Data Driven Charge Transfer Atlas Provides Topological View of Electronic Structure Properties for Arbitrary Proteins Complexes

Fang Liu¹, Hongwei Wang¹, Dongju Zhang², Likai Du¹*,

¹Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, 430070, P. R. China
²Institute of Theoretical Chemistry, Shandong University, Jinan, 250100, P. R. China

*To whom correspondence should be addressed.
Likai Du: dulikai@mail.hzau.edu.cn;

Abstract

Due to the highly complex chemical structure of biomolecules, the extensive understanding of the electronic information for proteomics can be challenging. Here, we constructed a charge transfer database at residue level derived from millions of electronic structure calculations among 20×20 possible amino acid side-chains combinations, which were extracted from available high-quality structures of thousands of protein complexes. Then, the data driven network (D²Net) analysis was proposed to quickly identify the critical residue or residue groups for any possible protein complex. As an initial evaluation, we applied this model to scrutinize the charge transfer networks for two randomly selected proteins, which highlighted the most critical residues with large node degrees as network hubs. This work may provide us a promising computational protocol for topologically understanding the electronic structure information in the growing number of high-quality experimental proteins structures, with minor computational costs.
Introduction

Charge transfer reactions take place in a wide range of biological processes, including photosynthesis, respiration, and signal transduction of biology, enzymatic reactions, gene replication and mutation and so on. In biological systems, electron or hole transfer reaction can occur between donors and acceptors separated by a long distance, for example across protein-protein complexes. Superexchange theory (or electron tunneling) and the hopping model are commonly used to describe electron and hole transfer processes. However, the issue of charge transfer in the entire proteomics is still intriguing and challenging due to the complicated structures of realistic proteins.

In recent years, the growing amount of high-quality experimental (X-ray, NMR, cryo-EM) structures have opened space to improve our theoretical understanding of biological charge transfer reactions in the foreseen big data scenario. The building blocks of proteins are only the twenty L-amino acids, which are distinguished by their different side-chain structures and chemical compositions. However, the physical interactions among residues at the local and overall level are quite complicated in contrast to periodic material systems. Bioinformatics scientists have paid much effort to the classification of protein structures in the last decades. And a large number of biological databases were constructed to depict the structural significance of protein complexes. In parallel to these exciting developments in structural biology and bioinformatics, it becomes increasingly important to incorporate our available structural knowledge, such as the significance of amino acid preference in proteins, into our physical understanding of charge transfer reactions.

It is interesting to capture the electron or hole transfer chain or network in a human-accessible, topological picture. The complex network analysis is an appeal and popularity to obtain qualitative insights, which has been widely used in various fields of chemical and biological studies, i.e. protein/protein interactions, identification of targets for drugs, chemical reaction network, metabolic engineering, and so on. There are also many works discussing how to represent the electron or hole transfer pathways connecting electron donating and accepting cofactors in biological and disordered material systems. For example, Beratan et. al. suggested to use the graph theory to search and identify tunneling pathways or pathway families in biomolecules. However, the electronic couplings were empirical in their work, for which electronic couplings along a given pathway were written as a product of a hypothetical
closest contact terms, involving covalent, hydrogen bond, and van der Waals interactions. Nevertheless, such simple coarse-grained models remain of considerable interest for exploring charge transfer in biological systems\textsuperscript{49}.

The charge transfer electronic coupling parameter is an important component for biological charge transfer reaction, which can be derived from various empirical or semi-empirical models\textsuperscript{16, 44–45, 49, 55–57} and from direct electronic structure calculations\textsuperscript{58–63}. Nowadays, in the era of high performance computing, the sophisticated models are becoming increasingly possible to obtain these electronic couplings for ensembles of structures in biomolecules. Therefore, it is possible to directly derive the charge transfer coupling parameters for millions of molecular fragments, which sufficiently represent most possible occurrences in proteins database.

In this work, we propose the data driven network (D\textsuperscript{2}Net) analysis tools to obtain the topological charge transfer features in any possible protein complex. First, we presented a computational protocol to construct the charge transfer database, which provided an overall view of charge transfer coupling atlas among millions of amino acid side-chain conformations. The sophisticated charge transfer parameters in the database could be reused with minor computational costs. Thus, this charge transfer database as a powerful look-up table was applied to construct complex charge transfer networks for realistic protein complexes. This is also our first attempt to simulate charge transfer networks in realistic proteins by incorporating sufficient structural information. The global topology highlights the most critical residues in charge transfer reactions act as network hubs. In spite of its simplifications, the complex network analysis shows the unique ability to place different charge transfer mechanisms on the same footing under a reasonable graph metric.

2. Methods and Theoretical Details

2.1 Mega Data Sets for the Charge Transfer Atlas

The initial step to develop any data driven or informatics based model is the data collection procedure. Here, we extracted the structural data set from an improved web version\textsuperscript{64} of the “Atlas of Protein Side-Chain Interactions”, which were derived from thousands of unique structures of protein complexes solved by X-ray crystallography to a resolution of 2.0 Å or better. As of June 2017, the Atlas comprised 482555 possible amino acid side-chain conformations for 20×20 sets of
amino acid contacts. And the snapshot of the database on June 2017 can also be obtained from us upon request.

**Figure 1.** Examples for spatial distribution of amino acid side-chain interactions and their associated clusters are visualized. The Thr/Ser (a) and His/Glu (b) pairs are given to illustrate the distinct geometric distributions for each type of amino acid combinations. Note that, the hydrogen atoms are not shown for clarity.

For each amino acid pair, the atlas shows how one amino acid side chain is distributed with respect to the other in the three-dimensional space. The preferred interaction geometric patterns are revealed by clusters in the distributions. As shown in Figure 1, the amino acid pairs have preferred interaction patterns, indicating their packing is not entirely random. The three atoms for the central amino acid are used to define the frame of reference (Figure S1), which is summarized in the supporting information. Each dimer, consisting of one amino acid side-chain pair is transformed to utilize the same frame of reference, with respect to one amino acid, which yields 20 distributions of amino acid residues around each one. The procedure to extract each dimer has been described in the work of Singh and Thornton, for which the clustering algorithm can be summarized as follows.

The root-mean-square deviation (RMSD) between atom positions was calculated for all pairs of amino acid side-chains using the same frame of reference. Any side chain with an RMSD of less than 1.5 Å from the selected side-chain is considered a “neighbor”. The side-chain structure with the largest number of neighbors is taken to be the cluster representative. This side-chain and all its
neighbors form the largest cluster, which are then removed from the original data set. And the calculation is repeated to obtain the cluster representative for the second largest cluster, etc. The clustering procedure was repeated up to six times for each type of amino acid dimer, depending on the size of the geometric cluster. For each cluster, the most representative side-chain is the one with the minimum RMSD to all of the other side-chains in the same cluster.

2.2 Charge Transfer Couplings from the Bio-molecules Tight Binding Method

The tight binding model is an effective approach to study complicated molecular systems. The single-electron motion equation for biomolecule systems can be derived according to the idea of tight-binding approximation. Here, we only provide a brief introduction of the tight binding method for biomolecules (bioTB), following the previous work of Liu and co-workers. Further theoretical details are given in the supporting information.

The building blocks of bio-molecules are just some repeated structural units, i.e. amino acids for proteins, nucleotides for DNA and so on, and the many-electron Hamiltonian in this physical view can be expressed as,

\[
H = \sum_i -\frac{1}{2} \nabla_i^2 + \sum_{i,L} \sum_{a \neq L} \frac{-Z_a}{r_{ai}} + \frac{1}{2} \sum_{i,L} \sum_{j \neq i} \frac{1}{r_{ij}} + \sum_{L} \sum_{a \neq b} \sum_{L} \frac{Z_a Z_b}{R_{ab}} + \sum_{L} \sum_{a \neq L} \sum_{b \neq L} \frac{Z_a Z_b}{R_{ab}}.
\] (1)

In Eq. 1, L and M refer to the repeated structural units or sites. The last term refers to the repulsion between the nuclei, which has no effect on the electronic structure of the bio-molecules.

The Eq. 1 can be rewritten as the sum of one-electron operators,

\[
H = \sum_i \left[ -\frac{1}{2} \nabla_i^2 + \sum_L V_L(i) \right]
\] (2)

And

\[
V_L(i) = \sum_{a \neq L} \frac{-Z_a}{r_{ai}} + \frac{1}{2} \sum_{j \neq i} \frac{1}{r_{ij}}
\] (3)

The corresponding many-electron Schrödinger equation can be solved via the one-electron Schrödinger equation,

\[
\hbar \psi = \varepsilon \psi
\] (4)

Whereas, \( \psi \) is the one-electron wave function for the bio-molecules, and the one-electron
Hamiltonian can be written as,

$$h_i = -\frac{1}{2} \nabla_i^2 + \sum_L V_L(i)$$  \hspace{1cm} (5)

The first step to solve the Eq. 4 is to calculate the one-electron Schrödinger equation of the isolated structural unit at the non-perturbation state.

$$\left[ -\frac{1}{2} \nabla_i^2 + V_L \right] \phi_i = \varepsilon_i^0 \phi_i$$  \hspace{1cm} (6)

In Eq. 6, $\phi_i$ is the molecular orbital of one structural unit L, and $\varepsilon_i^0$ is the corresponding orbital energy. The Eq. 6 can be solved by available electronic structure methods, such as the Hartree-Fock or DFT method. And the electronic orbital for the entire bio-molecules can be expanded as the linear combination of site orbitals for each repeated structural unit.

$$\psi = \sum_i c_i \phi_i$$  \hspace{1cm} (7)

Substituted the above formula into Eq. 4, we can derive the formulas for the on-site energy and transfer integral,

$$\varepsilon_n = \langle \phi_n | h | \phi_n \rangle = \langle \phi_n | -\frac{1}{2} \nabla^2 + \sum_L V_L | \phi_n \rangle$$  \hspace{1cm} (8)

$$t_{n,n+1} = -\langle \phi_n | h | \phi_{n+1} \rangle = -\langle \phi_n | -\frac{1}{2} \nabla^2 + \sum_L V_L | \phi_{n+1} \rangle$$  \hspace{1cm} (9)

The summation runs over all possible structural units L, however, only the neighboring units are required to be considered in the tight binding approximation. The transfer integral describes the ability to perform the charge transfer among neighbor sites, meanwhile, the on-site energy describes the ability to move or inject an electron from a specific site.

Thus, the on-site energy for unit n only needs the potential information of unit n and its closest neighboring units C, and the formula of on-site energy can be simplified as,

$$\varepsilon_n \approx \langle \phi_n | -\frac{1}{2} \nabla^2 + V_n + \sum_{L \in C} V_L | \phi_n \rangle$$

$$= \varepsilon_n^0 + \langle \phi_n | \sum_{L \in C} V_L | \phi_n \rangle$$  \hspace{1cm} (10)

The transfer integral only requires the potential of structural units n and $n+1$, that is

$$t_{n,n+1} \approx -\langle \phi_n | -\frac{1}{2} \nabla^2 + V_n + V_{n+1} | \phi_{n+1} \rangle$$  \hspace{1cm} (11)
Because the tight binding model is corresponding to the orthogonal basis, the Löwdin method is performed to minimize the orbital overlap. And the effective transfer integral can be transformed to,

\[
I_{\text{eff},n,n+1} = \frac{I_{n,n+1} - \frac{1}{2}(\epsilon_i + \epsilon_j) s_{n,n+1}}{1 - s_{n,n+1}^2}
\]  

(12)

In Eq. 12, \(s\) is the orbital overlap integral between units. This transformation shows minor effects on the on-site energy and therefore can be ignored in our code implementation. And one of the implementation of these formulas can be found in http://github.com/dulikai/bioX.

The on-site energy and transfer integral, as two basic variables, are the diagonal and off-diagonal elements of the tight binding Hamiltonian, which are basic physical variables in the study of DNA damage and respiration, photosynthesis and the design of molecular electronic device and charge transfer problems.

Once the tight binding Hamiltonian is constructed, we may directly solve the well-known eigenvalue equation (\(\mathbf{H}\mathbf{C} = \mathbf{E}\mathbf{C}\)) to obtain the electronic structure information of bio-molecules. However, we suggest that the topological understanding of the electronic properties for proteins can be possible by network analysis, without directly solving such electronic structure equation of bio-molecules, see below.

### 2.3 Data Driven Networks for Charge Transfer in Proteins

The charge transfer network is the generalization of charge transfer chains/pathways, which consists sets of substrates in their reduced and oxidized forms. In order to construct the protein charge transfer network, each residue is represented by a vertex in the graph, and the edge represents the strength of the charge transfer coupling among residues. The charge transfer rate is proportional to the square of electron transfer coupling strengths. Note that, the charge transfer rates depend on electronic coupling elements, reorganization energies, and driving forces. However, the exact evaluation of these contributions in realistic proteins is computationally prohibitive, which also significantly complicates the network analysis. Therefore, it is more straightforward to use the strength of charge transfer couplings in our network analysis.

As we are primarily concerned with the network topology, the undirected graph is considered with the following adjacency matrix, and the edge can be only possible to be 0 or 1.
The value of the edge is assigned to be 1, only if the pairwise side-chain structure $x$ is classified into a known cluster ($V$) with significant charge transfer strength in the charge transfer atlas. The selection of the threshold value is nontrivial, and we mainly consider this issue on two aspects. First, the threshold value should be relatively larger than the background value or the average value of the entire database. This is because our D$^2$Net model is statistic or informatics based in nature, and we need to filter the noisy data, in order to highlight the most significant features. And the background value of the database is in the range of 0.01~0.02 eV. Second, the threshold value should not be very large to ignore much valuable electronic structure information. After a few attempts in the network analysis, the threshold of significant charge transfer coupling is set to be 0.02 eV in this work.

As shown in Scheme 1, we suggest a procedure to construct the data driven network (D$^2$Net) model by reusing the charge transfer coupling atlas. This procedure is based on our experience on a few model systems$^{76-77}$, and we could reasonably estimate the charge transfer couplings for any possible amino acid conformations in realistic proteins, if our protein charge transfer atlas is large enough.

**Scheme 1.** The charge transfer atlas is constructed from thousands of PDB structures, along with QM calculations. Once the charge transfer atlas is constructed, the network model and other statistics can be derived by a variety of means.
In order to predict the geometric dependent fluctuations, one possible problem with our data driven approach is the determination of how far one unknown molecular structure can move away from the data set in geometric space before the estimation approach fails. Actually, we use the relative error to monitor the general distances between an arbitrary structure and a set of known structures. The relative error (RE) is defined as follows,

$$RE(k) = \left| \frac{X - X(k)}{X(k)} \right| \times 100\%$$  \hspace{1cm} (14)

In the above equation, $k$ runs over the known molecular structures. And $X$ and $X(k)$ are molecular coordinates as vectors, which are used as fingerprints to search the available charge transfer atlas with millions of side-chain pairs. And the unknown dimer structure can be classified to be similar with one of the known structures. If the calculated RE is larger than 20% for any known structures in the known data sets, the assignment for such unknown structure is said to be failed.

Then, an ab initio calculation of the charge transfer coupling is called to deal with the incompleteness of this charge transfer atlas, meanwhile, this unknown structure is added into our charge transfer atlas for future similarity assignment. Therefore, this process is boot-strapped from the available charge transfer atlas along with minor number of ab initio calculations. The entire computational protocol has been automated in a series of Python codes.

As an initial evaluation of the $D^2$Net model, we randomly selected two crystal structures of proteins from the PDB. One is the HIV-1 integrase core domain (PDB code 1qs4), and the other is the complex between the human H-Ras protein and the Ras-binding domain of C-Raf1 (Ras-Raf), which is central to the signal transduction cascade. The starting structures for the simulations of the human unbound proteins and complexes were taken from the PDB database (PDB codes: 1rrb, 121p, 1gua) and modified to achieve consistency with respect to the biological source and the number of amino acids. For demonstrative purposes, the molecular structure was only minimized in the molecular simulations. The mega data sets in our charge transfer atlas are large enough to represent most possible amino acid side-chains orientations in realistic protein structures. And the failure of the predictions is below 0.1% in both systems.
Results and Discussions

At first, we established a vocabulary to describe how the conformation ensemble influence electronic couplings for millions of amino acid side-chain combinations. Although, previous studies\(^{66-68}\) have revealed the relative abundance of various modes of amino acid contacts (van der Waals contacts, hydrogen bonds), relatively little is known about the qualitative charge transfer coupling terms of these noncovalent interactions. This database is helpful to unravel the richness of biological charge transfer coupling in realistic proteins, which would evolve within fluctuating biomolecules structures.

In fact, we have calculated the electronic couplings of the frontier orbitals of amino acid pairs, i.e. HOMOs, LUMOs or HOMO/LUMO, under the tight binding approximation. However, a thorough comparison of the electronic couplings among various frontier orbitals is beyond the scope of this work, but, instead, we roughly classify them into two categories, namely the electron transfer and hole transfer couplings. The electronic couplings between HOMOs represent the hole transfer process, i.e. removing an electron from the HOMO of one amino acid fragment. The electronic couplings between HOMO and LUMO should also be important, which can be analyzed by the similar protocol. Thus, without losing any generality, we will restrict our study to electronic hole transfer couplings between HOMOs of each amino acid side-chain combinations.

The twenty amino acids can be divided into several groups according to the chemical compositions of their side-chains, that are the hydrophobic group (i.e., GLY, ALA, VAL, ILE, MET, and PHE), polar and neutral group (i.e., SER, THR, CYS, TYR, ASN, and GLN), acidic group (i.e., ASP and GLU) and basic group (i.e., LYS, ARG, and HIS). To simplify the statistical representation of the charge transfer atlas, we defined a resolution for the atlas, for which the unsigned charge transfer coupling below a threshold value was assigned to be zero. Therefore, we can provide a global view of charge transfer couplings among twenty natural amino acids (totally 20×20 combinations) under various resolutions from 0.01 eV to 0.1 eV in Figure 2.
Figure 2. The distribution (zero vs. one) of unsigned charge transfer integral for each type of amino acid side-chain combinations, averaged over the available structures. The resolution of the heat map is assigned to be 0.01, 0.02, 0.03, 0.04, 0.05 and 0.1 eV for a-f, and each amino acid is referred as one letter. There are only two possible values in the heat map, namely zero and one. The white color and grey color refer to the amino acids below and above the specific resolution.

By analyzing the charge transfer atlas at different resolutions, the first of the important results is that most charge transfer coupling lies below 0.05 eV, however, the charge transfer coupling is not zero in most cases. In general, averaged transfer couplings for the polar/polar or basic/acidic combinations (i.e. K/F, S/E) are greater than the pairwise hydrophobic combinations (i.e. A/G, V/F), with only minor exceptions. Note that, the heat maps are not symmetric because the inhomogeneous of protein structures and the distribution of one type of amino acid in the frame of another reference residue type is distinct. Therefore, we believe this asymmetric feature reflects the significant protein structures which are most probably based on certain geometry preferences between interacting amino acids. This heat map may be helpful as reference materials at hand to qualitatively understand the possible charge transfer feature in proteins.

Then, we tried to describe the charge transfer couplings population for a few selected amino acids pairs in the context of overall geometric distribution. The signed charge transfer couplings
distribution for distinct geometric clusters is given in Figure 3. The averaged charge transfer coupling parameters (Figure 2) for each type of amino acid side-chain combinations seem to be not very suitable to describe the distribution of overall geometric clusters. These findings indicate that one must pay close attention when dealing with the geometric ensemble of amino acid pairs in realistic proteins, and the appropriate transfer coupling parameters should be applied only after performing tests on similar geometric features.

Figure 3. The distribution of charge transfer couplings for a few selected amino acids pairs. Six clusters of Lys/Asp combination (a-f) and Ser/Thr (g-l) are shown, whereas Lys and Ser is the center fragment.

Figure 3 also suggests that the charge transfer coupling parameters are found to be very sensitive to the structural orientation of the amino acid pairs. Take the Lys/Asp pair as an example,
the charge transfer coupling is very significant between the basic Lys and acidic Asp fragments (Figure 2), however, the distribution varies much from -0.3 eV to 0.3 eV (a-f in Figure 3). The feature of broad distribution is also found for Ser/Thr as one of the polar and neutral combinations. This rather inhomogeneous of electron transfer couplings may be caused by various inter-molecular interactions (covalent, hydrogen bonds, van der Waals, ionic) within realistic proteins.\textsuperscript{79-82} One can therefore conclude that there are no uniform criteria or empirical formula to accurately estimate the electron transfer coupling parameters for a specific amino acid pair, since the board distribution makes the estimation to be validated only for a small fraction of the geometric ensemble.

Next, the protein side-chain charge transfer database is used to construct the electron transfer network for any protein structures, which refers as D\textsuperscript{2}Net model. Because the anisotropic features cannot reasonably be ignored in making estimation of charge transfer couplings among amino acids, we tried to extensively use the charge transfer atlas for our subsequent network analysis. Note that, the network is different from the assumption of a charge transfer chain, for which a given reduced substrate can only transfer its electron to the next substrate in a linear fashion, and in a network model, each reduced substrate is able, in principle, to transfer electrons to any oxidized substrate. The large number of possible residues in proteins leads to charge transfer networks with hundreds or thousands of species, thousands of pathways and an exponential number of possible pathways and mechanisms to be considered. Of course, we may choose to rule out some possible transfer routines, and thus a given network should have a particular topology, for which a charge transfer chain is one specific example.
Figure 4. Construction of charge transfer network among residues at the resolution of 0.02 eV for the HIV-1 integrase core domain. The graph representations of protein charge transfer network are given with CoSE (a) and circular (b) layout. (c) The node degree distribution is mapped to the protein structure. (d) The secondary structures as building blocks with higher node degree are shown. The red color refers to the “hot” residues with large node degrees, while the blue color means “cold” residue with much smaller node degrees.

Figure 4 shows the charge transfer network analysis of the HIV-1 integrase core domain. To facilitate our following discussions, the words “hot” and “cold” are used to indicate the nodes (residues) with high and low degree in the network. In Figure 4a, we applied the CoSE layout\textsuperscript{83} to visualize the topological feature of undirected compound graphs, and this algorithm highlights the most important nodes (“hot” residues) and their surrounding nodes in an intuitive way. The “hot” residues with deep red color may form a reduced version of the protein core i.e. residue number 7, 22, 24, 29 and so on. And the remainder of the “cold” protein residues (light red nodes) does not
substantially change transfer coupling when included. In Figure 4b, the circular layout is used to represent alternative view of the protein charge transfer network. The advantage of a circular layout in the biological applications is its neutrality, because none of the nodes (residues) is given a privileged position by placing all vertices at equal distances from each other as a regular polygon. The circular layout views in protein systems may provide useful insights on the topology relationship among the protein structures. In summary, the network representation provides us a way to understand the topology of charge transfer networks.28 These derived results may be helpful for the mutation experience to improve the electronic properties of proteins.

Figure 4c maps the node degree distribution in the charge transfer network on to the three dimensional protein structures. The “hot” residues with significant charge transfer contributions in the protein structure are rendered in red color. Qualitatively, these “hot” residue groups with red color should be referred as somewhat “charge transfer topology associated domain (ctTAD)”, which may be distinct from the structural domain in a physical model of the protein structures.

Figure 4d provides the possible “hot” secondary structures in the network. As the most common building blocks for local segments of proteins, both the α-helix and β-sheet could contribute to charge transfer network. The red nodes in the loops structure and random coils are also observed in a few cases. These results illustrate non-local topology properties of the charge transfer in proteins, which may be critical to understand the underlying protein charge transfer problems.
Figure 5. Construction of charge transfer network among residues at the resolution of 0.02 eV for a protein dimer (Ras-Raf). The graph representations of protein charge transfer network are given with CoSE (a) and circular (b) layout. (c) The node degree in the network is mapped to the three dimensional protein structures. (d) The secondary structures as building blocks with higher node degree are shown. The red color refers to the “hot” residues with large node degrees, while the blue color for “cold” residue with smaller node degree.

The D²Net model is also applied to a protein dimer, which are the signal transduction cascade complex of human H-Ras protein and the Ras-binding domain of C-Raf1 (Ras-Raf). Figure 5 shows the topology properties of charge transfer network for this protein dimer, which is relatively different from charge transfer topology of the protein monomer in Figure 4. Figure 5a and 5b highlight the distinct boundaries of the charge transfer network between two sub-units in the three dimensional protein structures. The atomic details of the most critical “hot” node (residue) along with its connected nodes (residues) are also given in Figure 5c. The “hot” residues in the protein
interface may be important for protein-protein interactions. Figure 5d suggests that the “hot” residues are usually spaced and insulated by “cold” residues in the same secondary structure. Therefore, these “hot” regions should prefer to undergo charge transfer among distinct secondary structures. The existence of ctTAD should also be carefully considered in the fragment based quantum chemistry calculations and future force field development, due to its strong possibility for charge flux.

![Image of molecular structures](image)

**Figure 6.** The molecular structures of several types of “hot” nodes (residues) at the resolution of 0.02 eV are extracted from the protein monomer and dimer. (a) The acidic and polar residue, (b) the basic and polar residue, (c) the neutral and polar residue, (d) the non-polar residue.

Finally, we present the atomic details of several “hot” nodes/residues along with their connected nodes in Figure 6. In the network analysis, these “hot” nodes indicate its importance role in the charge transfer reactions. Further model refinement should focus on these important motifs. As shown in Figure 6, the structural motif could also reflect the microscopic environment of the protein. The most common “hot” residues are acidic/basic and polar residues, such as aspartate
arginine in Figure 6a and 6b, which could form hydrogen bonds and even ionic bonds among amino acids. This is also consistent with our common sense in protein charge transfer reactions.84-87 The “hot” residues may also to be neutral and polar residues, such as threonine in Figure 6c. The values of charge transfer couplings among hydrophobic or non-polar amino acids are generally low, however, the charge transfer motif with non-polar amino acids as “hot” residues is also found. Figure 6d provides the atomistic view of the valine with its surrounding amino acids. Note, the charge transfer couplings between this valine node and its connected nodes are in the range of 0.02 — 0.05 eV, while the “hot” polar residues usually exhibit larger charge transfer coupling above 0.05 eV. Thus, such “hot” non-polar residues may disappear if we adjust the resolution of the network to 0.05 eV.

Conclusions

The charge transfer through biological matter is one of the key steps underlying cellular energy harvesting, storage, and utilization, enabling virtually all cellular activity. As the possible structural changes will influence the electrical properties of a protein, the reasonable description of transfer couplings beyond the empirical formulas is very necessary.

In this work, we proposed the data driven network (D²Net) analysis tools for understanding the electron transfer reactions in proteins. The charge transfer atlas derived from millions of amino acid conformations could be used to quickly estimate the charge transfer coupling terms, without any require of the knowledge of chemical intuition about the chemical interactions or empirical formulas. The application of D²Net model revealed that the “hot” residues for charge transfer reactions can be located at the different secondary structures, i.e. α-helix, β-sheet or random coils. Therefore, the predictions of the “hot” residue or residue groups can be made from complex network analysis.

In summary, the data driven model offers us an alternative approach for efficiently identifying the critical residue or groups of residues in charge transfer reactions, which might avoid wasting computational resources. Therefore, we can understand the possible charge transfer reactions in a more explicit and more intuitionistic fashion. Future work may be possible to enumerate the most common “hot” motif that is suitable for charge transfer reactions in proteins database.
Acknowledgements

This work is financially supported by the National Key R&D Program of China (Grant No. 2017YFB0203405). The authors thank the support by the National Natural Science Foundation of China (Nos. 21503249, 21373124). This work is also supported by Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program No.2015RC008) and Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase) under Grant No.U1501501.

Supporting information

The technical details of charge transfer integral derivation, the classification of the charge transfer atlas and the definition of molecular coordination system for amino acids are given in the supporting information.

References

1. Fleming, G. R.; Martin, J. L.; Breton, J. Rates of Primary Electron Transfer in Photosynthetic Reaction Centres and Their Mechanistic Implications. Nature 1988, 333, 190.
2. Gray, H. B.; Winkler, J. R. Electron Transfer in Proteins. Annu. Rev. Biochem. 1996, 65 (1), 537-561.
3. Sturm, G.; Richter, K.; Doetsch, A.; Heide, H.; Louro, R. O.; Gescher, J. A Dynamic Periplasmic Electron Transfer Network Enables Respiratory Flexibility Beyond a Thermodynamic Regulatory Regime. ISME J. 2015, 9, 1802.
4. Giovani, B.; Byrdin, M.; Ahmad, M.; Brettel, K. Light-induced Electron Transfer in a Cryptochrome Blue-Light Photoreceptor. Nat. Struct. Biol. 2003, 10, 489.
5. Hervás, M.; Navarro, J. A.; De la Rosa, M. A. Electron Transfer Between Membrane Complexes and Soluble Proteins in Photosynthesis. Acc. Chem. Res. 2003, 36 (10), 798-805.
6. Roitberg, A. E.; Holden, M. J.; Mayhew, M. P.; Kurnikov, I. V.; Beratan, D. N.; Vilker, V. L. Binding and Electron Transfer Between Putidaredoxin and Cytochrome P450cam. Theory and Experiments. J. Am. Chem. Soc. 1998, 120 (35), 8927-8932.
7. Cordes, M.; Giese, B. Electron Transfer in Peptides and Proteins. Chem. Soc. Rev. 2009, 38 (4), 892-901.
8. Finklea, H. O.; Hanshew, D. D. Electron-transfer Kinetics in Organized Thiol Monolayers with Attached Pentaammine(pyridine)ruthenium Redox Centers. J. Am. Chem. Soc. 1992, 114 (9), 3173-3181.
9. Gilbert Gatty, M.; Kahnt, A.; Esdaile, L. J.; Hutin, M.; Anderson, H. L.; Albinsson, B. Hopping versus Tunneling Mechanism for Long-Range Electron Transfer in Porphyrin Oligomer Bridged Donor–Acceptor Systems. J. Phys. Chem. B 2015, 119 (24), 7598-7611.
10. Long, Y.-T.; Abu-Irhayem, E.; Kraatz, H.-B. Peptide Electron Transfer: More Questions than Answers. Chem. Eur. J. 2005, 11 (18), 5186-5194.
11. Voityuk, A. A. Long-Range Electron Transfer in Biomolecules. Tunneling or Hopping? *J. Chem. Phys.* **B 2011**, *115* (42), 12202-12207.
12. Grozema, F. C.; Berlin, Y. A.; Siebbeles, L. D. A. Mechanism of Charge Migration Through DNA: Molecular Wire Behavior, Single-Step Tunneling or Hopping? *J. Am. Chem. Soc.* **2000**, *122* (44), 10903-10909.
13. Lambert, C.; Nöll, G.; Schelter, J. Bridge-Mediated Hopping or Superexchange Electron-transfer Processes in Bis(triarylamine) Systems. *Nat. Mater.* **2002**, *1*, 69.
14. Kim, S. R.; Parvez, M. K.; Chhowalla, M. UV-reduction of Graphene Oxide and Its Application as an Interfacial Layer to Reduce The Back-transport Reactions in Dye-sensitized Solar Cells. *Chem. Phys. Lett.* **2009**, *483* (1), 124-127.
15. Gray, H. B.; Winkler, J. R. Electron Tunneling Through Proteins. *Q. Rev. Biophys.* **2003**, *36* (3), 341-372.
16. Page, C. C.; Moser, C. C.; Chen, X.; Dutton, P. L. Natural Engineering Principles of Electron Tunnelling in Biological Oxidation–Reduction. *Nature* **1999**, *402*, 47.
17. Rossi, R. A.; Pierini, A. B.; Peñénfory, A. B. Nucleophilic Substitution Reactions by Electron Transfer. *Chem. Rev.* **2003**, *103* (1), 71-168.
18. de la Lande, A.; Gillet, N.; Chen, S.; Salahub, D. R. Progress and Challenges in Simulating and Understanding Electron Transfer in Proteins. *Arch. Biochem. Biophys.* **2015**, *582* (Supplement C), 28-41.
19. Prytkova, T. R.; Kurnikov, I. V.; Beratan, D. N. Coupling Coherence Distinguishes Structure Sensitivity in Protein Electron Transfer. *Science* **2007**, *315* (5812), 622.
20. Tuffery, P.; Etchebest, C.; Hazout, S.; Lavery, R. A New Approach to the Rapid Determination of Protein Side Chain Conformations. *J. Biomol. Struct. Dyn.* **1991**, *8* (6), 1267-1289.
21. Ponder, J. W.; Richards, F. M. Tertiary Templates for Proteins: Use of Packing Criteria in the Enumeration of Allowed Sequences for Different Structural Classes. *J. Mol. Biol.* **1987**, *193* (4), 775-791.
22. Dunbrack, R. L.; Karplus, M. Backbone-Dependent Rotamer Library for Proteins Application to Side-Chain Prediction. *J. Mol. Biol.* **1993**, *230* (2), 543-574.
23. Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The Penultimate Rotamer Library. *Proteins* **2000**, *40* (3), 389-408.
24. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235-242.
25. Maeyer, M. D.; Desmet, J.; Lasters, I. All in One: A Highly Detailed Rotamer Library Improves Both Accuracy and Speed in the Modelling of Sidechains by Dead-End Elimination. *Folding and Design* **1997**, *2* (1), 53-66.
26. Laskowski, R. A.; Hutchinson, E. G.; Michie, A. D.; Wallace, A. C.; Jones, M. L.; Thornton, J. M. PDBsum: A Web-Based Database of Summaries and Analyses of All PDB Structures. *Trends Biochem. Sci.* **1997**, *22* (12), 488-490.
27. Grisham, C. Protein Structure - New Approaches to Disease and Therapy by M. Perutz. *Acta Crystallogr. D* **1993**, *49* (3), 355.
28. Banaji, M.; Baigent, S. Electron Transfer Networks. *J. Math. Chem.* **2008**, *43* (4), 1355-1370.
29. Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; Alon, U. Network Motifs: Simple Building Blocks of Complex Networks. *Science* **2002**, *298* (5594), 824.
30. Cohen, R.; Havlin, S. Scale-Free Networks Are Ultrasmall. *Phys. Rev. Lett.* **2003**, *90* (5), 058701.
31. Alon, U. Biological Networks: The Tinkerer as an Engineer. Science 2003, 301 (5641), 1866.
32. Bock, J. R.; Gough, D. A. Whole-Proteome Interaction Mining. Bioinformatics 2003, 19 (1), 125-134.
33. Yu, H.; Braun, P.; Yildirim, M. A.; Lemmens, I.; Venkatesan, K.; Sahalie, J.; Hirozane-Kishikawa, T.; Gebreab, F.; Li, N.; Simonis, N.; Hao, T.; Rual, J.-F.; Dricot, A.; Vazquez, A.; Murray, R. R.; Simon, C.; Tardivo, L.; Tam, S.; Srzvika, N.; Fan, C.; de Smet, A.-S.; Motyl, A.; Hudson, M. E.; Park, J.; Xin, X.; Cusick, M. E.; Moore, T.; Boone, C.; Snyder, M.; Roth, F. P.; Barabási, A.-L.; Tavernier, J.; Hill, D. E.; Vidal, M. High-Quality Binary Protein Interaction Map of the Yeast Interactome Network. Science 2008, 322 (5898), 104.
34. de Las Rivas, J.; Fontanillo, C. Protein–Protein Interactions Essentials: Key Concepts to Building and Analyzing Interactome Networks. PLOS Comput. Biol. 2010, 6 (6), e1000807.
35. Uhrig, J. F. Protein Interaction Networks in Plants. Planta 2006, 224 (4), 771-781.
36. Xie, L.; Li, J.; Xie, L.; Bourne, P. E. Drug Discovery Using Chemical Systems Biology: Identification of the Protein-Ligand Binding Network to Explain the Side Effects of CETP Inhibitors. PLOS Comput. Biol. 2009, 5 (5), e1000387.
37. Kanehisa, M.; Goto, S.; Furumichi, M.; Tanabe, M.; Hirakawa, M. KEGG for Representation and Analysis of Molecular Networks Involving Diseases and Drugs. Nucleic Acids Res. 2010, 38 (suppl_1), D355-D360.
38. Mizutani, S.; Pauwels, E.; Stoven, V.; Goto, S.; Yamanishi, Y. Relating Drug–Protein Interaction Network With Drug Side Effects. Bioinformatics 2012, 28 (18), i522-i528.
39. Simm, G. N.; Reiher, M. Context-Driven Exploration of Complex Chemical Reaction Networks. J. Chem. Theory Comput. 2017, 13 (12), 6108-6119.
40. Ulissi, Z. W.; Medford, A. J.; Bligaard, T.; Nørskov, J. K. To Address Surface Reaction Network Complexity Using Scaling Relations Machine Learning and DFT Calculations. Nat. Commun. 2017, 8, 14621.
41. Francke, C.; Siezen, R. J.; Teusink, B. Reconstructing the Metabolic Network of a Bacterium From Its Genome. Trends Microbiol. 2005, 13 (11), 550-558.
42. Duarte, N. C.; Becker, S. A.; Jamshidi, N.; Thiele, I.; Mo, M. L.; Vo, T. D.; Srivasa, R.; Palsson, B. Ø. Global Reconstruction of the Human Metabolic Network Based on Genomic and Bibliomic Data. Proc. Natl. Acad. Sci. USA 2007, 104 (6), 1777-1782.
43. Henry, C. S.; Delongh, M.; Best, A. A.; Frybarger, P. M.; Linsay, B.; Stevens, R. L. High-Throughput Generation, Optimization and Analysis of Genome-Scale Metabolic Models. Nat. Biotechnol. 2010, 28, 977.
44. Beratan, D. N.; Onuchic, J. N.; Hopfield, J. J. Electron Tunneling through Covalent and Noncovalent Pathways in Proteins. J. Chem. Phys. 1987, 86 (8), 4488-4498.
45. Beratan, D. N.; Betts, J. N.; Onuchic, J. N. Protein Electron Transfer Rates Set by the Bridging Secondary and Tertiary Structure. Science 1991, 252 (5010), 1285-1288.
46. Newton, M. D. Quantum Chemical Probes of Electron-Transfer Kinetics: the Nature of Donor-Acceptor Interactions. Chem. Rev. 1991, 91 (5), 767-792.
47. Nitzan, A. Electron Transmission through Molecules and Molecular Interfaces. Annu. Rev. Phys. Chem. 2001, 52 (1), 681-750.
48. Warshel, A. Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights. Annu. Rev. Biophys. Biomol. Struct. 2003, 32 (1), 425-443.
49. Balabin, I. A.; Hu, X.; Beratan, D. N. Exploring Biological Electron Transfer Pathway Dynamics with
the Pathways Plugin for VMD. J. Comput. Chem. 2012, 33 (8), 906-910.
50. Rühle, V.; Kirkpatrick, J.; Andrienko, D. A Multiscale Description of Charge Transport in Conjugated Oligomers. J. Chem. Phys. 2010, 132 (13), 134103.
51. Regan, J. J.; Risser, S. M.; Beratan, D. N.; Onuchic, J. N. Protein Electron Transport: Single Versus Multiple Pathways. J. Phys. Chem. 1999, 97 (50), 13083-13088.
52. Vehoff, T.; Chung, Y. S.; Johnston, K.; Troisi, A.; Yoon, D. Y.; Andrienko, D. Charge Transport in Self-Assembled Semiconducting Organic Layers: Role of Dynamic and Static Disorder. J. Phys. Chem. C 2010, 114 (23), 10592-10597.
53. Jackson, N. E.; Savoie, B. M.; Chen, L. X.; Ratner, M. A. A Simple Index for Characterizing Charge Transport in Molecular Materials. J. Phys. Chem. Lett. 2015, 6 (6), 1018-1021.
54. Baumeier, B.; Stenzel, O.; Poelking, C.; Andrienko, D.; Schmidt, V. Stochastic Modeling of Molecular Charge Transport Networks. Phys. Rev. B 2012, 86 (18), 184202.
55. de la Lande, A.; Babcock, N. S.; Řezáč, J.; Sanders, B. C.; Salahub, D. R. Surface Residues Dynamically Organize Water Bridges to Enhance Electron Transfer between Proteins. Proc. Natl. Acad. Sci. USA 2010, 107 (26), 11799-11804.
56. Hammi, E. E.; Houee-Levin, C.; Rezac, J.; Levy, B.; Demachy, I.; Baciou, L.; de la Lande, A. New Insights into the Mechanism of Electron Transfer within Flavohemoglobins: Tunnelling Pathways, Packing Density, Thermodynamic and Kinetic Analyses. Phys. Chem. Chem. Phys. 2012, 14 (40), 13872-13880.
57. de la Lande, A.; Salahub, D. R. Derivation of Interpretative Models for Long Range Electron Transfer from Constrained Density Functional Theory. J. Mol. Struct. THEOCHEM 2010, 943 (1), 115-120.
58. Balabin, I. A.; Onuchic, J. N. Connection between Simple Models and Quantum Chemical Models for Electron-Transfer Tunneling Matrix Element Calculations: A Dyson’s Equations-Based Approach. J. Phys. Chem. 1996, 100 (28), 11573-11580.
59. Medvedev, D. M.; Daizadeh, I.; Stuchebrukhov, A. A. Electron Transfer Tunneling Pathways in Bovine Heart Cytochrome c Oxidase. J. Am. Chem. Soc. 2000, 122 (28), 6571-6582.
60. Balabin, I. A.; Onuchic, J. N. Dynamically Controlled Protein Tunneling Paths in Photosynthetic Reaction Centers. Science 2000, 290 (5489), 114.
61. Kubas, A.; Hoffmann, F.; Heck, A.; Oberhofer, H.; Elstner, M.; Blumberger, J. Electronic Couplings for Molecular Charge Transfer: Benchmarking CDT, FODFT, and FODFTB against High-Level Ab Initio Calculations. J. Chem. Phys. 2014, 140 (10), 104105.
62. Gillet, N.; Berstis, L.; Wu, X.; Gajdos, F.; Heck, A.; de la Lande, A.; Blumberger, J.; Elstner, M. Electronic Coupling Calculations for Bridge-Mediated Charge Transfer Using Constrained Density Functional Theory (CDT) and Effective Hamiltonian Approaches at the Density Functional Theory (DFT) and Fragment-Orbital Density Functional Tight Binding (FODFTB) Level. J. Chem. Theory Comput. 2016, 12 (10), 4793-4805.
63. Kubas, A.; Gajdos, F.; Heck, A.; Oberhofer, H.; Elstner, M.; Blumberger, J. Electronic Couplings for Molecular Charge Transfer: Benchmarking CDT, FODFT and FODFTB against High-Level Ab Initio Calculations. II. Phys. Chem. Chem. Phys. 2015, 17 (22), 14342-14354.
64. Laskowski, R. A. M. S.; Thornton, J. M. Atlas of Side-chain Interactions. http://www.biochem.ucl.ac.uk/bsm/sidechains (accessed 2017-06-01).
65. Singh, J.; Thornton, J. M. SIRIUS: An Automated Method for the Analysis of the Preferred Packing Arrangements between Protein Pairs. J. Mol. Biol. 1990, 211 (3), 595-615.
66. Banerjee, R.; Sen, M.; Bhattacharya, D.; Saha, P. The Jigsaw Puzzle Model: Search for Conformational Specificity in Protein Interiors. J. Mol. Biol. 2003, 333 (1), 211-226.
67. Misura, K. M. S.; Morozov, A. V.; Baker, D. Analysis of Anisotropic Side-chain Packing in Proteins and Application to High-resolution Structure Prediction. J. Mol. Biol. 2004, 342 (2), 651-664.
68. Chakrabarti, P.; Bhattacharyya, R. Geometry of Nonbonded Interactions Involving Planar Groups in Proteins. Prog. Biophys. Mol. Bio. 2007, 95 (1), 83-137.
69. Zheng, B.; Wu, J.; Sun, W.; Liu, C. Trapping and Hopping of Polaron in DNA Periodic stack. Chem. Phys. Lett. 2006, 425 (1), 123-127.
70. Cui, P.; Zhang, D.; Liu, Y.; Yuan, S.; Li, B.; Gao, J.; Liu, C. Tight-Binding Model Method and its Applications in DNA Molecules. Scientia Sinica Chimica 2011, 41 (4), 748.
71. Cui, P.; Wu, J.; Zhang, G.; Liu, C. Hole Polarons in poly(G)-poly(C) and poly(A)-poly(T) DNA Molecules. Sci. China Chem. 2008, 51 (12), 1182-1186.
72. Löwdin, P. O. On the Non - Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. J. Chem. Phys. 1950, 18 (3), 365-375.
73. Valeev, E. F.; Coropceanu, V.; da Silva Filho, D. A.; Salman, S.; Brédas, J.-L. Effect of Electronic Polarization on Charge-Transport Parameters in Molecular Organic Semiconductors. J. Am. Chem. Soc. 2006, 128 (30), 9882-9886.
74. Savoie, B. M.; Kohlstedt, K. L.; Jackson, N. E.; Chen, L. X.; Olvera de la Cruz, M.; Schatz, G. C.; Marks, T. J.; Ratner, M. A. Mesoscale Molecular Network Formation in Amorphous Organic Materials. Proc. Natl. Acad. Sci. USA 2014, 111 (28), 10055-10060.
75. Jackson, N. E.; Chen, L. X.; Ratner, M. A. Charge Transport Network Dynamics in Molecular Aggregates. Proc. Natl. Acad. Sci. USA 2016, 113 (31), 8595-8600.
76. Liu, F.; Du, L.; Lan, Z.; Gao, J. Hydrogen Bond Dynamics Governs the Effective Photoprotection Mechanism of Plant Phenolic Sunscreens. Photochem. Photobiol. Sci. 2017, 16 (2), 211-219.
77. Liu, F.; Du, L.; Zhang, D.; Gao, J. Direct Learning Hidden Excited State Interaction Patterns from Ab Initio Dynamics and Its Implication as Alternative Molecular Mechanism Models. Sci. Rep. 2017, 7 (1), 8737.
78. Gohlke, H.; Case, D. A. Converging Free Energy Estimates: MM-PB(GB)SA Atudies on the Protein–Protein Complex Ras–Raf. J. Comput. Chem. 2004, 25 (2), 238-250.
79. Issa, J. B.; Krogh-Jespersen, K.; Isried, S. S. Conformational Dependence of Electronic Coupling Across Peptide Bonds: A Ramachandran Map. J. Phys. Chem. C 2010, 114 (48), 20809-20812.
80. Kolář, M. H.; Kubaň, T. Reaction Path Averaging: Characterizing the Structural Response of the DNA Double Helix to Electron Transfer. J. Phys. Chem. B 2017, 121 (7), 1520-1532.
81. Wolfgang, J.; Risser, S. M.; Priyadarshy, S.; Beratan, D. N. Secondary Structure Conformations and Long Range Electronic Interactions in Oligopeptides. J. Phys. Chem. B 1997, 101 (15), 2986-2991.
82. Brédas, J. L.; Calbert, J. P.; da Silva Filho, D. A.; Cornil, J. Organic Semiconductors: A Theoretical Characterization of the Basic Parameters Governing Charge Transport. Proc. Natl. Acad. Sci. USA 2002, 99 (9), 5804-5809.
83. Dogrusoz, U.; Giral, E.; Cetintas, A.; Civril, A.; Demir, E. A Layout Algorithm for Undirected Compound Graphs. Inform. Sciences 2009, 179 (7), 980-994.
84. Skourtis, S. S.; Beratan, D. N. Electron Transfer Contact Maps. J. Phys. Chem. B 1997, 101 (7), 1215-1234.
85. Gruschus, J. M.; Kuki, A. Ellipsoidal Delocalization of Tunneled Electrons in Long-Range Electron Transfer Proteins. J. Phys. Chem. B 1999, 103 (51), 11407-11414.
86. Improta, R.; Barone, V.; Newton, M. D. A Parameter-Free Quantum-Mechanical Approach for Calculating Electron-Transfer Rates for Large Systems in Solution. *ChemPhysChem* **2006**, *7* (6), 1211-1214.

87. Migliore, A.; Corni, S.; Di Felice, R.; Molinari, E. First-Principles Density-Functional Theory Calculations of Electron-Transfer Rates in Azurin Dimers. *J. Chem. Phys.* **2006**, *124* (6), 064501.
Supporting Information

Data Driven Charge Transfer Atlas Provides Topological View of Electronic Structure Properties for Arbitrary Proteins Complexes

Fang Liu¹, Hongwei Wang¹, Dongju Zhang², Likai Du¹*

¹Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, 430070, P. R. China
²Institute of Theoretical Chemistry, Shandong University, Jinan, 250100, P. R. China

*To whom correspondence should be addressed.

Likai Du: dulikai@mail.hzau.edu.cn;
Technical Details for Constructing Charge Transfer Atlas

The procedure to extract each dimer complex has been described in the work of Singh and Thornton. Two types of model subsystem were defined based on the hydrogenated amino acids: the first contained only the amino acid side chain starting from the Cβ atom (Cβ model), while the second consisted of the amino acid side chain plus the backbone Cα atom (Cα model).

Because the initial structures in the amino acid side-chain atlas contains only the coordinates of heavy atoms, the missing hydrogens were added using the tleap module in AmberTools package. The point of cutting covalent bond was saturated with hydrogen atoms (i.e., either the Cα or Cβ atom). In order to eliminate any potential nonspecific interactions, the positions of hydrogen atoms were optimized for each dimer at the semiempirical PM6 method with subsequent B3LYP/6-31G* calculations as implemented in Gaussian 09 package. The coordinates of the heavy atoms were kept fixed during the optimization procedure. The optimized structures are used for our subsequent construction of charge transfer atlas and network analysis.

The transfer network for the protein structure was constructed on the basis of the charge transfer atlas. First, we iteratively searched all the residues within 10 Å of each residue, which were supposed to have significant charge transfer couplings. Then, the inter-residue degrees of freedom for each residue pair were calculated. And each pairwise interaction could be represented by a vector, which was used as fingerprint to search the available charge transfer atlas with millions of side-chain pairs. We used the RMSD value as a criterion to evaluate the structural similarity between the unknown amino acid pair and the amino acid side-chain database.

If the assignment for such unknown structure is failed, an ab initio calculation of the charge transfer coupling is triggered; and this unknown structural pattern is automatically added into our charge transfer atlas for future similarity assignment. The electronic couplings for each amino acid side-chain dimer were derived from the RHF/6-31G* level according to the idea of tight-binding approximation. The quantum chemistry calculations are performed with Gaussian 09 package. After the assignment, the potential electron transfer networks for arbitrary protein are constructed.
Derivation of Tight Binding Method for Biomolecules (bioTB)

The tight binding model is an effective approach to study complicated molecular systems with large size. The single-electron motion equation for biomolecule systems is derived according to the idea of tight-binding approximation. Here, we provide a detailed summary of the tight binding method for biomolecules (bioTB), following the previous work of Liu and co-workers. The parameters for the on-site energy and transfer integral in the bioTB model are directly derived from ab initio calculations, which can be applied to the electronic structure calculations of biomolecules, such as DNA, proteins and so on.

The time independent Schrödinger equation for the many-electron problem of large biomolecules can be written as,

$$H \Psi = E \Psi$$  \hspace{1cm} (S1)

And H is the Hamiltonian operator for the biomolecular systems of nuclei and electrons, described by position vectors $R_a$ and $r$, respectively. In atomic units, the Hamiltonian for $n$ electrons and $m$ nuclei can be written as,

$$H = \sum_{i=1}^{n} -\frac{1}{2} \nabla_i^2 + \sum_{a=1}^{m} -\frac{1}{2M_a} \nabla_a^2 + \sum_{i=1}^{n} \sum_{a=1}^{m} -\frac{Z_a}{r_{ai}} + \sum_{i=1}^{n} \sum_{j<i}^{n} r_{ij} + \sum_{a=1}^{m} \sum_{b>a}^{m} \frac{Z_a Z_b}{R_{ab}}$$  \hspace{1cm} (S2)

In Eq. S2, $M_a$ is the ratio of the mass of nucleus $a$ to the mass of an electron, and $Z_a$ is the atomic number of nucleus $a$. The Laplacian operators $\nabla_i^2$ and $\nabla_a^2$ involve differentiation with respect to the coordinates of the $i$th electron and the $a$th nucleus. The first term in the equation is the operator for the kinetic energy of the electrons; the second term is the operator for the kinetic energy of the nuclei; the third term represents the coulomb attraction between electrons and nuclei; the fourth and fifth terms represent the repulsion between electrons and between nuclei, respectively.

The kinetic energy of the nuclei can be neglected within the Born-Oppenheimer approximation. The remaining terms in Eq. S2 are called the electronic Hamiltonian or Hamiltonian describing the motion of $n$ electrons in the field of $m$ point charges,

$$H = \sum_{i}^{n} -\frac{1}{2} \nabla_i^2 + \sum_{a,i} Z_a \frac{1}{r_{ai}} + \sum_{i<j} r_{ij} + \sum_{a<b} \frac{Z_a Z_b}{R_{ab}}$$  \hspace{1cm} (S3)
The building block of biomolecules is simply a few repeated structure units, i.e. amino acids for proteins, nucleotides for DNA and so on. Thus, the many electron Hamiltonian in this physical picture can be written as,

\[ H = \sum_i -\frac{1}{2} \nabla_i^2 + \sum_{L} \sum_i \frac{Z_i}{r_{ai}} + \frac{1}{2} \sum_{L,M} \sum_{j \neq i} \frac{1}{r_{ij}} \]

\[ + \sum_{L} \sum_{a,b \neq L} \frac{Z_a Z_b}{R_{ab}} + \sum_{L,M} \sum_{a \neq L, b \neq M} \frac{Z_a Z_b}{R_{ab}} \]

(S4)

In the above equation, L and M refers to the repeated structure units. The last term refers to the repulsion between the nuclei, which can be considered to be constant. Any constant added to an operator has no effect on the electronic structure of the biomolecules. Therefore, we can rewritten the Eq. S4 as the sum of one-electron operator,

\[ H = \sum_i \left[ -\frac{1}{2} \nabla_i^2 + \sum_{L} V_L(i) \right] \]

(S5)

And

\[ V_L(i) = \sum_{a \neq L} \frac{Z_a}{r_{ai}} + \frac{1}{2} \sum_{j \neq i} \frac{1}{r_{ij}} \]

(S6)

Thus, the corresponding many electrons Schrödinger equation can be solved via the one-electron Schrödinger equation,

\[ \hat{h} \psi = \varepsilon \psi \]

(S7)

whereas, \( \psi \) is the one-electron wavefunction for the biomolecules, and the one-electron Hamiltonian can be expressed as

\[ h = -\frac{1}{2} \nabla^2 + \sum_{L} V_L(i) \]

(S8)

The first step to solve the the Eq. S7 is to obtain the one-electron Schrödinger equation at the non-perturbation and isolated site.

\[ \left[ -\frac{1}{2} \nabla^2 + V_L \right] \phi_L = \varepsilon_L^0 \phi_L \]

(S9)

whereas, \( \phi_L \) is the molecular orbital of site L, and \( \varepsilon_L^0 \) is the corresponding orbital energy. The Eq. S9 can be solved by the Hartree-Fock or DFT method. Here, the site potential is derived from the self-consistent Hartree-Fock matrix of the amino acid side-chain dimer.
And the electronic orbital for the entire biomolecules can be expanded as the linear combination of site orbitals.

\[ \psi = \sum_i c_i \phi_i \]  

(S10)

Substituted the above equation into Eq. S8, the following equation is obtained.

\[ \sum_i c_i \langle \phi_m | \phi_i \rangle \phi_i = \varepsilon \sum_i c_i \langle \phi_m | \phi_i \rangle \]  

(S11)

According to the tight-binding approximation, the electrons in this model are tightly bound to the site to which they belong and they should have limited interaction with states and potentials on surrounding sites. And the potential of the neighboring sites can be treated as perturbations. The tight binding Hamiltonian is usually written as,

\[ H_{TB} = \sum_n \varepsilon_n c_n^+ c_n - t_{n,n+1} \left( c_{n+1}^+ c_n + c_n c_{n+1}^+ \right) \]  

(S12)

whereas, \( c \) and \( c^+ \) are the annihilation and creation operators for electrons or holes, \( \varepsilon \) is the on-site energy and \( t \) is the transfer integral between sites.

In the tight-binding approximation, the electron has limited interaction with the non-neighboring sites. Thus, we can get the following expression,

\[ \langle \phi_m | h | \phi_m \rangle c_m + \langle \phi_{m-1} | h | \phi_m \rangle c_{m-1} + \langle \phi_{m+1} | h | \phi_m \rangle c_{m+1} = \varepsilon_m c_m \]  

(S13)

After comparing the Eq. S12 and S13, the formulas for the on-site energy and transfer integral can be given as,

\[ \varepsilon_n = \langle \phi_n | h | \phi_n \rangle = \langle \phi_n | -\frac{1}{2} \nabla^2 + \sum_L V_L | \phi_n \rangle \]  

(S14)

\[ t_{n,n+1} = -\langle \phi_n | h | \phi_{n+1} \rangle = -\langle \phi_n | -\frac{1}{2} \nabla^2 + \sum_L V_L | \phi_{n+1} \rangle \]  

(S15)

The summation runs over all possible site L, however, only the neighboring sites are required to be considered in the tight binding approximation. Therefore, the on-site energy for site \( n \) only needs the potential information of site \( n \) and its neighboring site \( n-1 \) and \( n+1 \) (linear molecules), the related formula of on-site energy can be simplified as,

\[ \varepsilon_n \approx \langle \phi_n | -\frac{1}{2} \nabla^2 + V_n + V_{n-1} + V_{n+1} | \phi_n \rangle \]  

\[ = \varepsilon_n^0 + \langle \phi_n | V_{n-1} + V_{n+1} | \phi_n \rangle \]  

(S16)

In Eq. S16, the on-site energy is not fully equal to the orbital energy of the site \( n \), at least, the
first neighboring sites should be considered. For non-linear molecules, such as proteins, the on-site energy for unit \( n \) only needs the potential information of unit \( n \) and its closest neighboring units \( C \), and the formula of on-site energy can be simplified as,

\[
\epsilon_n \approx \langle \phi_n | -\frac{1}{2} \nabla^2 + V_n + \sum_{l \in C} V_l | \phi_n \rangle = \epsilon_n^0 + \langle \phi_n | \sum_{l \in C} V_l | \phi_n \rangle
\]

(S17)

The use of amino acid dimers is sufficient for the construction of charge transfer coupling matrix. Thus, the transfer integral only require to the potential of site \( n \) and \( n+1 \), that is

\[
I_{n,n+1} \approx -\langle \phi_n | -\frac{1}{2} \nabla^2 + V_n + V_{n+1} | \phi_{n+1} \rangle
\]

(S18)

Because the tight binding model is corresponding to the orthogonal basis, the Löwdin method is perform to minimize the orbital overlap. And the transfer integral can be transformed to,

\[
I_{n,n+1}^{\text{eff}} = \frac{I_{n,n+1} - \frac{1}{2}(\epsilon_n + \epsilon_{n+1})}{1 - s_{n,n+1}^2}
\]

(S19)

In Eq. S19, \( s \) is the orbital overlap integral between sites. This transformation shows minor effects on the on-site energy and therefore ignored in our code implementation.

As two basic variables, the on-site energy and transfer integral are the diagonal and sub-diagonal elements of the tight binding Hamiltonian, which are basic physical variables in the study of DNA damage and respiration, photosynthesis and the design of molecular electronic device and charge transfer problems. The charge transfer integral describes the ability to perform the charge transfer among neighbor sites, meanwhile, the on-site energy describes the ability to move or inject an electron from a specific site.

Reference

1. Singh, J.; Thornton, J. M. SIRIUS: An automated method for the analysis of the preferred packing arrangements between protein groups. J. Mol. Biol. 1990, 211 (3), 595-615.

2. D.A. Case; D.S. Cerutti; T.E. Cheatham, I.; T.A. Darden; R.E. Duke; T.J. Giese; H. Gohlke; A.W. Goetz; D. Greene; N. Homeyer; S. Izadi; A. Kovalenko; T.S. Lee; S. LeGrand; P. Li; C. Lin; J. Liu; T. Luchko; R. Luo; D. Mermelstein; K.M. Merz; G. Monard; H. Nguyen; I. Omelyan; A. Onufriev; F. Pan; R. Qi; D.R. Roe; A. Roitberg; C. Sagui; C.L. Simmerling; W.M. Botello-Smith;
J. Swails; R.C. Walker; J. Wang; R.M. Wolf; X. Wu; L. Xiao; D.M. York; Kollman, P. A. AmberTools, University of California, San Francisco, AmberTools, 2017.

3. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnener, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Gaussian, Inc.: Wallingford, CT, USA, Gaussian 09, 2009.

4. Cui, P.; Zhang, D.; Liu, Y.; Yuan, S.; Li, B.; Gao, J.; Liu, C. Tight-binding model method and its applications in DNA molecules. Scientia Sinica Chimica 2011, 41 (4), 748.

5. Cui, P.; Wu, J.; Zhang, G.; Liu, C. Hole polarons in poly(G)-poly(C) and poly(A)-poly(T) DNA molecules. Science in China Series B: Chemistry 2008, 51 (12), 1182-1186.
**Figure S1.** Definition of the spherical coordinate system, and the $r$, $\theta$, $\phi$ is shown for the Pro side-chain as an example. Each amino acid pairs are transformed this coordinate system for subsequent clustering or RMSD calculations.

The coordinate system is defined as follows for each amino acid, which given by the Cartesian coordinates in PDB format.

**Ala**

| Atom | Number | Type | Chain | X  | Y  | Z  | Temperature | Density |
|------|--------|------|-------|-----|-----|-----|-------------|---------|
| ATOM | 2698   | N    | F     | -1.198 | 0.835 | 0.000 | 1.00 | 10.00 |
| ATOM | 2699   | CA   | F     | 0.000  | 0.000 | 0.000 | 1.00 | 10.00 |
| ATOM | 2700   | CB   | F     | 1.251  | 0.872 | 0.000 | 1.00 | 10.00 |

**Arg**

| Atom | Number | Type | Chain | X  | Y  | Z  | Temperature | Density |
|------|--------|------|-------|-----|-----|-----|-------------|---------|
| ATOM | 1165   | NH1  | F     | -1.149 | 0.668 | 0.000 | 1.00 | 10.00 |
| ATOM | 1166   | CZ   | F     | 0.000  | 0.000 | 0.000 | 1.00 | 10.00 |
| ATOM | 1167   | NH2  | F     | 1.147  | 0.667 | 0.000 | 1.00 | 10.00 |
| ATOM | 1168   | NE   | F     | 0.015  | -1.329 | 0.000 | 1.00 | 10.00 |
| ATOM | 1169   | CD   | F     | -1.171 | -2.175 | -0.001 | 1.00 | 10.00 |

**Asn**

| Atom | Number | Type | Chain | X  | Y  | Z  | Temperature | Density |
|------|--------|------|-------|-----|-----|-----|-------------|---------|
| ATOM | 880    | OD1  | F     | -1.082 | 0.593 | 0.000 | 1.00 | 10.00 |
| ATOM | 881    | CG   | F     | 0.000  | 0.000 | 0.000 | 1.00 | 10.00 |
| ATOM | 882    | ND2  | F     | 1.165  | 0.639 | 0.000 | 1.00 | 10.00 |
| ATOM | 883    | CB   | F     | 0.057  | -1.514 | -0.001 | 1.00 | 10.00 |

**Asp**
| Atom | Type | Residue | Chain | X  | Y  | Z  | Temperature | Usage |
|------|------|---------|-------|----|----|----|-------------|-------|
| 1324 | OD1  | ASP     | F     | -1.094 | 0.605 | 0.000 | 1.00        | 10.00 |
| 1325 | CG   | ASP     | F     | 0.000  | 0.000 | 0.000 | 1.00        | 10.00 |
| 1326 | OD2  | ASP     | F     | 1.096  | 0.606 | 0.000 | 1.00        | 10.00 |
| 1327 | CB   | ASP     | F     | 0.012  | -1.515 | -0.004 | 1.00        | 10.00 |
| 496  | CA   | CYS     | F     | -1.282 | 0.835 | 0.000 | 1.00        | 10.00 |
| 497  | CB   | CYS     | F     | 0.000  | 0.000 | 0.000 | 1.00        | 10.00 |
| 498  | SG   | CYS     | F     | 1.515  | 0.987 | 0.000 | 1.00        | 10.00 |
| 676  | OE1  | GLN     | F     | -1.083 | 0.593 | 0.000 | 1.00        | 10.00 |
| 677  | CD   | GLN     | F     | 0.000  | 0.000 | 0.000 | 1.00        | 10.00 |
| 678  | NE2  | GLN     | F     | 1.164  | 0.638 | 0.000 | 1.00        | 10.00 |
| 892  | OE1  | GLU     | F     | -1.099 | 0.600 | 0.000 | 1.00        | 10.00 |
| 893  | CD   | GLU     | F     | 0.000  | 0.000 | 0.000 | 1.00        | 10.00 |
| 894  | OE2  | GLU     | F     | 1.099  | 0.600 | 0.000 | 1.00        | 10.00 |
| 2644 | N    | GLY     | F     | -1.210 | 0.806 | 0.000 | 1.00        | 10.00 |
| 2645 | CA   | GLY     | F     | 0.000  | 0.000 | 0.000 | 1.00        | 10.00 |
| 2646 | C    | GLY     | F     | 1.262  | 0.841 | 0.000 | 1.00        | 10.00 |
| 547  | CD2  | HIS     | F     | -1.085 | 0.813 | 0.000 | 1.00        | 10.00 |
| 548  | CG   | HIS     | F     | 0.000  | 0.000 | 0.000 | 1.00        | 10.00 |
| 549  | ND1  | HIS     | F     | 1.102  | 0.826 | 0.000 | 1.00        | 10.00 |
| 550  | CB   | HIS     | F     | 0.105  | -1.492 | -0.002 | 1.00        | 10.00 |
| 551  | CE1  | HIS     | F     | 0.698  | 2.087 | -0.001 | 1.00        | 10.00 |
| 552  | NE2  | HIS     | F     | -0.624 | 2.107 | -0.001 | 1.00        | 10.00 |
| 1360 | CG1  | ILE     | F     | -1.262 | 0.870 | 0.000 | 1.00        | 10.00 |
| 1361 | CB   | ILE     | F     | 0.000  | 0.000 | 0.000 | 1.00        | 10.00 |
| 1362 | CG2  | ILE     | F     | 1.259  | 0.868 | 0.000 | 1.00        | 10.00 |
| 1096 | CD1  | LEU     | F     | -1.252 | 0.867 | 0.000 | 1.00        | 10.00 |
| 1097 | CG   | LEU     | F     | 0.000  | 0.000 | 0.000 | 1.00        | 10.00 |
| 1098 | CD2  | LEU     | F     | 1.252  | 0.867 | 0.000 | 1.00        | 10.00 |
| Residue | Atom  | X    | Y    | Z    | Temp |occup.|
|---------|-------|------|------|------|------|------|
| LEU     | CB    | -0.003 | -0.930 | 1.213 | 1.00 | 10.00 |
| LYS     | CD    | -1.259 | 0.854  | 0.000 | 1.00 | 10.00 |
| LYS     | CE    | 0.000  | 0.000  | 0.000 | 1.00 | 10.00 |
| LYS     | NZ    | 1.235  | 0.838  | 0.000 | 1.00 | 10.00 |
| LYS     | CB    | -3.117 | 0.992  | 0.006 | 1.00 | 10.00 |
| LYS     | CG    | 1.235  | 0.838  | 0.000 | 1.00 | 10.00 |
| MET     | CG    | -1.391 | 1.155  | 0.000 | 1.00 | 10.00 |
| MET     | SD    | 0.000  | 0.000  | 0.000 | 1.00 | 10.00 |
| MET     | CE    | 1.373  | 1.139  | 0.000 | 1.00 | 10.00 |
| PHE     | CD1   | -1.196 | 0.709  | 0.000 | 1.00 | 10.00 |
| PHE     | CG    | 0.000  | 0.000  | 0.000 | 1.00 | 10.00 |
| PHE     | CD2   | 1.196  | 0.708  | 0.000 | 1.00 | 10.00 |
| PHE     | CE1   | -0.736 | 2.302  | 0.074 | 1.00 | 10.00 |
| PHE     | CE2   | 1.200  | 2.101  | -0.001| 1.00 | 10.00 |
| PHE     | CB    | 0.004  | -1.505 | -0.001| 1.00 | 10.00 |
| PHE     | CZ    | 0.000  | 2.796  | -0.001| 1.00 | 10.00 |
| PRO     | N     | -1.149 | 0.910  | 0.000 | 1.00 | 10.00 |
| PRO     | CA    | 0.000  | 0.000  | 0.000 | 1.00 | 10.00 |
| PRO     | CB    | 1.200  | 0.950  | 0.000 | 1.00 | 10.00 |
| PRO     | CD    | -0.736 | 2.302  | 0.074 | 1.00 | 10.00 |
| PRO     | CG    | 0.683  | 2.180  | 0.046 | 1.00 | 10.00 |
| SER     | CA    | -1.260 | 0.869  | 0.000 | 1.00 | 10.00 |
| SER     | CB    | 0.000  | 0.000  | 0.000 | 1.00 | 10.00 |
| SER     | OG    | 1.167  | 0.805  | 0.000 | 1.00 | 10.00 |
| THR     | OG1   | -1.163 | 0.838  | 0.000 | 1.00 | 10.00 |
| THR     | CB    | 0.000  | 0.000  | 0.000 | 1.00 | 10.00 |
| THR     | CG2   | 1.237  | 0.891  | 0.000 | 1.00 | 10.00 |
| TRP     | CD1   | -1.094 | 0.820  | 0.000 | 1.00 | 10.00 |
| TRP     | CG    | 0.000  | 0.000  | 0.000 | 1.00 | 10.00 |
| TRP     | CD2   | 1.146  | 0.859  | 0.000 | 1.00 | 10.00 |
| TRP     | CE2   | 0.672  | 2.189  | 0.000 | 1.00 | 10.00 |
|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ATOM | 701 | CE3 | TRP | F  | 1   | 2.529 | 0.635 | -0.001 | 1.00 | 10.00 |
| ATOM | 702 | CH2 | TRP | F  | 1   | 2.880 | 3.041 | 0.000  | 1.00 | 10.00 |
| ATOM | 703 | CZ2 | TRP | F  | 1   | 1.531 | 3.291 | 0.001  | 1.00 | 10.00 |
| ATOM | 704 | CZ3 | TRP | F  | 1   | 3.384 | 1.731 | 0.000  | 1.00 | 10.00 |
| ATOM | 705 | NE1 | TRP | F  | 1   | -0.698| 2.137 | 0.000  | 1.00 | 10.00 |

**Tyr**

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ATOM | 865 | CD1 | TYR | F  | 1   | -1.195| 0.716 | 0.000  | 1.00 | 10.00 |
| ATOM | 866 | CG  | TYR | F  | 1   | 0.000 | 0.000 | 0.000  | 1.00 | 10.00 |
| ATOM | 867 | CD2 | TYR | F  | 1   | 1.194 | 0.715 | 0.000  | 1.00 | 10.00 |
| ATOM | 868 | OH  | TYR | F  | 1   | 0.002 | 4.170 | -0.002 | 1.00 | 10.00 |
| ATOM | 869 | CE1 | TYR | F  | 1   | -1.200| 2.107 | -0.001 | 1.00 | 10.00 |
| ATOM | 870 | CE2 | TYR | F  | 1   | 1.201 | 2.106 | -0.001 | 1.00 | 10.00 |
| ATOM | 871 | CZ  | TYR | F  | 1   | 0.001 | 2.793 | -0.001 | 1.00 | 10.00 |

**Val**

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ATOM | 1477| CG1 | VAL | F  | 1   | -1.253| 0.870 | 0.000  | 1.00 | 10.00 |
| ATOM | 1478| CB  | VAL | F  | 1   | 0.000 | 0.000 | 0.000  | 1.00 | 10.00 |
| ATOM | 1479| CG2 | VAL | F  | 1   | 1.253 | 0.870 | 0.000  | 1.00 | 10.00 |