

# Quantum Chemical Associations Ligand–Residue: Their Role to Predict Flavonoid Binding Sites in Proteins

Alberto Rolo-Naranjo,<sup>†</sup> Edelsys Codorniu-Hernández,<sup>†,‡</sup> and Noel Ferro<sup>\*,§</sup>

Department of Molecular Design and Synthesis, Higher Institute of Technologies and Applied Sciences, Havana, Cuba, Department of Chemistry, University of Calgary, Calgary, Alberta, Canada, and Institute for Physical and Theoretical Chemistry, University of Bonn, Bonn, Germany

Received September 18, 2009

A novel approach is applied for the prediction of potential binding sites in ligand–protein interactions. This methodology introduces an integral strategy based on the calculation of protein geometrical parameters and the use of a quantum mechanical descriptor, Binding Local Site ( $B_{LS}$ ). A screening of the most likely cavities in the protein crystal structure is carried out where the analysis of geometric cavities is performed, and the virtual centers for binding (VCB) are located. The VCB surrounding amino acid residues (AA) are evaluated through the calculation of the  $B_{LS}$  by using the theoretical affinity order between the ligand and each AA. It includes a quantum scoring function based on the ligand–AA association energies and entropies. A contribution to the understanding of flavonoid–protein interactions is provided as well. The new bioinformatic strategy makes good predictions for flavonoid ligands. The calculated binding sites are quite in agreement with the crystal binding sites of 10 flavonoid binding proteins. This is a contribution of quantum mechanics in some phases of in silico drug design.

## I. INTRODUCTION

The analysis of protein–ligand interactions plays a crucial role in the discovery of new drug candidates. Data of biomacromolecules primarily determined experimentally by X-ray crystallography set significant progress in three-dimensional (3D) macromolecular structure. Several algorithms<sup>1–9</sup> and databases<sup>10–20</sup> have been created to support the analysis of intermolecular interactions. However, no consensus about a universal docking tool is available at the moment.<sup>5,21,22</sup> Molecular docking remains as a highly active area of research<sup>23–31</sup> focused essentially on two main failures: insufficient conformational sampling and scoring functions.<sup>5</sup>

The current docking scoring functions<sup>5,21,23,32,33</sup> rely almost exclusively on molecular force fields and empirical potentials. These approaches have failed to describe interactions of enzymes inhibitor systems.<sup>34</sup> Therefore, the inclusion of quantum mechanic (QM) energy terms is needed.<sup>32</sup> Its use in all phases of in silico drug design is a chimera but a logical step in the evolution of this field. Even when the computational power has been extraordinarily increased in recent years,<sup>34,35</sup> the cost of central processing unit (CPU) time and storage space are still a bottleneck for QM approaches in macromolecules.

At this state, an alternative way to apply QM information to analyze recognition between small molecules and proteins is presented herein. As a small ligand has nonsignificant contribution to the global energy of the protein during the binding, the local interactions of amino acid (AA)–ligand are a cornerstone to predicting interactions. Therefore, a QM

methodology predicts potential sites in a protein for specific ligand binding by taking into account local ligand–AA interactions. This QM analysis considers the dynamic behavior of ligand–AA associations with minimal time consumption. The approach of multiple minimum hypersurface<sup>36</sup> (MMH) was the focus on intermolecular interactions in 2005 MMH-1<sup>37</sup> and MMH-2.<sup>38</sup> MMH combines the calculation of QM energies of intermolecular interactions clusters with statistical thermodynamics for the selection of the most important configurations. In the present methodology two main outcomes are obtained: (i) the AAs with more affinity for the interactions with the ligand, and (ii) the potential regions for binding in the proteins with a ranking order.

Flavonoids are the selected ligands for this study because, first, these natural products are commonly found in plants as well as flavonoids and their derivatives show a diverse range of biological activities (e.g., anticarcinogenic effect).<sup>39</sup> Second, they have a well-known molecular action mechanism related to the recognition of the protein surface with large experimental data.<sup>39</sup> Third, a theoretical affinity order (TAO) among flavonoids and the 20 AAs of proteins was recently published based on the values of QM association energy for each flavonoid–AA complex.<sup>37,40</sup>

Next, a 3D screening of the AA population with stronger associations in flavonoid-binding proteins is the research aim of this approach. Those protein regions with simultaneous geometrical and chemical capacities for hosting the ligand shall be selected as potential sites for ligand binding. The chemical capacity of a protein cavity is referred to the capacity of having inside of it those AA that have the highest associations with the ligand, following QM information. This paper is focused on the comparison of the predicted binding sites by this novel approach with those known experimental sites for 10 flavonoid binding proteins. New information

\* Corresponding author: Telephone: +49(0) 228 73-3332. E-mail: ferro@thch.uni-bonn.de.

<sup>†</sup> Higher Institute of Technologies and Applied Sciences.

<sup>‡</sup> University of Calgary.

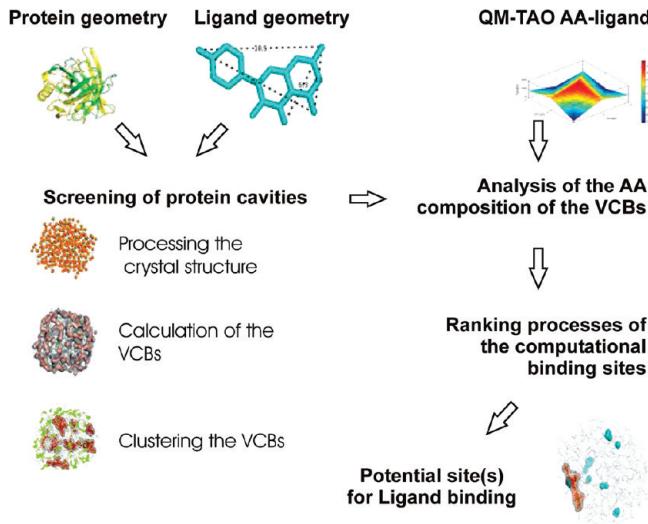
<sup>§</sup> University of Bonn.

about the general features of flavonoid–protein interactions is provided. The application of this approach is a new bioinformatic tool to predict potential binding sites of small molecules.

## II. THEORY AND PROCEDURES

The methodology has been developed in order to apply QM to the study of ligand–protein interactions. The procedure is able to predict the potential binding sites for small ligands. The crystallographic structure of each protein under study, the ligand dimensions, and the QM affinity order about ligand–AA interactions are the input data for the performance of this approach. The ligand position inside the predicted binding site can be improved, henceforth (i.e by molecular dynamic simulations), without to explore the whole protein structure. Additionally, this study allows the understanding of the most significant factors involved in ligand binding, which will be discussed in the paper for the case of flavonoids.

**Scheme 1.** General Methodology for the Prediction of Potential Sites for Ligand Binding, Applying Quantum Chemical Information



The methodology is outlined as (see Scheme 1):

- 1 Screening and analysis of the most likely cavities in the protein crystal structure.
- 2 Analysis and evaluation of the AA that occur in each selected receptor cavity, taking into account ligand–AA QM information.
- 3 Ranking process of the computed binding sites.

Detailed information about these general steps is exposed as follows:

**2.1. Screening of the Cavities in the Protein Crystal Structure.** In this step, the computational procedure explores the 3D protein structure. In order to select those protein cavities that could be suitable for hosting the ligand, the algorithm follows this order:

- 1 Preprocessing of the protein crystal structure.
- 2 Calculation of the virtual centers for binding (VCB).
- 3 Clustering process of the VCB.

**2.1.1. Preprocessing of the Protein Crystal Structure.** First, a particular reduction of the protein complexity is achieved in this step by calculation of the center of mass of each AA, instead of using the explicit atomic information. This solution reduces dramatically the computational time.

In addition, because the lateral chain atoms should be analyzed in detail, the calculation of the center of mass is ponderated by a positive weight (eq 1):

$$X_{CM}(AA) = \frac{\sum_{i=1}^4 x_i + \mu \sum_{j=5}^N x_j}{4 + \mu \sum_{j=5}^N j} \quad (1)$$

where  $x_i$  is the Cartesian coordinate of the backbone atoms in the protein (the first four atoms of each AA in the Protein Data Base [PDB]),  $x_j$  is the Cartesian coordinate of the other AA atoms,  $N$  is the total number of atoms for each AA reported in PDB, and  $\mu$  is a weight coefficient applied to those atoms of the lateral chain of each AA. This allows the displacement of the center of mass of each AA to the zone of the lateral chain of the AA. A similar equation can be obtained for the coordinates  $Y_{CM}(AA)$  and  $Z_{CM}(AA)$ .

Next, the procedure defines the 3D protein boundary for the calculation.  $X_{CM}$ ,  $Y_{CM}$  and  $Z_{CM}$  set up the coordinates of centers of mass for each AA:  $X_{CM} = \{x_1, x_2, x_3, \dots, x_N\}$ ,  $Y_{CM} = \{y_1, y_2, y_3, \dots, y_N\}$ , and  $Z_{CM} = \{z_1, z_2, z_3, \dots, z_N\}$ , where  $N$  is the number of AA. The calculated centers of mass are inserted in a box with dimensions  $(d_x, d_y, d_z)$ , where  $d_x = \text{lmax}\{X_{CM}\} - \text{min}\{X_{CM}\} + 2\xi$ ,  $d_y = \text{lmax}\{Y_{CM}\} - \text{min}\{Y_{CM}\} + 2\xi$ , and  $d_z = \text{lmax}\{Z_{CM}\} - \text{max}\{Z_{CM}\} + 2\xi$ . The term “ $\xi$ ” is a positive value that allows the calculations in the surface of the protein.

**2.1.1.2. Identification of VCB.** The VCB are suitable cavities in the previously defined box  $(d_x, d_y, d_z)$ , which have general and basic conditions for hosting the ligand. The calculation of all the potential VCB starts with the partition of the 3D boundary of the protein (Section 2.1.1) in small cubes with dimensions  $\lambda = 1 \text{ \AA}$ . This partition takes into account the van der Waal (vdW) limits among AAs in the binding site as well as the maximum dimensions of the specific ligand to study. This excludes those cavities without proper geometrical dimensions for hosting this ligand. For the current analysis, the ligands (flavonoids experimentally elucidated) were removed from both the binding site and the protein structures.

The VCB, located in the vertex of the small cubes, should fulfill the following conditions: each VCB should be surrounded by a predefined number of AA inside the volume between two hyperspheres with radius  $[R_{\max}, R_{\min}]$ . In this research, for all processed proteins, the optimum value of the radius was obtained in the interval  $[4, 6] \text{ \AA}$  and the number of AA between 4 and 5. These parameters were defined taking into account the minimal distance between two centers of mass calculated for all the proteins and considering geometrical distribution of AA around an isolated VCB. As a result of this step, a set of VCB is obtained, where  $NV$  is the number of virtual centers.

**2.1.1.3. Clustering of VCB.** This step defines the main concentration of VCB in the protein as data to be used in the last step of the methodology (the QM step). A set of  $V_{CB} = \{v_1, v_2, v_3, \dots, v_{NV}\} \in \mathcal{R}^3$  (calculated in Section 2.1.1.2) is considered as input data. The cluster analysis (also called segmentation analysis) creates groups of VCB (or clusters)  $C = \{c_1, c_2, c_3, \dots, c_{NC}\} \in \mathcal{R}^3$ , and each cluster conforms a family of virtual centers which fulfill the following condition:

( $d(i, j) \leq \varepsilon$ , where  $i, j \in C \forall i \neq j$ ). The threshold  $\varepsilon$  (in Å) is taken as a maximal line segment connecting two nonadjacent vertices of the small cube used for the partition of the volume.

The new information as a set of clusters ( $C$ ) allows to carry out a number of automatic operations to deduce potential receptor cavities for hosting the ligand. These operations are summarized as follows:

Condition 1: The computed VCB clusters that resemble the ideal VCB cluster for hosting the ligand shall be selected, if  $NV_C > 0.3NV_{CL}$ .  $NV_C$  is the number of VCB in each cluster of set  $C$ , and  $NV_{CL}$  is the number of VCB that should have an ideal cluster (CL) calculated for the specific ligand. CL is calculated in an approximate way, taking into account the geometry of the ligand and the partition step  $\lambda$  (Section 2.1.1.2). This condition gives a subset of  $C S = \{s_1, s_2, s_3, \dots, s_N\} \in \mathcal{R}^3$ , where  $N$  is the number of selected clusters, and  $NV_s$  can be defined as a number of VCB for each selected cluster.

Condition 2: This step compares the maximum dimension of the ligand with the maximum distance between the VCB for each cluster. The calculation of  $d_{max}(i, j)_s$ ,  $i \neq j$  for the selected VCB clusters (after Condition 1) is carried out, where  $d_{max}(i, j)_s$  is the maximum distance between the VCB that belongs to the selected clusters and  $d_L$  is the maximum characteristic length of the ligand. All clusters which fulfill this condition are considered for the next calculation. Nevertheless, those clusters which belong to the interval:  $[d_L - 3\lambda, d_L + 3\lambda]$  are possible cavity(ies) for binding geometrically. As a partial result so far, the number of clusters to be analyzed as probable cavity(ies) for binding is reduced. Condition 3: This compares the linear density of all the VCB clusters (as the number of VCB of each selected VCB cluster) with respect to the maximum ligand dimension. This VCB linear density calculation for the selected clusters (after Condition 1) and for the ideal cluster as a function of the ligand dimension is performed. For a selected cluster:  $\delta_s = NV_s/d_{max}$ , ( $d_{max} = \max\{(d_s(i, j)), i \neq j\}$ ),  $NV_s$  is the number of virtual centers of the selected cluster. The linear density for the ideal cluster is calculated as  $\delta_L = NV_{CL}/d_{Lmax}$ , where  $d_{Lmax}$  is the maximum calculated distances between the virtual centers in the ideal cluster ( $d_{Lmax} = \max\{(d_{VC(L)}(i, j))\}$ ). So far, those calculated clusters with  $\delta_s \geq 0.4\delta_L$  state an additional circumstance to consider a cavity for hosting the ligand.

Condition 4: *The application of this last condition guarantees a proper distribution of VCB around the center of mass of all the selected clusters.* The area and the VCB population distribution ( $P_{S,i}$ ) around the center of mass (CM<sub>s</sub>) of all the selected clusters is considered as hyperspheres  $R_i = (R_{max} - R_{min}/2d_L)i$ , where  $i = 1, 2, \dots, n$  and  $n$  is the number of hyperspheres that belong to the interval  $R_{min} = 0$ ,  $R_{max} \leq d_L$ .  $R_{max}$  is taken as the maximum length of the studied ligand. It defines the distribution function, and therefore, it is possible to calculate its area as  $S_{VC}(s) = (\sum_{i=1}^{Np} P_{S,i}\lambda)/Np$ , where  $Np$  is the number of concentric hyperspheres or partitions. In case of the ligand, a similar equation is obtained. As a result of this condition, the selected clusters are considered as potential cavities for binding if:  $S_{VC}(s) \geq 0.7S_{VC}(\text{CL})$ .

As you can observe, the conditions have additive characteristics (2, 3, and 4 are complements of Condition 1) and

contribute to a proper definition of the potential cavities for a binding site.

**2.2. QM Analysis of the Ligand-AA Interaction for the Selected Cavities.** At this instant, the QM information of the interaction ligand-AA is explicitly considered. The calculation of the thermodynamic association magnitudes (energies and entropies) related to the molecular association between each ligand and each of the 20 AA of proteins should be performed and represented by the TAO. The current study (with flavonoids) is based on the application of MMH-1<sup>37</sup> or MMH-2<sup>38</sup> procedures, improved by the inclusion of the entropic influence.

There are several local minima produced in the intermolecular interaction configuration space. Each ligand-AA pair was studied through the calculation of hundreds of random configurations (clusters) of the AA surrounding the ligand with MMH procedures.<sup>37</sup> Such clusters are fully optimized by a standard gradient path minimization, using the AM1 semiempirical Hamiltonian in our case (proved accurate results for calculation in flavonoids<sup>37,40</sup> and prediction of local minima in crystal structure<sup>38</sup>). The energy,  $\varepsilon_i$ , of every cluster is, thus, obtained. Next, a Boltzmann distribution is used in order to calculate the thermally averaged state of the typical macroscopic system at room temperature. A partition function should be calculated, assuming the appropriate energy scale with respect to a reference value. The conventional state for calculating such reference energy is chosen as a state where the interaction energies, among the constituent molecules in a given cluster, are excluded. Then, the translational, rotational, and vibrational states of molecules in the clusters are considered identical to those in the reference states. Therefore, the association process is considered to be isothermal. In this case, the reference state is taken as a set with the same number of noninteracting molecules of the same kind, so that the sum of their total energies is taken as the reference value on our energy scale. In the reference state, the partition function  $q = 1$ . Accordingly, the cluster energy, with respect to the new reference scale,  $\Delta\varepsilon_i$  is  $\Delta\varepsilon_i = \varepsilon_i - \varepsilon^{\text{ref}}$ , where  $\varepsilon^{\text{ref}} = \varepsilon_{\text{tot}}(\text{ligand}) + n\varepsilon_{\text{tot}}(\text{AA})$ . Consequently, the molecular partition function in such a reference scale is  $q^* = \sum_i g_i e^{-(\Delta\varepsilon_i/RT)} = q e^{\varepsilon^{\text{ref}}/RT}$ , where  $g_i$  defines the process of degeneration for the studied configurations.

Using common statistical formulas, it is possible to calculate important thermodynamic magnitudes (eqs 2–4) such as the ligand-AA association energy values ( $E^{\text{assoc}}$ ), the association entropies ( $S^{\text{assoc}}$ ), and the free Helmholtz association energies ( $A^{\text{assoc}}$ ).  $E^{\text{assoc}}$  has been used as an important magnitude for the selection of the most stable ligand-AA complexes and for the generation of a entropic corrected TAO.<sup>37,40</sup>

However, in the comparison step of the theoretical data with the crystallographic data of flavonoid–protein interactions is demonstrated that those AA with higher entropy values in the interactions with the flavonoid ligand are crucial in the stabilization of the ligands inside the binding site. The explanation lies in the fact that the values of  $E^{\text{assoc}}$  (eq 2) for each complex do not consider the history of energy values for each cluster. But the values of  $S^{\text{assoc}}$  (eq 3) have a direct relation with the number of degenerated configurations in the system. Therefore, high values of  $S^{\text{assoc}}$  imply the existence of several different configurations of local minima.

Then, the  $E^{\text{assoc}}$  value for certain ligand–AA complexes provides information about the thermodynamic stability of the complex, while the  $S^{\text{assoc}}$  value announces whether the interaction is highly selective (low local minima and low  $S^{\text{assoc}}$  value) or nonselective (several local minima and high  $S^{\text{assoc}}$  value).

$$E^{\text{assoc}} = E - E^{\text{ref}} = RT^2 \frac{q^{*'}}{q^*} \quad (2)$$

$$S^{\text{assoc}} = S - S^{\text{ref}} = R \ln q^* + \frac{E^{\text{assoc}}}{T} \quad (3)$$

$$A^{\text{assoc}} = A - A^{\text{ref}} = -RT \ln q^* \quad (4)$$

Next, the descriptor binding local site ( $B_{\text{LS}}$ ) is calculated for each VCB (eq 5), where  $j = 1, 2, \dots, NV_s$ ,  $NV_s$  is the number of VCB of the set  $V_{\text{CB}} = \{v_1, v_2, v_3, \dots, v_{NV}\} \in \mathcal{R}^3$ . The algorithm considers the AAs surrounding the VCB and calculates the affinity average of the vicinity for the specific ligand.

$$B_{\text{LS}}(j) = \frac{\sum_{i=1}^N \text{TAO}(i)}{N_{\text{part}}} \quad (5)$$

$\text{TAO}(i)$  is the amplitude of each AA  $i$  given by the values of association energies and entropies previously calculated in their interactions with the ligand,  $N_{\text{part}}$ , it is a positive number that defines the number of partitions between two concentric spheres that define the interval  $[R_{\max}, R_{\min}]$  with a center on each calculated VCB (Section 2.1.1.2). In this calculation, only the centers of mass of those AA that belong to the defined interval are considered. Therefore, each VCB will have its own  $B_{\text{LS}}$  value. The distribution of  $B_{\text{LS}}$  vs VCB allows to define maximum values or more probable site(s) for binding.

**2.3. Ranking Process of the Computed Binding Sites.** In the last step, a ranking process of the computed binding sites is performed through the combination of the result of Conditions (1–4) analyzed in the clustering calculation (Section 2.1) and those sites with higher values of  $B_{\text{LS}}$  (Section 2.2). The clusters of VCB, which includes VCB with higher values of  $B_{\text{LS}}$ , will be selected as potential binding sites and ranked by the  $B_{\text{LS}}$  values. This step guarantees good results in the prediction of the potential site(s) for ligand binding. The general methodology is summarized in Scheme 1.

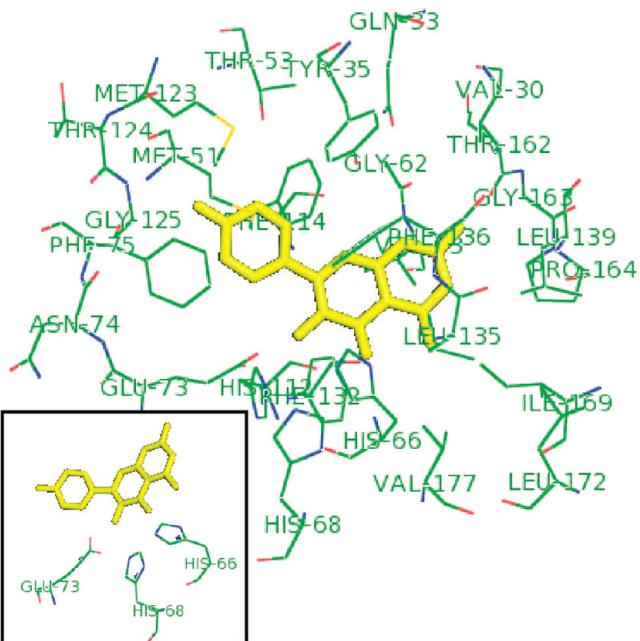
Pymol<sup>41</sup> and Chimera<sup>42</sup> were the softwares used for the preparation of the analyzed proteins. The potentialities of Pymol were particularly used as a 3D platform in the presentation of the results.

### III. RESULTS AND DISCUSSION

#### 3.1. QM Crosstalk between Flavonoids and Proteins.

The crystallographic structure of 10 flavonoid binding proteins were collected from PDB (1H1M,<sup>43</sup> 1JNQ,<sup>44</sup> 1SG0,<sup>45</sup> 1X7J,<sup>46</sup> 2HCK,<sup>47</sup> 2JJ2,<sup>48</sup> 2O3P,<sup>49</sup> 2O65,<sup>49</sup> 2UXU,<sup>50</sup> 3BPT,<sup>51</sup> and 3CKL<sup>52</sup>). The structures of solvents, ligands and other molecules were removed before the application of the methodology.

The highest catechin<sup>37</sup> and quercetin association energy values were predicted for the hydrophilic AA (Lys, Glu, Asp,



**Figure 1.** Binding site for 1H1M; evidence that Glu 73, His 66, and His 68 interacting with the flavonoid ligand.

His, and Arg). The exploration of 10 flavonoid binding proteins confirms that the stronger interactions hydrogen bonds (H-bonds) among the ligand and the AA are established with the hydrophilic AA (Figure 1). In addition, other hydrophobic residues are also important for the stabilization of the ligand in the binding sites.

The flavonoid–AA interactions analysis based on QM calculations as well as the AA composition of the 10 flavonoid binding sites suggest that the association entropy plays an important role in this substrate binding recognition. This thermodynamic magnitude becomes important because this methodology produces a set of local minima in the intermolecular interaction configuration space.

In the case of flavonoids, the hydrophilic AA (Lys, His, Arg, Glu, and Asp) have the higher  $E^{\text{assoc}}$  but the lower  $S^{\text{assoc}}$  values. This implies a very high selective association between the flavonoid molecules and the AA. As we demonstrated before (Figure 1), these AAs in the binding site are extremely important because they have the stronger interactions with the flavonoid molecule. However, from the comparisons of the crystallographic data, it is deduced that the other AA which have higher  $S^{\text{assoc}}$  values in the interaction with the flavonoid ligand are important for the stabilization of this ligand in the protein cavity. Therefore, protein–flavonoid interactions are dominated by those AA that do not require any specific conformation or functional group for the association with the flavonoid ligand. Higher entropies in these associations mean that many probable positions are favoring these interactions. These data contribute with the understanding of flavonoid interactions in the protein binding sites.

From the analysis based on theoretical and experimental knowledge, it is possible to conclude that flavonoids need two or three hydrophilic AAs in the binding site (i.e., Lys, His, Arg, Glu, and Asp) for H-bond interactions, and the other AA are used to stabilize the ligand through stacking interactions in the binding site.

**Table 1.** AA Abundance (%) in the Known Binding Sites of the Studied Proteins<sup>a</sup>

PDB code	group I	group II	group III
1H1M	31.97	52.32	15.69
1JNQ	24.97	48.29	26.73
1SG0	23.91	52.61	23.48
1X7J	20.85	54.46	24.68
2HCK	24.71	46.91	28.37
2JJ2	19.96	54.48	25.18
2O3P	22.99	47.89	29.19
2O65	22.99	47.89	29.19
2UXU	24.56	52.42	33.00
3BPT	22.16	49.71	28.12
3CKL	22.97	45.95	31.08

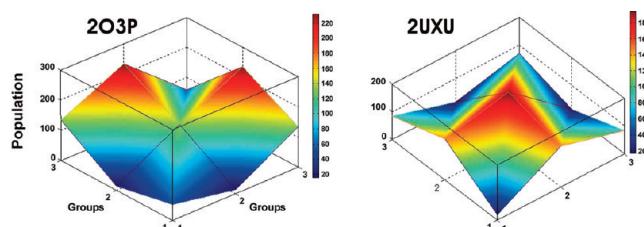
<sup>a</sup> Normalized to the number of AA in each group.

Following this criteria, the 20 AAs were divided into three groups. The AAs from Group III have the higher association energy in the interactions with flavonoids. The AAs from Group II have the higher association entropy values and should be the majority in the binding sites, and the AA from the Group I have the lower interactions.

Group III: Lys (K), His (H), Arg (R), Glu (E), and Asp (D). Group II: Ile (I), Thr (T), Val (V), Trp (W), Gly (G), Leu (L), Met (M), Phe (F), and Ala (A).

Group I: Gln (Q), Ser (S), Pro (P), Tyr (Y), Asn (N), and Cys (C).

Table 1 shows that the AAs from Group II are the majority in the studied binding sites. Figure 2, additionally, evidences that the AAs from Group II are in the vicinity of the other AA groups, which reinforces our assumption. This result confirms the hydrophobic characteristics of the flavonoid binding site, where quantum chemical analysis must be performed to analyze the effect of weak interactions, like VdW.<sup>53</sup> Therefore, the created algorithm, in order to predict the potential flavonoid binding sites in proteins, was trained to perform the ranking process for the created VCB looking for those binding sites that have two or three AAs from the

**Figure 2.** Vicinity of AA groups for two of the analyzed proteins.

Group III and the majority from the Group II. Experimental data of interactions between human serum albumina (HSA) and flavonoids confirm very outstanding the predictions, as they report that Tyr residues are not involved in a direct interaction with the flavonoid.<sup>54</sup> However AA residues Asp, Ala, Trp, Lys, and Gln are directly involved in the binding.<sup>54</sup>

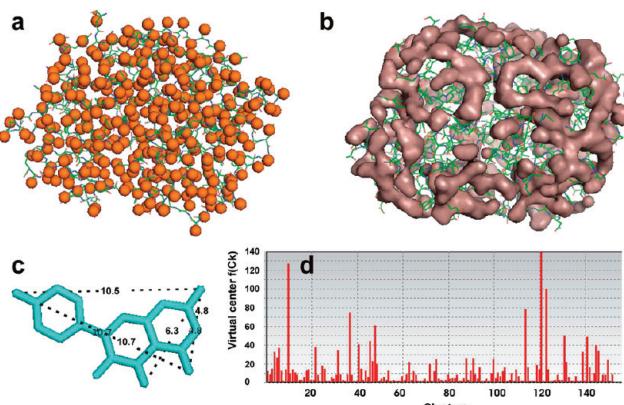
**3.2. General Results in the Prediction of the Potential Sites for Flavonoid Binding.** Table 2 shows the AAs present in the binding site per protein (column three) and those predicted as the most probable binding site (column two), following Scheme 1. This is the major validating proof of our hypothesis about the general features that should have a protein cavity for flavonoid binding. The comparison between these columns evidence that the predicted AA residues are the majority of the AA presented in the known binding sites of these proteins (note that the coincident AAs are highlighted in column three). This was also confirmed with the calculation of the percent of predicted AA in column four.

Other cavities (around 5–7 per protein) are also predicted as potential sites for flavonoid binding, but in this paper, only the most probable site for binding is shown, which agrees with the experimental findings. This is the result of the ranking process which combines the selection of those proper VCB that have simultaneously higher values of  $B_{LS}$  (as a QM descriptor, eq 5) and have proper results in the clustering calculation. It is important to remark that the results

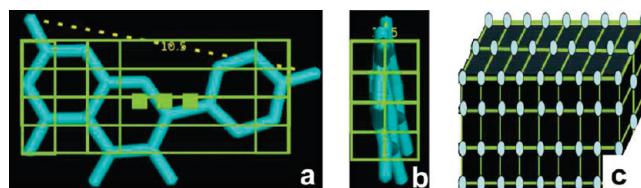
**Table 2.** Comparison between the AA in the Predicted and Known Binding Sites<sup>a</sup>

PDB code	predicted site	known site	predicted AA (%)
1H1M	V 63.A, H 66.A, H 68.A, E 73.A, F 75.A, H 112.A, F 114.A, F 132.A, L 135.A, F 136.A, V 177.A	H 112.A, M 51.A, V 63.A, H 66.A, H 68.A, E 73.A, F 75.A, H 112.A, F 114.A, G 125.A, F 132.A, L 135.A, F 136.A, L 139.A, G 163.A, P 164.A, V 177.A	75
1JNQ	L 515.A, H 518.A, W 519.A, H 523.A, I 557.A, L 565.A, V 566.A, L 773.A, I 857.A	N 514.A, L 515.A, H 518.A, W 519.A, H 523.A, I 557.A, L 565.A, V 566.A, I 572.A, F 576.A, H 709.A, Q 716.A, V 769.A, L 773.A, I 857.A	69
1X7J	E 305.A, L 339.A, R 346.A, I 376.A, H 475.A, L 476.A, M 479.A	M 295.A, L 298.A, L 301.A, E 305.A, L 339.A, R 346.A, I 373.A, I 376.A, H 475.A, L 476.A, M 479.A	64
2HCK	L 273.B, G 274.B, A 275.B, T 338.B, L 393.B, A 403.B, D 404.B	L 273.B, G 274.B, A 275.B, V 281.B, A 293.B, T 338.B, E 339.B, M 341.B, G 344.B, D 348.B, L 393.B, A 403.B, D 404.B	54
2JJ2	R 291.B, A 256.G, T 259.G, K 260.G, I 263.G, E 264.G, E 292.C, V 279.F	R 291.B, G 292.B, E 292.C, V 279.F, A 256.G, T 259.G, K 260.G, I 263.G, E 264.G	89
2O3P	F 49.A, A 65.A, K 67.A, I 104.A, L 120.A, E 121.A, R 122.A, I 185.A, D 186.A	L 44.A, F 49.A, A 65.A, K 67.A, E 89.A, I 104.A, L 120.A, E 121.A, R 122.A, P 123.A, V 126.A, L 174.A, I 185.A, D 186.A	64
2O65	F 49.A, A 65.A, K 67.A, I 104.A, L 120.A, E 121.A, R 122.A, I 185.A, D 186.A	L 44.A, F 49.A, A 65.A, K 67.A, E 89.A, I 104.A, L 120.A, E 121.A, R 122.A, P 123.A, V 126.A, L 174.A, I 185.A, D 186.A	64
2UXU	L 66.A, H 67.A, A 74.A, L 92.A, V 96.A, H 114.A, F 168.A, I 175.A	L 66.A, H 67.A, A 74.A, L 92.A, L 93.A, V 96.A, N 110.A, H 114.A, I 141.A, M 167.A, F 168.A, V 71.A, I 175.A	61
3BPT	F 56.A, A 59.A, G 98.A, D 99.A, L 350.A	F 56.A, L 57.A, A 59.A, G 97.A, G 98.A, D 99.A, I 100.A, P 168.A, I 172.A, V 349.A, L 350.A	45
3CKL	T 76.B, M 81.B, M 84.B, R 90.B, K 107.B, H 109.B, F 143.B, L 149.B, L 244.B, V 248.B	T 21.B, F 24.B, T 76.B, M 81.B, M 84.B, R 90.B, K 107.B, H 109.B, F 143.B, L 149.B, Q 150.B, Y 170.B, L 244.B	77

<sup>a</sup> Column four: predicted AAs in %.



**Figure 3.** Clustering of 1H1M protein. (a) Center of mass representation for residues; (b) Reconstructed surface for all calculated cluster; (c) Flavonoid ligand (KMP); and (d) plot  $N_{VC}$  vs Clusters.

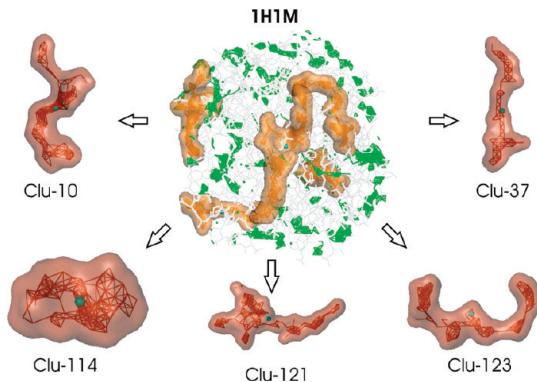


**Figure 4.** Definition of the ideal cluster geometry for the ligand KMP. (a) Frontal and (b) lateral views of the ligand. (c) Virtual centers in the ideal configuration for the ligand KMP (each virtual center is shown as blue spheres).

obtained for the proteins, 2O3P and 2O65, are a good evidence of the reproducibility of this methodology. They were treated as different proteins, and the result was the same. A deeper analysis of both structures showed out that the proteins are the same and that the differences lie in the bonded ligand, in one case a quercetin ligand and in the second case a pentahydroxyflavanone molecule. An additional potentiality of this methodology is the possibility to predict the access way of the ligand to the binding site. It shall be discussed as well.

**3.3. Examples of Some Studied Cases (Crystallographic Structures 1H1M and 3CKL).** This section shows some examples of the applied procedure for two of the studied proteins. Protein 1H1M has a reported binding site for flavonoids inside the protein core,<sup>23</sup> which makes the cavity location difficult. Other analyzed proteins such as 1X7J, 3CKL, and 1JNQ have a similar situation. The application of this stepwise procedure (Scheme 1) provides a fast way to achieve the location of these internal cavities.

Figure 3a shows the position of the center of mass of all the residues, and Figure 3 b represents all the obtained VCB clusters  $C = \{c_1, c_2, c_3, \dots, c_{NC}\} \in \mathcal{R}$ . The 157 clusters for this protein



**Figure 5.** Geometry of the selected VCB clusters of protein 1H1M in the ranking process and their centers of mass.

(Figure 3d, plot  $N_{VC}$  vs Cluster) shows a huge diversity of geometries, and it is not evident to select the suitable host cluster for the ligand. Conditions 1–4 (Section 2.1.1.3) were successfully applied. KMP ligand dimensions were used as a reference for the calculation. Figure 4 shows the ideal geometry considered for the KMP ligand geometry (1H1M ligand).

Five clusters (1H1M-Clu-10, 1H1M-Clu-37, 1H1M-Clu-114, 1H1M-Clu-121, 1H1M-Clu-123) were selected as optimum geometrical cavities for hosting the ligand, after the application of Condition 1. Table 3 shows the values of the calculated descriptors (Conditions 1–4) for the selected clusters. Considering the reported values in this table, 1H1M-Clu-114 seems to be the potential site for binding flavonoids in this protein. It fulfills all the imposed Conditions 1–4. Nevertheless, all the obtained clusters are analyzed in details for further comparisons. Figure 5 shows the geometry of the selected clusters and their centers of mass. The geometry of these clusters is in agreement with the descriptors presented in Table 3.

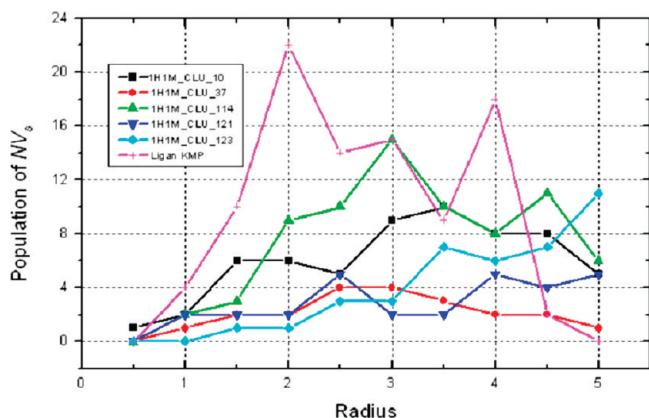
Figure 6 shows the population distribution of  $N_{VC}$  (virtual sites for each selected cluster) as a function of  $i$  hyperspheres, with centers in the center of mass of each selected cluster. The plot for the ligand KMP is the ideal behavior. As it can be observed, the plot of 1H1M-Clu-114 has a very similar behavior to the ligand ideal plot, which is a good indicator for the potential binding site definition. Cluster 1H1M-Clu-114 seems to be the most likely site for KMP binding so far, after the correlation of the presented results so far.

The final step in the ranking process is the analysis of the superimposed structures of the selected clusters with those VCB that have higher values of the  $B_{LS}$  descriptor (QM information). Figure 7 evidences a good overlap between the cluster 1H1M-Clu-114 and the maximum values of the  $B_{LS}$  descriptor. Cluster 1H1M-Clu-114 has been selected as the most likely site for flavonoid binding in this protein

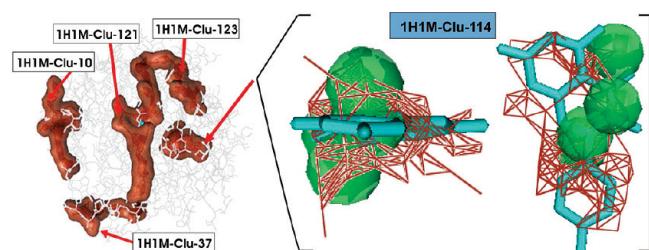
**Table 3.** Calculated Indicators (Conditions 1–4, Section III) for the Five Selected Clusters and the Ligand (KMP)<sup>a</sup>

cluster	$N_{VC}$	$d_{max}, \text{ \AA}$	linear density, $1/\text{\AA}$	$S_{VC}(ck)$
1H1M-Clu-10	126	<b>21.21</b>	5.94	6
1H1M-Clu-37	74	<b>22.63</b>	<b>3.27</b>	<b>2.1</b>
1H1M-Clu-114	77	10.44	7.37	7.4
1H1M-Clu-121	139	27.46	5.09	<b>3.9</b>
1H1M-Clu-123	99	<b>19.42</b>	5.10	<b>2.9</b>
KMP	93	9.43	9.86	9.40
	( $N_{VC} > 65$ )	( $d_{max} \geq 9.43 \pm 3$ )	( $\delta_{ck} \geq 3.94$ )	( $S_{VC}(s) \geq 6.58$ )

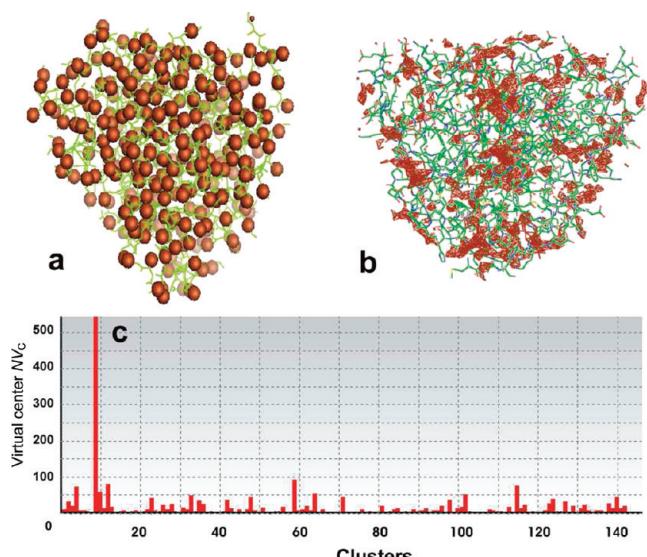
<sup>a</sup> The highlighted values are the clusters that do not fulfill the conditions.



**Figure 6.** Population distributions of NVs as a function of radius, with the center of mass of each selected cluster.



**Figure 7.** Cluster 1H1M–Clu-114 has the highest value of BLS (green spheres, QM information) among all the analyzed VCB clusters.



**Figure 8.** Clustering of 3CKL protein. (a) Center of mass representation for residues, (b) reconstructed surface for all calculated clusters, and (c) plot  $NV_c$  vs clusters.

structure, and this result is in agreement with the experimental evidence.

**Protein 3CKL.**<sup>34</sup> The geometrical characterization of the 3CKL cavities considers the dimensions of the experimental ligand STL as reference (Figure 8). The 146 clusters were obtained for this protein (Figure 8c, plot  $NV_c$  vs cluster). Such as in the previous case of the 1H1M protein, from this information it is not evident to select those suitable clusters for hosting the ligand. Therefore, Conditions 1–4 were applied, considering the STL (experimental ligand code) dimensions as a reference.

The application of Conditions 1–4 selected five clusters (3CKL–Clu-4, 3CKL–Clu-9, 3CKL–Clu-12, 3CKL–Clu-59, and 3CKL–Clu-115) from a total of 146. Table 4 shows the values of the calculated descriptors. Once the conditions are fulfilled (Table 4), further analysis must be carried out. The population distribution of  $NV_s$  (virtual sites for each selected cluster), as a function of  $i$  hyperspheres with centers in the center of mass of each selected cluster, is shown as well (Figure 9). The plot for the ligand STL presents the best behavior. The population distributions of clusters 3CKL–Clu-4, 3CKL–Clu-12, and 3CKL–Clu-115 have a smooth asymmetric behavior, which is an indication of a small area for hosting the ligand (Figure 9). In the case of cluster 3CKL–Clu-59, its behavior is close to the STD model. Therefore, taking into account this result, it is possible to infer that cluster 3CKL–Clu-59 can be preselected as a good candidate for a likely site for STL binding. Cluster 3CKL–Clu-9 has a growing trend of this function, which could be an indication that in this long structure there is not a suitable cavity for a defined ligand in the vicinity of the center of mass. Figure 10 shows, additionally, the geometry of the selected clusters and the centers of mass, which have a coincidence with the previous analysis.

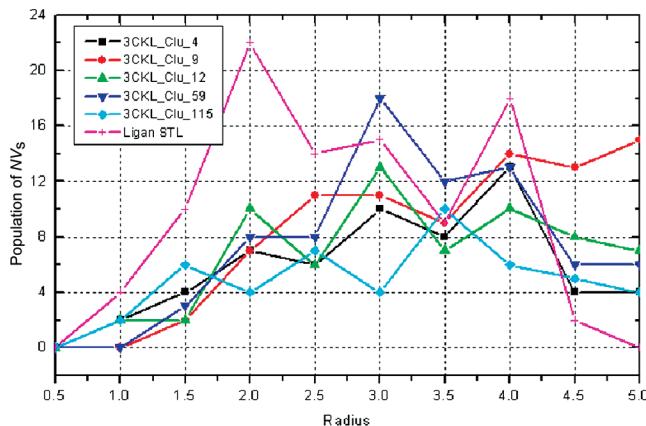
After the ranking process, clusters 3CKL–Clu-9, 3CKL–Clu-59, and 3CKL–Clu-115 were found as likely sites for STL binding. Clusters 3CKL–Clu-4 and 3CKL–Clu-12 were excluded because in the vicinity of these clusters does not exist any site with a high value of  $B_{LS}$ . These last clusters were analyzed in detail in order to select the most likely site for STL binding. The clusters 3CKL–Clu-115 and 3CKL–Clu-59 were selected as the most improvable sites for binding because even when they have maximum values of  $B_{LS}$  descriptor, the clusters are located on the protein surface and not enough space is located around these VCB for hosting the ligand (Figure 11). The characterization of the 3CKL protein predicts the Clu-9 as the most likely site for flavonoids binding in this protein. Figure 12 show details of this cluster, which covers a big surface (deduced from Figure 10) as well as has two areas with high values of  $B_{LS}$  (areas I and II, Figure 12b). Segmenting the cluster, area II has the possibility to host the ligand due to local geometrical considerations (Figure 12c1 and c2). Therefore, after the analysis of the AA that are surrounding the center of mass of this cluster and the AA that occur in the experimental binding site (Table 2), it is possible to conclude that this result is in agreement with the experimental evidence.

**3.4. Ability to Detect the Possible Protein Area for the Ligand Access to the Binding Site.** If the protein has a direct space (with ligand dimensions), from the surface to the binding site or predicted cavity, the algorithm will be able to detect it. This suggests how the ligand can access to the binding site. The analyzed protein 3CKL is a good example to this statement. Among the five clusters presented in Table 4, one of them has a bigger quantity of VCB: cluster 3CKL–Clu-9. As it was explained above, the most likely cavity for flavonoid binding is located in one of these areas, but the dimension of this cluster is considerably large and can reach 546 virtual centers. The maximum distance of this cluster is 40.73 Å, which represents 75.09% of the maximal length (54.24 Å) founded between two center of masses of all the protein AAs. For this kind of protein (globular behavior), the calculated length suggests that this cluster

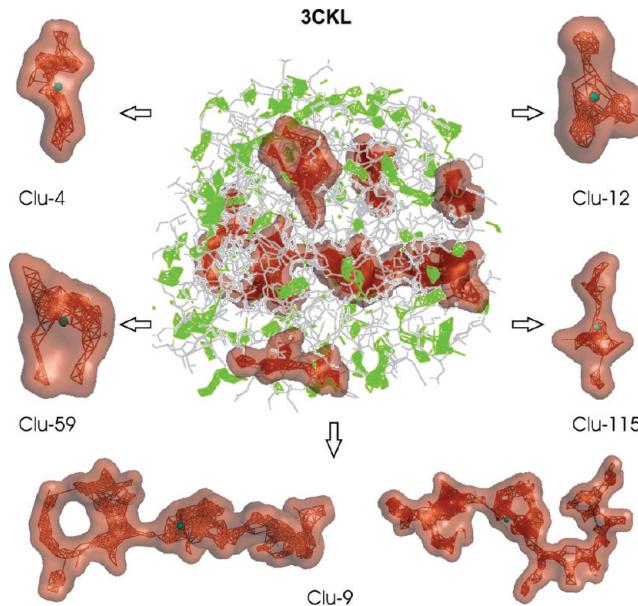
**Table 4.** Calculated Indicators (Conditions 1–4, Section III) for the Five Selected Clusters and the Ligand (STL)<sup>a</sup>

cluster	$NV_s$	$d_{\max}$ , Å	Linear density, 1/Å	$S_{VC}(s)$
3CKL_Clu_4	70	13.42	5.22	5.80
3CKL_Clu_9	546	40.73	13.41	8.20
3CKL_Clu_12	78	11.75	6.664	6.50
3CKL_Clu_59	89	13.19	6.75	7.40
3CKL_115	73	16.55	4.41	4.80
STL	93	9.43	9.86	9.40
	( $NV_s > 65$ )	( $d_{\max} \geq 9.43 \pm 3$ )	( $\delta_{ck} \geq 3.94$ )	( $S_{VC}(ck) \geq 6.58$ )

<sup>a</sup> The model for ideal geometry of STL coincide with KMP ligand.

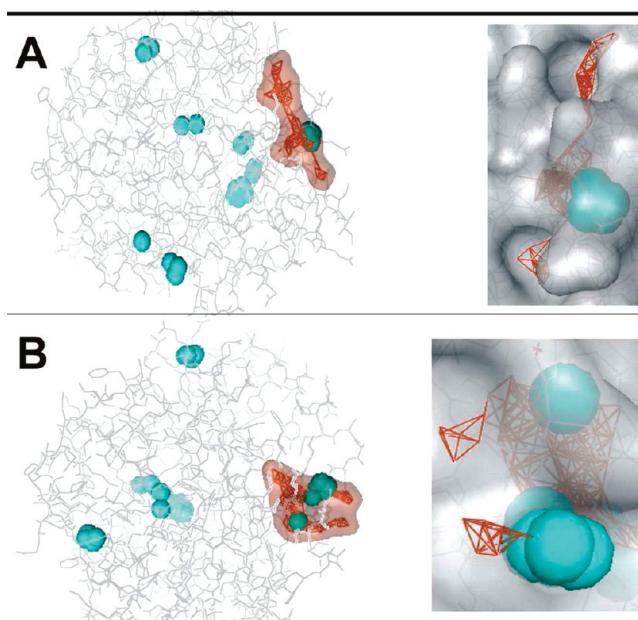


**Figure 9.** Population distribution of NVs as a function of radius, with the center of mass of each selected cluster.

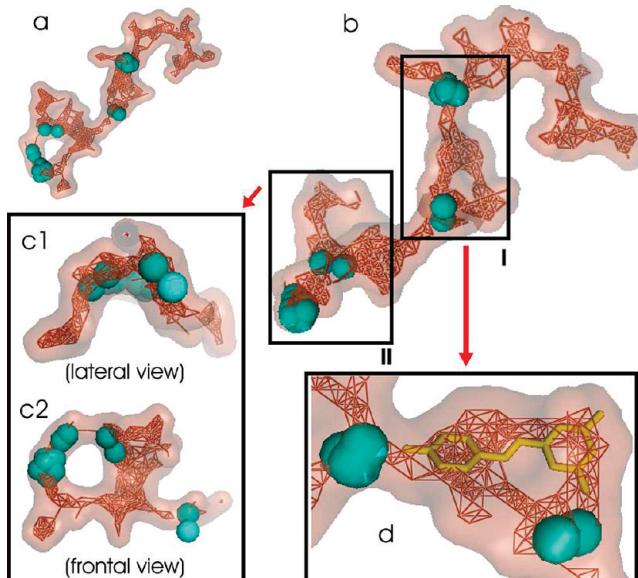


**Figure 10.** Geometry of the selected VCB clusters of protein 3CKL in the ranking process and their centers of mass.

travels through all over the protein structure, which can be corroborated from Figure 10. The presence of this kind of cluster can be taken as a possible path of access of the ligand to the core of protein. We propose, that in the presence of a cluster like 3CKL–Clu-9, the access of the ligand to the cavity would be initiated from the entrance that does not have a VCB with higher values of  $B_{LS}$  descriptor. In the case of Figure 12, it should be through the left (Figure 12, a and b) where there are not predicted sites with high values of  $B_{LS}$  (blue spheres). In this way, the ligand can be transported from the beginning of the cavity until the site where important AA for flavonoids binding will establish the



**Figure 11.** Combination of geometrical positions of cluster (A) Clu-115 and (B) Clu-59 in 3CKL and the maximum values of the  $B_{LS}$  descriptor.



**Figure 12.** Combination of geometrical position of cluster Clu-9 in the protein 3CKL as well as the maximum values of the  $B_{LS}$  descriptor.

stronger associations. It is important to remark that in the previous published paper<sup>26</sup> it was stated that all AA can form a stable complex with flavonoids. The difference lies in the strength of these associations which allows to use the TAO

and to propose, in this case, that flavonoids can “walk” until the end of the cavity, finally finding those AA to establish the stronger associations. This assumption is logically supported from steric effects and quantum chemical information.

## CONCLUSIONS AND REMARKS

The approach described above introduces predictions of potential sites for ligand binding, taking into account quantum chemical information of ligand–AA interactions. The predicted cavities per protein always agree with the binding site already proposed by X-ray experiments. The 50–90% of the amino acid residues occurring in the experimental binding site were properly predicted. Protein cavities which includes mainly either of the AAs: Ile, Thr, Val, Trp, Gly, Leu, Met, Phe, and Ala and includes only two or three of any hydrophilic AA, such as: Lys, His, Arg, Glu, and/or Asp will be favored as a potential site for binding flavonoids ligands. This new information about the general features of flavonoid–protein interactions contributes to the understanding of the bioactive behavior of these natural compounds. The novel approach is also able to detect the possible protein region for the ligand access from the protein surface to the binding site. This suggests a new bioinformatic strategy for binding site analysis phases of in silico drug design.

## ACKNOWLEDGMENT

Authors acknowledge Deutscher Akademischer Austausch Dienst (DAAD), German Academic Exchange Service, for the financial support of a three-month research grant for one author (A.R.-N.) in the Institute for Plant Genetics from the University of Hannover, Germany.

## REFERENCES AND NOTES

- (1) Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5* (4), 375–382.
- (2) Baroni, M.; Cruciani, G.; Scialoba, S.; Perruccio, F.; Mason, J. S. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for ligands and proteins (FLAP): Theory and application. *J. Chem. Inf. Model.* **2007**, *47* (2), 279–294.
- (3) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided. Mol. Des.* **2001**, *15* (5), 411–428.
- (4) Gane, P. J.; Dean, P. M. Recent advances in structure-based rational drug design. *Curr. Opin. Struct. Biol.* **2000**, *10* (4), 401–404.
- (5) Halperin, I.; Ma, B. Y.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Protein: Struct., Funct., Genet.* **2002**, *47* (4), 409–443.
- (6) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.
- (7) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Protein: Struct., Funct., Genet.* **1999**, *37* (2), 228–241.
- (8) Moreno, E.; Leon, K. Geometric and chemical patterns of interaction in protein-ligand complexes and their application in docking. *Protein: Struct., Funct., Genet.* **2002**, *47* (1), 1–13.
- (9) Waszkowycz, B.; Perkins, T. D. J.; Sykes, R. A.; Li, J. Large-scale virtual screening for discovering leads in the postgenomic era. *IBM Syst. J.* **2001**, *40* (2), 360–376.
- (10) Babaoglu, K.; Simeonov, A.; Lwin, J. J.; Nelson, M. E.; Feng, B.; Thomas, C. J.; Cancian, L.; Costi, M. P.; Maltby, D. A.; Jadhav, A.; Ingles, J.; Austin, C. P.; Shoichet, B. K. Comprehensive mechanistic analysis of hits from high-throughput and docking screens against beta-lactamase. *J. Med. Chem.* **2008**, *51* (8), 2502–2511.
- (11) Chen, Y. Z.; Zhi, D. G. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Protein: Struct., Funct., Genet.* **2001**, *43* (2), 217–226.
- (12) Gold, N. D.; Jackson, R. M. SitesBase: a database for structure-based protein-ligand binding site comparisons. *Nucleic Acids Res.* **2006**, *34*, D231–D234.
- (13) Gunther, J.; Bergner, A.; Hendlich, M.; Klebe, G. Utilising structural knowledge in drug design strategies: Applications using relibase. *J. Mol. Biol.* **2003**, *326* (2), 621–636.
- (14) Hendlich, M.; Bergner, A.; Gunther, J.; Klebe, G. Relibase: Design and development of a database for comprehensive analysis of protein-ligand interactions. *J. Mol. Biol.* **2003**, *326* (2), 607–620.
- (15) Kleywegt, G. J.; Jones, T. A. Databases in protein Crystallography. *Acta Crystallogr., D: Biol. Crystallogr.* **1998**, *54*, 1119–1131.
- (16) Laskowski, R. A. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.* **2001**, *29* (1), 221–222.
- (17) Powers, R.; Copeland, J. C.; Germer, K.; Mercier, K. A.; Ramanathan, V.; Revesz, P. Comparison of protein active site structures for functional annotation of proteins and drug design. *Proteins* **2006**, *65* (1), 124–135.
- (18) Puvanendampillai, D.; Mitchell, J. B. O. L/D Protein Ligand Database (PLD): additional understanding of the nature and specificity of protein-ligand complexes. *Bioinformatics* **2003**, *19* (14), 1856–1857.
- (19) Shin, J. M.; Cho, D. H. PDB-ligand: a ligand database based on PDB for the automated and customized classification of ligand-binding structures. *Nucleic Acids Res.* **2005**, *33*, D238–D241.
- (20) Smith, R. D.; Hu, L. G.; Falkner, J. A.; Benson, M. L.; Neroth, J. P.; Carlson, H. A. Exploring protein-ligand recognition with binding MOAD. *J. Mol. Graph. Model.* **2006**, *24* (6), 414–425.
- (21) David, L.; Nielsen, P. A.; Hedstrom, M.; Norden, B. Scope and limitation of ligand docking: methods, scoring functions and protein targets. *Curr. Comp.-Aided Drug Des.* **2005**, *1* (3), 275–306.
- (22) Korb, O.; Stutzle, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49* (1), 84–96.
- (23) Andrusier, N.; Mashiach, E.; Nussinov, R.; Wolfson, H. J. Principles of flexible protein-protein docking. *Proteins* **2008**, *73* (2), 271–289.
- (24) Chen, Y.; Shoichet, B. K. Molecular docking and ligand specificity in fragment-based inhibitor discovery. *Nat. Chem. Biol.* **2009**, *5* (5), 358–364.
- (25) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49* (23), 6789–6801.
- (26) Irwin, J. J. Community benchmarks for virtual screening. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3–4), 193–199.
- (27) Kowalsman, N.; Eisenstein, M. Inherent limitations in protein-protein docking procedures. *Bioinformatics* **2007**, *23* (4), 421–426.
- (28) Lang, P. T.; Aynechi, T.; Moustakas, D.; Shoichet, B.; Kuntz, I. D.; Brooijmans, N.; Oshiro, C. Molecular docking and structure-based design: New Frontiers in the Post-Genomic Era. In *Drug Discovery Research*; Huang, Z., Ed. Wiley: Hoboken, NJ, 2007; pp 3–23.
- (29) McGaughey, G. B.; Sheridan, R. P.; Bayly, C. I.; Culberson, J. C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J. F.; Cornell, W. D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.* **2007**, *47* (4), 1504–1519.
- (30) Sheridan, R. P.; McGaughey, G. B.; Cornell, W. D. Multiple protein structures and multiple ligands: effects on the apparent goodness of virtual screening results. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3–4), 257–265.
- (31) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49* (20), 5912–5931.
- (32) Raha, K.; van der Vaart, A. J.; Riley, K. E.; Peters, M. B.; Westerhoff, L. M.; Kim, H.; Merz, K. M. Pairwise decomposition of residue interaction energies using semiempirical quantum mechanical methods in studies of protein-ligand interaction. *J. Am. Chem. Soc.* **2005**, *127* (18), 6583–6594.
- (33) Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, B. S.; Johnson, A. P. eHITS: An innovative approach to the docking and scoring function problems. *Curr. Prot. Pep. Science* **2006**, *7* (5), 421–435.
- (34) Kellenberger, E.; Muller, P.; Schalon, C.; Bret, G.; Foata, N.; Rognan, D. sc-PDB: an annotated database of druggable binding sites from the protein data bank. *J. Chem. Inf. Model.* **2006**, *46* (2), 717–727.
- (35) Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; WollaCott, A. M.; Westerhoff, L. M.; Merz, K. M. The role of quantum mechanics in structure-based drug design. *Drug Discovery Today* **2007**, *12* (17–18), 725–731.
- (36) Montero, L. A.; Esteva, A. M.; Molina, J.; Zapardiel, A.; Hernandez, L.; Marquez, H.; Acosta, A. A theoretical approach to analytical properties of 2,4-diamino-5-phenylthiazole in water solution. Tautomerism and dependence on pH. *J. Am. Chem. Soc.* **1998**, *120* (46), 12023–12033.
- (37) Codorniu-Hernandez, E.; Mesa-Ibirico, A.; Hernandez-Santiesteban, R.; Montero-Cabrera, L. A.; Martinez-Luzardo, F.; Santana-Romero, J. L.; Borrmann, T.; Stohrer, W. D. Essential amino acids interacting with flavonoids: A theoretical approach. *Int. J. Quantum Chem.* **2005**, *103* (1), 82–104.

- (38) Codorniu-Hernandez, E.; Boese, A. D.; Schauerte, C.; Rolo-Naranjo, A.; Miranda-Quintana, R.; Montero-Cabrera, L. A.; Boese, R. MMH-2 as a new approach for the prediction of intermolecular interactions: the crystal packing of acetamide. *CrysEngComm* **2009**, *11* (11), 2358–2370.
- (39) Nandy, J. P.; Prakesch, M.; Khadem, S.; Reddy, P. T.; Sharma, U.; Arya, P. Advances in Solution- and Solid-Phase Synthesis toward the Generation of Natural Product-like Libraries. *Chem. Rev.* **2009**, *109* (5), 1999–2060.
- (40) Codorniu-Hernández, E.; Rolo-Naranjo, A.; Montero-Cabrera, L. A. Theoretical affinity order among flavonoids and amino acids: An approach to understand flavonoid protein-interactions. *J. Mol. Struct.: THEOCHEM* **2007**, *819* (1–3), 121–129.
- (41) DeLano, W. L. *The PyMOL Molecular Graphics System*, DeLano Scientific: San Carlos, CA, 2002.
- (42) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.
- (43) Steiner, R. A.; Kalk, K. H.; Dijkstra, B. W. Anaerobic enzyme-substrate structures provide insight into the reaction mechanism of the copper-dependent quercetin 2,3-dioxygenase. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (26), 16625–16630.
- (44) Skrzypczak-Jankun, E.; Zhou, K. J.; Jankun, J. Inhibition of lipoxygenase by (-)-epigallocatechin gallate: X-ray analysis at 2.1 angstrom reveals degradation of EGCG and shows soybean LOX-3 complex with EGC instead. *Int. J. Mol. Med.* **2003**, *12* (4), 415–422.
- (45) Buryanovskyy, L.; Fu, Y.; Boyd, M.; Ma, Y. L.; Hsieh, T. C.; Wu, J. M.; Zhang, Z. T. Crystal structure of quinone reductase 2 in complex with resveratrol. *Biochemistry* **2004**, *43* (36), 11417–11426.
- (46) Manas, E. S.; Xu, Z. B.; Unwalla, R. J.; Somers, W. S. Understanding the selectivity of genistein for human estrogen receptor-beta using X-ray crystallography and computational methods. *Structure* **2004**, *12* (12), 2197–2207.
- (47) Sicheri, F.; Moarefi, I.; Kuriyan, J. Crystal structure of the Src family tyrosine kinase Hck. *Nature* **1997**, *385* (6617), 602–609.
- (48) Gledhill, J. R.; Montgomery, M. G.; Leslie, A. G. W.; Walker, J. E. Mechanism of inhibition of bovine F-1-ATPase by resveratrol and related polyphenols. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104* (34), 13632–13637.
- (49) Holder, S.; Zemskova, M.; Zhang, C.; Tabrizizad, M.; Bremer, R.; Neidigh, J. W.; Lilly, M. B. Characterization of a potent and selective small-molecule inhibitor of the PIM1 kinase. *Mol. Cancer Ther.* **2007**, *6* (1), 163–172.
- (50) Alguel, Y.; Meng, C. X.; Teran, W.; Krell, T.; Ramos, J. L.; Gallegos, M. T.; Zhang, X. D. Crystal structures of multidrug binding protein TtgR in complex with antibiotics and plant antimicrobials. *J. Mol. Biol.* **2007**, *369* (3), 829–840.
- (51) Pilka, E. S.; Phillips, C.; King, O. N. F.; Guo, K.; von Delft, F.; Pike, A. C. W.; Arrowsmith, C. H.; Weigelt, J.; Edwards, A. M.; Oppermann, U., Crystal structure of human beta-hydroxyisobutyryl-coa hydrolase in complex with quercetin. To be published. DOI:10.2210/pdb3bpt/pdb.
- (52) Pan, P. W.; Tempel, W.; Dong, A.; Loppnau, P.; Kozieradzki, I.; Edwards, A. M.; Arrowsmith, C. H.; Weigelt, J.; Bountra, C.; Bochkarev, V.; Min, V., Crystal structure of human cytosolic sulfotransferase sult1b1 in complex with pap and resveratrol. To be published, DOI:10.2210/pdb3ck1/pdb.
- (53) Fiorucci, S.; Golebiowski, J.; Cabrol-Bass, D.; Antonczak, S. Molecular simulations bring new insights into flavonoid/quercetinase interaction modes. *Proteins* **2007**, *67* (4), 961–970.
- (54) Jurasekova, Z.; Marconi, G.; Sanchez-Cortes, S.; Torreggiani, A. Spectroscopic and Molecular Modeling Studies on the Binding of the Flavonoid Luteolin and Human Serum Albumin. *Biopolymers* **2009**, *91* (11), 917–927.

CI900358Z