DiffBond: A Method for Predicting Intermolecular Bond Formation

Justin Tam[1], Talulla Palumbo[2], Julie M. Miwa[2] and Brian Y. Chen[1†]
Dept. Computer Science and Engineering[1], Dept. Biological Sciences[1],
Lehigh University, Bethlehem, PA 18015, USA

***Abstract*—Many tools that explore models of protein complexes are also able to analyze interactions between specific residues and atoms. A comprehensive exploration of these interactions can often uncover aspects of protein-protein recognition that are not obvious using other protein analysis techniques. This paper describes DiffBond, a novel method for searching for intermolecular interactions between protein complexes while differentiating between three different types of interaction: hydrogen bonds, ionic bonds, and salt bridges. DiffBond incorporates textbook definitions of these three interactions while contending with uncertainties that are inherent in computational models of interacting proteins. We used it to examine the barnase-barstar, Rap1a-raf, and Smad2-Smad4 complexes, as well as a subset of protein complexes formed between three-finger toxins and nAChRs. Based on electrostatic interactions established by previous experimental studies, DiffBond was able to identify ionic and hydrogen bonds with high precision and recall, and identify salt bridges with high precision. In combination with other electrostatic analysis methods, DiffBond can be a useful tool in helping predict influential amino acids in protein-protein interactions and characterizing the type of interaction.**

I. Introduction

Deducing the role of chemical bonds is a crucial part of understanding how protein-protein complexes achieve selective binding. In structural biology, this effort occurs frequently after the structure of a protein-protein complex is determined. First, bonds are identified with the application of the appropriate chemical and geometric constraints. Next, hypotheses are developed about what role certain bonds play in stabilizing particular parts of the complex. Finally, mutational experiments that remove or alter specific bonds can begin to test these hypotheses by establishing the resulting change in binding affinity. Once the effect of those mutations are evaluated, new mutational experiments can be devised, until the role of bonding in the apparatus of recognition is explained.

Unfortunately, a high resolution structure of most protein complexes is unavailable, so the closest alternative is to develop hypotheses from computational models of the interacting proteins. This approach must contend with additional uncertainties: First, the presence of intermolecular bonds will be constrained by the limits of bond geometry, which have been carefully measured in the chemical literature [1]–[3]. Second, the parts of a bond that form in vivo might be separated distantly enough in the model that their potential for assembly might not be discovered. Finally, intermolecular bonds may exhibit trends in length and angle that are atypical of the general bonds surveyed in the literature. There are few tools that perform an in-depth search of intermolecular bonds while attributing a biochemical reason to the interaction. To perform such a search while contending with existing constraints, this paper aims to assess the predictability of intermolecular bonds based on standard descriptions in the chemical literature.

Our approach is to treat the textbook chemical measurements of salt bridges, ionic interactions, and hydrogen bonds as a predictor for the presence of intermolecular bonds, and verify these predictors against experimentally established findings. In the case of hydrogen bonds, standard bond angles and bond lengths are extremely well defined [1], but in the case of ionic bonds, it is far less so. Coulomb's law defines attraction or repulsion between charged atoms at any distance, but the degree of attraction or repulsion is modulated by the presence and geometry of the dielectric between them. High dielectric aqueous environments attenuate the electric field, whereas low dielectric environments within a protein enhance it [4]. Thus, the concept of ionic bond lengths must always be an approximation based on assumptions of a biological environment, as several groups have done [2], [5]. Salt bridges, being the co-occurrence of both a hydrogen bond and an ionic bond, must also exist in the presence of these assumptions [5], [6]. Following these conventional definitions of bond geometry, we created DiffBond, a basic classifier for predicting the presence of salt bridges, ionic bonds, and hydrogen bonds.

Many methods deduce influential amino acids from computational models and predict the effects of mutation. Some provide information about the stability of a mutation by analyzing heuristic energy changes, rigidity-based mutation analysis, or molecular dynamic simulations [7]–[9]. Other methods compute and predict interactions from the computational model and infer mutation stability based on these interactions, such as hydrogen bond location prediction [1]. Like hydrogen bond location prediction, this method predicts existence and location of interactions between complexes. However, DiffBond is the first to predict the formation and location of intermolecular salt bridges and ionic bonds, and analyze these results in conjunction with hydrogen bond predictions. This bottom-up approach not only provides enough specificity to classify each interaction, but also is flexible enough to merge with other methods to improve prediction of influential amino acids.

We evaluated the classifier on a small dataset of protein complexes with well documented chemical bonds, measuring how frequently the classifier agreed with the authors' findings

†Correspondence: chen@cse.lehigh.edu

on each complex. While such a dataset cannot be representative of all intermolecular bonds within the space of protein complexes, it can determine whether intermolecular bonds infringe conventional norms.

II. Methods

DiffBond identifies a list of all bonds that have the geometric and electrostatic capacity to make intermolecular bonds and affect binding affinity between two proteins that form a protein complex. First, we identify pairs of residues that are able to form three types of intermolecular interactions: salt bridge, ionic bond, and hydrogen bond. Identified residue pairs must satisfy geometric and electrostatic criteria for forming any of these three interactions based on textbook and literature measurements. We then form lists of bond predictions for each type of interaction as output. Finally, the resulting lists can be interpreted to help identify significant residues or aid in experimental design of mutations. This paper explores our software which encompasses two methods for identifying protein mutations that affect binding affinity, and discusses prospects for applying them in conjunction to mutation testing.

A. Scanning for Bond Formation

We outline a method to scan for intermolecular bonds between two proteins that form a protein complex, especially at the interface. We first generate a fully connected graph between all atoms from one protein to all atoms from the other protein; to identify specific and significant connections within the graph, we filter the connected graph using biochemical criteria like distance, residue charge, and contacting atoms.

A list of possible hydrogen bonds were compiled using HBPlus [1], which takes two hydrogen coordinates and looks at several criteria to decide if a hydrogen bond is possible and likely; criteria include minimum bond angles between atoms at 90°, maximum distance depending on the type of bond with 3.9Å for donor-acceptor pairs and 2.5Å for hydrogen-acceptor pairs, and minimum covalent separation of 3 covalent bonds. In addition, we also searched for amino-aromatic hydrogen bonds. These criteria values have been used in previous hydrogen bond interaction studies [1], [10], [11].

A list of ionic bonds were compiled by searching within a distance constraint for oppositely charged amino acids, namely interactions between arginine, histidine, and lysine with aspartate or glutamate [12]. In this study, we define an ionic bond as residues whose charged atoms, namely a positive N (nitrogen) in basic residues or negative O (oxygen) in acidic residues, are less than a cutoff distance [3]. Our software allows variable distance as an input parameter, but for the purposes of this paper, we use 5 angstroms. The 5Å cutoff is strict enough to yield only amino acids that are biochemically likely to form a bond, while the N-O atom pairs make sure the residue side chains are oriented towards each other. Ionic bonds, rarely, can form over long distances between 5-10Å in length [3] and so we also provide ionic bond predictions at 7.5Å and 10Å in Supplemental Text S1.

We also compile a list of salt bridges in the same way that we compute ionic bonds. Barlow and Thornton define salt bridges as a pair of oppositely charged residues whose side chain N-O atoms are within a cutoff distance of 4Å [2]. Fig 1 shows an example of oppositely charged side chains, glutamate and arginine of a Barnase-Barstar protein complex, within 4Å distance of each other. A cutoff at 4Å is a well defined distance that only considers "good" salt bridge geometries [13]. This 4Å measurement of salt bridge length aligns well with the textbook definition where salt bridges are a co-occurrence of both a hydrogen bond and an ionic bond [5], [6]. At less than 4Å, oppositely charged atoms are likely to interact; this also creates an environmental condition where water molecules cannot fit between the interacting residues. This implies the formation of a hydrogen bond which agrees with textbook measurements [5], [14].

Our bond scanning method consists of searching for neighboring amino acids that satisfy specific electrostatic and distance criteria; this design allows us to not only search within dimers, but also scan among higher oligomers consisting of many subunits.

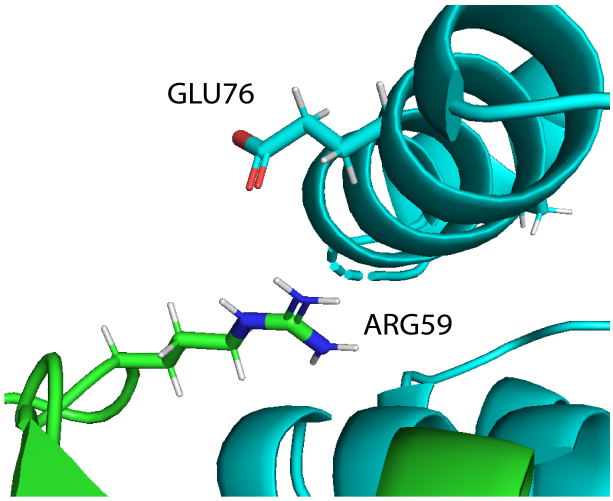

Fig. 1. Sidechain visualization of Arg59 on Barnase (green) and Glu76 on Barstar (teal). Arg59 and Glu76 are within 4Å and are oppositely charged amino acids, so they are predicted to form a salt bridge by DiffBond.

B. Computing Electrostatic Isopotential Surfaces

An electrostatic isopotential surrounding a protein, at some potential p, is a subset of the electrostatic isopotential field that has potential equal to p. The electrostatic isopotential is a surface that outlines an area within an electrostatic potential at some threshold electrostatic potential k (kT/e). This isopotential threshold creates a surface where one side of the surface has isopotentials less than k and the other side has isopotentials greater than k. When the surface does not have infinite dimensions, it can be said to describe a geometric solid with measurable volume.

To generate an electrostatic isopotential surface, we first solve the overall potential field of a protein using DelPhi,

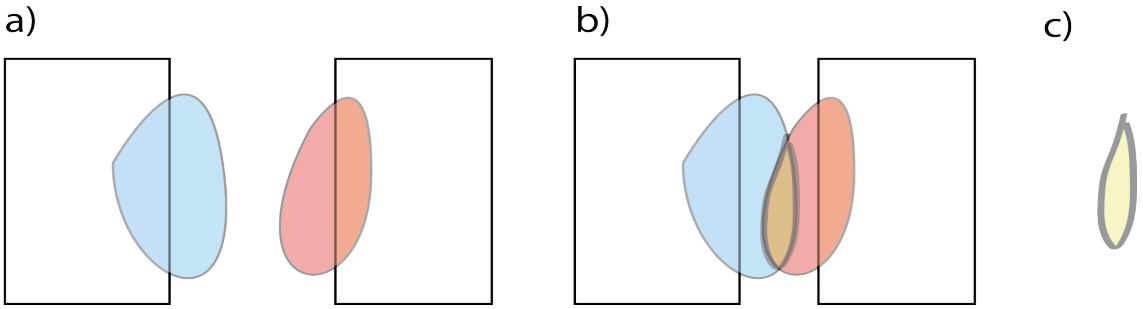

Fig. 2. **Intersection using CSG** a) Two proteins with oppositely charged electrostatic fields. b) When the proteins are in complex, the oppositely charged fields overlap forming an intersection region shown in orange. c) The intersection region represents the degree to which the field of one protein complements the field of the other.

an application that takes a 3-D coordinate molecule as input, computes solutions to the Poisson-Boltzmann equation for the input molecule, and outputs the potential field of the whole system [15], [16]. An algorithm called Marching Cubes then takes the potential field and some isopotential threshold k and generates an approximation of the isopotential surface representing the protein at k [17].

VASP-E is a tool that implements Marching Cubes and is able to manipulate, and calculate the volume of electrostatic isopotentials [18]. As implemented in VASP-E, marching cubes takes the potential field from DelPhi's output and first aligns it to a lattice grid. We then approximate the isopotential by determining which grids the isopotential intersects within the lattice. We use the intersections to further improve the isopotential approximation [17]. This technique produces a high-resolution approximation of the electrostatic potential surface of a protein [18].

1) Interface Field Comparison: VASP-E extends the method for manipulating electrostatic surfaces to perform a comparison of interface fields; VASP-E allows us to manipulate isopotential surfaces using constructive solid geometry (CSG) [19]. Based on VASP-E implementation, CSG can calculate the union, intersection, and difference of volumetric objects [18]. In this method, we use a series of CSG operations to perform a comparison of the region of electrostatic interaction between two proteins, called the interface field. We first generate isopotential surfaces for two proteins that form a protein complex. For one protein, the isopotential surface is generated at +k kT/e. On the other protein, a surface is generated at -k. The intersection between the +k and -k surfaces is the interface region where the positively charged region of one protein overlaps with the negatively charged region of the other protein as seen in Fig. 2. In other words, the intersection of +k and -k represent the degree to which the field of one protein complements the field of the other; the greater the volume of this intersection region, the more electrostatically complementary the proteins are.

2) Nullification: DelPhi is able to solve the potential field of a molecule while ignoring the charge contribution by an amino acid in a process called nullification [20]. Nullification of amino acids will affect the overall electrostatic surface of a protein. For example, if a large positive charge is normally observed in a potential surface, a lack of this positive charge by nullification will result in volume decrease of the potential. Important to note, the nullification of an amino acid outside of the protein-protein interface region will usually result in no surface volume change since volume is only measured around the interface region.

In interface field comparison, nullification allows us to generate volume differences for each amino acid. We use the interface field comparison method to generate differences in volume between an un-nullified interface field and a nullified interface field. An example of the effect of nullification on volume of an electrostatic surface can be seen in Fig. 3. Nullification at residue 59 removed a large electrostatic region at the interface between barnase and barstar.

C. Interpreting Data

We introduced DiffBond, a method that outputs lists of bonds for three interactions: salt bridges, ionic bonds, and hydrogen bonds. We also discuss a method for identifying residue mutations that are likely to change the electrostatic complementarity between proteins. Nullifying a residue and comparing the interface field outputs an intersection region representing the complementarity of two proteins. We discuss how we can interpret these outputs to identify residues that are significant to electrostatic interactions and can be strong candidates for mutation testing.

1) Nullification Graphs: Using a similar method design to VASP-E, we apply nullification to each amino acid in a complex and perform wildtype-mutant comparison on each to generate a volume difference for each amino acid nullification [18]. We define two conservative prediction thresholds to identify electrostatically influential amino acid interactions. We first find the two amino acids, i and j, that maximize or minimize volume difference. For maximum volume difference Ω at i, we assign an upper prediction threshold of $P = \Omega/2$; likewise, for minimum volume difference ω at j, we assign a lower prediction threshold of $p = \omega/2$. If the nullification of an amino acid x increases the volume difference above P, then x is predicted to reduce complimentarity of the complex and reduce affinity; a decrease in volume difference by amino acid y below p aligns with a prediction that y improves

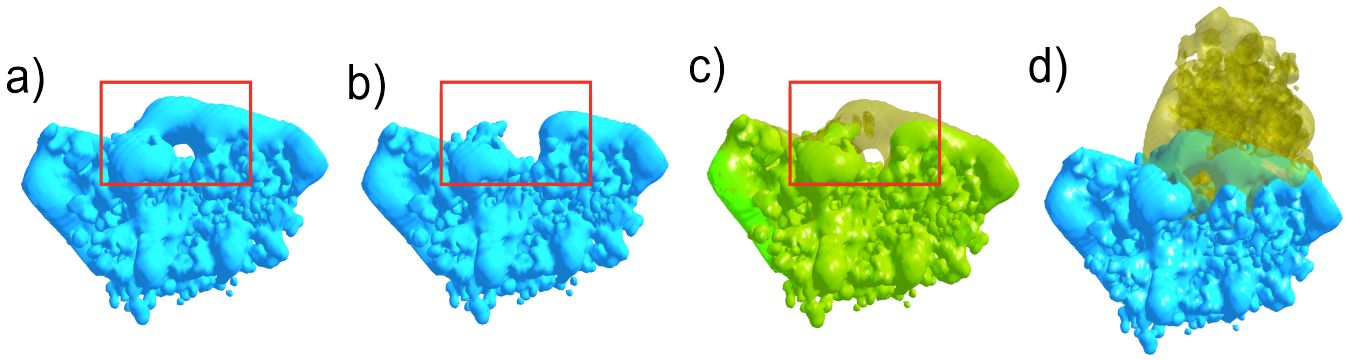

Fig. 3. **Effect of Nullification on Barnase-barstar** a) Wildtype barnase electrostatic surface at isopotential of +1 kT/e. b) Barnase nullified at residue 59, electrostatic surface at isopotential of +1 kT/e. c) Overlap of wildtype (transparent yellow) and nullified barnase (green) surfaces. a,b,c) The red square encompasses the main difference in isopotential surface due to nullification. d) Wildtype barnase (blue) in complex with barstar (transparent yellow).

complementarity and therefore increases affinity. If more than 10% of amino acids exceed the prediction threshold P, it is likely the case that no amino acids contributed a significant decrease to electrostatic complementarity, or that the data contains a large amount of noise. In either case, we ignore the prediction threshold and predict that there are no amino acids that reduce complementarity. Similarly, if 10% of amino acids are less than p, then we predict that no amino acids increase electrostatic complementarity.

2) Nullification and DiffBond Mutation Prediction: Electrostatic complementarity predictions in nullification graphs can be analyzed in conjunction with bond formation data by cross-referencing influential amino acids with bonds formed. Based on nullification, we can predict whether mutating an amino acid will increase or decrease electrostatic complementarity. Similarly, knowing what intermolecular bonds can be formed by an amino acid can help predict changes in complementarity when mutated. For example, mutating one end of an ionic bond to an uncharged or like-charged residue would cause the pair to lose the attraction. Although predictions from nullification and intermolecular bond prediction have not yet been shown to be related, both methods predict behaviors in electrostatic complementarity. Predictions from two separate sources can help point to significant mutation candidates.

D. Data Set Construction

DiffBond was designed to identify electrostatic influences and bonds and, with VASP-E, form predictions of mutations that would either increase or decrease affinity for protein-protein interaction. Because of this design, we validate DiffBond using several families of protein for which specific bond formations that are highly involved in protein electrostatics are well documented. The three-finger toxin family (pdb: 1yi5, 4hqp, 2qc1, 1kc4), barnase-barstar complex (pdb: 1brs), rap1A-RAF complex (pdb: 1c1y), and smad2-smad4 complex (pdb: 1u7v) were selected for validating DiffBond because they have all been extensively studied for specific electrostatically influential amino acids and bonds that affect binding affinity to their corresponding binding partners.

1) Barnase-Barstar: Barnase is an extracellular RNase of Bacillus amyloliquefaciens that is often co-expressed with its inhibitor barstar; without the concurrent expression of barnase in complex with barstar, barnase can be lethal to the cell [21]. Barstar inhibits barnase by forming a tight complex with many intermolecular steric and electrostatic interactions at binding site residues [22]–[24]. As a result, mutating residues involved in these intermolecular interactions often results in enhanced or diminished electrostatic complementarity between barnase and barstar.

2) Rap1a-Raf: Ras is a family of GTPase that transmits signals via protein-protein interaction to regulate many biological systems, like cell cycle progression, cell division, apoptosis, lipid metabolism, DNA synthesis, and cytoskeletal organization [25]. While little is known about ras structure in complex with its effector ligands, rap1a has a similar structure to ras; it has an almost identical binding interface, and binds competitively to ras effectors like raf, an oncogene involved in ERK 1/2 signaling [26], [27]. Like ras, the binding interface of rap1a-raf consists of a few crucial intermolecular bond interactions whose mutations alter binding affinity [26], [28], [29].

3) Smad2-Smad4: Smads is a family of structurally similar proteins that act as main signal transducers for TGF-B receptors, a super family of proteins that help regulate cell development and growth [30], [31]. R-Smad proteins, like Smad2, direct the TGF-B signaling while Smad4 help mediate the formation of the heterometic complex between R-Smads and Smad4 [30]. This dataset uses the trimer consisting of one Smad4 and two Smad2 subunits whose binding interface are well studied for electrostatic interactions and mutation effect [32].

4) Three-finger Toxin Family: Three-finger toxins are a protein superfamily consisting of many small and structurally similar toxin proteins from elapid snake venom [33], [34]. Their distinct structure consist of three beta strand loops emanating from a cysteine rich core, which facilitates interaction with many receptor or channel proteins; neurotoxin members of the family, such as α-bungarotoxin [35]–[37] and

Total	Ionic Bond	Hydrogen Bond	Salt Bridge
True Positive	14	35	6
False Positive	2	5	1
False Negative	3	12	6
True Negative	Unknown	Unknown	Unknown
Precision	87.5%	87.5%	85.7%
Recall	82.4%	74.5%	50.0%

TABLE I
PRECISION AND RECALL OF THE BOND LIST FOR PREDICTING THE FORMATION OF BONDS.

α-cobratoxin [38] interact with neuronal and muscle nicotinic acetylcholine receptors (nAChRs) while other members can interact with muscarinic acetylcholine receptors (mAChRs) or different neuronal nAChR subtypes [34]. The interaction between many members of this toxin family and nAChRs are well studied with known interactions across the interface.

The protein complexes formed by barnase-barstar, rap1a-raf, and smad2-smad4 are commonly used protein complexes for studying electrostatic interaction. They are comprehensively tested and reviewed in literature on both sides of the interface, and provide clear descriptions of bond formation between residues and altered binding affinity due to mutation. Although some three-finger toxin members are not as well studied as the three protein complexes mentioned prior, α-bungarotoxin and α-cobratoxin are both studied extensively in mutation testing.

III. RESULTS

DiffBond is the first method for identifying ionic bonds and salt bridges, and it uses HBPlus, one of the only methods for identifying hydrogen bonds. Consequently, DiffBond cannot be compared to HBPlus and there are no other existing methods it can be compared against for identifying salt bridges and ionic bonds. Instead, we compare predictions from DiffBond to experimental data.

A. Bond Prediction Validation

We validate the DiffBond bond prediction method by comparing bond predictions with known interactions published in experimental findings. Intermolecular bond formations were gathered from published journal papers rather than from a database because we not only want to confirm the existence of a bond formed between a pair of amino acids, but we also want to attribute a reason for bond formation that came from experts in biology through mutation testing or in-depth crystallographic analysis. By collecting the predictions as a set of ionic bonds, hydrogen bonds, and salt bridges, we can measure the prediction performance of predicted bonds on each type of bond separately.

We start by counting true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs). A bond prediction is defined as a TP if our findings predict that a bond forms between an amino acid pair and the literature agrees for the same specific amino acid pair. Similarly, hydrogen bonds are TPs only if experimental findings conclude a hydrogen bond exists between the pair. Finally, salt bridges are TPs only if studies either state the bond contains both a hydrogen bond and electrostatic interaction or explicitly states a salt bridge exists. FPs are bond predictions that were found but are not considered to form the predicted bond by experimental findings. TNs are bond predictions that we predicted would not occur and that experimental findings agree would not occur. FNs are bonds that DiffBond did not predict would occur, but experimental findings found those bonds to form.

We cannot fully count TNs because no studies specifically discuss and analyze the electrostatic influence of every amino acid in a protein. However, we can evaluate the prediction accuracy of DiffBond without TNs; we compute precision and recall to verify accuracy. Precision is the fraction of correctly predicted bonds among all bonds verified in experimental findings, and recall (sensitivity) is the fraction of correctly predicted bonds among all true interactions.

When searching literature for intermolecular interactions, if a bond was generated by our bond list but was not mentioned in literature, we considered this prediction to be a FP. This strict criteria verifies that the precision we report is a lower limit, with the possibility that any unmentioned bond predictions may be validated in the future.

The precision and recall for individual protein complexes and the cumulative accuracy statistics are reported in Table I. Although the total number of ionic predictions were low at 16, ionic bond prediction exhibited high precision and recall at 87.5% and 82.4% respectively. Hydrogen bond predictions showed both strong precision and recall at 87.5% and 74.5% over a large set of predictions at n=40 . Salt bridge prediction had a high precision but a low recall at 85.7% and 50% respectively, over a small sample size of n=7.

Of the 3 false negatives from ionic bond prediction, all 3 pairs were correctly predicted by expanding distance threshold to 7.5Å instead of 5Å. Similarly, of the 6 false negatives for salt bridge prediction, 3 of the 6 were correctly predicted by a distance threshold of 5Å instead of our chosen threshold of 4Å. Although hydrogen bond prediction had 12 false negatives, HBPlus uses many more constraints, including an additional minimum distance cutoff that may create false negatives. However, we did not measure a larger threshold range for hydrogen bonds to determine whether distance played a large role in lowering recall.

B. Nullification Graph Predictions

Using VASP-E, we performed a nullification over each amino acid of protein complexes and calculated the volume difference between wildtype and nullified surfaces. The volume difference for each nullification in the barnase-barstar complex can be seen in Fig. 4, represented by the colored lines.

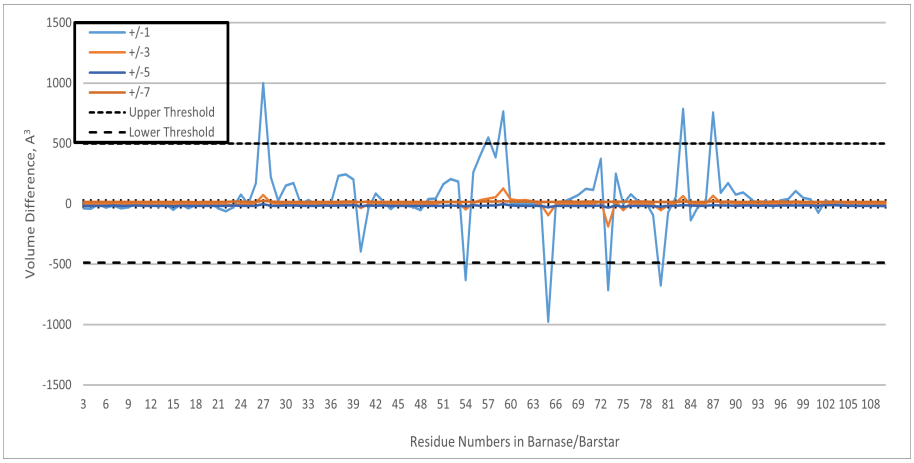

Fig. 4. Volume difference between wildtype and mutant barnase-barstar complex when nullifying barnase amino acids at k= +/-1, +/-3, +/-5, and +/-7. Significant residue nullifications are those that surpass either the upper or lower threshold.

We present all predictions from nullification in Supplemental Table S2 alongside predictions from the bond list.

IV. Discussion

We have presented DiffBond, a method for identifying significant bonds in protein-protein interactions and predicting influential amino acids for mutation testing. There are few tools that perform an in-depth search of intermolecular bonds while attributing a biochemical reason to the interaction. To our knowledge, none have paired this search with a volumetric electrostatic analysis to further inform possible mutation testing. In experimental settings, the design of mutational studies can be difficult because there are many amino acids to consider, and picking a misguided mutation candidate can be time consuming, expensive and unproductive. DiffBond was designed to guide mutation testing by gathering additional structural and electrostatic information that may not be apparent to support experimental designs.

Scanning for salt bridges, ionic bonds, and hydrogen bonds is a novel approach introduced in this paper and has demonstrated promising capability in identifying intermolecular bonds in protein complexes with high precision and recall. Bond prediction was able to predict 14 ionic bonds and their paired partners out of 17 bonds known in literature. Similarly, salt bridge prediction maintained a high precision. Recall for salt bridge prediction was slightly lower which was expected due to the very strict distance cutoff when defining a salt bridge. Hydrogen bond prediction also demonstrated high precision and recall. When using textbook defined measurements for each type of interaction, we found that these definitions are strong predictors for each bond type.

One criteria in collecting data that reduced precision was that any bond predictions that had no mention in literature were considered false positives. This supports that our precision statistic is a lower limit and that future studies may verify bonds that we considered false predictions. To our knowledge, no experimental results have established that predictions we considered false positive do not occur, and so we present it as an open prediction.

Furthermore, we expect DiffBond parameters like distance threshold or bond list interpretation to vary on a case-by-case basis. The assumption for a 4Å distance threshold for salt bridges, or a 5Å distance threshold for ionic bonds may not always hold true; low resolution structures can introduce a margin of error for amino acid spatial placement. Even with high resolution structures, we cannot always account for side chain flexibility especially when residues are not sequestered. Some studies may be interested in gathering more data at the cost of precision. For example, a study may be interested in finding more salt bridges, but the conventional salt bridge definition yields few residues due to its strict distance criteria for a good salt bridge geometry. Expanding the distance threshold means fewer true salt bridges are missed, but more false positives are identified. Incorporation of dynamic information might also be helpful in improving predictions of bond locations generally, and will be considered in future work.

Similar to scanning for intermolecular bonds, VASP-E and the use of nullification presents another method for identifying mutation candidates and mutation prediction. In conjunction, nullification and intermolecular bonds offer two perspectives on electrostatic interactions that can compliment each other without necessarily being related. If both methods point to a residue for mutation prediction, we can consider two hypotheses: First, mutating this residue may affect binding affinity when mutated, either increasing electrostatic complementarity if the peak points upward or decreasing electrostatic complementarity if the peak points downward. Second, mutating this residue to one that no longer forms the predicted bond may break the bond. For example, mutating from a charged acidic residue like lysine to an uncharged residue like glycine likely means any predicted ionic interactions before are no longer possible.

Outside of selecting for high quality mutation candidates,

the additional information provided by these methods offers high value to researchers in experimental design. Pointing out possible interactions and predicting electrostatic complementarity can inform how researchers design a mutation; for example, they can test each of the hypotheses above with mutations to different residues.

DiffBond has the potential to be extended to different applications. Although we have not assumed any correlation between mutation predictions from nullification and from intermolecular bond list, we can assimilate other predictor techniques to begin an artificial reasoning process. Nullification as a first step implies change in electrostatic complementarity by mutation, but does not provide a biochemical reason. The intermolecular bond list interprets three possible interactions when explaining nullification peaks. Adding more prediction techniques for different interactions can form a decision tree of reasoning for mutation prediction.

ACKNOWLEDGEMENTS

This work was funded in part by NIH Grant R01GM123131 to Brian Y. Chen, Julie M. Miwa, and Talulla Palumbo. This work was also funded in part by NIH Grant DA043567 to Julie M. Miwa.

REFERENCES

[1] I. K. McDonald and J. M. Thornton, "Satisfying hydrogen bonding potential in proteins," *Journal of molecular biology*, vol. 238, no. 5, pp. 777–793, 1994.

[2] D. J. Barlow and J. Thornton, "Ion-pairs in proteins," *Journal of molecular biology*, vol. 168, no. 4, pp. 867–885, 1983.

[3] H.-X. Zhou and X. Pang, "Electrostatic interactions in protein structure, folding, binding, and condensation," *Chemical reviews*, vol. 118, no. 4, pp. 1691–1741, 2018.

[4] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig, "Focusing of electric fields in the active site of cu-zn superoxide dismutase: Effects of ionic strength and amino-acid modification," *Proteins: Structure, Function, and Bioinformatics*, vol. 1, no. 1, pp. 47–59, 1986.

[5] J. E. Donald, D. W. Kulp, and W. F. DeGrado, "Salt bridges: geometrically specific, designable interactions," *Proteins: Structure, Function, and Bioinformatics*, vol. 79, no. 3, pp. 898–915, 2011.

[6] S. Pylaeva, M. Brehm, and D. Sebastiani, "Salt bridge in aqueous solution: Strong structural motifs but weak enthalpic effect," *Scientific reports*, vol. 8, no. 1, pp. 1–7, 2018.

[7] R. Dehghanpoor, E. Ricks, K. Hursh, S. Gunderson, R. Farhoodi, N. Haspel, B. Hutchinson, and F. Jagodzinski, "Predicting the effect of single and multiple mutations on protein structural stability," *Molecules*, vol. 23, no. 2, p. 251, 2018.

[8] C. Lee and M. Levitt, "Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core," *Nature*, vol. 352, no. 6334, pp. 448–451, 1991.

[9] M. Prevost, S. J. Wodak, B. Tidor, and M. Karplus, "Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the ile-96—-ala mutation in barnase," *Proceedings of the National Academy of Sciences*, vol. 88, no. 23, pp. 10880–10884, 1991.

[10] J. M. Thornton, M. W. MacArthur, I. K. McDonald, D. T. Jones, J. B. Mitchell, C. L. Nandi, S. L. Price, and M. J. Zvelebil, "Protein structures and complexes: what they reveal about the interactions that stabilize them," *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, vol. 345, no. 1674, pp. 113–129, 1993.

[11] E. N. Baker and R. E. Hubbard, "Hydrogen bonding in globular proteins," *Progress in biophysics and molecular biology*, vol. 44, no. 2, pp. 97–179, 1984.

[12] K. K. Lee, C. A. Fitch, and B. García-Moreno E, "Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein," *Protein science*, vol. 11, no. 5, pp. 1004–1016, 2002.

[13] S. Kumar and R. Nussinov, "Salt bridge stability in monomeric proteins," *Journal of molecular biology*, vol. 293, no. 5, pp. 1241–1255, 1999.

[14] S.-Y. Sheu, D.-Y. Yang, H. Selzle, and E. Schlag, "Energetics of hydrogen bonds in peptides," *Proceedings of the National Academy of Sciences*, vol. 100, no. 22, pp. 12683–12687, 2003.

[15] L. Li, C. Li, S. Sarkar, J. Zhang, S. Witham, Z. Zhang, L. Wang, N. Smith, M. Petukh, and E. Alexov, "Delphi: a comprehensive suite for delphi software and associated resources," *BMC biophysics*, vol. 5, no. 1, pp. 1–11, 2012.

[16] C. Li, Z. Jia, A. Chakravorty, S. Pahari, Y. Peng, S. Basu, M. Koirala, S. K. Panday, M. Petukh, L. Li, *et al.*, "Delphi suite: new developments and review of functionalities," *Journal of computational chemistry*, vol. 40, no. 28, pp. 2502–2508, 2019.

[17] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *ACM siggraph computer graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[18] B. Y. Chen, "Vasp-e: Specificity annotation with a volumetric analysis of electrostatic isopotentials," *PLoS computational biology*, vol. 10, no. 8, p. e1003792, 2014.

[19] H. B. Voelcker and A. A. Requicha, "Geometric modeling of mechanical parts and processes," *Computer*, vol. 10, no. 12, pp. 48–57, 1977.

[20] W. Rocchia, E. Alexov, and B. Honig, "Extending the applicability of the nonlinear poisson- boltzmann equation: multiple dielectric constants and multivalent ions," *The Journal of Physical Chemistry B*, vol. 105, no. 28, pp. 6507–6514, 2001.

[21] M. Jucovic and R. W. Hartley, "Protein-protein interaction: a genetic selection for compensating mutations at the barnase-barstar interface," *Proceedings of the National Academy of Sciences*, vol. 93, no. 6, pp. 2343–2347, 1996.

[22] R. W. Hartley, "Directed mutagenesis and barnase-barstar recognition," *Biochemistry*, vol. 32, no. 23, pp. 5978–5984, 1993.

[23] A. M. Buckle, G. Schreiber, and A. R. Fersht, "Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-. ang. resolution," *Biochemistry*, vol. 33, no. 30, pp. 8878–8889, 1994.

[24] G. Schreiber and A. R. Fersht, "Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering," *Biochemistry*, vol. 32, no. 19, pp. 5145–5150, 1993.

[25] A. D. Cox and C. J. Der, "The dark side of ras: regulation of apoptosis," *Oncogene*, vol. 22, no. 56, pp. 8999–9006, 2003.

[26] T. Devanand, P. Venkatraman, and S. Vemparala, "Phosphorylation promotes binding affinity of rap-raf complex by allosteric modulation of switch loop dynamics," *Scientific reports*, vol. 8, no. 1, pp. 1–15, 2018.

[27] J. M. Kyriakis, H. App, X.-f. Zhang, P. Banerjee, D. L. Brautigan, U. R. Rapp, and J. Avruch, "Raf-1 activates map kinase-kinase," *Nature*, vol. 358, no. 6385, pp. 417–421, 1992.

[28] N. Nassar, G. Horn, C. A. Herrmann, A. Scherer, F. McCormick, and A. Wittinghofer, "The 2.2 å crystal structure of the ras-binding domain of the serine/threonine kinase c-raf1 in complex with rapla and a gtp analogue," *Nature*, vol. 375, no. 6532, pp. 554–560, 1995.

[29] I. Muegge, T. Schweins, and A. Warshel, "Electrostatic contributions to protein–protein binding affinities: application to rap/raf interaction," *Proteins: Structure, Function, and Bioinformatics*, vol. 30, no. 4, pp. 407–423, 1998.

[30] J. Xu and L. Attisano, "Mutations in the tumor suppressors smad2 and smad4 inactivate transforming growth factor β signaling by targeting smads to the ubiquitin–proteasome pathway," *Proceedings of the National Academy of Sciences*, vol. 97, no. 9, pp. 4820–4825, 2000.

[31] K.-i. Miyazono, Y. Ohno, H. Wada, T. Ito, Y. Fukatsu, A. Kurisaki, M. Asashima, and M. Tanokura, "Structural basis for receptor-regulated smad recognition by man1," *Nucleic acids research*, vol. 46, no. 22, pp. 12139–12153, 2018.

[32] B. M. Chacko, B. Y. Qin, A. Tiwari, G. Shi, S. Lam, L. J. Hayward, M. De Caestecker, and K. Lin, "Structural basis of heteromeric smad protein assembly in tgf-β signaling," *Molecular cell*, vol. 15, no. 5, pp. 813–823, 2004.

[33] J. M. Miwa, K. R. Anderson, and K. M. Hoffman, "Lynx prototoxins: Roles of endogenous mammalian neurotoxin-like proteins in modulating nicotinic acetylcholine receptor function to influence complex biological processes," *Frontiers in pharmacology*, vol. 10, p. 343, 2019.

[34] G. Faure, I. Shelukhina, D. Porowinska, M. Shulepko, E. Lyukmanova, D. Dolgikh, E. Spirova, I. Kasheverov, Y. N. Utkin, J.-P. Corringer, *et al.*, "Interaction of three-finger proteins from snake venoms and from mammalian brain with the cys-loop receptors and their models," in

Doklady Biochemistry and Biophysics, vol. 468, pp. 193–196, Springer, 2016.

[35] S. Li, K. Cheng, R. Gomoto, N. Bren, S. Huang, S. Sine, and L. Chen, "Structural principles for alpha-neurotoxin binding to and selectivity among nicotinic receptors," *BE Publ*.

[36] L. Moise, A. Piserchio, V. J. Basus, and E. Hawrot, "Nmr structural analysis of α-bungarotoxin and its complex with the principal α-neurotoxin-binding sequence on the α7 subunit of a neuronal nicotinic acetylcholine receptor," *Journal of Biological Chemistry*, vol. 277, no. 14, pp. 12406–12417, 2002.

[37] M. Zouridakis, A. Papakyriakou, I. A. Ivanov, I. E. Kasheverov, V. Tsetlin, S. Tzartos, and P. Giastas, "Crystal structure of the monomeric extracellular domain of α9 nicotinic receptor subunit in complex with α-conotoxin rgia: Molecular dynamics insights into rgia binding to $\alpha 9\alpha 10$ nicotinic receptors," *Frontiers in pharmacology*, vol. 10, p. 474, 2019.

[38] Y. Bourne, T. T. Talley, S. B. Hansen, P. Taylor, and P. Marchot, "Crystal structure of a cbtx–achbp complex reveals essential interactions between snake α-neurotoxins and nicotinic receptors," *The EMBO journal*, vol. 24, no. 8, pp. 1512–1522, 2005.

TABLE S1: Bond predictions and validation by literature. [22]–[24], [26], [28], [29], [32], [35]–[38]

Protein/PDB	Bond Type/Distance	Interacting Amino Acid Pair	Shown by Literature
1YI5	**Ionic Bond**		
	5A	F ASP27 / B LYS34	N
	7.5A	F ARG33 / A GLU149	N
		F ARG33 / A ASP194	N
		F ARG36 / A ASP194	N
	10A	F ASP8 / A ARG148	N
		F ASP8 / A LYS180	N
		F ARG33 / A ASP85	N
		F ARG33 / A GLU190	N
		F LYS35 / B GLU163	N
		F LYS49 / B GLU163	N
		F ARG68 / B ASP108	N
		F ARG68 / B GLU110	N
	Hydrogen Bond	F ASP27 / A TYR185	Y
		F LYS35 / A SER186	Y
		F ASP27 / TYR185	
	Salt Bridge	F ASP27 / B LYS34	N
	Missed	None	
4HQP	**Ionic Bond**		
	5A	I ASP30 / A ARG182	Y
		I ARG36 / A ASP193	N
		I LYS38 / A GLU185	Y
	7.5A	I ARG25 / A GLU185	N
		I LYS38 / B GLU185	N
	10A	I ARG36 / A ASP87	N
		I ARG36 / A GLU151	N
		I LYS38 / B ASP160	N
		I GLU41 / A ARG182	N
		I HIS68 / A GLU185	Y
		I HIS68 / A GLU189	N
		I LYS70 / A GLU185	Y
	Hydrogen Bond	I VAL40 / A PHE183	Y
		I ASP30 / A TYR184	Y
		I LYS38 / A GLU185	Y
	Salt Bridge	I ASP30 / A ARG182	Y
	Missed	ARG36 / TYR91	Hydrogen Bond
		ARG36 / TRP145	Hydrogen Bond
		LYS38/GLU185	Salt Bridge
2QC1	**Ionic Bond**		
	5A	None	
	7.5A	A ARG36 / B ASP152	N
		A LYS52 / B GLU129	N
	10A	A ASP30 / B LYS145	N
		A ARG36 / B ASP89	N
		A ARG36 / B ASP200	N
		A GLU41 / B LYS145	N

Table S1 continued from previous page

Protein/PDB	Bond Type/Distance	Interacting Amino Acid Pair	Shown by Literature
		A GLU41 / B HIS186	N
	Hydrogen Bond	A ASP30 / B TYR190	Y
		A ARG36 / B CYS192	Y
		A ARG36 / B THR148	Y
		A ARG36 / B ARG149	Y
		A LYS38 / B SER191	Y
		A VAL40 / B PHE189	Y
		A HIS68 / B SER191	N
		A LYS70 / B CYS192	N
	Salt Bridge	None	
	Missed	None	
Barnase/Barstar	**Ionic Bond**		
	5A	A ARG59 / D GLU76	Y
		A ARG59 / D ASP35	Y
		A HIS102 / D ASP39	Y
	7.5A	A LYS27 / D GLU80	N
		A ARG59 / D GLU80	N
		A ARG59 / D ASP39	Y
		A GLU60 / D HIS17	N
		A ARG83 / D ASP39	Y
		A ARG87 / D ASP39	Y
	10A	A LYS27 / D ASP39	Y
		A LYS39 / D GLU46	N
		A LYS62 / D ASP35	N
		A HIS102 / D ASP35	Y
	Hydrogen Bond	A LYS27 / D THRE42	Y
		A ARG59 / D ASP35	Y
		A ARG59 / D GLU76	Y
		A GLU60 / D ASP35	N
		A GLU60 / D LEU34	Y
		A ARG83 / D TYR29	Y
		A ARG83 / D ASP39	Y
		A ARG83 / D GLY43	Y
		A ARG87 / D ASP39	Y
		A HIS102 / D ASN33	Y
		A HIS102 / D GLY31	Y
		A HIS102 / D ASP39	Y
	Salt Bridge	ARG59 / D GLU76	Y
	Missed	ARG83/ASP39	Hydrogen Bond + Electrostatic
		ARG87/ASP39	Hydrogen Bond + Electrostatic
		ASN84 / TYR29	Hydrogen Bond
Rap1a/Raf	**Ionic Bond**		
	5A	A GLU3 / B LYS65	Y
		A ASP33 / B LYS84	Y
		A GLU37 / B ARG67	Y
		A ASP38 / B ARG89	Y
	7.5A	A GLU37 / B ARG59	Y
		A GLU54 / B LYS65	N
		A GLU54 / B ARG67	N
	10A	A GLU30 / B LYS87	N

Table S1 continued from previous page

Protein/PDB	Bond Type/Distance	Interacting Amino Acid Pair	Shown by Literature
		A GLU37 / B ARG89	N
		A ASP38 / B ARG67	N
		A ASP38 / B LYS84	N
		A ASP57 / B LYS84	N
		A ASP57 / B ARG89	N
	Hydrogen Bond	A ASP33 / B LYS84	Y
		A GLU37 / B VAL69	Y
		A GLU37 / B ARG59	Y
		A ASP38 / B THR68	Y
		A ASP38 / B ARG89	Y
		A SER39 / B ARG67	Y
		A SER39 / B ARG89	Y
		A ARG41 / B ASN64	Y
	Salt Bridge	ASP33 / B LYS84,	Y
	Missed	ASP38 / ARG89	Salt Bridge
		GLU37/ARG67	Salt Bridge
		GLU37 / ARG59	Salt Bridge
		ARG41 / GLN 66	Hydrogen Bond
1KC4	**Ionic Bond**		
	5A	A LYS38 / B GLU188	Y
		A HIS68 / B GLU188	Y
	7.5A	A ARG25 / B GLU188	N
		A ARG36 / B GLU192	N
	10A	A ARG36 / B GLU184	N
		A ARG36 / B GLU188	Y
	Hydrogen Bond	A TRP28 / B TYR194	N
		A ARG36 / B GLU188	Y
	Salt Bridge	None	
	Missed	Arg36 / Phe186	Hydrogen Bond
		Arg36 / Tyr 194	Hydrogen Bond
Smad2/Smad4	**Ionic**	B chain with C chain	
	5A	B LYS340 / C GLU288	Y
		B ASP493 / C ARG321	Y
		B ASP493 / C ARG329	Y
	7.5A	B HIS317 / C ASP304	N
		B ASP493 / C ARG330	N
		B ASP494 / C ARG329	N
		B ASP494 / C ARG330	N
		B ARG496 / C GLU281	N
		B ARG496 / C GLU326	N
		B ARG497 / C GLU326	N
		B ASP537 / C ARG310	Y
		B HIS541 / C ASP300	N
		B ASP547 / C ARG330	N
		B ASP547 / C HIS331	N
	10A	B GLU337 / C ARG285	N
		B GLU337 / C HIS291	N
		B ASP494 / C ARG321	N
		B ARG502 / C GLU326	N
		B LYS519 / C GLU281	N

Table S1 continued from previous page

Protein/PDB	Bond Type/Distance	Interacting Amino Acid Pair	Shown by Literature
		B GLU526 / C ARG321	Y
		B HIS528 / C GLU288	N
		B HIS530 / C GLU288	N
		B HIS530 / C ASP304	N
		B ASP537 / C ARG329	N
		B HIS541 / C ASP304	N
	Hbond	B THR338 / C GLU288	N
		B LEU533 / C THR303	Y
		B ASP537 / C ARG310	Y
		B GLU526 / C SER317	Y
		B ASP493 / C ARG321	Y
	Salt bridge	B LYS340 / C GLU288	Y
		B ASP493 / C ARG321	Y
		B ASP493 / C ARG329	Y
	Missed	LYS340 / GLU288	Hydrogen Bond
		ASP332 / ASN320	Hydrogen Bond
		HIS528 / SER318	Hydrogen Bond
		GLN534 / ASP304	Hydrogen Bond
		ASP537 / THR303	Hydrogen Bond
		ASP537 / ASP304	Hydrogen Bond

TABLE S2: Nullification predictions and agreement with bond list.

PDB ID	Nullification	Peak Direction	Agrees with Bond List?
1YI5			
	ASP27	-	Y
	ALA28	-	N
	ARG33	+	Y
	ARG36	+	Y
4HQP			
	ASP30	-	Y
	SER34	+	N
	SER35	+	N
	ARG36	+	Y
	GLY37	+	N
	LYS38	+	Y
	GLU56	-	N
	LYS70	+	Y
2QC1			
	ASP30	-	Y
	SER35	+	N
	ARG36	+	Y
1brs			
	LYS27	+	Y
	ASP54	-	N
	SER57	+	N
	ARG59	+	Y
	GLY65	-	N
	GLU73	-	N
	SER80	-	N

Table S2 continued from previous page

PDB ID	Nullification	Peak Direction	Agrees with Bond List?
	ARG83	+	Y
	ARG87	+	Y
1c1y			
	GLU3	+	Y
	5	-	N
	16	-	N
	ASP33	+	Y
	36	-	N
	GLU37	+	Y
	ASP38	+	Y
	SER39	-	Y
	ARG41	-	Y
	GLU54	+	Y
	ASP57	+	Y
1KC4			
	ARG36	-	Y
	GLY37	+	N
SMAD2/SMAD4			
	VAL492	-	N
	ASP493	+	Y
	ARG496	-	Y
	ARG497	-	N
	ARG502	-	Y
	LYS519	-	N
	ASP537	+	Y
	ASP547	+	Y