J Comput Aided Mol Des (2009) 23:185–194
DOI 10.1007/s10822-008-9250-3

Interaction force diagrams: new insight into ligand-receptor binding

Hooman Shadnia · James S. Wright · James M. Anderson

Received: 18 July 2008 / Accepted: 16 October 2008 / Published online: 7 November 2008

Abstract A method is described to calculate and visualize the interaction forces of ligand-receptor complexes. Starting from an X-ray crystallographic structure, a "thawing" procedure results in a force-field energy-minimized geometry which is close to the crystallographic starting point. By subtracting non-bonded interactions of the ligand with each amino acid residue and using the resulting force vectors to describe the slope of the remaining potential, two types of interaction force diagrams are created; the first shows the direction of the force vectors in 3D and the second shows the magnitude of the force vectors. The latter representation leads to definition of an 'Interaction Force Fingerprint' (IFFP) which is characteristic of the ligand-receptor binding. IFFPs are used to discuss ligand binding in the human estrogen receptors ERα and ERβ, and provide new insight into ligand selectivity between receptor isoforms.

Keywords Force vectors · Interaction Force Fingerprints · IFFP · Ligand-receptor · Estrogen receptor · Selectivity

Introduction

As of 2008, more than 8000 structures per year are becoming available through the protein databank [1]. It is now possible to find structures of many proteins with different ligands, including ligands with different affinities or different functions (agonist—antagonist, or enzyme inhibitors). With all this data becoming available, there is substantial demand for a fast and an efficient method for comparative studies of related structures. Currently there are two main types of algorithms used to perform such comparative studies; geometry-based algorithms and potential-energy based algorithms.

Geometry-based methods rely only on geometrical descriptors (distances, angles, etc.). Examples of these methods include Structural Interaction Fingerprint method (SIFt) [2], Relibase [3], the algorithm for Comparison of Protein Active Site Structures (CPASS) [4], the protein—ligand interaction fingerprints module in MOE [5] and the Ligand-Protein Contact algorithm and Contacts of Structural Units algorithm (LPC-CSU) [6]. These methods rarely require energy optimization of the protein geometry, as they do not involve calculation of any energy criteria. As a result, some of these programs use 'similarity functions' based on atom types and charges in order to classify the favorable and unfavorable interactions. However, abandoning the known potential energy terms brings a set of problems to these methods. For example, without an energy criterion, it is not possible to identify whether the interaction between two given hydrophobic groups in close proximity is attractive or repulsive. This is because in reality existence of any partial charges on the two groups adds an electrostatic term to the potential energy, and alters the position of the minimum energy geometry. Such important details are not modeled in geometry-based methods which use coarse cutoffs to identify the hydrophobic interactions. Similarly, the effects of other major potentials including the polar interactions and intramolecular (conformational) energies are only loosely approximated using these geometry-based algorithms.

Electronic supplementary material The online version of this article (doi:10.1007/s10822-008-9250-3) contains supplementary material, which is available to authorized users.

H. Shadnia (✉) · J. S. Wright · J. M. Anderson
Department of Chemistry, Carleton University, Ottawa Canada K1S 5B6
e-mail: hooman@shadnia.com

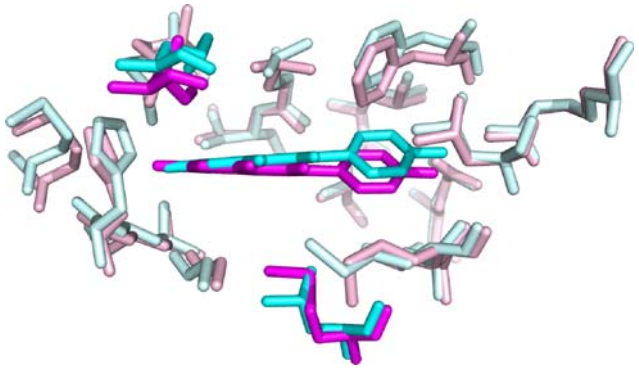


Fig. 1 Genistein in active sites of ERα (cyan) versus ERβ (pink). Raw coordinates are from crystal structures 1X7R, 1X7J. The ligands and the subtype-specific residues are highlighted. The geometrical differences between the two complexes are subtle (RMSD = 0.78), and unrelated to the position of the subtype-specific residues and ligands

The crudeness of geometry-based approaches is highlighted when these methods are used for comparative study of protein complexes. The noise in the crystal coordinates can be much larger than the geometrical differences related to biological activity. To illustrate this, crystal geometries[1] of complexes of genistein with human estrogen receptor subtypes α and β are compared in Fig. 1. Genistein is an ERβ-selective compound (IC50 = 360 nM for ERα, 9 nM for ERβ [7]). As can be seen in the picture, the geometrical differences of ligand and active site residues are subtle (RMSD = 0.87) and difficult to explain. In fact, the subtype-specific residues which are highlighted in this picture (Met 421, Leu 384 for ERα vs. Leu373, Met336 in ERβ) are positioned above and below the ligand plane, so the distances of ligand atoms to these residues are very similar in both receptor subtypes. Thus the preference of this ligand for ERβ cannot be explained via any geometry descriptors.

The second type of comparative algorithms are potential energy-based. The existing methods in this category use different potential functions for pairs of molecular fragments (for example the ligand and individual residues in the active site) mostly by using empirical functions. Examples of such programs include Ligplot [8] and COMBINE [9, 10]. By using energy functions, these methods are superior to the geometry based-methods, and they usually require some degree of preparation and energy minimization on the protein geometry. Ligplot uses coarse and almost semi-qualitative potential energy functions namely the atom clash energy $\mathbf{E}_a$ and the bond overlap energy $\mathbf{E}_b$, along with an experimental weight factor to generate the final energy function. The most significant interactions produced in this way are schematically represented on 2D diagrams of the protein-ligand complex. COMBINE on the other hand, uses a more elaborate approach. The original algorithm of COMBINE [9] partitions the calculated binding energy into *pair-wise terms* for all the interacting fragments of ligand and receptor. It then further decomposes each pair-wise term to the *basic components*, i.e. the ten terms describing the bonded and non-bonded potentials. Next, COMBINE uses a chemometrical analysis to analyze thousands of descriptors created, to find the 'signal', i.e. the terms that correlate with activity. The more recent versions of the algorithm exclude the internal (conformational) potential terms, since they were considered to be a source of noise, cumbersome to compute, and their inclusion did not improve the regression models significantly on the systems that have been used for validation [10]. Some variations also include the solvation terms in the set of descriptors [11–14]. The partitioning of the potential function into pair-wise terms can be justified, since it may address possible inaccuracies in geometries and energetics of particular residues. But the additional decomposition of the potential energy function into basic components seems unjustified and counterproductive, since it is well known that these individual components are empirically defined and modified to yield reasonable overall potentials and geometries. At the same time, the decomposition increases the dimensions of an already large dataset by 10 times, thus necessitating the use of a variable selection method and forcing the algorithm to make limited choices between the components. As a result the following chemometrical analysis can result in multiple models which recruit different sets of variables from the large pool of available descriptors [9]. All these can make the validation and interpretation of results difficult. Nonetheless, it is possible to link the corresponding models to physicality, especially when a small number of variables can explain the majority of the variance [15]. By comparison, in the method we shall describe we maintain the integrity and physicality of the pair-wise potential energy terms by avoiding such decomposition and arbitrary elimination of its components.

Even when the potential terms are defined properly, some intrinsic features of these terms limit their application, representation, and usefulness as descriptors. Further improvements can be obtained by using the derivatives of force-field potentials, i.e. the interaction forces. While molecular forces have been essentially described ever since the force-fields were available, they have rarely been used for the analysis of structures and interactions. However as we will show in several examples, there are certain advantages in using the molecular forces instead of the potentials. First, unlike the potentials which are scalar numbers, the forces are 3D vectors and can be graphically represented and overlaid onto the geometry, in an understandable fashion. Force vectors pointing towards a fragment will indicate pro-binding interactions and vice versa. Second, the well defined physicochemical interaction forces are understandable for a

[1] PDB entries 1X7R, 1X7J [7].

single complex and there is no need for comparisons to interpret the outcome. Furthermore, as we will demonstrate, a statistical treatment is not necessary when a comparison between two or more complexes is to be made using the interaction forces. These features expand the applicability of the method to virtually any single complex that is appropriately modeled in-silico. Moreover, a partitioned potential term for a ligand-residue pair needs to be subtracted by its baseline value, i.e. the equivalent term for the free reactants, while the partitioned force terms all have the same baseline, i.e. a zero length vector. This even eliminates the need to model the free reactants for study of a single complex. It is also foreseeable that when used along with a chemometrical training system, data of molecular forces should be superior over the potentials, as they do not require individual offset values.

In this paper we define a method to partition the native force-field potentials and derive the net interaction forces. We use the well established crystal structures of estrogenic ligands for the analysis and demonstrate the intricate level of details which are revealed by using the method. We demonstrate how the method can be used to describe an individual complex, and to compare complexes of different ligands to the same receptor. Also, complexes of some selective estrogenic ligands are analyzed, leading to new insight into the origin of selectivity, which can be used for the purpose of drug design.

Methods

The PDB structures were subjected to a structure preparation and energy minimization procedure as follows. First, hydrogen atoms were added to the protein and ligand structures. Using MOE2007 software energy minimization was performed using MMFF94s force-field [16, 17] and the Born solvation model [18]. Since the hydrogen atoms were initially added in default orientations, they can cause huge steric clashes. Direct energy minimization at this point alters the geometry too much from crystal coordinates. To address this problem, we use an in-house 'thawing' algorithm which uses tether forces to limit the movements of atoms of interest. In the MOE potential model, tether forces are quadratic forces applied to each atom to pin the atom close to its original coordinates. Thus, the energy minimization procedure will be biased towards solutions which move atoms selectively with no or smaller tether forces. After selection of the ligand atoms, the thawing algorithm defines four shells of amino acids with tether constants which diminish with distance from the ligand, and then gradually reduces the tether constants during step-by-step energy minimization. To validate this, Table 1 compares the effects of thawing versus direct energy minimization on geometries and energies of two receptor chains from PDB entry 1GWR [19]. Table 1a shows that the RMSD values for the thawed complexes are lower than those which were directly minimized (e.g. For 1GWR.B, 1.00 RMSD thawed, versus 1.25 RMSD direct). Table 1b shows that ligand—receptor interaction energies are more negative and more consistent (-34.2 ± 0.4 kcal/mol) for the thawed complexes, in comparison with energy minimized complexes (-28.8 ± 2.0 kcal/mol).

Table 1 Effect of thawing: (a) Geometrical comparison: pair-wise RMSD values for superposition of heavy atoms of protein chains from the PDB entry 1GWR. Values are in angstroms, (b) Energy comparison: force field interaction energies of ligand-receptor chains. Values are in kcal/mol

	1GWR.A	1GWR.B
(a)		
Crystal vs. thawed	1.19	1.00
Crystal vs. minimized	1.25	1.25
(b)		
Thawed	−33.83	−34.63
Minimized	−26.78	−30.88

After preparation as described above, the geometry of the ligand-receptor complex which results is very similar to the original PDB structure, and is a potential energy minimum in the chosen force field. At the minimum the slope in the potential is zero, so the sum of all the bonded and non-bonded interaction forces on each particular atom in the system is zero. An important requirement at this step is to partition or decompose the molecular forces on each atom, to create terms describing the pair-wise interactions between ligand and receptor atoms. There are several ways to perform this decomposition. Here we describe one method that is more suitable for the purpose of study of ligand-receptor interactions.

Consider a simple example, the water dimer, shown in Fig. 2 at its potential minimum. Here one water molecule $\boldsymbol{w_1}$ is (arbitrarily) chosen as ligand, $\boldsymbol{w_2}$ as receptor. The force field energy terms are $V = V_1 + V_2 + V_{12}$, where V_1 contains the bond stretching, angle bending and stretch-bend interaction terms for ligand (**L**) $\boldsymbol{w_1}$, V_2 contains the equivalent terms for receptor (**R**) $\boldsymbol{w_2}$, and V_{12} consists of

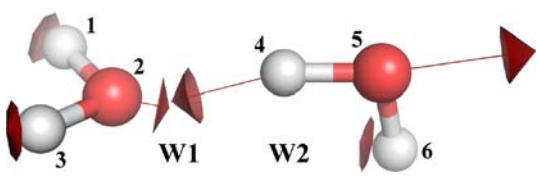


Fig. 2 Atomic forces on water dimer. Despite being in an energy-minimized geometry, the method sums non-bonded interactions between the two residues which are non-zero

the non-bonded interaction terms between w_1 and w_2. Numbering the atoms (1,2,3) on **L** and (4,5,6) on **R**, the latter are pair-wise van der Waals interactions between 1–4, 1–5, 1–6, 2–4,…3–6, as well as pair-wise coulombic interactions between partial charges $\Delta q_1 \Delta q_4 / R_{14}$, etc. *Now suppose that we remove all the non-bonded interaction terms contained in* V_{12}. In that case we are no longer at a potential minimum and there are forces present which represent the contribution from the non-bonded interactions. These forces are the opposite of the sum of the disabled forces. By reversing their direction, we will produce the correct sum of the disabled forces. Figure 2 shows that there are force vectors between atoms 2 (oxygen on **L**) and 4 (hydrogen on **R**) bringing the sub-units together, i.e. a hydrogen bond. However, there is also a force vector on atom 5 (oxygen on **R**) pulling the sub-units apart; this is caused by repulsion between partial negative charges on the two oxygen atoms. The small resultant vectors on atoms 2 and 4 pulling the subunits together show that there is an attractive interaction, which is usually identified as a hydrogen bond.

The water dimer example shows that this type of analysis brings added insight into how the ligand is bound to the receptor. Now let us generalize to the case where there are multiple amino acid residues surrounding a ligand. Figure 3 shows how this can be done. Here there are three residues $\mathbf{R_1}$, $\mathbf{R_2}$ and $\mathbf{R_3}$ surrounding ligand **L**. The force-field energy will contain the local terms V_L, V_{R1}, V_{R2} and V_{R3}, as well as all the non-bonded interaction terms such as $V_{L-R1}, \ldots V_{R2-R3}$. All of the interaction terms between the residues are retained. Figure 3 highlights the interactions of a particular atom of ligand with residue R_1. For the indicated atom on L, all the non-bonded interaction terms with R_1 are removed. The same is done for the rest of the atoms on L interacting with R_1. However, all other pair-wise non-bonded interactions are retained. The resulting force vectors are reversed as explained above. The focus then shifts to L–R_2 and then L–R_3, and the process is repeated. Thus, the final potential energy sum omits all non-bonded ligand-residue terms but retains all other terms. As in the water dimer, there will be remaining forces which can be helpful in discussing ligand-residue interactions.

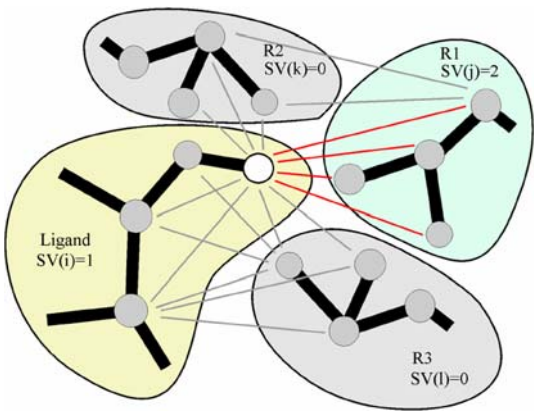


Fig. 3 Diagram of calculation method. By using different non-zero *state values* for the atoms of the ligand and target residue (R_1) and using a scale factor of zero for 'unlike' interactions ($T_{unlike} = 0$), the non-bonded interactions between ligand and R_1 (shown in red) will be zeroed. The sum of atomic forces at this point is equal to sum of disabled forces, in the opposite direction

To avoid re-defining or approximating the true force-field potentials, we used the program MOE2007, which allows access to the internal functions. We use MOEs potential model to scan through all the atomic interactions. In MOE, all the non-bonded potentials are scaled by a subset-dependent *scale factor* SF_{ij}. Associated with each atom *i*, there is a *state value* SV_i which is an integer value (default = 0) and can be set arbitrarily for any subset of atoms. For any given pair of atoms (i,j), based on their *state values* (SV_i, SV_j) the *scale factor* SF_{ij} will be equal to one of three constant values of T_{like}, $T_{wildcard}$, $T_{unlike;}$

$$\begin{aligned}
&\text{if } SV_i = SV_j \quad \text{then } SF_{ij} = T_{like}\\
&\text{if } SV_i \neq SV_j \text{ and } (SV_i = 0 \text{ or } SV_i = 0) \quad \text{then } SF_{ij} = T_{wildcard}\\
&\text{Otherwise} \quad SF_{ij} = T_{unlike}
\end{aligned}$$

As shown in Fig. 3, by setting different non-zero *state values* on atoms of ligand and those of the amino acid of interest ($\mathbf{R_1}$), the non-bonded interactions between the the two groups will fall into the 'unlike' category and the scaling factor SF_{ij} for them will be set equal to T_{unlike}. By setting T_{unlike} equal to zero, these interactions will be disabled. Even though the complex is in energy minimized geometry, at this point the sum of all the forces on each atom of the ligand and $\mathbf{R_1}$ will not be equal to zero. Instead they will be equal to the sum of the disabled forces, but in the opposite direction. The values of these forces which are scalar derivatives of force-field potentials $(-dU/dx, -dU/dy, -dU/dz)$ can be calculated using MOE's *Potential[]* function, and may be shown as a 3D vector centered on the corresponding atoms. By switching the direction of these vectors, the resulting vectors show the 'overall non-bonded forces' of each object (ligand—R_1) on the other. The algorithm then continues to the next amino acid in the active site and repeats the procedure until it scans all the amino acids within a user defined distance of the ligand (usually 6Å). MOE's *Potential[]* function uses kcal/mol Å as the unit for force. When drawing the force vectors, our algorithm uses user-defined scale factors for graphical representation of these forces.[2]

[2] This program is freely available to MOE users at http://svl.chemcomp.com.

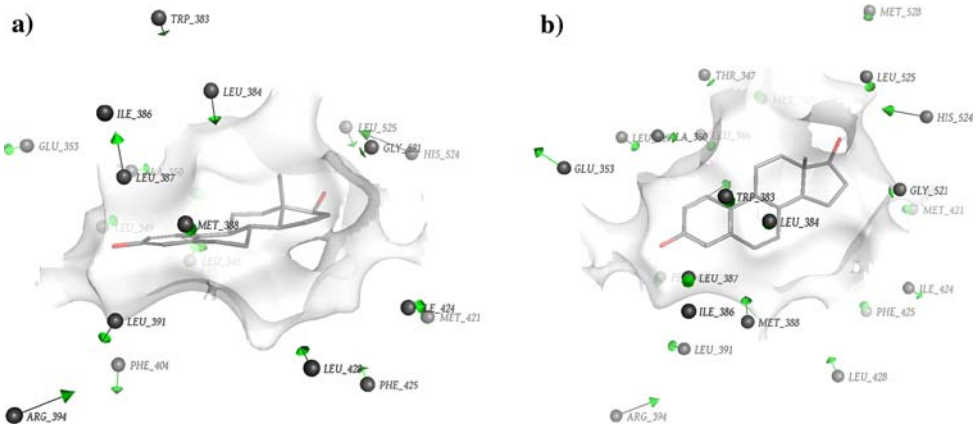
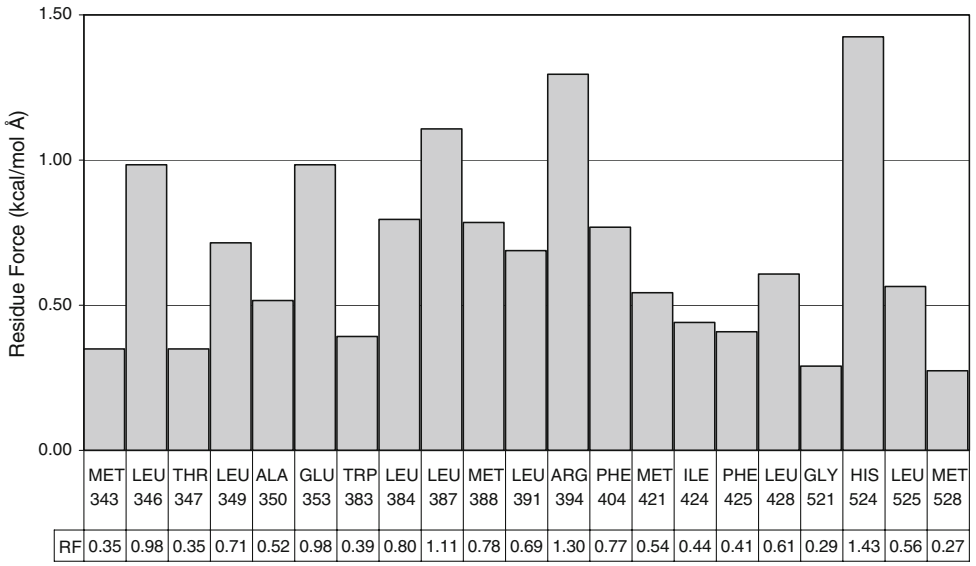


Fig. 4 Residue forces of estradiol in hERα. **a** Side view and **b** Top view. Residue forces are shown on the alpha carbons. The cavity surface is rendered in semi-transparent gray

Chart 1 Interaction Force Fingerprint of estradiol in the active site hERα



	MET 343	LEU 346	THR 347	LEU 349	ALA 350	GLU 353	TRP 383	LEU 384	LEU 387	MET 388	LEU 391	ARG 394	PHE 404	MET 421	ILE 424	PHE 425	LEU 428	GLY 521	HIS 524	LEU 525	MET 528
RF	0.35	0.98	0.35	0.71	0.52	0.98	0.39	0.80	1.11	0.78	0.69	1.30	0.77	0.54	0.44	0.41	0.61	0.29	1.43	0.56	0.27

Using this method on several complexes of estrogenic ligands in the estrogen receptor, we will demonstrate that the direction and magnitude of the force vectors are useful and informative about the nature of interactions in these systems, and can be used to make comparisons.

Results and discussion

The first example for the application of this method in studying a single complex is shown in Fig. 4 where we examine the forces between the ligand estradiol and the cavity of human estrogen receptor α from PDB entry 1GWR. Here we make use of a more compact method to illustrate the interaction forces. Only alpha carbons of active site residues are shown, and the atomic force vectors on each residue are summed (in 3D) to create 'residue forces'.[3] To display these vectors in a more compact format, one can display only their magnitude (vector length) in a graph (Chart 1). Since these diagrams describe the unique pattern of interactions in the complex, we designate these representations as 'Interaction Force Fingerprints' or IFFPs. The residues with forces smaller than 0.1 kcal/mol Å are not shown on this graph. As can be seen, the three important pharmacophore amino acids of estrogen receptor (Glu353, Arg394, His524) are among those with the largest residue forces, as would be expected. However, Leu346 and Leu387 also show strong residue forces and hence contribute to determining the strength of ligand binding. As shown, weaker but non-negligible forces persist for more than 10 other residues. This implies that the energetics of optimal ligand binding to a protein receptor are more complex than normally imagined.

[3] Since it is difficult to visualize the 3D details of force vectors using the 2D images printed here, the original files for the figures discussed in this manuscript are provided as supplemental information. These files can be viewed with PyMol [20] which is freely available from www.pymol.org.

One problem with the IFFP diagrams is that they do not show the direction of interactions with respect to the ligand (attractive or repulsive). By referring to the 3D picture, we can see that within the top five amino acids named here, overall forces on Glu353 and Leu387 are repulsive, as their force vectors point away from the ligand. One can easily trace the origin of these repulsions by referring to a detailed picture of atomic forces on these residues (provided as supplemental data); e.g. the carbon atom of the side chain carboxylate of Glu353 which has a point charge of +0.906 is repelled by the hydroxyl hydrogen of the ligand. Also, one of the hydrogens on the beta carbon of Leu387 has a steric clash with the phenol ring on the ligand. Nevertheless, important information is instantly revealed from IFFPs, including particularly the relative magnitude of the interactions. For example, the interactions of the two leucines mentioned above stand out from all the other hydrophobic residues in the active site. Also, IFFPs reveal the true extent of interactions since they do not use arbitrary distance-based threshold values for inclusion or exclusion of residues.

Our first example of the application of IFFPs for the *functional* comparison of two complexes examines the binding of estradiol (E2) in the human estrogen receptor alpha (hERα) vs. the derivative 4-methyl E2. In this case it is known that there is not enough room in the narrow cavity around the A-ring (phenol) to accommodate the methyl group, as shown in the IC_{50} values (0.5 nM for E2, 125.9 nM for 4-methylE2 [21]). After adding the methyl group at the 4 position and performing an energy minimization, the 3D force vectors were calculated for both ligands and shown in Fig. 5. The IFFPs are also shown in Chart 2 for both ligands. As can be seen, the 4-methyl group causes a significant increase in the forces on residues on two different sites in the cavity. The first site is in direct contact with the 4-methyl group (Leu387, Leu391, Phe404, Ala350, Glu353) and the second site is on the opposite side of the cavity (Met343, His524,

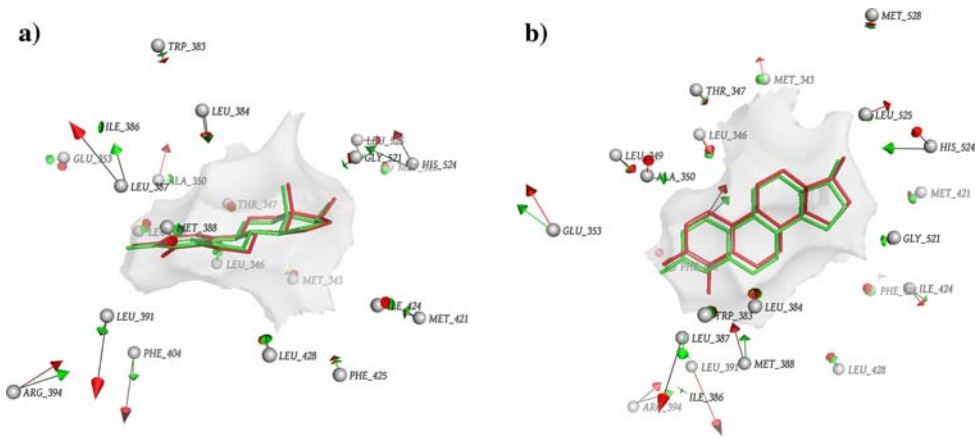
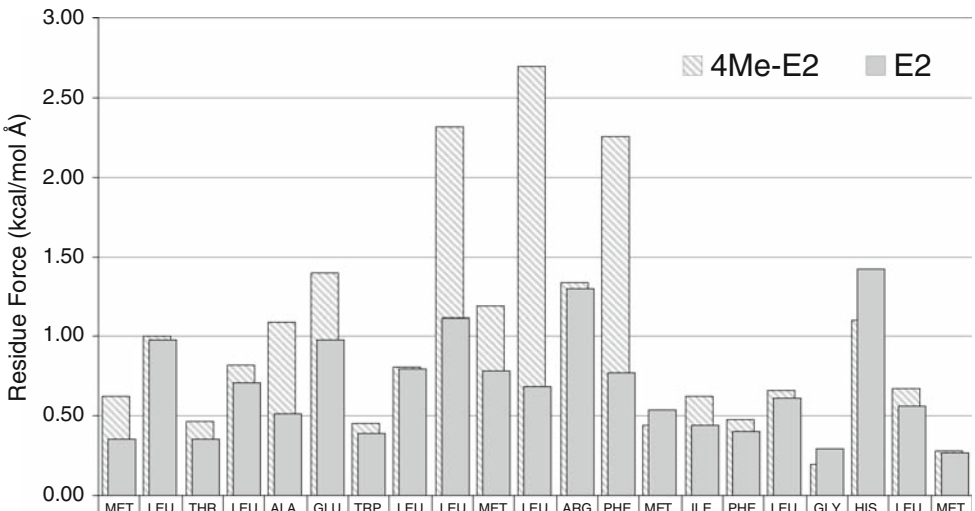


Fig. 5 Comparison of residue forces of complexes of estradiol (green) and 4-methylestradiol (red) in hERα. **a** Side view and **b** Top view. Residue forces of each complex are shown using similar colors. The major differences are on two sites, one close to the 4-methyl group, and the other on the opposite side of the cavity



	MET 343	LEU 346	THR 347	LEU 349	ALA 350	GLU 353	TRP 383	LEU 384	LEU 387	MET 388	LEU 391	ARG 394	PHE 404	MET 421	ILE 424	PHE 425	LEU 428	GLY 521	HIS 524	LEU 525	MET 528
4Me-E2	0.63	1.01	0.46	0.82	1.10	1.39	0.46	0.81	2.31	1.19	2.70	1.33	2.25	0.44	0.62	0.48	0.66	0.19	1.10	0.68	0.28
E2	0.35	0.98	0.35	0.71	0.52	0.98	0.39	0.80	1.11	0.78	0.69	1.30	0.77	0.54	0.44	0.41	0.61	0.29	1.43	0.56	0.27

Chart 2 Comparison of IFFPs of two different ligands with hERa

Leu525). This means that the ligand is re-adjusting its position due to the steric clashes initiated from the extra methyl group. While the geometrical variations of the ligands and amino acid side chains between the two complexes are subtle and difficult to describe, the changes in residue forces are clear.

There are some interesting details in these results which suggest how this type of analysis can be used can be used for the purpose of drug design. For example, Met388 shows favorable forces in both complexes (arrows point towards the ligand in both), and the force vectors are larger (more attractive) for the 4-MeE2 than those of the estradiol complex. Close examination of the geometries reveals the reason: the shortest distance between the two elements (ligand and Met388) are of the two hydrogens which have a distance of 3.53 Å for the estradiol complex and 2.63 Å for the 4-MeE2 complex. Using the identical fragments in a model experiment (data not shown), the optimum H–H distance is calculated to be 2.72 Å. Thus, despite many unfavorable steric clashes caused by the 4-methyl group, at least one interaction is improved. This level of detail can be helpful in explanation of results of structure based design studies and point mutation experiments.

The protein geometries of the two complexes in the previous example were taken from the same source (1GWR. PDB). When using coordinates from different PDB entries, variations of the geometries are larger, and can affect force-field potentials and forces. These variations include small errors in atomic positions resulting from use of different procedures on large structures including specific conditions of the crystallization procedure, including choice of counter-ions, or crystal packing effects can affect the protein geometry. Also, uncertainties in solving the electron density maps introduce noise into atomic coordinates. To illustrate the magnitude of these variations, an example is given in Table 2. Here we compare the unmodified crystal geometries of 8 instances of the ligand binding domain of human receptor alpha (hERα-LBD) from two PDB entries; 1ERE [22], 1GWR. RMSD values show that the geometries are far from 'identical'. Comparison of the two monomer chains from the more recent crystal coordinates (1GWR.A,B) which also has higher resolution (2.4 Å vs. 3.1 Å for 1ERE) shows that the differences are not negligible within the same PDB entry (RMSD = 0.6) and are even larger when comparing different entries (RMSD = 0.74). While these differences are within commonly 'acceptable' limits for the methods, for the purpose of energy calculations they are significant. Since the RMSD values are averaged over the number of atoms, for large complexes, the geometrical differences add up and result in considerable energy differences. For the two ERα-LBD monomers from 1GWR with more than 1800 heavy atoms, the difference is about 250 kcal/mol (Table 3). Obviously, since the geometrical variances are randomly distributed among all atoms, when focusing on a small part of the structure, the magnitude of the differences is smaller. This can be seen in a comparison of the force-field energies of the active site residues of the two ERα-LBD chains (Table 3); which are only different by 6 kcal/mol (151.34 kcal/mol for 1GWR.A vs. 157.34 kcal/mol for 1GWR.B). Thus, to study the ligand-receptor interactions it is helpful to break down the potential energy or other energy related descriptors into smaller local or even per-residue terms.

To examine the effects of these variations on the results of force analysis, complexes of estradiol with the ligand binding domain of hERα from two different PDB entries (1ERE, 1GWR) were used. 1ERE contains three dimers and 1GWR contains one dimer. After the preparation and energy minimization of each dimer, the residue forces of ligand—active site for all monomers where calculated and the 2D-IFFPs were compared (Chart 3). The average standard deviations of residue forces are 18.8% for this graph which allows us to use the technique to compare forces on related residues from different crystal structures. When comparing different chains from the same PDB entry, the variances are slightly smaller (15%, data not shown).

Subtype selectivity of some estrogenic ligands was selected as a test case for application of the method for a comparative analysis of ligand—receptor binding. The

Table 2 Comparison of unprepared crystal geometries of ligand binding domains of hERα from PDB entries 1ERE and 1GWR

	1	2	3	4	5	6	7
1:1ERE.A							
2:1ERE.B	0.13						
3:1ERE.C	0.13	0.07					
4:1ERE.D	0.08	0.13	0.13				
5:1ERE.E	0.07	0.13	0.13	0.07			
6:1ERE.F	0.13	0.07	0.06	0.13	0.13		
7:1GWR.A	0.73	0.72	0.72	0.73	0.73	0.72	
8:1GWR.B	0.74	0.74	0.74	0.74	0.73	0.74	0.60

Pair-wise superposition RMSD values are in Angstrom

Table 3 Comparison of energies of energy minimized geometries of two ligand binding domains of hERα from the same PDB entry (1GWR)

	1GWR.A		1GWR.B	
	E	N	E	N
Total	−5378.49	1868	−5609.68	1874
Active site	151.34	178	157.34	178

E is the force field energy in kcal/mol; N is number of heavy atoms. Active site includes amino acids having at least one atom within 4.5 Å from ligand

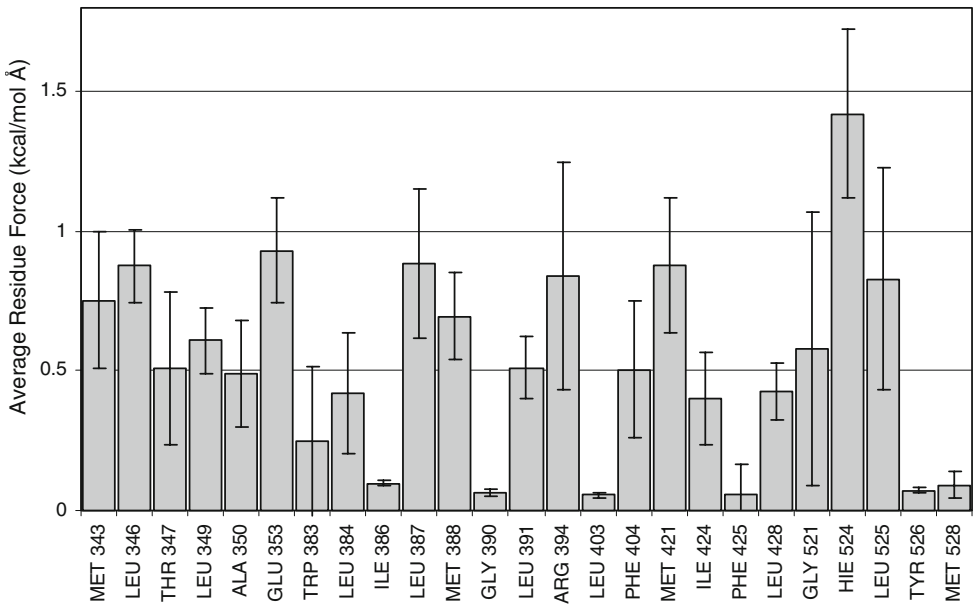


Chart 3 Comparison of IFFPs of 8 monomers of estradiol in hERα from two different crystal structures. Error bars are ±SD values

Table 4 ERβ selective compounds used for study

Ligand	ERα			ERβ			Ref.
	IC50	PDB	E_{int}	IC50	PDB	E_{int}	
Genistein	360	1X7R	−29.93	9.0	1X7J	−30.66	3
WAY-244	1062	1X7E	−30.60	75.8	1X78	−31.68	12
BP-2b	11	2QE4	−32.66	0.28	2JJ3	−35.16	13

IC50 values are in nM. E_{int} is the force-field interaction energy of ligands for the prepared complexes in kcal/mol

basic idea is to compare the patterns of changes in force vectors for complexes of ligands in the estrogen receptors α and β, to help understand the selectivity. Table 4 shows selectivity data for the three ligands; genistein, WAY-244, and a benzopyrene ligand BP-2b [8, 23, 24]. The table gives the PDB entry codes and the IC50 values, which determine the selectivity. It can be seen that all three ligands are selective for the receptor ERβ (lower IC50 values than for ERα). Additionally, the force-field interaction energies of the ligand-receptors are also given. Understandably, without a regression and inclusion of solvation terms and internal (conformational) energies, the interaction energies do not follow the trend of binding affinities. The force vector analysis was applied to help understand the selectivity data.

The structures were prepared and energy minimized as previously explained. To make direct comparison of force vectors possible, the sequences of receptor chains were then aligned using MOE-Align with default parameters[4] and the coordinates of complexes were also superposed. Next the residue forces for each monomer complex were calculated (Table 1 in supplemental data). To avoid the misleading nature of using the vector lengths, the 3D-components of force vectors (x, y, z) were used. This generates a large amount of data since for each receptor subtype there are three ligands and three components per residue. To simplify the analysis, we averaged the components of force vectors per receptor subtype, which created *average 3D-IFFPs* shown in Table 5. Next, the average forces on β subtype were subtracted from those of the α subtype, shown as x, y, z components of the $\alpha - \beta$ *vector* in Table 5. These results reflect the significant changes on each component of the force vectors between two receptor subtypes, which are common between all three ligands. Data in this table is sorted by length of the $\alpha - \beta$ *vector* which can be interpreted as the magnitude of the differences.

Table 5 shows that for residues 1–5 and 7–9 the amino acids are equivalent for both receptors, however substitutions appear at residues 6 and 10 which are therefore subtype-specific and expected to play an important role in determining ligand selectivity. But the maximum difference in the $\alpha - \beta$ vector occurs for the pair Leu(α:387, β:339), where the magnitude of the difference is 1.05 kcal/mol Å. Interestingly, the subtype specific pairs (shown in bold fonts) are not the most important residues. According to the $\alpha - \beta$ criterion, the most significant residues between the two receptor subtypes belong to an unexpected set of amino acid pairs, namely Leu(α:387, β:339), Glu(α:353, β:305), Ala(α:350, β:302), Leu(α:346, β:298).

To visualize the position of these amino acids, Fig. 6 shows alpha carbons of the top 5 residues in Table 5 from crystal structures 2QE4 and 2II3. ERα atoms are shown in

[4] Substitution Matrix: blosum 62 [25], penalties for gap start and gap length 7 and 1 respectively, iteration limit: 100 and failure limit: 10, Tree based method was used for initial estimate build up.

Table 5 Average 3D-IFFPs of complexes of ligands in ERα and ERβ, along with components of their differences (α − β)

ERα				ERβ				α − β			
Residue	x	y	z	Residue	x	y	z	x	y	z	Length
LEU 387	−0.35	−0.36	0.32	LEU 339	−1.15	−0.83	0.81	0.80	0.47	−0.49	1.05
GLU 353	0.23	−0.50	−0.87	GLU 305	0.22	0.23	−1.25	0.01	−0.73	0.38	0.83
ALA 350	0.00	0.13	1.18	ALA 302	0.04	0.20	0.39	−0.04	−0.07	0.79	0.80
LEU 346	0.00	0.19	0.39	LEU 298	−0.53	−0.20	0.30	0.53	0.39	0.08	0.66
PHE 404	0.04	−0.06	0.00	PHE 356	0.27	0.17	−0.43	−0.23	−0.23	0.42	0.54
MET 421	−0.26	0.07	0.49	**ILE 373**	0.08	0.12	0.18	−0.34	−0.05	0.31	0.46
HIS 524	0.22	−1.14	0.27	HIS 475	0.28	−1.49	0.54	−0.06	0.34	−0.26	0.44
MET 343	0.19	−0.01	0.14	MET 295	−0.06	−0.25	0.02	0.25	0.24	0.12	0.37
PHE 425	0.30	0.26	−0.21	PHE 377	0.09	0.02	−0.09	0.21	0.24	−0.12	0.34
LEU 384	0.61	0.21	−0.48	**MET 336**	0.39	0.19	−0.60	0.22	0.02	0.12	0.25

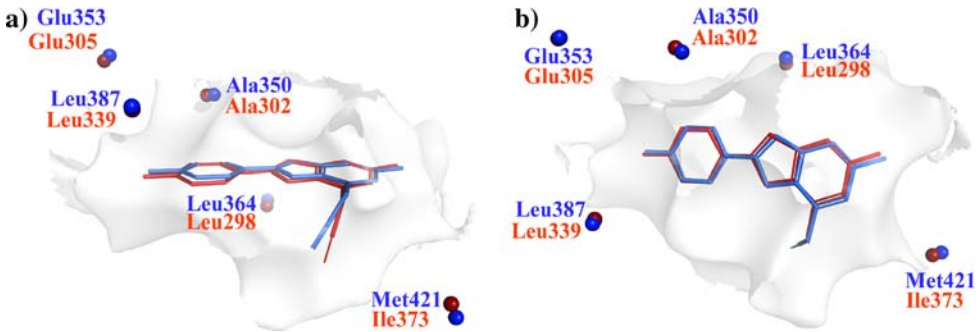


Fig. 6 Residues with most significant changes of forces between ERα and ERβ The co-crystallized ligand (WAY-244) is shown using the stick model

blue and ERβ atoms are in displayed in red. The receptor cavity is rendered in transparent gray and the co-crystallized ligands are shown using a stick model. It can be seen from the picture that the nitril group on the ligand WAY-244 was specifically designed to cause specific steric clashes with ERα:Met421 but not with its substitute in ERβ (Ile373). But the data suggests that the ligands readjust their location and cause some steric clashes with the amino acids on the opposite side of the cavity. For this reason, the amino acids carrying the biggest differences in forces are all on the opposite side of the cavity, related to the subtype specific pair α:Met421, β:Ile373. Thus, a functional description of subtype selectivity can be given as: "ER β selective ligands cause increased repulsions in ERα versus ERβ on the subtype specific pair (α:Met421, β:Ile373) and on the residues of the opposite side, including Leu(α:387, β:339), Glu(α:353, β:305), Ala(α:350, β:302), Leu(α:346, β:298)." The novelty of this definition is that it is not geometry-based, and that it emphasizes the crucial role of identical amino acids in interaction with selective ligands.

Conclusions

A method is described which partitions the native force-field potentials and derives the residue-specific interaction forces of the ligand-receptor complex. These forces, which can be displayed as 3D vectors based on atomic coordinates, can also be summed per residue, producing residue forces, which can be located on the alpha-carbons of amino acids. Residue forces allow one to visualize the overall interaction of each particular amino acid with the ligand, and show features characteristic of hydrogen bonding, steric clashes, and other aspects of molecular interactions. The lengths of these vectors can be used to create IFFPs, which represent the ligand-receptor interactions only by their magnitudes. We showed how IFFPs can be applied to the important case of ligand selectivity for different receptor isoforms. Thus, the force vector analysis and IFFPs provide a new tool to the computational chemist, which can be applied to the process of pharmaceutical drug design. Future work will include more quantitative comparisons of this method with potential energy based methods, along with its application in analysis of docking results.

References

1. The Research Collaboratory for Structural Bioinformatics www.rcsb.org
2. Deng Z, Chuaqui C, Singh J (2004) J Med Chem 47:337–344. doi:10.1021/jm030331x

3. Hendlich M, Bergner A, Günter J, Klebe G (2003) J Mol Biol 326:607–620. doi:10.1016/S0022-2836(02)01408-0
4. Powers R, Copeland JC, Germer K, Mercier KA, Ramanathan V (2006) Funct Bioinformatics 65:124–135. doi:10.1002/prot.21092
5. Molecular Operating Environment (MOE) (2007) Chemical Computing Group, Montréal, QC
6. Sobolev V, Sorkine A, Prilusky J, Abola EE, Edelman M (1999) Bioinformatics 15:327–332. doi:10.1093/bioinformatics/15.4.327
7. Manas ES, Xu ZB, Unwalla RJ, Somers WS (2004) Structure 12:2197–2207. doi:10.1016/j.str.2004.09.015
8. Wallace AC, Laskowski RA, Thornton JM (1995) Protein Eng 18:127–134. doi:10.1093/protein/8.2.127
9. Ortiz AR, Pisabarro MT, Gago F, Wade RC (1995) J Med Chem 38:2681–2691. doi:10.1021/jm00014a020
10. Wade R C, Ortiz A R, Gago F (1998) Prespect Drug Discov Design 9–11:19–34
11. Arkawa M, Hasegawa K, Funatsu K (2008) Chemom Intell Lab Syst 92:145–151. doi:10.1016/j.chemolab.2008.02.004
12. Lozano JJ, Pastor M, Cruciani G, Gaedt K, Centeno NB, Gago F, Sanz F (2000) J Comput Aided Mol Des 14(4):341–353. doi:10.1023/A:1008164621650
13. Tomić S, Nilsson L, Wade RC (2000) J Med Chem 43:1780–1792. doi:10.1021/jm9911175
14. Tomić S, Kojić-Prodić B (2002) J Mol Graph Model 21:241–252. doi:10.1016/S1093-3263(02)00148-1
15. Rodríguez-Barrios F, Gago F (2004) J Am Chem Soc 126:2718–2719. doi:10.1021/ja038893t
16. Halgren TA (1999) J Comput Chem 20:720 10.1002/(SICI)1096-987X(199905)20:7<720::AID-JCC7>3.0.CO;2-X
17. Halgren TA (1999) J Comput Chem 20:730 10.1002/(SICI)1096-987X(199905)20:7<730::AID-JCC8>3.0.CO;2-T
18. Bashford D, Case D (2000) Annu Rev Phys Chem 51:129. doi:10.1146/annurev.physchem.51.1.129
19. Warnmark A, Treuter E, Gustafsson JA, Hubbard RE, Brzozowski AM, Pike AC (2002) J Biol Chem 277:21862–21868. doi:10.1074/jbc.M200764200
20. PyMOL 2007, DeLano Scientific LLC, Palo Alto, CA
21. Zhu BT, Han GZ, Shim JY, Wen Y, Jiang XR (2006) Endocrinology 147:4132–4150. doi:10.1210/en.2006-0113
22. Brzozowski AM, Pike AC, Dauter Z, Hubbard RE, Bonn T, Engstrom O, Ohman L, Greene GL, Gustafsson JA, Carlquist M (1997) Nature 389:753–758. doi:10.1038/39645
23. Manas ES, Unwalla RJ, Xu ZB, Malamas MS, Miller CP, Harris HA, Hsiao C, Akopian T, Hum WT, Malakian K, Wolfrom S, Bapat A, Bhat RA, Stahl ML, Somers WS, Alvarez JC (2004) J Am Chem Soc 126:15106–15119. doi:10.1021/ja047633o
24. Norman BH, Richardson TI, Dodge JA, Pfeifer LA, Durst GL, Wang Y, Durbin JD, Krishnan V, Dinn SR, Liu S, Reilly JE, Ryter KT (2007) Bioorg Med Chem Lett 17:5082–5085. doi:10.1016/j.bmcl.2007.07.009
25. Henikoff S (1992) Proc Natl Acad Sci USA 89:10915–10919. doi:10.1073/pnas.89.22.10915