torsional, and loop-closing components. In addition, solvation energy is also considered based on solvent accessible areas. The detailed mathematical model and potential functions used are described in sections 3.1 and 3.3.

Stage 3 Having a mathematical model which accounts for all the inter and intra interactions of the specific pocket and the considered naturally occurred amino acid, in this stage we formulate the global optimization problem which minimizes the total potential energy as it is explained in detail in section 3.4.

Stage 4 A deterministic global optimization method, α**BB**, [3, 4, 24, 25] is adopted at this stage for the solution of the resulting nonconvex mathematical model of stage 3. This stage requires the connection of the α**BB** global optimization method with the conformation energy program **PACK**, [37], which utilizes ECEPP/3, [36], and the program **MSEED**, [33], that supplies the solvation contribution as described in detail in section 4.

Stage 5 In this stage, we introduce an energetic-based criterion that allows for the comparison of the binding between a given pocket and each naturally occurring amino acid. This measure, which is denoted as ΔE, corresponds to the difference of (i) the global minimum total potential energy that is obtained in stage 4 and which is indicated as E_{Total}, and (ii) the global minimum potential energy of the considered naturally occurring amino acid when it is far away from the pocket and which is denoted by E^o_{Res}:

$$\Delta E \;=\; E_{Total} - E^o_{Res} \tag{1}$$

Note here that the energies E_{Total} and E^o_{Res} include the consideration of the solvation energy as it will be discussed in more detail in section 3.3. This criterion represents a measure of the binding affinity of each amino acid to the given pocket, in the sense that it quantifies the tendency of an amino acid to bind with the pocket of the HLA-DR1 molecule. The amino acid that exhibits the least ΔE corresponds to the one with the best possible binding to that pocket of the HLA-DR1 protein.

Stage 6 In this stage we repeat the previous stages for each naturally occurring amino acid and hence create a rank ordered list for the binding of each of them to the specific pocket. The detailed results are shown in section 5.

3 Mathematical Modeling

3.1 Protein representation

The geometry of a protein can be fully described by defining the relative cartesian coordinates of each atom. Instead of specifying the coordinate vector for all atoms in a protein, one can specify all bond lengths, covalent bond angles and dihedral angles. Under biological conditions, the bond lengths and bond angles are fairly rigid and thus can be assumed to be fixed at their equilibrium values. Under this assumption, the dihedral angles determine the geometric shape of the folded protein. The names of the dihedral angles of a folded protein chain follow a standard nomenclature as shown in Figure 2.

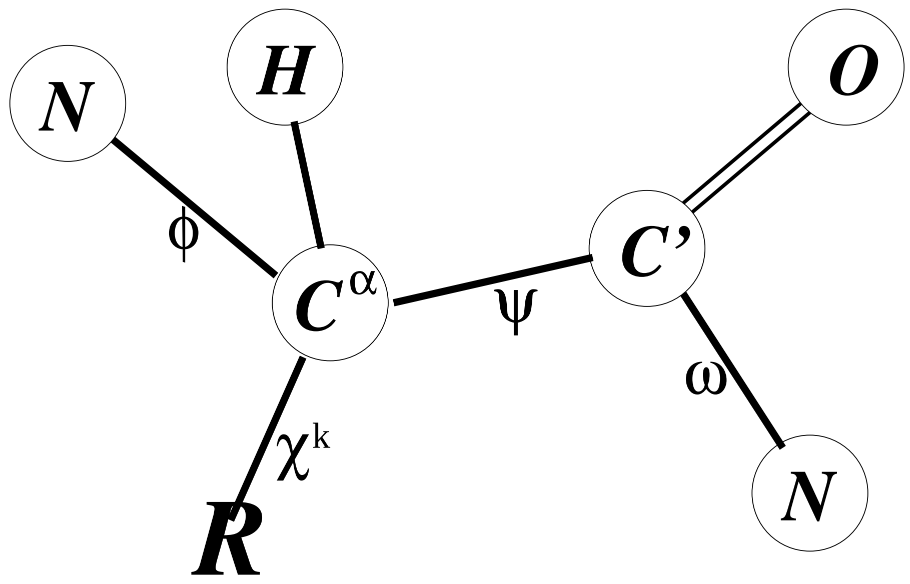

Figure 2: **Dihedral angles of a standard amino acid**

If more than one polypeptide is involved then the relative orientations, and locations of these different chains must be defined. This can most easily be accomplished by defining a translation vector and a rotation matrix. The translation is achieved through the cartesian coordinates of the initial nitrogen atom of each independent chain. The Euler angles specify the rotations necessary to orient a particular polypeptide and are defined as the angles between the coordinate axes defined by the initial hydrogen, nitrogen, and alpha carbon of each residue. The detailed determination of the euler angles is given in the Appendix A.

3.2 Pocket definition

The relative energies of minimization of each of the five protein pockets is mainly determined by the residues that constitute these pockets. A Program for Pocket Definition, denoted as PPD, constructs these pockets through the selection of all residues that are within a radius $\mathcal{R}$ of the atoms of the crystallographic binder. A range of values for $\mathcal{R}$ have been evaluated in an attempt to discover a radius that realistically represents the pocket, while limiting the number of residues necessary to define the pocket. The information required by the user is provided in a file containing the coordinates of the HLA-DR1 molecule and a file with the coordinates of the influenza virus binding peptide, as well as the value of radius $\mathcal{R}$. Each of the five pockets of the HLA-DR1 molecule were run through PPD for three different radius lengths ($\mathcal{R}$ = 4.0, 4.5, 5.0 Å). The program filters through the α and β strands of the molecule and returns two output files one having a list of all atoms within a radius $\mathcal{R}$, as well as the exact distances and a PDB file with the coordinates of all residues hat define a given protein pocket for the specific radius. Table II presents the residues defining each of the protein pockets for the various radii considered. There are several important observations that should be made. First, there is the intuitive trend of the pockets becoming more complex with increased radius. Second, note that pocket 1 is composed of a significantly larger number of residues than any of the other pockets. Finally, note that in some cases an increase in radius does not cause the inclusion of additional amino acids. An example of this is pocket 7, where the pocket is identically defined across the entire range of radii.

The resulting PDB file is then translated to the internal coordinate system by a program denoted as ARAS (Amino acid Residue Angle Solver). The output file obtained by this program is the one required by the conformational energy program **PACK** to evaluate the potential energy as is described in detail in the next section.

3.3 Potential Energy model

Molecular dynamics calculations employ an empirical derived set of potential energy contributions for approximating the force field of the protein system. These energy functions are based upon specific types of interactions instead of being associated with a particular molecule. The parameters for these correlations have been determined to provide the best possible agreement with experimental data. Many different parameterizations have been proposed for approximating the *force field* in protein folding calculations. Some of the most popular ones are: ECEPP [26], MM2 [1], ECEPP/2 [30], CHARMM [6], DISCOVER [9], AMBER [41, 42], GROMOS [39], ENCAD [23], MM3 [2], and ECEPP/3

Table II: PPD Pocket Compositions for $\mathcal{R}$ = 4.0-5.0 $\AA$

Pocket	$\mathcal{R}$ =	**4.0**		**4.5**		**5.0**
1	ileα31	pheα32	ileα31	pheα32	ileα31	pheα32
	trpα43	alaα52	trpα43	alaα52	trpα43	alaα52
	serα53	pheα54	serα53	pheα54	serα53	pheα54
	valβ85	glyβ86	valβ85	glyβ86	valβ85	glyβ86
	pheβ89		pheβ89	pheα24	pheβ89	pheα24
			asnβ82		asnβ82	gluα55
					thrβ90	
4	glnα09	asnα62	glnα09	asnα62	glnα09	asnα62
	pheβ13	glnβ70	pheβ13	glnβ70	pheβ13	glnβ70
	argβ71	alaβ74	argβ71	alaβ74	argβ71	alaβ74
	tyrβ78		tyrβ78	gluα11	tyrβ78	gluα11
			leuβ26		leuβ26	
6	gluα11	asnα62	gluα11	asnα62	gluα11	asnα62
	valα65	aspα66	valα65	aspα66	valα65	aspα66
	leuβ11		leuβ11		leuβ11	pheβ13
					argβ71	
7	valα65	asnα69	valα65	asnα69	valα65	asnα69
	gluβ28	tyrβ47	gluβ28	tyrβ47	gluβ28	tyrβ47
	trpβ61	leuβ67	trpβ61	leuβ67	trpβ61	leuβ67
	argβ71		argβ71		argβ71	
9	ileα72	asnα69	ileα72	asnα69	ileα72	asnα69
	metα73	argα76	metα73	argα76	metα73	argα76
	trpβ09	aspβ57	trpβ09	aspβ57	trpβ09	aspβ57
	tyrβ60		tyrβ60	trpβ61	tyrβ60	trpβ61
					leuα70	

[29]. In this work the ECEPP/3 [29] detailed potential model is utilized. In this potential model, it is assumed that the covalent bond lengths and angles are fixed at their equilibrium values and the conformational energy is treated as the sum of electrostatic, nonbonded, hydrogen bonding, torsional and cystine contributions.

The potential function of ECEPP/3 includes the following terms :

$$
\begin{aligned}
E &= \sum_{(i,j)\in\mathcal{ES}} 332.0\frac{q_i q_j}{Dr_{ij}} \quad \text{(Electrostatic)} \\
&+ \sum_{(i,j)\in\mathcal{NB}} F\frac{A}{r_{ij}^{12}} - \frac{C}{r_{ij}^{6}} \quad \text{(Nonbonded)} \\
&+ \sum_{(hx)\in\mathcal{HX}} F\frac{A'}{r_{hx}^{12}} - \frac{B}{r_{hx}^{10}} \quad \text{(Hydrogen bonding)} \\
&+ \sum_{k\in\mathcal{TOR}} (\frac{E_o}{2})(1 \pm \cos n_k\theta_k) \quad \text{(Torsional)} \\
&+ \sum_{l\in\mathcal{LOOP}} B_L \sum_{i_l=1}^{i_l=3} (r_{i_l} - r_{i_o})^2 \quad \text{(Cystine Loop–Closing)} \\
&+ \sum_{l\in\mathcal{LOOP}} A_L (r_{4_l} - r_{4_o})^2 \quad \text{(Cystine Torsional)}
\end{aligned}
$$

In addition, the solvation energy is also considered through the utilization of the program MSEED [33] which supplies solvent accessible areas. Once these areas have been calculated, the following formula can be utilized to define the solvation potential:

$$E_{SOL} = \sum_{i=1}^{n} \sigma_{k(i)} \mathcal{A}_i \tag{2}$$

where n equals the total number of atoms in the molecule, $\sigma_{k(i)}$ is a coefficient dependent upon the atom type, and $\mathcal{A}_i$ is the solvent accessible area of the i^{th} atom. The σ coefficients were determined by the research performed in [40].

The solvent accessible area is determined by rolling a spherical test probe over the surface of the molecule (see Figure 3). The areas of direct contact between the molecule and the probe define the accessible surface. Additionally, the area of the bottom most part of the probe traces the surface in inaccessible cavities of the protein. The probe radius is equivalent to the van der Waals radius of a water molecule, which is equivalent to 1.4 Å. These empirical solvent accessible surface areas are calculated by the program MSEED. This program utilizes Connolly's analytical algorithm, which is described in [33]. Note that E_{SOL} is only added to this overall potential at local minima, and hence is not explicitly stated in the above equation. This is done because the parameters of the JRF set used in [40] were derived based on a set of tetrapeptide conformations that correspond

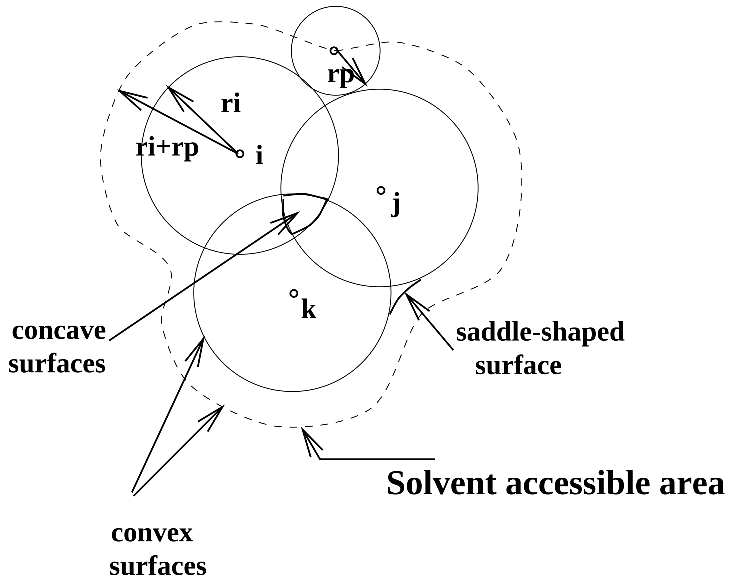

Figure 3: **Determination of solvent accessible area**

to local minima of the ECEPP potential energy, [38]. The total energy E_{Total} is then defined as:

$$E_{Total} = E + E_{SOL}$$

3.4 Problem Formulation

As it was described in section 3.1 a particular amino acid chain could be defined by a translation vector, a rotation matrix, and the corresponding set of dihedral angles. The translation vector will be defined as the coordinates of the nitrogen atom on the first residue of a chain, while the rotation matrix will be defined by the euler angles. Since the pocket is considered to be rigid, the only variables will be the amino coordinates, euler angles, and dihedral angles of the amino acid binder.

Let $k = 1, \ldots, K$, where K is the total number of side chain angles of the amino acid residue that attempts to bind the pocket. Then, the set of variable dihedral angles would include the backbone angles (ϕ, ψ, and ω), and the side chain angles (χ^k). The cartesian coordinates of the amino translation vector will be defined by the variables N_x, N_y, and N_z. Similarly, the cartesian coordinates of the backbone carboxyl carbon are represented by C'_x, C'_y, and C'_z. Finally, the euler angles will be represented by ε_1, ε_2, and ε_3. Utilizing the above definitions the potential energy minimization problem can be formulated as follows:

$$\min \quad E\left(\phi, \psi, \omega, \chi^k, N_x, N_y, N_z, \varepsilon_1, \varepsilon_2, \varepsilon_3\right) \tag{3}$$

$$
\begin{aligned}
\text{s.t.}\quad -\pi &\le \phi \le \pi && (4)\\
-\pi &\le \psi \le \pi && (5)\\
-\pi &\le \omega \le \pi && (6)\\
-\pi &\le \chi^k \le \pi, \quad k=1,\ldots,K && (7)\\
-\pi &\le \varepsilon_1 \le \pi && (8)\\
-\pi &\le \varepsilon_2 \le \pi && (9)\\
-\pi &\le \varepsilon_3 \le \pi && (10)\\
N_x^l &\le N_x \le N_x^u && (11)\\
N_y^l &\le N_y \le N_y^u && (12)\\
N_z^l &\le N_z \le N_z^u && (13)\\
C_x'^l &\le C_x'\left(\phi,\psi,\omega,\chi^k,N_x,N_y,N_z,\varepsilon_1,\varepsilon_2,\varepsilon_3\right) \le C_x'^u && (14)\\
C_y'^l &\le C_y'\left(\phi,\psi,\omega,\chi^k,N_x,N_y,N_z,\varepsilon_1,\varepsilon_2,\varepsilon_3\right) \le C_y'^u && (15)\\
C_z'^l &\le C_z'\left(\phi,\psi,\omega,\chi^k,N_x,N_y,N_z,\varepsilon_1,\varepsilon_2,\varepsilon_3\right) \le C_z'^u && (16)
\end{aligned}
$$

Note that the superscripts u and l denote upper and lower bounds, respectively, for the cartesian coordinates of both the amino nitrogen and the carboxyl carbon. In addition to the constraints on the amino nitrogen, note that in the above formulation there are additional constraints on the carboxyl carbon. It has been assumed due to the decomposition employed that enables the consideration of each pocket separately, that the conformational movements of the binding peptide are only constrained by the locations of these two atoms. In the original problem though, the binding residue is part of a longer antigen peptide. The rest of the binding peptide is assumed to bind normally so even if the binding residue is changed these backbone atoms will be relatively confined to their initial positions due to their peptidic linkages. Although the constraints of amino nitrogen can be directly considered in the above formulation since they correspond to problem variables, the C' coordinates are not explicit variables and consequently they must be defined as a function of the other variables (see Appendix B). Note that E is a nonconvex function involving numerous local minima that correspond to metastable states of the specific amino acid binding to the pocket 1. A single global minimum defines the energetically most favorable peptide conformation. In establishing a ranked-order list of binding peptides, one needs to identify rigorously the best conformation of (i) the binding residue far from the pocket and (ii) the complex of Pocket 1 with the binding residue. Consequently, there is a need for a method that can guarantee convergence to the global minimum potential energy conformation and which is capable of solving large scale constrained optimization problems. In this paper, the global optimization approach αBB [3, 4, 25], described in more detail in the next section, has been extended to peptide systems interacting with realistic atomistic potential energy models (e.g,, ECEPP/3 [36]), which include solvation contributions via surface accessible area as the solvent method (e.g., MSEED [33]).

4 Deterministic Global Optimization

4.1 Global Optimization Approach

The global optimization scheme αBB [3, 4, 25] is a deterministic branch and bound algorithm for locating the global optimum based on the construction of converging lower and upper bounds. Upper bounds can be simply obtained by minimizing E using local methods. Lower bounds can be evaluated by constructing the convex underestimator, L, of the original function E and evaluating the single global minimum of the resulting convex problem.

A convex lower bounding function L of potential energy function E can be defined by augmenting E using the ideas of the approach introduced in [25]:

$$
\begin{aligned}
L \;=\; E + \alpha \{ \quad & \left(\phi^l - \phi\right)\left(\phi^u - \phi\right) + \\
& \left(\psi^l - \psi\right)\left(\psi^u - \psi\right) + \\
& \left(\omega^l - \omega\right)\left(\omega^u - \omega\right) + \\
& \sum_{k=1}^{K} \left(\chi^{k,l} - \chi^k\right)\left(\chi^{k,u} - \chi^k\right) + \\
& \left(N_x^l - N_x\right)\left(N_x^u - N_x\right) + \\
& \left(N_y^l - N_y\right)\left(N_y^u - N_y\right) + \\
& \left(N_z^l - N_z\right)\left(N_z^u - N_z\right) + \\
& \left(\varepsilon_1^l - \varepsilon_1\right)\left(\varepsilon_1^u - \varepsilon_1\right) + \\
& \left(\varepsilon_2^l - \varepsilon_2\right)\left(\varepsilon_2^u - \varepsilon_2\right) + \\
& \left(\varepsilon_3^l - \varepsilon_3\right)\left(\varepsilon_3^u - \varepsilon_3\right) \}
\end{aligned}
$$

where α is a nonnegative parameter which must be greater or equal to the negative one half of the minimum eigenvalue of the hessian of E over the rectangular under consideration described by the lower and upper bounds of the involved variables defined by the superscripts l and u, respectively. The following properties of function L will enable the construction of a global optimization algorithm. These properties whose proof is given in [25] demonstrate that:

(i) L is always a valid underestimator of E;

(ii) L matches E at all corner points of the box constraints;

(iii) L is convex;

(iv) the maximum separation between L and E is bounded and proportional to α and to square of the diagonal of the current box constraints. This property ensures that an ϵ_f feasibility and ϵ_c convergence tolerances can be reached for a finite size partition element;

(v) the underestimators L constructed over supersets of the current set are always less tight than the underestimator constructed over the current box constraints for every point within the current box constraints.

These bounds are successively refined by iteratively partitioning the initial feasible region into smaller ones. The feasible region partition is achieved by subdivision of a rectangle into two subrectangles by halving along the longest side of the initial rectangle (bisection). At each iteration the lower bound would be the minimum over all the minima in every subrectangle composing the original domain. Therefore, a simple way to produce a nondecreasing sequence of lower bounds is to halve only the subrectangle responsible for the infimum of the minima. A nonincreasing sequence of upper bounds can also be produced by solving locally the nonconvex problem and selecting the minimum over the previously recorded upper bounds. Based on this procedure a fathoming step of the algorithm leads to no further consideration of a subrectangle where the minimum is greater than the current upper bound. Convergence proof to an ϵ-global solution in finite steps is given in [25].

4.2 Algorithmic Description

The proposed approach for the determination of the global minimum of E that corresponds to the peptide conformation binding to the pocket 1 of HLA-DR1 as posed in section 2, necessitates the development of an optimization interface that combines the global optimization program αBB, the conformational energy program PACK which utilizes ECEPP/3, the solvation program MSEED, and the local optimization solver NPSOL. Additional program files serve to link these programs. A schematic diagram of the interface between the used programs is shown in Figure 4.

The following steps are required in the calculation of the global minimum:

(1) The local solver (NPSOL [11]) obtains a local minimum of the potential function supplied by PACK in a domain (rectangle) defined by the original lower and upper bounds of the variables (bounds are supplied by αBB). PACK determines the energies of individual chains through repeated calls to ECEPP/3.

(2) The solvation energy at this local minimum is calculated by MSEED. This hydration energy is added to the potential function to yield E_{Total}, which will serve as an upper bound on the global minimum solution in the current rectangle.

(3) The current best upper bound is updated to be the minimum of those thus far stored.

(4) The current rectangle is partitioned by bisection along the longest side.

(5) The convex function L is minimized in each rectangle and the solvation energy is added at the minimum. If a solution is greater than the best upper bound it will be eliminated, otherwise it will be kept on the stack.

(6) The rectangle with the current minimum solution for L is selected for further partitioning.

(7) If the best upper and lower bounds are within ϵ the program will terminate, otherwise it will proceed to Step 1.

It should be noted that ECEPP/3 calculates the potential energy function for a polypeptide chain, [36], while PACK is a program that inputs multiple chain data and makes the appropriate calls to ECEPP/3 for calculation of the interaction energies, [37]. Special type penalty functions had to be added to the upper bound function, E, in order to implement the previously discussed constraints on C', [(15), (16), (17)]. The modified objective function takes then the following form:

$$\begin{aligned} E' \; = \; E + \beta\{ \;\; & \langle C_x'^l - C_x' \rangle + \langle C_x' - C_x'^u \rangle + \\ & \langle C_y'^l - C_y' \rangle + \langle C_y' - C_y'^u \rangle + \\ & \langle C_z'^l - C_z' \rangle + \langle C_z' - C_z'^u \rangle \; \} \end{aligned}$$

The $\langle\rangle$ function is defined as follows: $\langle \mathcal{A} \rangle$ equals $\mathcal{A}$ if $\mathcal{A}$ is greater than zero, otherwise $\langle \mathcal{A} \rangle$ equals zero. Thus, as long as the coordinates are within the defined bounds the objective function will not be modified. Yet, if a particular coordinate falls outside of the bounds, the function will be increased by the value of the transgression multiplied by the arbitrarily large constant β.

Since the pocket is assumed rigid, the optimization variables are the dihedral angles, the translation vector and the euler angles of the amino acid under consideration. These variables are partitioned into three sets. The first one (i.e., global variables) consists of the variables where branching occurs; the second set (i.e., local variables) consists of the variables where branching is not performed, and the third set (i.e., fixed variables) includes the variables for which there exists

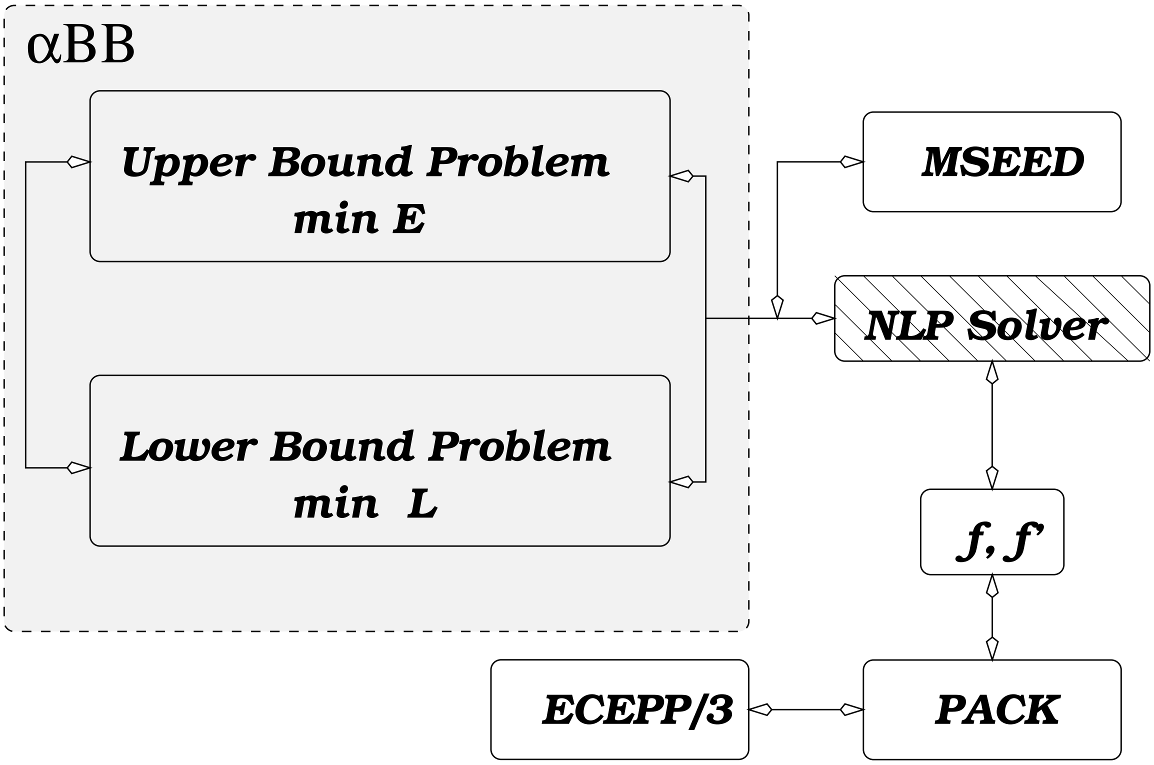

Figure 4: **The interface for global optimization.**

sufficient experimental evidence for keeping them fixed.

The information required by the user is provided in four different files. The first one is required by **PACK** and contains information about the different protein chains to be considered. The second one is needed by **ECEPP/3** and contains information about the sequence and number of the amino acid residues and the type of end groups. It also initializes the dihedral angles, translation vector and euler angles. The third file provides the bounds on the amino nitrogen and carboxyl carbon of the binding residue. These bounds are defined around the cartesian coordinates of these two atoms of the influenza virus peptide, [38]. Thus, for each pocket bounds were set in the x,y,z directions around the coordinates of the corresponding atoms of the influenza virus peptide presented by Stern et al. [38]. These bounds are given in Table III.

Table III: Bounds on N and C' for pockets 1 of HLA-DRB1

Bounds		*Lower*	*Upper*
N	x	-9.2	-8.4
	y	23.3	24.1
	z	17.3	18.1
C'	x	-10.0	-9.0
	y	25.0	27.0
	z	16.0	18.0

5 Computational Studies and Discussion

In this section, each one of the naturally occurring amino acids will be examined and accessed regarding its binding affinity with Pocket 1 of HLA-DR1 molecule. Prior to presenting the results obtained by the proposed approach, the following point regarding the amino acid polarity should be made. The aliphatic side chains the amino acids Ala, Val, Ile and Leu can be clearly considered as nonpolar ones, whereas, at the opposite end of the polarity scale are the charged residues Glu, Asp, Arg, Lys. Asn and Gly which have amide side chains, as well as the hydroxylic amino acids Ser and Thr are polar and expected to interact strongly with water and have high solubility. The polarity of the rest of the amino acids is more ambiguous. Cys and His have pK_a values close to 7 and may actually be charged in many proteins under physiological conditions. In our computational studies we consider as positively charged residues the Arg+, His+, Lys+

and as negatively charged residues the Asp- and Glu- using the parameters of **ECEPP/3** for eveluating their energy contributions. In Tyr the aromatic ring compensates for the hydroxyl and makes Tyr a nonpolar residue. Gly and Pro are special but from the way they behave in proteins can be considered as nonpolar and polar, respectively.

5.1 Individual Solvated Residues away from Pocket 1

As it has been mentioned in section 2, in order to evaluate the energy of interaction of the 20 naturally occurred amino acids within a pocket the intramolecular energy due to atomic interactions between the atoms of the single residue far away from the pocket has to be calculated. Thus, the global minimum energy for each residue in isolation is found with the consideration of the solvation contribution as described in section 3.3. The results obtained by applying the global optimization algorithm, α**BB**, [3, 4, 25] are shown in Table IV where based on the previous remark regarding amino acid polarity some of them are considered charged.

5.2 Complex of Pocket 1 and Binding Amino Acids

As mentioned earlier, ΔE, defined as the difference $E_{Total} - E^{o}_{Res}$, has been considered to represent the binding potential of a specific naturally occurring amino acid to the pocket considered. As shown in Table II, the number of amino acids included in pocket 1 increase as R increases. Specifically from 4.5 to 5.0 Å the amino acids that were added involve threonine and glutamic acid. Note that Glu is negatively charged and is an important factor for evaluating the interactions with positive charged residues as illustrated in Table VI.

The results for pocket 1 with R=5.0 Å are presented in Tables V,VI and in Figure 5a,b. Tyr, Phe and Trp are found to have the strongest binding affinities with interaction energies in a range of -20.00 to -16.95. At lower positions in the middle of the list there are the Leu, Ile and Val having interaction energies between -12.481 and -11.209. At the bottom of the list finally are the negative charged residues Glu- and Asp- with 40% smaller interaction energy than that of Val. An interesting result of the theoretical studies is the one obtained for the positive charged residues that appear to be the most unfavorable binders for this pocket.

A series of competitive binding assays was performed that involved analogs of the HA peptide (306-318) and the DRB1*0101 molecule [28]. Since the HA(306-318) peptide residue that interacts with Pocket 1 is Y(308) a number of analog peptides were synthesized that substituted the Y(308) with 11 different amino

Table IV: Standard Energies for Individual Solvated Residues

Residue	**Code**	$\mathbf{E^{o,S}_{Res}}$ (kcal/mol)
Ala	A	-42.143
Asn	N	-94.376
Cys	C	-74.667
Gln	Q	-86.964
Gly	G	-56.260
Ile	I	-17.074
Leu	L	-23.166
Met	M	-46.269
Phe	F	-160.850
Ser	S	-82.476
Thr	T	-69.423
Trp	W	-184.230
Tyr	Y	-178.950
Val	V	-25.055
Glu-	E-	-56.607
Asp-	D-	-67.416
His+	H+	-125.460
Arg+	R+	-105.800
Lys+	K+	-27.706

Table V: Relative Energies for Solvated Residues in Pocket 1 ($\mathcal{R} = 5.0$Å)

Residue	$\mathbf{E^{S}_{Total}}$ (kcal/mol)	$\mathbf{E^{o,S}_{Res}}$ (kcal/mol)	$\mathbf{\Delta\ E}$ (kcal/mol)
Tyr	-198.950	-178.950	-20.000
Phe	-180.475	-160.850	-19.625
Trp	-201.180	-184.230	-16.950
Gln	-102.360	-86.964	-15.396
Met	-60.212	-46.269	-13.943
Asn	-108.160	-94.376	-13.784
Thr	-82.718	-69.423	-13.297
Leu	-35.647	-23.166	-12.481
Ile	-29.539	-17.074	-12.465
Ser	-94.033	-82.476	-11.557
Cys	-85.947	-74.667	-11.280
Val	-36.264	-25.055	-11.209
Ala	-52.498	-42.143	-10.355
Gly	-66.351	-56.260	-10.091
Glu-	-64.531	-56.607	-7.744
Asp-	-69.847	-67.416	-2.431

Table VI: Relative Energies for Solvated Positive Charged Residues in Pocket 1 ($\mathcal{R} = 5.0$Å)

Residue	$\mathbf{E^{S}_{Total}}$ (kcal/mol)	$\mathbf{E^{o,S}_{Res}}$ (kcal/mol)	$\mathbf{\Delta\ E}$ (kcal/mol)
His+	-58.374	-124.870	+66.496
Lys+	+196.15	-27.706	+223.856
Arg+	+182.78	-105.640	+288.420

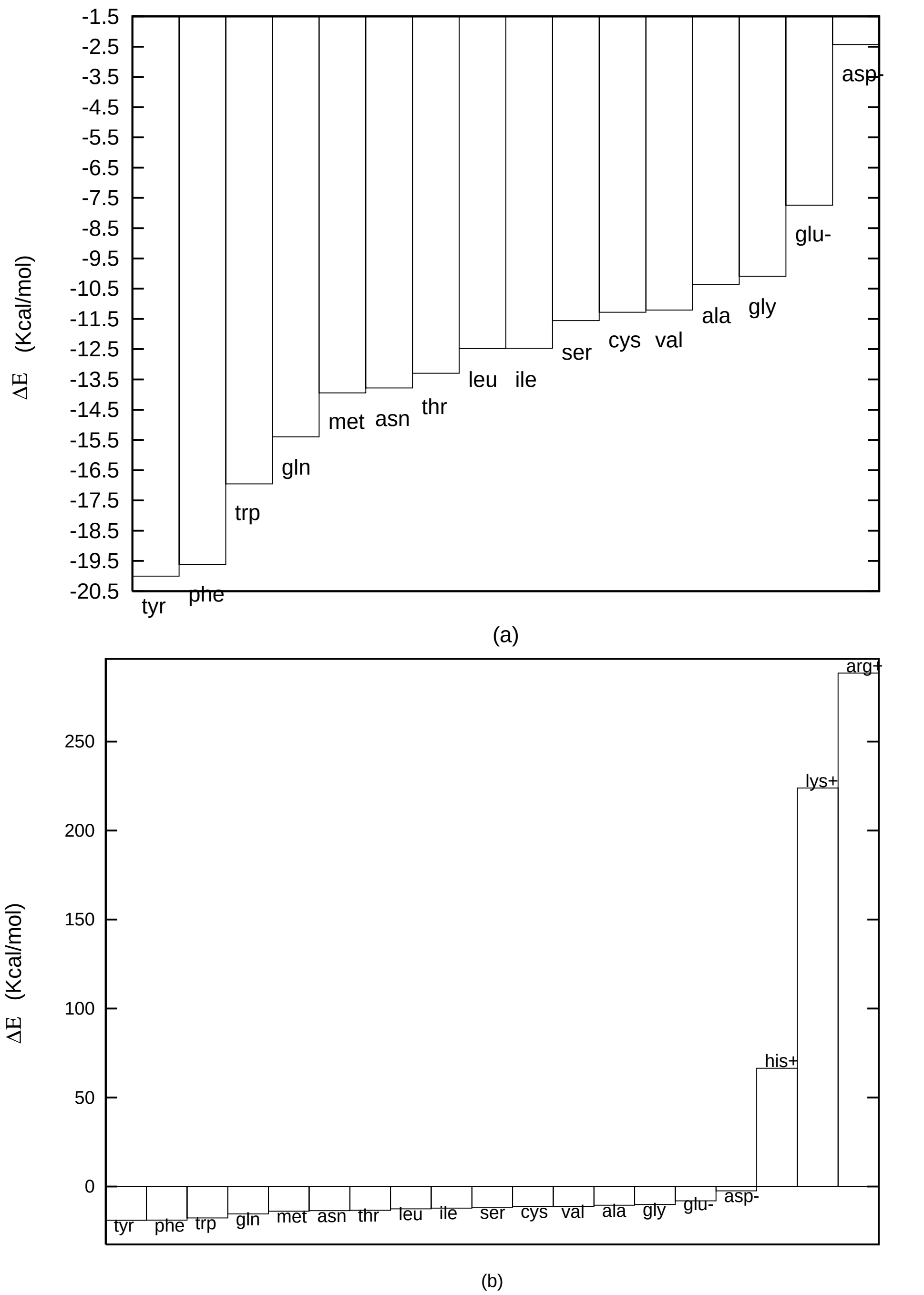

Figure 5: **ΔE (kcal/mol) of the naturally occurring amino acids**

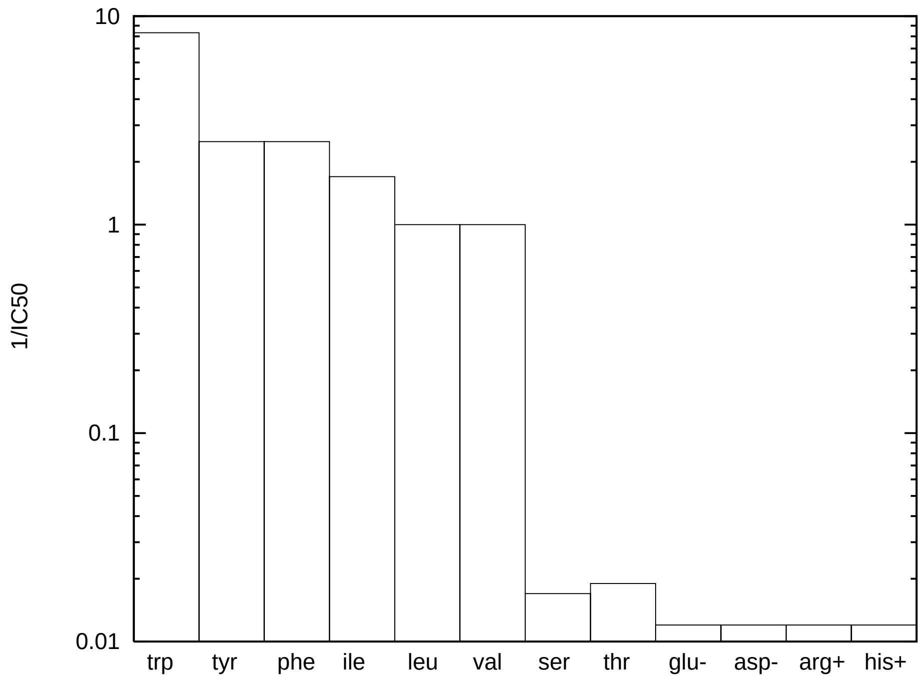

Figure 6: **Experimental data for the naturally occurring amino acids**

acids. The relative binding affinity was derived from the reciprocal of 50% inhibitory concentration (IC50) of each analog peptide in a logarithmic scale. Figures 5a,b and 6 show the results of both the theoretical and the experimental results, respectively.

Based on the competitive binding assays shown in Figure 6, three groups of binding affinities have been identified. The first group includes the amino acids Tyr, Phe, and Trp that are the residues with the highest affinity to DR1. The second group includes the amino acids Ile, Leu, and Val and are characterized by an intermediate level of affinity to DR1. The third group finally consists of low level affinity amino acids. This group involves the charged residues Asp-, Glu-, Arg+, His+ as well as the amino acids serine and threonine.

Based on the theoretical predictions, shown in Figure 5a,b, Tyr, Phe and Trp are at the top positions of the rank ordered list of the examined naturally occurred amino acids a result that is further supported from the strong preference of this pocket for large hydrophobic side chains. Furthermore, the amino acids Leu, Ile, and Val were found by the optimization studies to be characterized by potential energies that correspond to 7th, 8th and 11th position on the ordered list, respectively. The binding assays resulted in intermediate affinities for these amino acids. The ΔE value of -11.209 Kcal/mol for the binding of Val, reflects an approximate increase of 43% as compared to Tyr. Provided that an increase of 15% in potential energy defines the group of strong binders (tyr,phe,trp), an increase of up to 43% could very well reflect a group of intermediate level binders. At the bottom of table V, the global optimization studies put the charged residues which is also in agreement with experimental data. An increase of approximately 31% between ΔE values of Val and Glu- reflects the low affinity group of amino acids.

The laboratory studies present serine and threonine as relatively weak binders. The hydroxyl groups on both of these residues would favor interaction with polar molecules. Thus, weak interactions with the hydrophobic pocket 1 would be a predictable consequence. Although, valine appears to have comparable binding energy with these two residues again the optimization results support the observation of these residues being weaker binders than the small aliphatic residues.

Finally, there is a number of amino acids including Gln, Lys+, Met, Asn, Cys,Gly and Ala for which no analogs were synthesized. However, it has been reported in [18] and [13] that the peptides with Met are intermediate binders while peptides with Ala result in loss of peptide binding. The values of ΔE equal to -13.943, -10.355 for Met and Ala, respectively, found from our global optimization studies are consistent with the reported binding studies.

Therefore, the theoretical results are in excellent agreement with those obtained by the experimental approach of competitive binding assays, [28].

Moreover, since the optimization interface produces a PDB file of the coordinates of the minimum energy conformation of the binder a direct comparison with the crystallographic data can be made for the tyrosine residue that binds to pocket 1. Figure 7 shows the HA peptide binder (Tyrosine 308) in white and the minimum conformation of tyrosine for pocket 1 in grey. An almost identical orientation with 1.28 Å is observed. Figures 9 and 10 illustrate the orientation of plenylalanine and tryptophan in comparison to the virus peptide binding shown in Figure 11, suggest that these residues are in fact very strong binders.

The need for determining the global minimum conformation is illustrated in Figure 8 where a local minimum conformation of tyrosine corresponding to -196.637 kcal/mol, that is having only 1.16% difference from the global minimum of -198.95 kcal/mol is illustrated with blue whereas the global minimum conformation is shown with yellow. Note that the local minimum energy conformation of tyrosine is very different than the global minimum despite its proximity to the global minimum energy value.

For the charged residues, experimental data suggest their weak binding affin-

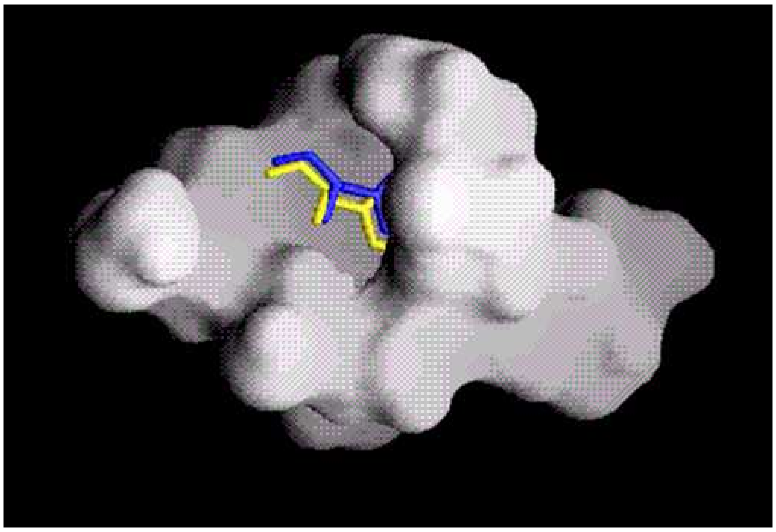

Figure 7: **Comparison of Tyrosine binding to Pocket 1**

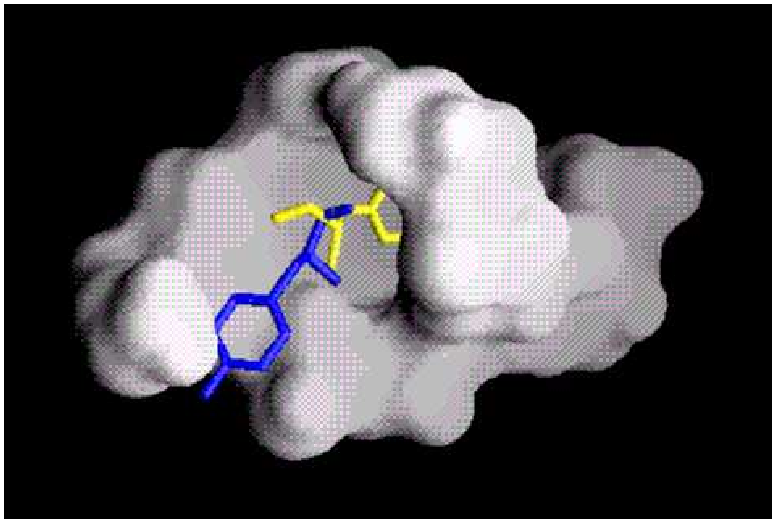

Figure 8: **Local vs Global minimum configuration of tyrosine**

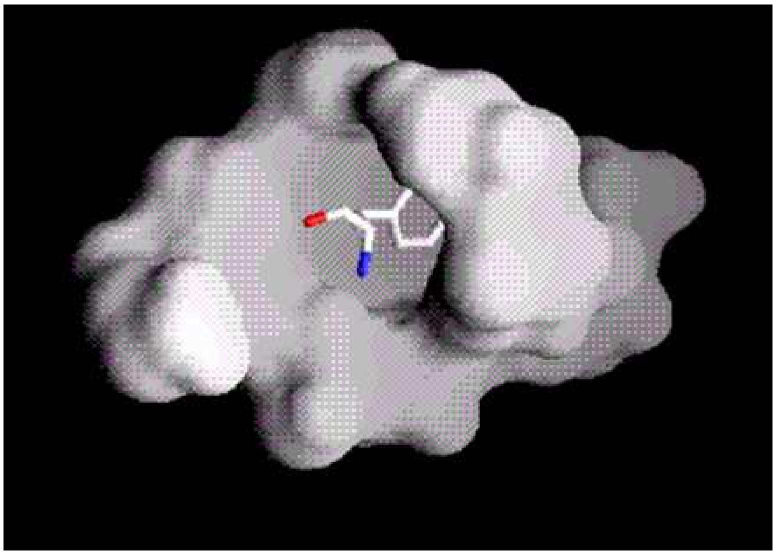

Figure 9: **Phenylalanine binding to pocket 1**

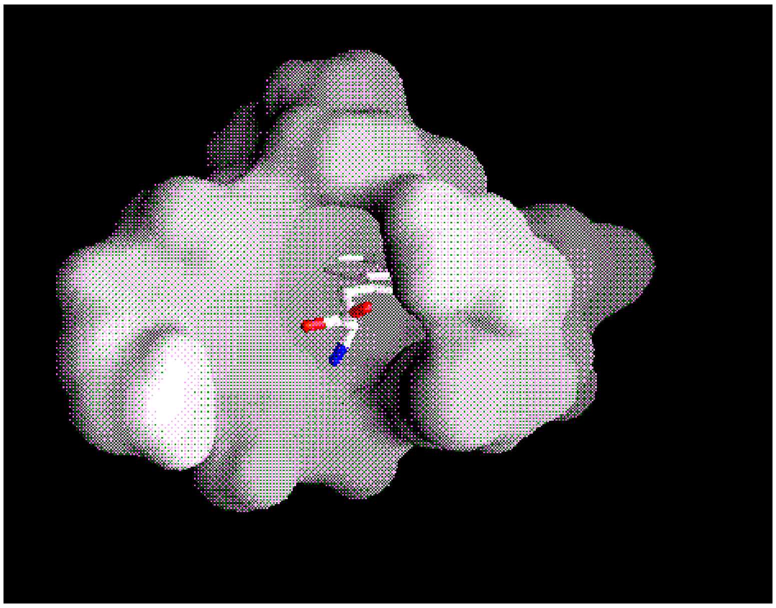

Figure 10: **Tryptophan binding to pocket 1**

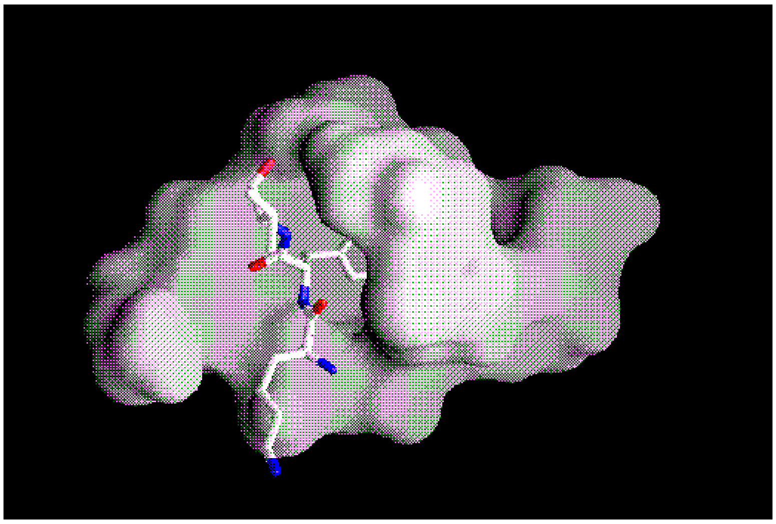

Figure 11: **Influenza Virus Peptide binding to pocket 1**

ity. The fact that pocket 1 is extremely hydrophobic region intuitively verifies this result. Charged residues are not stabilized by the weak van der Waals interactions that stabilize the conglomeration of hydrophobic residues. The hypothesis that results from this knowledge is that the inclusion of charge on the five aforementioned residues should greatly destabilize their interactions with pocket 1, which corresponds to an increase in their overall conformational energies (ΔE). The theoretical results obtained for the negative charged residues support this idea. Positive charged residues have large binding energy due to large electrostatic contribution from the interaction with the negative charged glutamic within the pocket which however does not suggest favorable binders as their orientation shown in Figures 12,13,14 for arginine+, histidine+, and lysine+, respectively. The enforcement of these residues inside the pocket gives rise to large (positive) energies (Table VI, Figure 5b) that indicate highly unfavorable residues (see Figures 15,16, 17).

6 Summary and Conclusions

In this paper, a novel predictive method is proposed for modeling and studying the binding affinity of different naturally occurring amino acids with pocket 1 of the HLA-DR1 protein. First, the composition of the pocket is identified together with the cartesian coordinates of the atoms that participate in each amino acid of the pocket 1 of the HLA-DR1 protein. Second, explicit relations for all the en-

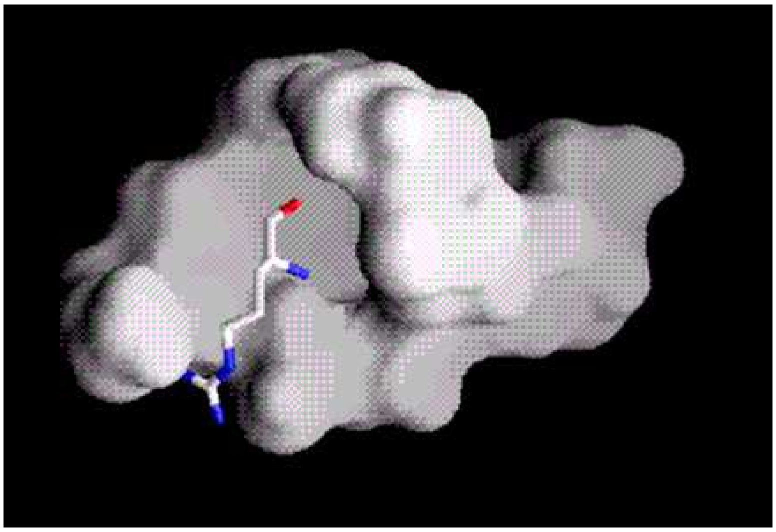

Figure 12: **Arginine+ binding to pocket 1**

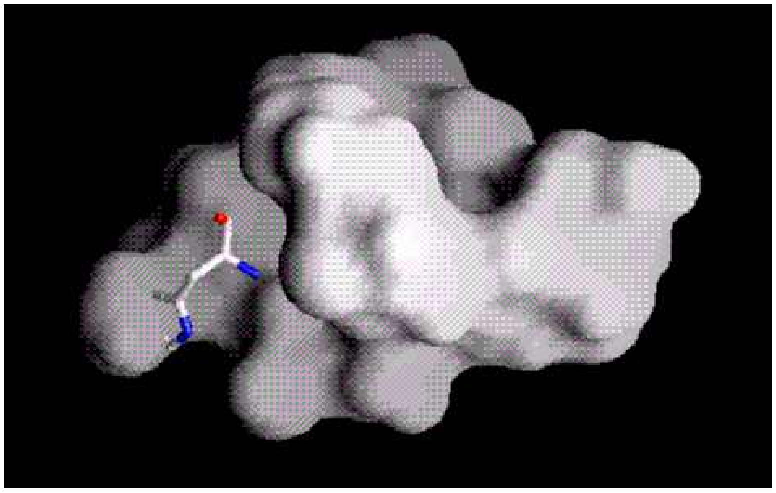

Figure 13: **Histidine+ binding to pocket 1**

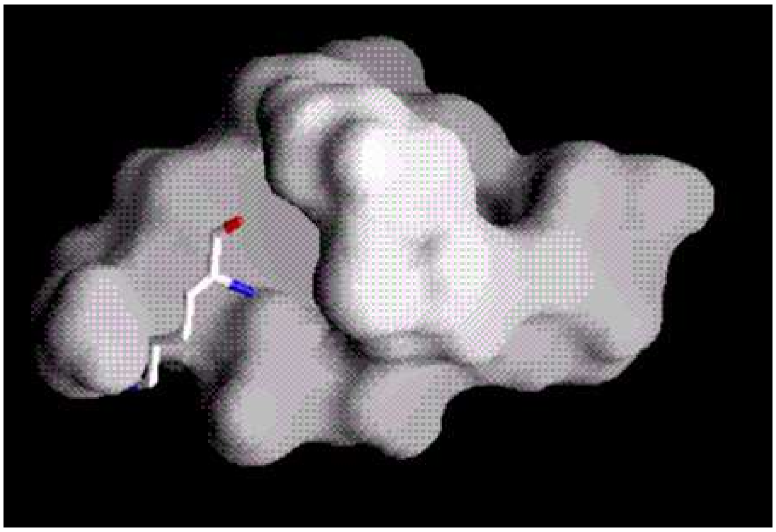

Figure 14: **Lysine+ binding to pocket 1**

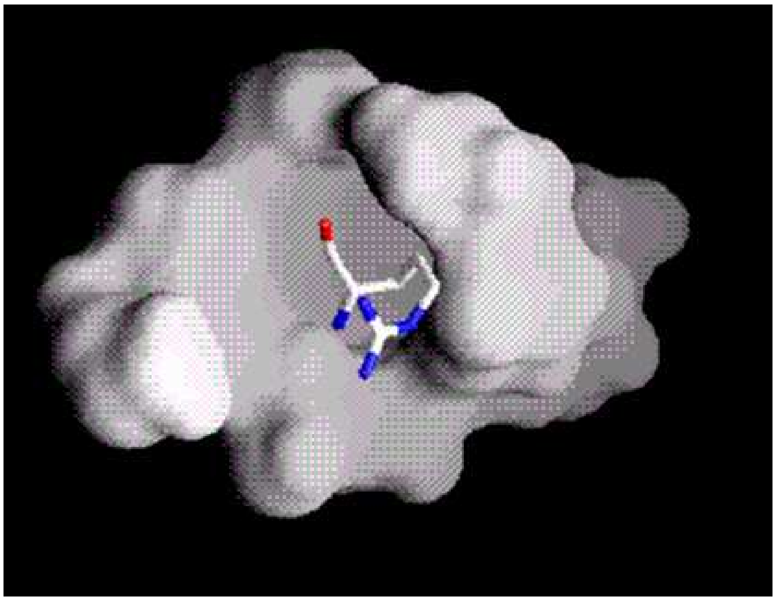

Figure 15: **Arginine+ forced within pocket 1**

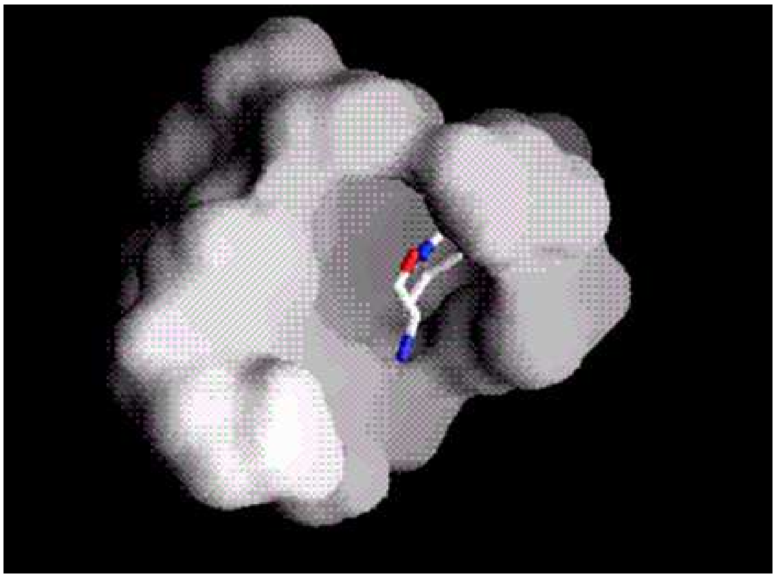

Figure 16: **Histidine+ forced within pocket 1**

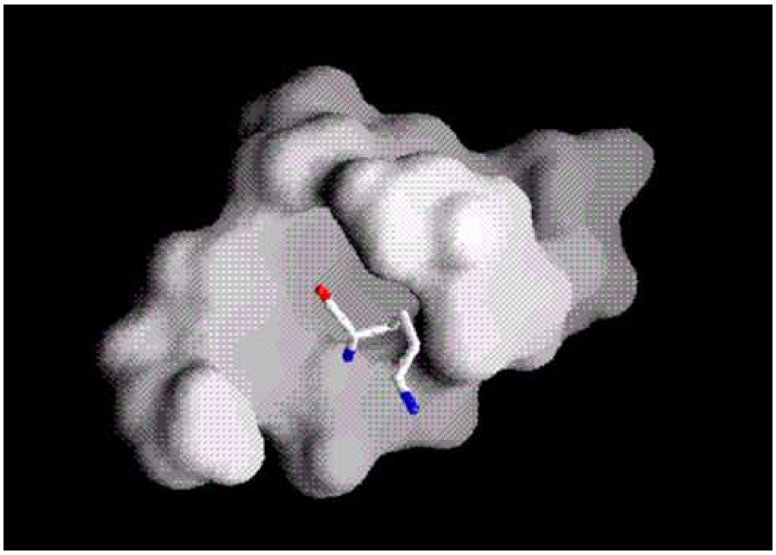

Figure 17: **Lysine+ forced within pocket 1**

ergetic inter and intra interactions between the atoms of the residues that define the pocket of the HLA-DR1 protein and the atoms of the considered naturally occurring amino acid were derived. Moreover, solvation energy was also taken into account based on solvent accessible area method. Then, the docking problem is formulated as a nonconvex optimization problem on a set of independent dihedral angles, euler angles and translation variables. The deterministic global optimum method, $\alpha\mathbf{BB}$, is then adopted for the solution of the resulting problem which is based on the generation of a sequence of converging upper and lower bounds found from the local solution of the nonconvex problem and the convex lower bounding problem which is constructed based on eigenvalue analysis of the nonconvex potential energy function. The final step of the proposed approach consists of evaluating the interaction energy for all naturally occurring amino acids and generate a ranked-order list.

The results of the proposed approach were found to agree very well with the experimental competitive binding assays. It should be emphasized that although in this paper only one of the binding sites of the HLA-DR1 protein is examined, the approach is applicable to predict the binding affinity of the amino acid residues in the different pockets. Our current work focuses on studying the binding information for the different pockets in order to be able to predict the binding of the whole peptide to the HLA-DR1 protein.

Acknowledgments
Financial Support from the National Science Foundation and the Air Force Office of Scientific Research is gratefully acknowledged.

References

[1] N.L. Allinger. Conformational Analysis. MM2: A Hydrocarbon Force Field Utilizing V_1 and V_2 Tortional terms. *J. Am. Chem. Soc.*, 99:8127, 1977.

[2] N.L. Allinger, Y.H. Yuh, and J.-H. Liu. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. *J. Am. Chem. Soc.*, 111:8551, 1989.

[3] I.P. Androulakis, C.D. Maranas, and C.A. Floudas. αBB: A Global Optimization Method for General Constrained Nonconvex Problems. *Journal of Global Optimization*, 7:337–363, 1995.

[4] I.P. Androulakis, C.D. Maranas, and C.A. Floudas. Prediction of Oligopeptide Conformations via Deterministic Global Optimization. *Journal of Global Optimization*, 1996. Accepted for publication.

[5] D.J. Bacon and J. Moult. Docking by least-square fitting of molecular surface patterns. *Jl. Mol. Biol.*, 225:849–858, 1992.

[6] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus. CHARM: A Program for Macromolecular Energy Minimization and Dynamics Calculation. *J. Comp. Chem*, 8:132, 1983.

[7] A. Calfisch, P. Niederer, and M. Anliker. Monte Carlo docking of oligopeptides to proteins. *Proteins*, 13:223–230, 1992.

[8] M.L. Connolly. Solvent-accesible surfaces of proteins and nucleic acids. *Science*, 221:709, 1983.

[9] P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, and A.T. Hagler. Structure and Energetics of Ligand Binding to Peptides: Escherichia coli Dihydrofolate Reductase–Trimethoprim, A Drug Receprtor System. *Proteins*, 4:31, 1988.

[10] X. Fu, C. Bono, S. Woulfe, C. Swearingen, N. Summers, F. Sinigaglia, A. Sette, B. Schwartz, and R.W. Carr. Pocket 4 of the HLA-DR Molecule is a Major Determinant of T cell Recognition of Peptide. *Journal of Experimental Medicine*, 181:915–926, 1995.

[11] P. Gill, W. Murray, M. Saunders, and M. Wright. *User's Guide for NPSOL (Version 4.0): A Fortran Package for Nonlinear Programming.* Stanford University Department of Operations Research, January 1986.

[12] D.S. Goodsell and A.J. Olson. Automated docking of substrates to proteins by simulated annealing. *Proteins*, 8:195–202, 1990.

[13] J. Hammer, C. Bolin, D. Papadopoulos, J. Walsky, J. Higelin, W. Danho, F. Sinigaglia, and Z.A. Nagy. High-affinity binding of short peptides to major histocompatibility complex class II molecules by anchor combinations. *Proc. Natl. Acad. Sci.*, 91:4456, 1994.

[14] J. Hammer, E. Bono, F. Gallazi, C. Belunis, Z. Nagy, and F. Sinigaglia. Precise Prediction of Major Histocompatibility Complex Class II-Peptide Interaction Based on Peptide Side Chain Scanning. *Journal of Experimental Medicine*, 180:2353–2358, 1994.

[15] J. Hammer, F. Gallazzi, E. Bono, R. Karr, J. Guenot, P. Valsasnini, and Z. Nagy. Peptide Binding Specificity of HLA-DR4 Molecules: Correlation with Rheumatoid Arthritis Association. *Journal of Experimental Medicine*, 181:1847–1855, 1995.

[16] T.N. Hart and R.J. Read. A multiple-start Monte-Carlo docking method. *Proteins*, 13:206–222, 1992.

[17] I.D.Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge, and T.E. Ferrin. A geometric approach to macromolecule-ligand interactions. *Jl. Mol. Biol.*, 161:269–288, 1982.

[18] T.S. Jardesky, J.C. Gorga, R. Bush, J. Rothbard, J.L. Strominger, and D.C. Wiley. Peptide binding to HLA-DR1: a peptide with most residues substituted to alanine retains MHC binding. *EMBO Jl*, 9:1797, 1990.

[19] F. Jiang and S.H.Kim. Soft docking:Matching of molecular surface cubes. *Jl. Mol. Biol.*, 219:79–102, 1991.

[20] J.L.Nauss, R.H. Reid, and S. Sadegh-Nasseri. Accuracy of a Structural Homology Model for a Class II Histocompatibility Protein, HLA-DR1: Comparison to the Crystal Structure. *Journal of Biomolecular Structure and Dynamics*, 12:1213–1233, 1995.

[21] K.Gulukota, S.Vajda, and C.Delisi. Peptide Docking Using Dynamic Programming. *Jl. Comp. Chem.*, 17:418–428, 1996.

[22] B. Lee and F.M. Richards. The interpetation of protein structures: Estimation of static accesibility. *Jl. Mol. Biol.*, 55:379–400, 1971.

[23] M. Levitt. Protein Folding by Restrained Energy Minimization and Molecular Dynamics. *J. Mol. Biol.*, 170:723, 1983.

[24] C.D. Maranas, I.P. Androulakis, and C.A. Floudas. A deterministic global optimization approach for the protein folding problem. In P.M. Pardalos, D. Shalloway, and G. Xue, editors, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, volume 23, pages 133–150. American Mathematical Society, 1995.

[25] C.D. Maranas and C.A. Floudas. Global minimum Potential Energy Conformations of Small Molecules. *Journal of Global Optimization*, 4:135–170, 1994.

[26] F.A. Momany, L.M. Carruthers, R.F. McGuire, and H.A. Scheraga. Energy Parametersin Polypeptides. VII. Geometric Parameters, Partial Atomic Charges, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occuring Amino Acids. *J. Phys. Chem.*, 79:2361, 1975.

[27] F.A. Momany, R.F. McGuire, A.W. Burgess, and H.A. Scheraga. Energy Parameters in Polypeptides. *Journal of Physical Chemistry*, 79(22), 1975.

[28] D. Monos, A. Soulika, E. Argyris, J. Gorga, L. Stern, V. Magafa, P. Cordopatis, I.P. Androulakis, and C.A. Floudas. *Proceedings of 12th International Histocompatibility World Conference HLA diversity:functional and medical implications*. D.Charron Ed. In press May 1997.

[29] G. Némethy, K.D. Gibson, K.A. Palmer, C.N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H.A. Scheraga. Energy Parameters in Polypeptides. 10. Improved Geometrical Parameters and Nonbonded Interactions to usein the ECEPP/3 Algorithms, with Applications to Proline–Containing Peptides. *J. Phys. Chem.*, 96:6472, 1992.

[30] G. Némethy, M.S. Pottle, and H.A. Scheraga. Energy Parameters in Polypeptides. 9. Updating of Geometricall Parameters, Nonbinded Interaction and Hydrogen Bond Ingteractions for the Naturally Occuring Amino Acids. *J. Phys. Chem.*, 89:1883, 1983.

[31] A. Nicholls. *GRASP: Graphical Representation and Analysis of Surface Properties*, October 1992.

[32] W.E. Paul. *Fundamental Immunology.* Raven Press, 1993.

[33] G. Perrot, B. Cheng, K.D. Gibson, K.A. Palmer J. Vila, A. Nayeem, B. Maigret, and H.A. Scheraga. MSEED: A Program for the Rapid Analytical Determination of Accessible Surface Areas and their Derivatives. *Journal of Computational Chemistry*, 13:1–11, 1992.

[34] D. Rogman, L. Scapozza, G. Folkers, and A. Daser. Molecular Dynamics Simulation oof MHC-Peptide Complexes as a Tool for Predicting Potential T Cell Epitopes. *Biochemistry*, 33:11476–11485, 1994.

[35] R. Rosenfeld, Q. Zheng, S. Vajda, and C. DeLisi. Computing the Structure of Bound Peptides. Applications to Antigen Recognition by Class I Major Histocompatibility Complex Receptors. *Journal of Molecular Biology*, 234:515–521, 1993.

[36] H.A. Scheraga. *ECEPP/3 USER GUIDE.* Cornell University Department of Chemistry, January 1993.

[37] H.A. Scheraga. *PACK: Programs for Packing Polypeptide Chains*, 1996. online documentation.

[38] L. Stern, J. Brown, T. Jardetzky, J. Gorga, R. Urban, L. Strominger, and D. Wiley. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature*, 368:215–221, 1994.

[39] W.F. van Gunsteren and H.J.C. Berendsen. GROMOS *Groningen Molecular Simulation.* Groningen, The Netherlands, 1987.

[40] J. Vila, R.L. Williams, M. Vasquez, and H.A. Scheraga. Empirical Solvation Models can be used to Differentiate Native from Non-Native Conformations of Bovine Pancreatic Trypsin Inhibitor. *Proteins*, pages 199–218, 1991.

[41] S. Weiner, P. Kollmann, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Weiner. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.*, 106:765, 1984.

[42] S. Weiner, P. Kollmann, D. Nguyen, and D. Case. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comp. Chem.*, 7:230, 1986.

[43] K.W. Wucherpfennig, A. Sette, S. Southwood, C. Oseroff, M. Matsui, J.L. Strominger, and D.A. Hafler. Structural Requirements for Binding of an Immunodominant Myelin Basic Protein Peptide to DR2 Isotypes and for its Recognition by Human T Cell Clones. *Journal of Experimental Medicine*, 179:279–290, 1994.

[44] K.W. Wucherpfennig and J.L. Strominger. Selective Binding of Self Peptides to Disease-associated MHC Molecules: A Mechanism for MHC-linked Susceptibility to Human Autoimmune Diseases. *Journal of Experimental Medicine*, 181:1597–1601, 1995.

[45] L. Zerva, B. Cizman, N.K. Mehra, S.K. Alahari, R. Murali, C.M. Zmijewski, M. Kamoun, and D.S. Monos. Arginine at positions 13 or 70-71 of HLA-DRB1 alleles confer sucseptibility to Tuberculoid Leprosy. *Journal of Experimental Medicine*, 183:829–836, 1996.

Appendix A

Determination of Euler Angles

The calculation of the Euler angles is complicated by the fact that the location of the hydrogen atom in the (x,y,z) space is not known. For this reason, a discussion of the method of hydrogen position determination will precede the Euler angle theory.

Determination of Hydrogen Location

The basic steps behind finding the location of a hydrogen is the following: (1) define a basis system, (2) find the relative position of the hydrogen in this system, and (3) translate and rotate this position to a new basis defined by a particular molecule. This will become more clear as the explanation proceeds.

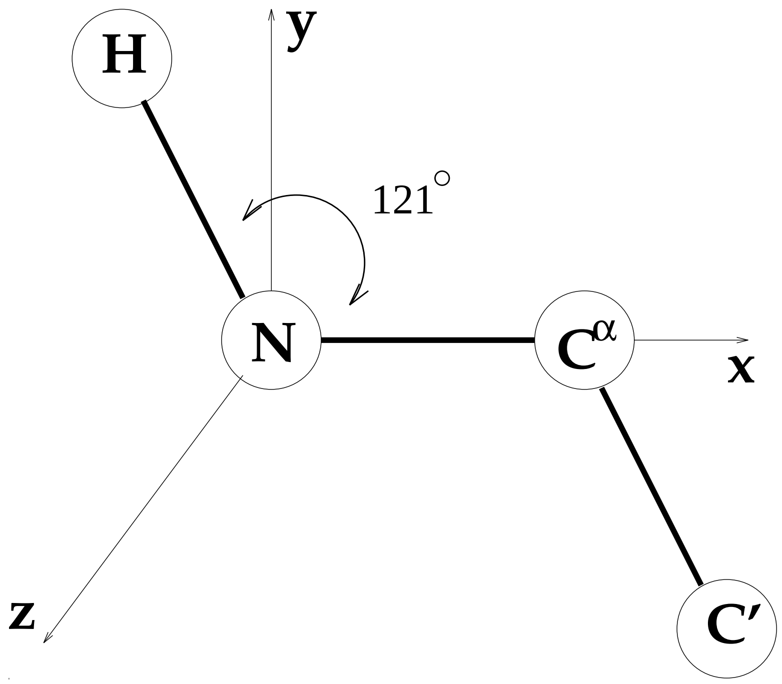

Figure 18: **Axes defined by the relative positions of C', N, and C^{α}.**

The diagram in Figure 18 shows how the axes are defined in relation to the prime carbon, nitrogen and alpha carbon. The phi angle is always approximately equal to 180 degrees. Hence, the N-C^{α} bond defines the x-axis, while the C^{α}-C'and N-H bonds lie in the XY plane.

The position of the hydrogen is initially unknown. The position of the nitrogen is taken to be at the origin. The position of the alpha carbon is known because the N-C bond length is approximately 1.435 Å. By this definition, the positions of the atoms are N(0,0,0), and C^{α}(1.435,0,0). The position of the hydrogen can be found by the knowledge that the H-N-C^{α} bond angle is 121 degrees. This angle is illustrated by the arrow in Figure 18. Since the hydrogen lies in the x,y plane then the hydrogen will lie in the quadrant defined by negative x and positive y. Hence the position is easily found remembering that the N-H bond length is approximately 1.0 Å. The x and y coordinates are just the cosine and sine of the angle of 121 degrees, respectively. Solving this system yields: H(-0.5150,0.8570,0), or since Nitrogen is defined as the origin in this basis, the following vector is defined:

$$\mathbf{NH} = -0.4226\mathbf{i} + 0.3791\mathbf{j} + -0.8232\mathbf{k} \tag{17}$$

where, **i**, **j**, and **k** are the defined unit vectors for the basis.

Now if N, C', and C^{α} are at a random orientation in space, the magnitudes (or distances from the defined origin, nitrogen) do not change, but the directions of the unit normals do change. Hence if **i**, **j**, and **k** are defined in the new orientation the vector will still be defined as above, but will have a different value due to the change in the unit vectors.

The final step is to define these unit vectors as the basis of the orientation of a particular molecule, where the coordinates of N, C', and C^{α} are known. The vectors $\mathbf{N}C^{\alpha}$, and $C^{\alpha}\mathbf{C'}$ are easily defined by subtraction of coordinates. Then the unit vectors in the axis directions can be defined as follows.

In the x-direction,

$$\mathbf{i} = \frac{\mathbf{NC}^{\alpha}}{|\mathbf{NC}^{\alpha}|} \tag{18}$$

In the z-direction,

$$\mathbf{z} = (\mathbf{C}^{\alpha}\mathbf{C}')\times(\mathbf{NC}^{\alpha}) \tag{19}$$

$$\mathbf{k} = \frac{\mathbf{z}}{|\mathbf{z}|} \tag{20}$$

In the y-direction,

$$\mathbf{j} = \mathbf{k}\times\mathbf{i} \tag{21}$$

Then for the particular orientation:

$$\mathbf{NH} = -0.5150\mathbf{i} + 0.8570\mathbf{j} + 0.00\mathbf{k} \tag{22}$$

Finally the exact coordinates of hydrogen can be determined by adding **NH** to the given coordinates of nitrogen for the system.

Determination of Euler Angles

The euler angles are found by comparing the angles between the unit normals defining the coordinate axes. The basis coordinate system is defined as follows and is subscripted with a 1 when mentioned later: N(0,0,0), C^{α}(1.435,0,0), H($-$0.515,0.857,0).

So initially for given coordinates of the H, N, and C^{α} atoms, the unit vectors defining the coordinate axes must be found. First, the vectors $\mathbf{N}C^{\alpha}$ and **NH** are found by subtracting the respective coordinates. These vectors define the axes as shown in Figure 18. Once again the N-C^{α} bond lies on the x-axis, and the N-H bond is defined as lying in the xy plane. This orientation is shown on the coordinate axes in Figure 19. The unit vectors on the axes are easily described by the following equations.

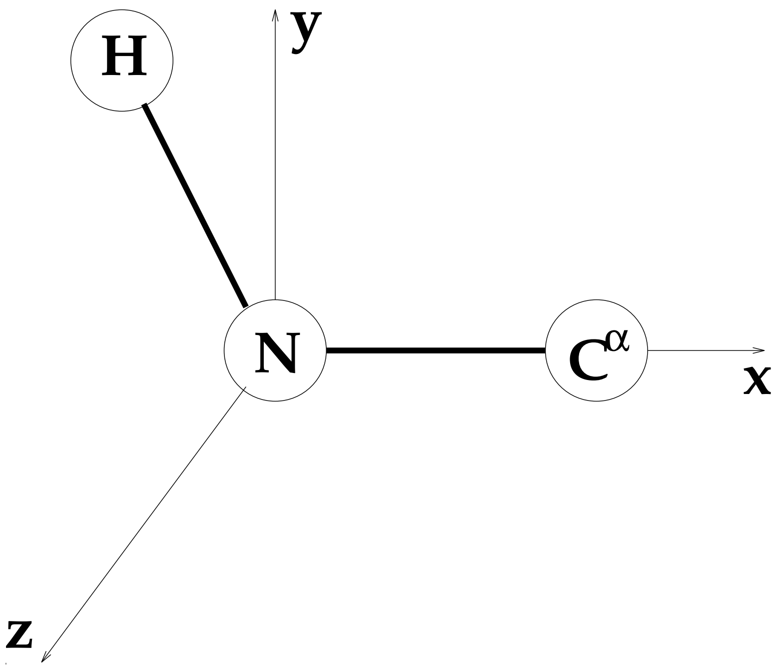

Figure 19: **Axes defined by the relative positions of H, N, and C^{α}.**

In the x-direction,

$$\mathbf{i} = \frac{\mathbf{NC^{\alpha}}}{|\mathbf{NC^{\alpha}}|} \tag{23}$$

In the z-direction,

$$\mathbf{k} = \frac{\mathbf{n}}{|\mathbf{n}|} \tag{24}$$

where,

$$\mathbf{n} = \mathbf{i} \times (\mathbf{NH}) \tag{25}$$

In the y-direction,

$$\mathbf{j} = \mathbf{k} \times \mathbf{i} \tag{26}$$

After performing the above operations on the initial basis and a residue there are two coordinate systems to compare. By definition the following relations hold for the euler angles in the basis system defined above:

$$\sin\theta_1 = -\mathbf{i_1} \cdot \mathbf{k_2} \tag{27}$$

$$\cos\theta_1 = -\mathbf{j_1} \cdot \mathbf{k_2} \tag{28}$$

$$\cos\theta_2 = \mathbf{k_1} \cdot \mathbf{k_2} \tag{29}$$

$$\sin\theta_3 = -\mathbf{k_1} \cdot \mathbf{i_2} \tag{30}$$

$$\cos\theta_3 = \mathbf{k_1} \cdot \mathbf{j_2} \tag{31}$$

$$\sin^2\theta_2 = 1 - \cos^2\theta_2 \tag{32}$$

The appropriate euler angles (θ_1, θ_2, θ_3) can easily be found by taking the arctangent of the ratio of sine to cosine. The signs on the euler angles are determined by using the signs of the sine and cosine of the angle to determine the quadrant where the angle is defined.

Appendix B

Determination of C' Location

As mentioned in section 3.4 the coordinates of the backbone carboxyl carbon have to be expressed as a function of other variables since they do not correspond to explicit optimization variables. This can be made using analytical geometry and linear algebra. In particular, for each amino acid the coordinates of C' atom can be expressed as a function of the coordinates of the N atom (i.e., the translation vector) denoted as n_x, n_y, n_z, the euler angles $\theta_1, \theta_2, \theta_3$ and the ϕ dihedral angle, and requires the knowledge of the following parameters the bond lengths NC^α and C^αC' as well as the value of the angle NC^αC' taken from the literature [27].
Given the above information and based on the graphic representation of protein shown in Figure 7, the following expressions are derived for the coordinates of C' atom:

$$\begin{aligned} x &= n_x + sc \times temp \times \sin(\theta_1) \times \sin(\theta_2) \\ &+ sa \times sc \times \cos(\phi) \times temp \times \frac{(-\cos(\theta_2) \times \cos(\theta_3) \times \sin(\theta_1) - \cos(\theta_1) \times \sin(\theta_3))}{\sqrt{(1-\cos^2(\phi))}} \\ &+ ((\mathbf{NC^\alpha}) + (\mathbf{C^\alpha C'}) \times \cos(\theta)) \times (\cos(\theta_1) \times \cos(\theta_3) - \cos(\theta_2) \times \sin(\theta_1) \times \sin(\theta_3)) \end{aligned}$$

$$\begin{aligned} y &= n_y - sc \times temp \times \cos(\theta_1) \times \sin(\theta_2) \\ &+ sa \times sc \times \cos(\phi) \times temp \times \frac{(\cos(\theta_1) \times \cos(\theta_2) \times \cos(\theta_3) - \sin(\theta_1) \times \sin(\theta_3))}{\sqrt{(1-\cos^2(\phi))}} \\ &+ ((\mathbf{NC^\alpha}) + (\mathbf{C^\alpha C'}) \times \cos(\theta)) \times (\sin(\theta_1) \times \cos(\theta_3) + \cos(\theta_1) \times \cos(\theta_2) \times \sin(\theta_3)) \end{aligned}$$

$$\begin{aligned} z &= n_z + sc \times temp \times \cos(\theta_2) + sa \times sc \times \cos(\phi) \times temp \times \cos(\theta_3) \times \frac{\sin(\theta_3)}{\sqrt{(1-\cos^2(\phi))}} \\ &+ ((\mathbf{NC^\alpha}) + (\mathbf{C^\alpha C'}) \times \cos(\theta)) \times (\sin(\theta_2) \times \sin(\theta_3)) \end{aligned}$$

where:

$$temp = \sqrt{\frac{(\mathbf{C^\alpha C'})^2 - (\mathbf{C^\alpha C'})^2 \times \cos^2(\theta)}{1 + \frac{\cos^2(\phi)}{1-\cos^2(\phi)}}}$$

$$sa = 1.0, sc = -1.0 \text{ if } \phi > 0$$

$$sa = -1.0, sc = 1.0 \text{ if } \phi \leq 0$$

and θ is the angle NC^αC'.