Characterising protein-protein interactions with the fragment molecular orbital method

Alexander Heifetz[1], Vladimir Sladek[2], Andrea Townsend-Nicholson[3] and Dmitri G. Fedorov[4*]*

[1]Evotec (UK) Ltd., 114 Innovation Drive, Milton Park, Abingdon, Oxfordshire OX14 4RZ, United Kingdom

[2]Institute of Chemistry, Centre for Glycomics, Slovak Academy of Sciences, Dubravska cesta 9, 84538 Bratislava, Slovakia

[3]Research Department of Structural & Molecular Biology, Division of Biosciences, University College London, London, WC1E 6BT, United Kingdom

[4]Research Center for Computational Design of Advanced Functional Materials (CD-FMat), National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

***Corresponding authors**

Keywords:

Protein-protein interactions, PPI, QM, Molecular recognition, Quantum Mechanics, FMO, Fragment Molecular Orbital, protein residue networks, subsystem analysis, binding energy, PIE-PRN, PRN, Desolvation penalty, Network analysis of protein residue networks, efficiency centrality.

Abstract

Proteins are vital components of living systems. They play many different roles, including serving as building blocks, molecular machines, enzymes, receptors, ion channels, sensors, transporters, and protein–protein interactions (PPIs) are a key part of their function. There are more than 645,000 reported disease-relevant PPIs in the human interactome but drugs have been developed for only 2% of these targets. The advances in PPI-focused drug discovery are highly dependent on the availability of structural data and accurate computational tools for analysis of this data. Quantum mechanical approaches are often too expensive computationally, but the fragment molecular orbital (FMO) method offers an excellent solution that combines accuracy, speed and the ability to reveal key interactions that would otherwise be hard to detect. FMO provides essential information for PPI drug discovery, namely identification of key interactions formed between residues of two proteins, including their strength (in kcal/mol) and their chemical nature (electrostatic or hydrophobic). In this chapter, we have demonstrated how three different FMO-based approaches (pair interaction energy analysis (PIE-analysis), subsystem analysis (SA) and analysis of protein residue networks (PRN)) have been applied to study PPI in three protein-protein complexes.

1. Introduction

1.1. Living organisms rely on the specific recognition of pairs of proteins in practically every biological process [1-3], hence, the importance of understanding molecular recognition and analysis of protein-protein interactions (PPIs) cannot be overstated. Analyses of experimentally determined structures indicate that intermolecular recognition is facilitated by a myriad of noncovalent interactions that, together, promote specificity at different levels.

1.2. There are more than 645,000 reported disease-relevant PPIs in the human interactome [4]. However, only 2% of these had been targeted with drugs by 2011. Most of the remaining disease relevant PPIs in protein complexes, such as transcription factors and many other signalling proteins, have been widely considered 'undruggable' and remain elusive, underexplored and yet to be fully understood [4-6]. Modulating disease-relevant protein-protein interactions using small-molecule inhibitors remains a challenging task owing to their highly dynamic and expansive interfacial areas (flat, featureless and relatively large). However, advances in PPI-focused drug discovery technology have been reported and a few drugs are already on the market, with some potential drug-like candidates already in clinical trials [7].

1.3. Technological progress has played a central role in the identification of small-molecule modulators of PPIs that have to date reached clinical production [7]. The use of structural biology to determine 'hotspots' in PPIs' binding interfaces has been an important strategy in discovering small-molecule modulators [4, 5]. Despite the large sizes of PPI interfaces, only a small subset of amino acid residues that comprise the hotspot contributes most of the binding free energy. These 'hotspot' regions are potential targets for drug discovery [4]. An experimental way of identifying these hotspots in PPIs has been the combination of alanine-scanning mutagenesis and X-

ray crystallography [8]. However even with the availability of high-resolution crystal structures, “visual inspection” and the force field-based molecular mechanics (MM) calculations typically used for structural exploration cannot explain the full complexity of the intramolecular interactions [9]. Recently, several notable reports have been published [9-12] that emphasize the crucial role of ‘underappreciated’ or non-obvious intramolecular interactions involved in biomolecular recognition. These interactions include CH/π [13, 14], halogen/π [15], cation/π [16] and non-classical hydrogen bonds [17], features that are often not properly parameterized in the currently available force fields (FF) [11]. Furthermore, hydrophobic interactions, vital for receptor stability[18], still have no reliable non-QM predictive method for their quantification [9, 19].

1.1. Quantum mechanical (QM) methods have always been considered to be a reliable approach for the exploration of molecular interactions.[20, 21] However, despite their many advantages, traditional QM approaches are generally not feasible for large biological systems like GPCRs, due to their high computational cost [22]. We therefore have employed the fragment molecular orbital (FMO) QM approach [14, 21, 23, 24] in the current study.

1.2. The fragment molecular orbital method [14, 21, 23] offers a considerable computational speed-up over traditional QM methods [25]. By performing QM calculations on fragments, one can make the computational cost scaling with respect to system size nearly linear.

1.3. A specific advantage of FMO is that it provides a quantitative breakdown of the interactions formed between pairs of fragments (residues), including their strength (in kcal/mol) and chemical nature (electrostatic or hydrophobic) [14]. FMO offers an

excellent solution that combines accuracy, speed and the ability to reveal key interactions that would be hard to detect otherwise [24].

1.4. The pair interaction energy (PIE, see Method section 2.1) between any two fragments calculated by FMO can be expressed as the sum of several energy terms such as electrostatic, exchange-repulsion, charge transfer, dispersion, and solvent screening accomplished in the pair interaction energy decomposition analysis (PIEDA) [26, 27]. The electrostatic and charge transfer terms are predominant in salt-bridge, hydrogen bond and polar interactions, whilst the dispersion term generally corresponds to interactions which are predominantly hydrophobic in nature. A strong electrostatic interaction is compensated for by the solvent screening, so that the sum of these two terms is often comparable to the dispersion interaction. The role of hydrophobic interactions is integral for biomolecular recognition but there is still no reliable predictive method for its quantification [9]. The exchange-repulsion term is a non-electrostatic QM term which quantifies the Pauli repulsion between electrons [22]. FMO is an extensively validated method for the structural exploration [28] of large biological systems.

1.5. As described in the literature [29] and as can be observed from Figure 1, protein interfaces are characterized by complementarity of the shape and the chemical character of the interacting surfaces. FMO provides detail analysis of the interactions formed by the residues located on these interfaces.

1.6. In this chapter, we demonstrate how FMO can be applied to characterise the strength and chemical nature of protein-protein interactions. We applied three different FMO-based approaches (pair interaction energy analysis (PIE-analysis), subsystem analysis (SA) and analysis of protein residue networks (PRN)) to explore PPIs

between 3 protein complexes (Figure 1): Trypisin/Inhibitor (PDB code 1AVW [30]), Barnase/Barstar (PDB code 1BRS [31]) and Subtilisin/Eglin-C (PDB code 2SEC [32]).

<Figure 1a here>

<Figure 1b here>

<Figure 1c here>

Figure 1. Overall architecture of the three tested systems: for trypsin-soy-bean inhibitor (PDB code 1AVW [30]) (**A**), barnase-barstar (PDB code 1BRS [31]) (**B**) and subtilisin/eglin-C (PDB code 2SEC [32]) (**C**). Left side: overall architecture of the protein-protein complex when structure of each protein represented as transparent complexes and its backbone as a ribbon. Right side: section through the protein's surfaces illustrating interface shape complementarity.

2.0 Methods

2.1 *Pair interaction energy analysis (PIE-analysis)*

2.1.1 For analysis of PPI, we used FMO as implemented [33] in the general atomic and molecular electronic structure system (GAMESS) [34]. In FMO calculations, a large biological system is partitioned into fragments [14, 21]. Each residue is a fragment, and the interaction energies can be sorted out so that they correspond to actual amino acid residues, not to residue fragments. The detailed description of the FMO methodology can be found in chapter 4 of this book and in the other published reports [14, 21, 26]. To speed up the PPI calculations we truncated the proteins

to retain only residues located on protein-protein interface using 'Contacts' module of MOE v.2018.01 (Computational chemistry group). Only those residues of one protein positioned within a radius of ≤ 4.5 Å from a residues of the second protein were defined as interface residues and included in our calculations.

2.1.2 We used the MP2 method (second order Møller-Plesset perturbation theory [35]) with the 6-31G* basis set as a compromise between speed and accuracy. To describe solvent, the conductor polarizable continuum solvent model (C-PCM) was used [36].

2.1.3 In PIEDA/MP2, the pair interaction energy ΔE_{IJ} or PIE between two fragments *I* and *J* is divided into the electrostatic (ES), exchange-repulsion (EX), charge transfer and mix terms (CT+mix), dispersion and remainder correlation (DI+RC) and solvent screening (SOLV) [27][37] as shown in Equation 1:

$$\Delta E_{IJ} = \Delta E_{IJ}^{\mathrm{ES}} + \Delta E_{IJ}^{\mathrm{EX}} + \Delta E_{IJ}^{\mathrm{CT+mix}} + \Delta E_{IJ}^{\mathrm{DI+RC}} + \Delta E_{IJ}^{\mathrm{SOLV}} \quad (1)$$

2.1.4 From these components one can define the pair interaction character (PIC) describing the chemical nature of a given interaction. In this work, PIC is defined differentiating between electrostatics and dispersion as:

$$f_{IJ} = \frac{\left|\Delta E_{IJ}^{\mathrm{ES}} + \Delta E_{IJ}^{\mathrm{SOLV}}\right|}{\left|\Delta E_{IJ}^{\mathrm{ES}} + \Delta E_{IJ}^{\mathrm{SOLV}}\right| + \left|\Delta E_{IJ}^{\mathrm{DI+RC}}\right|} \quad (2)$$

2.1.5 Because $\Delta E_{IJ}^{\mathrm{DI+RC}}$ is usually dominated by dispersion, f_{IJ} is equal to 1 and 0 for purely electrostatic and dispersive interactions, respectively. The remaining two components, $\Delta E_{IJ}^{\mathrm{EX}} + \Delta E_{IJ}^{\mathrm{CT+mix}}$, describing exchange-repulsion and charge transfer, typically have an opposite sign, and partially cancel each other [26], although the sum may still be substantial for very short contacts, representing the important non-electrostatic physical effect of short-distance repulsion. Here, for the sake of differentiating between electrostatics and dispersion, the two terms $\Delta E_{IJ}^{\mathrm{EX}} + \Delta E_{IJ}^{\mathrm{CT+mix}}$ were not considered in the definition of PIC.

2.1.6 These PIE maps in Figure 2 were made automatically using a Python script developed in this work. This functionality will be included into future releases of the ProGA (Protein Graph Analyser) package Protein Graph Analyser since 2016. Freely available upon e-mail request at sladek.vladimir@savba.sk]. This toolkit is intended for the analysis of protein residue networks based on distances or pair interaction energies. Based on previous reports [14], we considered any interaction with an absolute PIE ≥3.0 kcal/mol to be significant [24].

2.2 ***Subsystem analysis (SA)***

2.2.1 To analyse the binding directly, the subsystem analysis [38] was applied to the same truncated models as in the analysis of PIEs. It provides a complete integrated picture, in which interaction energies are added to the desolvation penalty, protein polarization energy and deformation energy. In this work, the repulsive deformation energy describing the energy cost of deforming isolated protein structures

during the complex formation was not considered. The binding energies in solution ΔE in SA are divided into fragment contributions as:

$$\Delta E = E(\text{complex}) - E(\text{protein 1}) - E(\text{protein 2}) = \sum_{I=1}^{\text{protein1}} \Delta E_I^{\text{bind}} + \sum_{I=1}^{\text{protein2}} \Delta E_I^{\text{bind}} \quad (3)$$

2.2.2 Here, ΔE_I^{bind} includes a sum over *J* of fragment-fragment interaction energies ΔE_{IJ} between the two proteins (divided by 2 to avoid double counting), and partial fragment energies. The latter incorporates the effect of the mutual polarization (of one protein by another) and desolvation (the loss of the solute-solvent energy because of the complexation). The effect of the intra-protein charge transfer is included in the partial fragment energies, whereas the effect of the inter-protein charge transfer is incorporated in inter-protein ΔE_{IJ}. Further details can be found in chapter 4 of this book.

2.3 ***Analysis of protein residue networks (PRN)***

2.3.1 The interface picture described above is an important but simplified model of PPI. For a more comprehensive analysis one has to consider whole proteins. In doing so, it is very useful to use the topological analysis of protein residue networks.

2.3.2 Protein-protein interactions are not limited to interfaces. In reality, proteins and protein-protein complexes are dynamic systems undergoing constant small conformational fluctuations. The formation of a protein-protein complex is usually a multistep process. It is often

the case that the initial binding area is relatively small and will not be identical to the binding hotspot as identified in the final complex. In contrast, upon initial binding some conformational, potentially long-reaching allosteric, changes are likely to take place. These, together with the mutual attraction cause a desolvatation of the binding hotspots and the arrangement of the interacting proteins into the final binding pose. One can think of these conformational changes as an initially localised structural (and energetic) perturbation that is subsequently propagated throughout the protein.

2.3.3 The formation of a protein-protein complex can be facilitated if certain domains containing the final binding hotspot residues are pre-oriented to fit the surface of the binding partner. It is shown below that 'hubs' are responsible for this local stability. This process involves whole proteins and not just the final interface, hence, hub identification should also take into account all PIEs within the proteins.

2.3.4 Protein residue networks (PRN) are a relatively new topic. In 1999, Kannan and Vishveshwara [39] developed an approach for identifying side-chain clusters in proteins by graph spectral methods. Since then the network theoretic analysis was applied to PRNs mainly for identification of important residues [40]. In the past, distance-based PRNs (D-PRN) were predominantly used, and the use of pair interaction energy (PIE) as a criterion for elucidating the residue contact is a relatively recent phenomenon [41]. The studies of PRN topology and related properties of proteins that have been conducted both for particular cases and generalised for proteins as a class of molecules [42] have focused primarily on D-PRN models.

2.3.5 Recently, a detailed analysis of PIE-based networks, PIE-PRN, was recently developed [43]. It was shown that the topology of these energy-based PRNs is in some regards quite different to that obtained from the distance-based D-PRN, particularly in regard to the small-world character of the network.

2.3.6 Global and/or local efficiencies do not reveal anything about the hubs in the PRN. In contrast, the efficiency centrality, a parameter quantifying the contribution of each residue to the global efficiency, has been found to be useful for identification of hubs that maintain the tertiary structure of the protein [44, 45].

2.3.7 In this chapter PIE-PRN is used to demonstrate that not only are residues in direct contact with another protein at the interface important, but also that residues embedded deeper within the protein may be crucial for quantifying PPI.

2.3.8 In PIE-PRN one building block of the network (usually called node or vertex) corresponds to one fragment as defined in the FMO calculation. Hence, the PIEs ΔE_{IJ} form FMO define the relationship (i.e. the connection) of the residues *I* and *J*. The total interaction energy for fragment *I* is defined as:

$$\Delta E_I^{\text{tot}} = \frac{1}{2} \sum_{J=1}^{\text{proteins 1 and 2}} \Delta E_{IJ} \qquad (4)$$

(where 1/2 is added to avoid double counting). The sum in ΔE_I^{tot} includes both inter- and intra-protein PIEs, thus the picture offered by total PIEs does not necessarily agree with the binding hotspot picture,

based on inter-protein PIEs only. In eq 4, index *I* runs over all residue fragments in both proteins (like index *J*).

2.3.9 As mentioned in paragraph 2.1.6, we used the criterion that the absolute PIE between residues is ≥3.0 kcal/mol in order for this interaction to be represented in the network. Within a network model of the protein it is possible to find a connection between two residues even if they do not interact directly with each other. In such a case, the connection will include some intermediate residues forming a path.

2.3.10 If two residues *I*, *J* interact directly, then the path length *I*↔*J* is one. However, if in order to reach residue *J* from *I* one has to pass through residue *K*, then the path has length 2 (*I*↔*K*↔*J*). Although *I* and *J* do not directly interact, but both interact with *K*. The distance between two residues *I*, *J*, called the shortest path length d_{IJ} in network terms, is in the PIR-PRN model the sum of all reciprocal values of PIEs of residues included in the path. In the above example, ΔE_{IK} and ΔE_{KJ}. The FMO results predict whether residues *I* and *J* interact directly. The network model predicts whether there is possible long-range, or allosteric, interaction/influence between the protein residues *I* and *J*.

2.3.11 It is useful to have a measure of how well a network is internally connected. The definition of the global efficiency by Latora and Marchiori [46] fits this purpose.

$$E_G = \frac{1}{N(N-1)} \sum_{I \neq J} d_{IJ}^{-1} \qquad (5)$$

Herein N is the number of residues in the network. The global efficiency is usually normalised between zero and one and it quantifies how efficiently messages can pass through the network. The closer the number is to one, the more efficient is the communication across the network. In other words, the protein or PPI complex is more prone to allosteric effects.

2.3.12 Additionally, one can define the local efficiency E_{loc}. This metric reveals to what extent the system is fault tolerant or, in other words, resilient to removal of residue interactions. Consider for each residue *I* a subset of the network consisting only of the direct contacts *J* of the *I*-th residue (i.e. only those residues, to which *I* has the path length equal to one, hence *I* is not in the subset.). The efficiencies for these subsets $\tilde{E}_I$, allow the definition of the local efficiency of the whole network E_{loc}. $\tilde{E}_I$ is calculated via equation 5, where one uses *N* equal to the number of residues in the subset.

$$E_{\text{loc}} = \frac{1}{N}\sum_{I}\tilde{E}_I \qquad (6)$$

PIE-PRN networks with larger local efficiency values are more fault tolerant to local residue removal. One can think of this as either a deletion of a residue from the sequence or its replacement (by e.g mutation) by a residue that has different interactions.

2.3.13 A small world network is usually used as a term describing a network in which paths between any random pair of *I* and *J* do not include many

intermediate connections. In other words, the paths are short and the distances are small. For the purpose of this work one can anticipate, that a PIE-PRN with a large E_G may be considered a small world network [44].

2.3.14 Lastly, the efficiency centrality [45, 46] is defined. This metric characterises each residue in the residue network by how much it contributes to the small world character of the PIE-PRN. It can be viewed as an index measuring how much the global efficiency changes, when the *I*-th residue is removed from the network ($E'_{G,I}$), compared to the efficiency with all residues included, E_G

$$C_I^{\text{eff}} = 1 - \frac{E'_{G,I}}{E_G} \qquad (7)$$

Larger values of C_I^{eff} indicate that the node is important with respect to the global efficiency in the PRN. This metric s used to identify both binding hotspots and hubs in PIE-PRN.

3.0 Notes

In this chapter, we demonstrated how three different FMO-based approaches (pair interaction energy analysis (PIE-analysis), subsystem analysis (SA) and analysis of protein residue networks (PRN)) can be applied to explore PPIs between 3 protein complexes: Trypsin/Inhibitor (PDB code 1AVW [30]), Barnase/Barstar (PDB code 1BRS [31]) and Subtilisin/Eglin-C (PDB code 2SEC [32]).

3.1 *Pair interaction energy analysis (PIE-analysis)*

3.1.1 PIE calculations (see Methods sections 2.1) provided a quantitative breakdown of the key interactions formed between residues of two proteins, including their strength (in kcal/mol) and chemical nature (electrostatic or hydrophobic). Typical raw FMO output as calculated for Trypsin/Inhibitor is shown in Table 1

<Table 1 here>

3.1.2 The raw data calculated by FMO as demonstrated in Table 1 can be converted to convenient PPI maps (see methods section 2.1.6) as shown in Figure 2.

<Figure 2a here>

<Figure 2b here>

<Figure 2c here>

Figure 2. PPI maps derived from FMO calculations for trypsin-soy-bean inhibitor (PDB code 1AVW [30]) (**A**), barnase-barstar (PDB code 1BRS [31]) (**B**) and subtilisin/eglin-C (PDB code 2SEC [32]) (**C**) The line between a pair of circles indicates that the interaction energy $\left|\Delta E_{IJ}\right| > 3.0$ kcal/mol, and the thickness of the line is proportional to $\left|\Delta E_{IJ}\right|$. The lines are coloured by their chemical character (PIC): dark blue (hydrophobic, driven by dispersion) and red (electrostatics). The dashed line indicates repulsion $\Delta E_{IJ} \geq 3.0$ kcal/mol and non-dashed line represent attraction $\Delta E_{IJ} \leq -3.0$ kcal/mol. The residues in the first protein in each complex are on the bottom, and those of the second residues are in the top row.

3.1.3 FMO detected that the proteins form a large number of interactions with each other (Figure 2), with some residues forming more than one interaction with residues of the other protein. For example, residue ARG563 of Trypsin inhibitor forms 7 interactions with the residues of Trypsin (Figure 2A), ASP39 of Barstar forms 3 interactions with the residues of Barnase (Figure 2B), and LEU59 of Eglin forms 4 interactions with residues of Subtilisin, see Figure 2C and Figure 3. Such information is essential for structure-based PPI drug design. Overall, the above interactions are dominated by electrostatics, although a hydrophobic (dispersion) contribution is also noticeable.

3.1.4 FMO also detected that backbone atoms play an important role in those PPIs that form a large number of backbone-backbone and backbone-sidechain interactions.

<Figure 3 here>

Figure 3: Example of interaction between Subtilisin and Eglin (PDB code 2SEC [32]) (**A**) Section through the proteins surfaces illustrating interface shape complementarity and four interactions formed between Leu59 of Eglin with four residues of Subtilisin. (**B**) Zoom figure when the key interactions detected by FMO shown as green dashed line.

3.2 **Subsystem analysis (SA)**

3.2.1 Subsystem analysis (SA) was used to analyse the PPI as described in Methods section 2.2. The results of this analysis are shown in Table 1. The values mean attraction, the top part of the list for each complex indicates potential hotspots, such as ASP39 for 1BRS. There is a good correlation between hotspots identified by FMO-PIEs in Figure 2, and potential hotspots in Table 1. However, the two are not identical; in the subsystem analysis large PIEs for charged and polar residues are corrected by the addition of partial fragment energies, and the scores are also conveniently evaluated fragment-wise. LEU59 in Eglin, for instance, has a substantial interaction energy with Subtilisin, however, its binding energy is repulsive due to the addition of contributions missing in the interaction energy.

<Table 2 here>

3.2.2 In addition to attractive hotspots, there are some repulsive spots, as PPI cannot please every residue in each protein, and there are points of dissent. ASN62 in 2SEC is an example of an unhappy fragment. There is more repulsion found in the binding energies than in the interaction

energies due to the desolvation and mutual polarization, both of which are usually repulsive when summed over all fragments (although individual fragment contributions may be attractive). In the above discussion, the protein polarization incorporated into the partial fragment energies means the repulsive destabilization component of the polarization according to the PIEDA picture, whereas the stabilization component of the polarization is a part of PIEs.

3.3 *Analysis of protein residue networks (PRN)*

3.3.1 As described in Methods section 2.3, one of the interesting properties of proteins is that they are adapted to perform relatively fast and precise "communication" across their structure, which is utilised in long-range conformational adaptations - the so-called allosteric effects. The concept of the global and local efficiency introduced by Latora and Marchiori [46] is useful in describing how well the protein is adapted to such long-range "communications". The global efficiency E_G is a value, usually normalised to be between zero and one, which tells how efficiently messages can pass through the network. The closer the number is to one, the more efficient is the communication across the network. As an example, in the subtilisin – eglin complex (PDB code 2SEC [32]) the global efficiency is $E_G = 0.82$, which is a fairly typical value for protein residue networks [43]. High global efficiency is characteristic for so-called small-world networks. Another typical feature is the presence of highly connected nodes, often called hubs, which facilitate efficient communication. The local efficiency of the network is a value quantifying the resilience of the network to the loss

of its capacity for efficient communication (its global efficiency) if one removes the nodes (i.e., the residues). The local efficiency is normally defined as one number for the whole network and should be understood as some average topological property, i.e. it is not used for identification of the hubs. Again, in the subtilisin – eglin complex the value is about 0.11 (see Methods section 2.3). From this one can deduce that potential removal of some residues, could break the ability of the protein for allosteric, long-range, communication and/or for formation of PPI complexes.

3.3.2 The structure of proteins have characteristic features at several levels: the primary amino acid sequence, the secondary structure, including helices, loops and sheets which are the key structural units of proteins. These secondary structures are stabilised by side-chain interactions. Moreover, the relative position of these secondary structure units is also governed by side-chain interactions, creating a tertiary structure. Residues responsible for maintaining this tertiary structure may be considered hubs in the framework of PRN (see Method section 2.3).

3.3.3 The difference between a hub and a hotspot is that a hotspot, in the context of PPI, is directly responsible for binding between the two proteins. In the binding energy picture, potential hotspots are identified based on integrated fragment energies including contributions from all residues, not just those at the interface. If PIEs are used as a guide, a hotspot is a residue in one of the proteins forming the PP complex strongly interacting with, potentially multiple, residues within the other protein in the complex. Hence, hotspots describe only immediate contacts. On the other hand, the network analysis enables one to

identify the hubs, the important central residues which do not necessarily mediate the binding. but are indispensable in maintaining the stability of the binding hotspot neighbourhood.

3.3.4 To summarise, potential hotspot residues are typically located within a secondary structure moiety, e.g. a loop. The position and orientation of the loop can be stabilised by a hub. To identify the hubs, the whole network of pair interactions both within and between proteins is considered (Figure 4). Hubs, therefore, describe residues that can take part in the inter-protein energy transfer, but are more responsible for intra-protein redistribution of the energy. The hotspot view is simpler as it is, in some sense “one-dimensional” (see the row of interactions in Figure 2 or values in Table 1). The hub view is “two-dimensional”, as demonstrated in Figure 5.

<Figure 4 here>

Figure 4. Structure of the complex of Subtilisin and Eglin (PDB code 2SEC [32]). (**A**) Eglin is in green, Subtilisin is blue. (**B**) Colour heat map based on the total interaction energy (see Method section 2.3, Equation 4) of each residue, ΔE_I^{tot} (some potential hotspots are shown in the stick representation). Red and blue colours indicate strong and weak binding, respectively. (**C**) Colour heat maps based on efficiency centraility C_I^{eff} ; red and blue colours represent high and low ranking, respectively. Some domain hubs identified by the efficiency centrality in PIE-PRN are shown in the stick representation.

3.3.5 Asp60 is an example of a potential hotspot with a highly attractive total interaction (denoted in Figure 4B, in red). Other examples are ARG62 and ARG67 of Eglin. These three residues are also identified as

hotspots in Figure 2C. However, ARG65 is not a hotspot, despite having a very strong total attractive interaction – might it be a candidate for a domain hub? We answer this below.

3.3.6 Because of the extra dimensionality, topological plots can be more difficult to interpret visually, as they reflect the complexity of the protein residue network. Global network values are characteristic for the whole protein network. Examples of such values are the global and local efficiencies. These can be useful for comparisons between networks / proteins. On the other hand, “centralities” are properties that are determined for each node (residue) and can be used to compare the contribution of individual residues to the topology of the protein.

3.3.7 It was found recently that the efficiency centrality C_I^{eff} correlates well with the total interaction energy ΔE_I^{tot} [44, 45]. Additionally, residues ranking highly in C_I^{eff} tend to mediate contact with the second protein [43]. From this, it follows that they are important in the spreading of conformational perturbations and that in the equilibrated state they often maintain favourable positions of the binding hotspots. Residues with high efficiency centrality are residues that are responsible for the inter-protein energy transfer and intra-protein redistribution of the energy – they are the hubs. Figure 5 displays the PIE-PRN of the Subtilisin – Eglin complex

<Figure 5 here>

Figure 5. The PIE-PRN of the complex of Subtilisin (blue) and Eglin (green) (PDB code 2SEC [32]). Nodes with $\left|\Delta E_{IJ}\right|$ >3.0 kcal/mol were connected (the sign of ΔE_{IJ} is not used in PRN). The line thickness is proportional to $\left|\Delta E_{IJ}\right|$. The bottom-left scheme highlights the cluster in which some residues are binding hotspots (ASP60, LEU61, ARG62) and ARG65 (a hub, but not a hotspot).

3.3.8 An example of a residue with a high efficiency centrality is ARG65 of Eglin in the Subtilisin – Eglin complex. ARG65 is not a binding hotspot; it has no direct contact to Subtilisin residues, yet Figure 4 shows that it has a large ΔE_I^{tot} value. A further investigation of the PIE-PRN reveals why it scores highly in the C_I^{eff} ranking, i.e. why it is a domain hub. ARG65 has strong interactions with two key binding hotspots, ASP60 and LEU61. Furthermore, it interacts strongly with GLY84 which, in turn, interacts strongly with Arg62, another hotspot. The interaction with ARG67 is indirect, but takes place over a strongly interacting path of two additional nodes. Hence, ARG65 is in close contact with four hotspots and can be considered crucial for the stability of this small cluster of hotspots. Figure 4C shows that there are more such hubs (coloured red, orange or green depending on their ranking). Several of them are, however, quite distant from the binding hotspot and, presumably, are likely responsible for maintaining the stability of these distant secondary structure units. Most frequently, the strong stabilising interactions are electrostatic in nature.

4.0 Conclusions

4.1 In this chapter, we described how three different FMO-based approaches (FMO-PPI) can be applied for PPI analysis of protein-protein complexes. The results produced by these methods demonstrated significant correlations in results but differences were also observed.

4.2 Further improvement of the predictive power of FMO-PPI can be achieved if the conformational averaging in molecular dynamics [47] would be take into account temperature and the possible multiple minima on the energy surface.

4.3 In this stage the further development and optimisation of these methods require experimental validation where it will be cleared when and how to apply each of these pioneering approaches in structure-based PPI drug design.

Acknowledgements

A.H. and A.T.N. would like to acknowledge the support of EU H2020 CompBioMed project (http://www.compbiomed.eu/, 675451). V.S. acknowledges the grants VEGA 2/0035/16 and 2/0031/19 by the Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences.

References

1. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA (1992) Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques. *Proc Natl Acad Sci U S A*. 89(6):2195-2199.
2. Snider J, Kotlyar M, Saraon P, Yao Z, Jurisica I, Stagljar I (2015) Fundamentals of protein interaction network mapping. *Molecular systems biology*. 11(12):848-848.
3. Heifetz A, Katchalski-Katzir E, Eisenstein M (2002) Electrostatics in protein-protein docking. *Protein Sci*. 11(3):571-587.
4. Gan Z, Jessica A, Guillermo G-N (2018) Peptidomimetics Targeting Protein-Protein Interactions for Therapeutic Development. *Protein & Peptide Letters*. 25(12):1076-1089.
5. Robertson NS, Spring DR (2018) Using Peptidomimetics and Constrained Peptides as Valuable Tools for Inhibiting Protein(-)Protein Interactions. *Molecules*. 23(4).
6. Díaz-Eufracio BI, Naveja JJ, Medina-Franco JL: Chapter Three - Protein–Protein Interaction Modulators for Epigenetic Therapies. In: *Advances in Protein Chemistry and Structural Biology.* Edited by Donev R, vol. 110: Academic Press; 2018: 65-84.
7. Stevers LM, Sijbesma E, Botta M, MacKintosh C, Obsil T, Landrieu I, Cau Y, Wilson AJ, Karawajczyk A, Eickhoff J *et al* (2018) Modulators of 14-3-3 Protein-Protein Interactions. *J Med Chem*. 61(9):3755-3778.
8. Moreira IS, Fernandes PA, Ramos MJ (2007) Hot spots—A review of the protein–protein interface determinant amino-acid residues. *Proteins: Structure, Function, and Bioinformatics*. 68(4):803-812.
9. Bissantz C, Kuhn B, Stahl M (2010) A medicinal chemist's guide to molecular interactions. *J Med Chem*. 53(14):5061-5084.
10. Tong Y, Mei Y, Li YL, Ji CG, Zhang JZ (2010) Electrostatic polarization makes a substantial contribution to the free energy of avidin-biotin binding. *J Am Chem Soc*. 132(14):5137-5142.
11. Raha K, Peters MB, Wang B, Yu N, Wollacott AM, Westerhoff LM, Merz KM, Jr. (2007) The role of quantum mechanics in structure-based drug design. *Drug Discov Today*. 12(17-18):725-731.
12. Beratan DN, Liu C, Migliore A, Polizzi NF, Skourtis SS, Zhang P, Zhang Y (2015) Charge transfer in dynamical biosystems, or the treachery of (static) images. *Acc Chem Res*. 48(2):474-481.
13. Ozawa T, Okazaki K, Kitaura K (2011) CH/pi hydrogen bonds play a role in ligand recognition and equilibrium between active and inactive states of the beta2 adrenergic receptor: an ab initio fragment molecular orbital (FMO) study. *Bioorg Med Chem*. 19(17):5231-5237.
14. Fedorov DG, Nagata T, Kitaura K (2012) Exploring chemistry with the fragment molecular orbital method. *Phys Chem Chem Phys*. 14(21):7562-7577.
15. Lu Y-X, Zou J-W, Wang Y-H, Yu Q-S (2007) Substituent effects on noncovalent halogen/π interactions: Theoretical study. *International Journal of Quantum Chemistry*. 107(6):1479-1486.
16. Gallivan JP, Dougherty DA (1999) Cation-pi interactions in structural biology. *Proc Natl Acad Sci U S A*. 96(17):9459-9464.
17. Johnston RC, Cheong PH (2013) C-H...O non-classical hydrogen bonding in the stereomechanics of organic transformations: theory and recognition. *Org Biomol Chem*. 11(31):5057-5064.
18. Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, Shirley BA, Hendricks MM, Iimura S, Gajiwala K, Scholtz JM *et al* (2011) Contribution of hydrophobic interactions to protein stability. *Journal of molecular biology*. 408(3):514-528.
19. Popov P, Peng Y, Shen L, Stevens RC, Cherezov V, Liu ZJ, Katritch V (2018) Computational design of thermostabilizing point mutations for G protein-coupled receptors. *Elife*. 7.
20. Yu N, Li X, Cui G, Hayik SA, Merz KM, 2nd (2006) Critical assessment of quantum mechanics based energy restraints in protein crystal structure refinement. *Protein Sci*. 15(12):2773-2784.
21. Fedorov DG, Kitaura K (2007) Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J Phys Chem A*. 111(30):6904-6914.
22. Phipps MJ, Fox T, Tautermann CS, Skylaris CK (2015) Energy decomposition analysis approaches and their evaluation on prototypical protein-drug interaction patterns. *Chem Soc Rev*. 44(10):3177-3211.
23. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M (1999) Fragment molecular orbital method: an approximate computational method for large molecules. *Chemical Physics Letters*. 313(3–4):701-706.

24. Heifetz A, Chudyk EI, Gleave L, Aldeghi M, Cherezov V, Fedorov DG, Biggin PC, Bodkin MJ (2016) The Fragment Molecular Orbital Method Reveals New Insight into the Chemical Nature of GPCR-Ligand Interactions. *J Chem Inf Model*. 56(1):159-172.
25. Alexeev Y, Mazanetz MP, Ichihara O, Fedorov DG (2012) GAMESS as a free quantum-mechanical platform for drug research. *Curr Top Med Chem*. 12(18):2013-2033.
26. Fedorov DG, Kitaura K (2007) Pair interaction energy decomposition analysis. *J Comput Chem*. 28(1):222-237.
27. Fedorov DG, Kitaura K (2012) Energy Decomposition Analysis in Solution Based on the Fragment Molecular Orbital Method. *The Journal of Physical Chemistry A*. 116(1):704-719.
28. Chudyk EI, Sarrat L, Aldeghi M, Fedorov DG, Bodkin MJ, James T, Southey M, Robinson R, Morao I, Heifetz A (2018) Exploring GPCR-Ligand Interactions with the Fragment Molecular Orbital (FMO) Method. *Methods Mol Biol*. 1705:179-195.
29. Ben-Shimon A, Eisenstein M (2010) Computational mapping of anchoring spots on protein surfaces. *J Mol Biol*. 402(1):259-277.
30. Song HK, Suh SW (1998) Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from Erythrina caffra and tissue-type plasminogen activator. *J Mol Biol*. 275(2):347-363.
31. Buckle AM, Schreiber G, Fersht AR (1994) Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. *Biochemistry*. 33(30):8878-8889.
32. McPhalen CA, James MN (1988) Structural comparison of two serine proteinase-protein inhibitor complexes: eglin-c-subtilisin Carlsberg and CI-2-subtilisin Novo. *Biochemistry*. 27(17):6582-6598.
33. Fedorov DG, Kitaura K (2004) The importance of three-body terms in the fragment molecular orbital method. *The Journal of Chemical Physics*. 120(15):6832-6840.
34. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S *et al* (1993) General atomic and molecular electronic structure system. *Journal of Computational Chemistry*. 14(11):1347-1363.
35. Fedorov DG, Kitaura K (2004) Second order Moller-Plesset perturbation theory based upon the fragment molecular orbital method. *J Chem Phys*. 121(6):2483-2490.
36. Li H, Fedorov DG, Nagata T, Kitaura K, Jensen JH, Gordon MS (2010) Energy gradients in combined fragment molecular orbital and polarizable continuum model (FMO/PCM) calculation. *J Comput Chem*. 31(4):778-790.
37. Fedorov DG (2019) Solvent Screening in Zwitterions Analyzed with the Fragment Molecular Orbital Method. *Journal of Chemical Theory and Computation*.
38. Fedorov DG, Kitaura K (2016) Subsystem Analysis for the Fragment Molecular Orbital Method and Its Application to Protein–Ligand Binding in Solution. *The Journal of Physical Chemistry A*. 120(14):2218-2231.
39. Kannan N, Vishveshwara S (1999) Identification of side-chain clusters in protein structures by a graph spectral method. *J Mol Biol*. 292(2):441-464.
40. Salamanca Viloria J, Allega MF, Lambrughi M, Papaleo E (2017) An optimal distance cutoff for contact-based Protein Structure Networks using side-chain centers of mass. *Sci Rep*. 7(1):2838.
41. Vijayabaskar MS, Vishveshwara S (2010) Interaction Energy Based Protein Structure Networks. *Biophysical Journal*. 99(11):3704-3715.
42. Estrada E, Hatano N, Benzi M (2012) The physics of communicability in complex networks. *Physics Reports*. 514(3):89-119.
43. Sladek V, Tokiwa H, Shimano H, Shigeta Y (2018) Protein Residue Networks from Energetic and Geometric Data: Are They Identical? *J Chem Theory Comput*. 14(12):6623-6631.
44. Wang S, Du Y, Deng Y (2017) A new measure of identifying influential nodes: Efficiency centrality. *Communications in Nonlinear Science and Numerical Simulation*. 47:151-163.
45. Sladek V (2018) A note on the interpretation of the efficiency centrality. *Communications in Nonlinear Science and Numerical Simulation*. 61:225-229.
46. Latora V, Marchiori M (2001) Efficient behavior of small-world networks. *Phys Rev Lett*. 87(19):198701.
47. Fedorov DG, Kitaura K (2018) Pair Interaction Energy Decomposition Analysis for Density Functional Theory and Density-Functional Tight-Binding with an Evaluation of Energy Fluctuations in Molecular Dynamics. *J Phys Chem A*. 122(6):1781-1795.

Table 1. Typical raw FMO output for PIE analysis, calculated for Trypisin/Inhibitor (PDB code 1AVW [30]) when F.name1 are residues Trypsin inhibitor and F.name2 are residues of Trypsin. In **bold** we marked potential hotspot. PIEtotal (total PIE as calculated by Equation 1), PIEes (electrostatic term), PIEex (exchange repulsion term), PIEct (charge-transfer term), PIEdisp (dispersion), PIEsolv (solvation term) and PIC (pair interaction character as calculated by equation 2)

F.name1	F.name2	PIEtotal	PIEes	PIEex	PIEct	PIEdisp	PIEsolv	PIC
ASP501	LYS60	-15.16	-53.91	8.37	-3.72	-4.56	38.65	0.77
TYR562	HIS57	-6.41	-5.53	7.15	-2.36	-8.45	2.77	0.25
TYR562	GLY96	-13.39	-19.99	9.25	-2.88	-3.59	3.83	0.82
TYR562	LEU99	-3.22	-1.03	5.50	-2.73	-4.97	0.01	0.17
TYR562	TRP215	-7.16	-4.06	6.69	-4.32	-6.25	0.78	0.34
PRO561	GLY216	-8.30	-13.65	3.47	0.70	-2.40	3.58	0.81
ARG563	**ASP189**	-35.17	-134.46	28.62	-10.57	-7.49	88.73	0.86
ARG563	**SER190**	-6.71	-17.25	2.58	-1.42	-3.24	12.63	0.59
ARG563	**GLN192**	-4.42	-7.90	6.11	-2.65	-5.95	5.97	0.25
ARG563	**SER195**	-11.34	-16.36	8.23	-2.34	-7.59	6.71	0.56
ARG563	**SER214**	-9.06	-13.80	4.89	-1.87	-3.78	5.50	0.69
ARG563	**GLY219**	-10.63	-31.43	11.52	-3.29	-5.55	18.13	0.71
ARG563	**GLY193**	-8.43	-17.75	12.11	-3.19	-3.45	3.85	0.80
ARG565	HIS40	-11.96	-28.45	7.78	-3.85	-5.28	17.84	0.67
ARG565	PHE41	-6.86	-10.35	2.88	-0.52	-2.85	3.98	0.69
ARG565	TYR151	-7.29	-3.08	3.53	-2.24	-6.27	0.76	0.27
HIS571	HIS57	-7.89	-14.20	4.38	-1.64	-2.66	6.23	0.75
PRO572	GLY96	-4.90	-4.11	2.43	-1.39	-3.69	1.86	0.38
TRP617	ASN97	-3.32	-5.79	0.52	-0.90	-1.34	4.20	0.54

Table 2. Potential hotspots in PPI using subsystem analysis (score is equal to ΔE_I^{bind}, in kcal/mol, and only values $\left|\Delta E_I^{\text{bind}}\right| > 3.0$ are shown).

fragment	protein	score	fragment	protein	score	fragment	protein	score
1BRS[a]			1AVW[b]			2SEC[c]		
ASP39	2	-18.2	ILE564	2	-11.8	GLY63	1	-10.2
ARG59	1	-11.5	ASP189	1	-10.6	ASP60	2	-6.3
ASN84	1	-9.4	TYR562	2	-7.6	ARG62	2	-5.4
GLU60	1	-8.8	CYS220	1	-5.7	THR58	2	-3.9
ASP35	2	-8.5	PHE41	1	-5.6	TYR104	1	-3.8
HIS102	1	-7.6	TRP215	1	-5.5	LEU126	1	-3.6
GLU76	2	-7.6	CYS58	1	-3.2	PRO56	2	4.9
ARG83	1	-7.4	VAL227	1	4.0	ASN62	1	5.9
ARG87	1	-7.2	CYS636	2	4.2			
TYR103	1	-6.1						
TYR29	2	-5.9						
THR42	2	-4.4						
ASN58	1	-3.3						
ASN33	2	-3.2						
LYS27	1	-3.0						
SER38	1	3.1						
GLY43	2	4.2						

[a] 1 is barnase, 2 is barstar

[b] 1 is trypsin, 2 is trypsin inhibitor

[c] 1 is subtilisin, 2 is eglin