**Chemical properties that affect binding of**

**enzyme-inhibiting drugs to enzymes**

Introduction

The production of new drugs requires time for development and testing, and can result in large prohibitive costs if done in vitro. Advances in computer technology allow for the computational testing and development of drugs before any wet-lab experiments are conducted, thus saving time and money. Programs for simulating how proteins bind to each other have been developed, and can be categorized as predicting either the 3D shape complementarity of proteins (1), or predicting whether proteins likely to bind based on chemical properties (2). These simulations prove to be more efficient in the drug-development-pipeline, but generally require some initial user input in order to begin working (such as maximum distance between atoms) (3), thus making them inaccurate for predicting all possible protein types. Accuracy can be improved by developing reference databases from which developers can compare successful binding techniques, or by experimentally determining a set of models that allow simulation programs to predict protein-protein (or protein-ligand) affinity based on the specific type of molecule being assessed (4).

Molecular complex prediction models are developed by assessing the chemical properties of the molecules’ atomic compositions. Differing atomic compositions mean some binding patterns rely more on hydrophobicity, while others depend more on acid/base electrostatics (5). These chemical properties include solvent accessible surface area, hydrophobicity, electrostatics, van der waals forces, residue pair potential, desolvation energies, atomic contact energies, complementary determining regions, etc... (4) (6). These properties can be tested mathematically by evaluating atom type, atomic distances, and neighboring atoms, thus they can be placed in an equation that produces an overall positive or negative score (called an affinity score), indicating whether the two atoms will have a favorable or unfavorable interaction.

*bij = ai aj Si Sj Tij Rij + rij*

**Equation 1. The HINT Equation**

* *i & j* : the atoms being compared
* *b* : the affinity score for i & j
* *a* : atomic contact energy
* *S* : solvent accessible surface area
* *T* : electrostatics (acid/base chem.)
* *R* : atomic distance
* *r* : Lennard-Jones potential

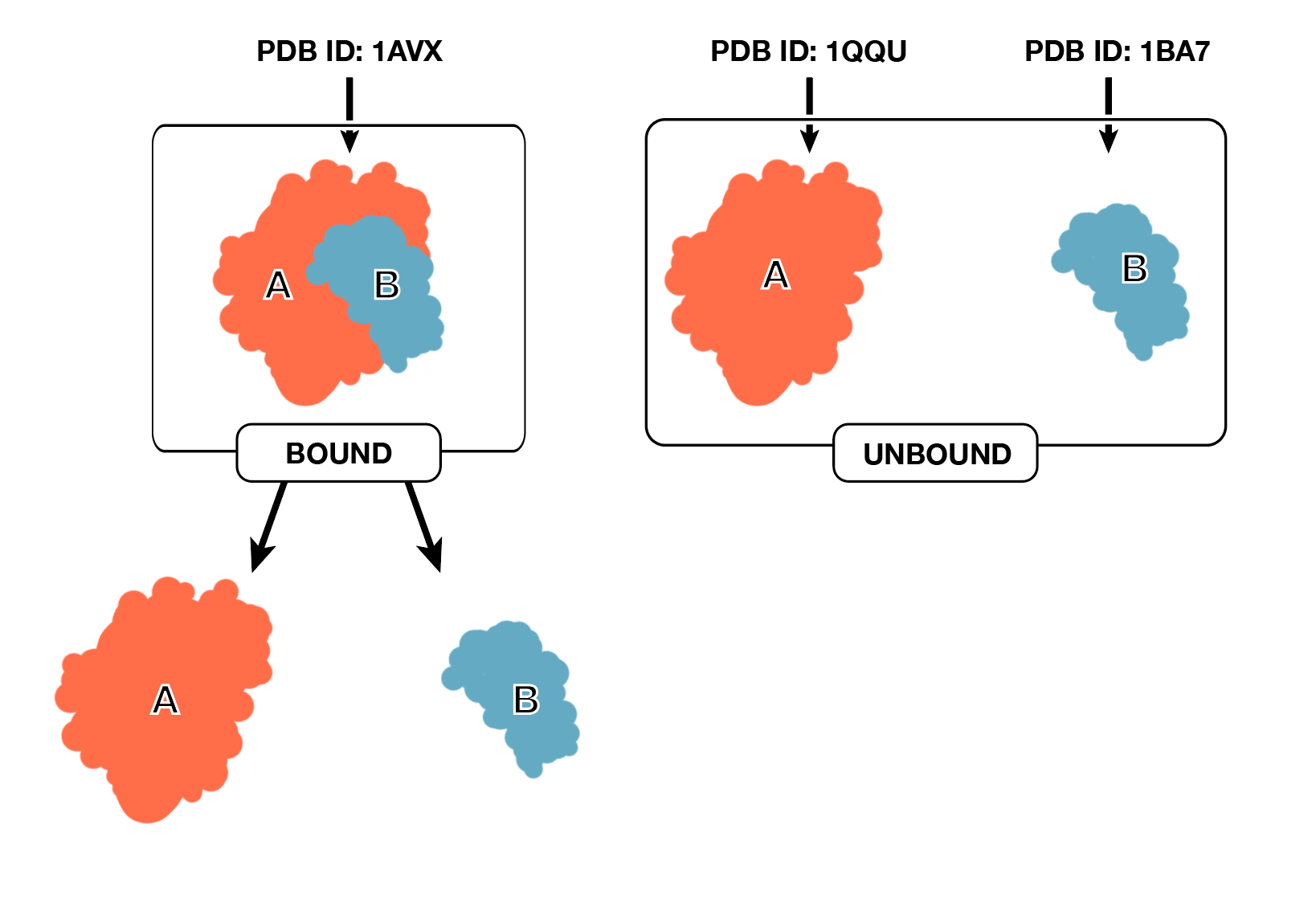
The model of interest in this experiment will be the Hydropathic INTeractions (HINT) model developed at Virginia Commonwealth University (6) (7). The HINT equation, along with the chemical properties it evaluates are shown in equation 1. The equation produces scores on an atom by atom basis, and sums them up at the end, thus producing a score for the molecules as a whole.

Li et al (2007) found that “weighing” different chemical properties allows for the simulation model to produce a more type-specific affinity score, based on the properties of the protein complex being evaluated. A summary of their results shown in table 1 shows that once processed through their weighted equation, the success of the binding simulation improved (4).

**Table 1**

*Post-weighted equation results from Li et al (2007)*

|  |  |
| --- | --- |
| **Name** | **Success Ratio** |
| Protease/Inhibitor | 16/17 |
| Enzyme/Inhibitor | 6/6 |
| Antibody/Antigen | 18/19 |
| Other | 11/15 |

With these results in mind, this experiment seeks to find whether weighing the HINT algorithm through exponentiation of the variables *a, S, T, R, and r* will allow for the discovery of which chemical properties play the greatest role in the binding of enzyme/inhibitor complexes.

**#1**

**#2**

**#3**

True-complex

Methods

Enzyme/inhibitor complexes will be taken from the Benchmark 5 (8), a list of PDB files commonly used to test molecular docking software, curated at the Massachusetts Institute of Technology. PDB files are lists of every primary, secondary, tertiary, and quaternary structure in a protein or protein-complex, presented as a list of atomic coordinates. Figure 1 shows the process for how the PDB files will be used by the software and explains the significance of the bound/unbound terminology. As of now, there are 46 enzyme/inhibitor complexes on the Benchmark 5, thus 46 will be used for the experiment (shown in table 5 at the end of this document).

**Figure 1. Use of Benchmark 5 PDB files**

The first PDB file is the true experimentally discovered complex. The second and third are the constituents of the true complex, discovered through separate experiments. The true-complex is computationally split into the constituents, thus in the end there are two copies of the complex constituents ready for processing. The “bound” ones contain important conformational changes, and the unbound ones don’t.

Conformational changes are small alterations that happen in atomic structure when two molecules bind. This is due to repulsive and attractive forces in the atomic structure.

Those 46 bound and unbound complexes will be processed through an initial 3D shape-complementarity software called FTDock (9), to produce a large list of possible docked complexes. That large list will be culled through a ligand root mean square deviation (L\_RMSD) comparison to the true-complex. The L\_RMSD comparison is an overlapping of the simulated complex and the true-complex, wherein atomic coordinates are compared in angstroms. Thus, the higher an L\_RMSD value is, the more distant the atomic compositions of the molecules are, and the lower it is, the more similar (and closer to the true-complex) they are. After doing this to the large list of complexes, only the top 20 with the lowest L\_RMSD score will be selected for each of the 46 complexes (producing a list of 920 complexes)

Next, all 920 of the complexes will be processed through the modified hint algorithm. Each of the 5 variables *a, S, T, R, and r* will be exponentiated at either 0, 0.5, 1, 1.5, or 2, thus 25 tests will be conducted for each of the 920 complexes. Exponentiation by 1 will serve as the control.

The output of the HINT algorithm will produce 23,000 HINT affinity scores. The highest 200 of those scores and their corresponding simulated complexes will be processed through a second L\_RMSD comparison to the true-complex, and the lowest scores from that test will reveal which weights produced the most favorable simulated complex. Thus, a conclusion could be made about which chemical properties, and what proportion of those properties are important to the simulated binding of enzyme/inhibitor complexes.

Possible Results

**Table 2**

*Possible Results Indicating Importance of Surface Accessible Surface Area amplified by 1.5*

|  |  |  |  |
| --- | --- | --- | --- |
| **Complex** | **Final L\_RMSD Score** | **Weighing Used** | **Significant Chemical Property** |
| **#1 Bound** | 4 Å | *ai aj (Si Sj)1.5 Tij Rij + rij* | **Solvent Accessible Surface Area** |
| **#1 Unbound** | 6 Å | *ai aj Si Sj (Tij)2 Rij + rij* | Electrostatics |
| **#2 Bound** | 2 Å | *ai aj Si Sj Tij (Rij)0.5 + rij* | Atomic Distance |
| **#2 Unbound** | 4 Å | *(ai aj)1.5 Si Sj Tij Rij + rij* | Atomic Contact Energy |
| **#3 Bound** | 3 Å | *ai aj (Si Sj)1.5 Tij Rij + rij* | **Solvent Accessible Surface Area** |
| **#3 Unbound** | 5 Å | *ai aj Si Sj (Tij)0 Rij + rij* | Electrostatics |
| **…** | … | … | … |
| **#46 Bound** | 2 Å | *ai aj (Si Sj)1.5 Tij Rij + rij* | **Solvent Accessible Surface Area** |
| **#46 Unbound** | 6 Å | *(ai aj)0.5 Si Sj Tij Rij + rij* | Atomic Contact Energy |

Table 2 indicates the importance of solvent accessible surface area (SASA) since it appears three times in the sample results, and all three times it shows a 1.5 exponent on the SASA variable.

**Table 3**

*Alternative Results Showing Importance of Electrostatics*

|  |  |  |  |
| --- | --- | --- | --- |
| **Complex** | **Final L\_RMSD Score** | **Weighing Used** | **Significant Chemical Property** |
| **#1 Bound** | 4 Å | *ai aj (Si Sj)1.5 Tij Rij + rij* | Solvent Accessible Surface Area |
| **#1 Unbound** | 6 Å | *ai aj Si Sj (Tij)2 Rij + rij* | **Electrostatics** |
| **#2 Bound** | 2 Å | *ai aj Si Sj Tij (Rij)0.5 + rij* | Atomic Distance |
| **#2 Unbound** | 4 Å | *ai aj Si Sj (Tij)1.5 Rij + rij* | **Electrostatics** |
| **#3 Bound** | 3 Å | *ai aj (Si Sj)1.5 Tij Rij + rij* | Solvent Accessible Surface Area |
| **#3 Unbound** | 5 Å | *ai aj Si Sj (Tij)1.5 Rij + rij* | **Electrostatics** |
| **…** | … | … | … |
| **#46 Bound** | 2 Å | *ai aj Si Sj (Tij)2 Rij + rij* | **Electrostatics** |
| **#46 Unbound** | 6 Å | *(ai aj)0.5 Si Sj Tij Rij + rij* | Atomic Contact Energy |

The following shows the importance of Electrostatics since there are 4 instances in which is produced a favorable L\_RMSD score. In this case the exponents were 2, 0.5, 1.5, 1.5, and 2. There is no clear number that is most favorable, but one could conclude that an increase in the weight of electrostatics is important to enzyme/inhibitor binding.

**Table 4**

*Results In Which No Clear Conclusion Can Be Reached*

|  |  |  |  |
| --- | --- | --- | --- |
| **Complex** | **Final L\_RMSD Score** | **Weighing Used** | **Significant Chemical Property** |
| **#1 Bound** | 4 Å | *ai aj (Si Sj)1 Tij Rij + rij* | Solvent Accessible Surface Area |
| **#1 Unbound** | 6 Å | *ai aj Si Sj (Tij)2 Rij + rij* | Electrostatics |
| **#2 Bound** | 2 Å | *ai aj Si Sj Tij (Rij)0.5 + rij* | Atomic Distance |
| **#2 Unbound** | 4 Å | *ai aj Si Sj (Tij)1.5 Rij + rij* | Electrostatics |
| **#3 Bound** | 3 Å | *ai aj (Si Sj)1.5 Tij Rij + rij* | Solvent Accessible Surface Area |
| **#3 Unbound** | 5 Å | *ai aj Si Sj (Tij)0.5 Rij + rij* | Electrostatics |
| **…** | … | … | … |
| **#46 Bound** | 2 Å | *ai aj (Si Sj)0 Tij Rij + rij* | Solvent Accessible Surface Area |
| **#46 Unbound** | 6 Å | *(ai aj)0.5 Si Sj Tij Rij + rij* | Atomic Contact Energy |

Table 4 prevents any conclusion from being made. There is an even split between SASA and electrostatics, so one may say those are both important in the binding of enzyme/inhibitor complexes, though the weighing used does not show a definite preference for increasing or decreasing the variable (split between exponent that are <1 and >1), thus no definite conclusion can be made. It seems unlikely that no conclusion will be reached given a large enough sample size, though it is a possibility.

If results indicate the most favorable weight is 1 (the control), this would mean that the HINT model works best unchanged for that complex.

Discussion

Use of the weighted HINT equation may provide insight into which chemical properties are most significant in the binding of enzyme/inhibitor complexes. Based on the results from Li et al (2007) (4), the use of weighted variables plays a role in finding which chemical properties are most important to the binding of a specific molecular complex. Thus, further work, perhaps using different simulation models and types of complexes may allow for further specialization of docking software, and allow for more efficient and accurate experimentation for drug development.

**Table 5**

*Benchmark 5 PDB Files To Be Used*

References

1. Chen, R., Li, L. & Weng, Z., 2003. ZDOCK : An Initial-Stage Protein-Docking Algorithm. Proteins: Structure, Function and Genetics, 87(November 2002), pp.80–87.
2. Dominguez, C., Boelens, R. & Bonvin, A.M.J.J., 2003. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. Journal of the American Chemical Society, 125(7), pp.1731–1737.
3. Jiang, F. & Kim, S.H., 1991. “Soft docking”: Matching of molecular surface cubes. Journal of Molecular Biology, 219(1), pp.79–102.
4. Li, C.H. et al., 2007. Complex-type-dependent scoring functions in protein-protein docking. Biophysical Chemistry, 129(1), pp.1–10.
5. Jackson, R.M., 1999. Comparison of protein-protein interactions in serine protease-inhibitor and antibody-antigen complexes: implications for the protein docking problem. Protein science : a publication of the Protein Society, 8, pp.603–613.
6. Eugene Kellogg, G. & Abraham, D.J., 2000. Hydrophobicity: Is LogP(o/w) more than the sum of its parts? European Journal of Medicinal Chemistry, 35(7-8), pp.651–661.
7. Kellogg, G.E., Burnett, J.C. & Abraham, D.J., 2001. Very empirical treatment of solvation and entropy: A force field derived from Log Po/w. Journal of Computer-Aided Molecular Design, 15, pp.381–393.
8. Vreven, T. et al., 2015. Updates to the integrated protein-protein interaction benchmarks: Docking benchmark version 5 and affinity benchmark version 2. J. Mol. Biol., 427(19), pp.3031–3041.
9. Gabb, H.A., Jackson, R.M. & Sternberg, M.J.E., 1997. Modeling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information. J. Mole. Biol., 272, pp.106–120.