Influential Mutations in the Smad4 Trimer Complex can be Detected from Disruptions of Electrostatic Complementarity

BRIDGET E. NOLAN[1], EMILY LEVENSON[1] and BRIAN Y. CHEN[1*]

ABSTRACT

This paper examines three techniques for rapidly assessing the electrostatic contribution of individual amino acids to the stability of protein-protein complexes. Whereas the energetic minimization of modeled oligomers may yield more accurate complexes, we examined the possibility that simple modeling may be sufficient to identify amino acids that add to or detract from electrostatic complementarity. The three methods evaluated were (a) the elimination of entire sidechains (e.g. glycine scanning), (b) the elimination of the electrostatic contribution from the atoms of a sidechain, called nullification, and (c) sidechain structure prediction using SCWRL4. These techniques generate models in seconds, enabling large-scale mutational scanning. We evaluated these techniques on the SMAD2/SMAD4 heterotrimer, whose formation plays a crucial role in antitumor pathways. Many studies have documented the clinical and structural effect of specific mutations on trimer formation. Our results describe how glycine scanning yields more specific predictions, though nullification may be more sensitive, and how sidechain structure prediction enables the identification of uncharged-to-charge mutations.

[1]Department of Computer Science and Engineering, Lehigh University

[*]Current address: Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA, USA 18015

Key words: Protein structure comparison, Molecular Electrostatics, binding specificity.

1. INTRODUCTION

Electric fields provide essential data for analyzing molecular interactions. In protein complexes, the field can permeate the interface between oligomers, creating patterns of localized attraction and repulsion that finely constrain binding preferences (e.g. Schreiber and Fersht (1993); Nassar et al. (1996)). Examining these patterns, visually or computationally, can reveal conserved characteristics of electric fields that point to similarities in protein function Record et al. (1978); Warshel and Russell (1984); Matthew (1985); Honig et al. (1986); Rogers (1986); Harvey (1989); McLaughlin (1989); Sharp and Honig (1990); Honig and Nicholls (1995); Nakamura (1996); Kinoshita and Nakamura (2003); Zhang et al. (2006) and binding specificity Polticelli et al. (1999); Hendsch and Tidor (1994); Sindelar et al. (1998); O'Shea et al. (1992); Kangas and Tidor (2001); Lee and Tidor (2001). When analyzing the geometry of electrostatic isopotentials, computationally detected differences in electrostatic complementarity and non-complementarity can reveal variations in specificity and insights into the role of individual amino acids on recognition Chen (2014). To better anticipate the electrostatic effects of mutation, many methods analyze similarities and variations in electrostatic fields.

Predicting the electrostatic influences of specific mutations is challenging in part because mutations simultaneously create both structural and electrostatic changes. Replacing a charged sidechain with an uncharged one alters the electrostatic field and it can also alter the pattern of sidechain packing and solvent exposure nearby. Changes in the solvent exposure of a charged group can locally decrease the dielectric of the polar solvent, enhancing nearby electrostatic potentials through

electrostatic focusing Klapper et al. (1986); Polticelli et al. (1999); Rohs et al. (2009); Blumenthal et al. (2013). While the computationally intensive minimization of mutant structures could better approximate realistic packing, we hypothesized that computationally inexpensive modifications may be sufficient to approximate the effect of most mutations and thus rapidly screen for electrostatically influential amino acids.

This paper evaluates three computationally inexpensive ways to modify protein structures for the purpose of evaluating the electrostatic influence of specific amino acids on protein-protein recognition. The first method is glycine scanning, where the sidechain of each amino acid is systematically removed without modifying the rest of the structure. The second method, which we call "nullification", systematically eliminates the electrostatic contribution of each amino acid within each member of an oligomer. Nullification maintains the position of all atoms, ensuring that the spatial arrangement of dielectrics remains identical to the original structure and thus that the shielding influence of structure does not change with the field. Finally, the third approach applies the program SCWRL Krivov et al. (2009) to predict the conformations of all 20 possible amino acids at significant positions along the backbone. The effect of each method on electrostatic complementarity was evaluated using VASP-E, a method we described earlier Chen (2014). Our comparison revealed several interesting observations about how these techniques can predict amino acids that have an electrostatic influence on binding and specificity.

We evaluated these predictors on the two heteromeric interfaces of the SMAD2-SMAD4 trimer Chacko et al. (2004), which is part of a signalling pathway with antitumor effects Hardwick et al. (2008); Xu et al. (2006); Ding et al. (2011); Zhang et al. (2010); Itatani et al. (2013); Hernanda et al. (2015). The formation of this trimer, from one SMAD4 and two SMAD2 subunits, is sup-

ported in part by strong electrostatic interactions across its heteromeric interfaces Chacko et al. (2004). Mutations that eliminate certain charge-charge interactions are frequently associated with degradations of trimer formation and hereditary disease. Careful data gathering by researchers in the field (e.g. Wooderchak et al. (2010)) has revealed that many missense mutations of SMAD4 are experimentally associated with juvenile polyposis syndrome (JPS) Howe et al. (1998); Schwenter et al. (2012), a hereditary disorder characterized by the formation of polyps in the rectum, colon or GI tract, and hereditary hemorrhagic telangiectasia (HHT) Gallione et al. (2004, 2006), a different hereditary disorder that causes recurring nosebleeds, the dialation of capillaries in the mouth, face, hands, and GI tract, and venous malformation. We used charge-eliminating SMAD4 missense mutants as a benchmark to evaluate the accuracy of hypothetical mutation testing.

2. METHODS

2.1. Measuring Changes in Electrostatic Complementarity

We use VASP-E Chen (2014) to measure electrostatic complementarity. We paraphrase this technique here for convenience and to clarify the experimental design. Beginning with a complex of two proteins, A and B, we first prepare each structure by stripping existing hydrogens and modeling their positions using MolProbity Chen et al. (2009). Next, we use Delphi Honig and Nicholls (1995); Rocchia et al. (2001) to solve the electrostatic potential field for each protein in isolation. We then use VASP-E to produce electrostatic isopotential surfaces at a given range of isopotential thresholds $R = \{a1, a2, a3 \ldots\}$. We refer to the isopotential of some structure X, at potential threshold a as $p(X, a)$.

The isopotentials generated by VASP-E are represented as three dimensional geometric solids.

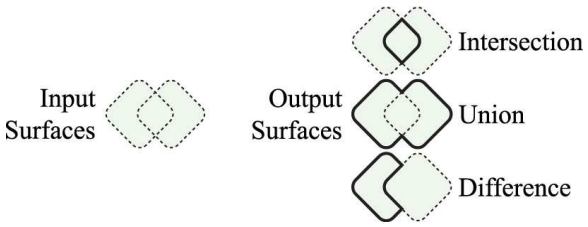

Fig. 1: Boolean operations on geometric solids. Input surfaces are shown in grey with dotted lines. Output surfaces for three basic operations are shown in grey with solid lines.

Using Boolean set operations (Fig. 1), VASP-E can compute the intersection region where two or more isopotentials overlap. We denote boolean intersections with $\cap$. VASP-E can also precisely measure the volume occupied a given region X in cubic angstroms. We denote the volume of X as $v(X)$.

To measure electrostatic complementarity between oligomers A and B at a given isopotential threshold a, we use VASP-E to compute Boolean intersections between $p(A, a)$ and $p(B, -a)$ (Fig. 2d). We say that the volume $C = v(p(A, a) \cap p(B, -a))$ of the intersection represents the degree of electrostatic complementarity at the threshold a, because it measures the degree of geometric overlap between oppositely charged isopotentials. This measurement is repeated for each threshold in the given range R. We interpret these measurements as relative indicators of electrostatic complementarity at different levels of electrostatic potential. When comparing two similar complexes at the same potential threshold, larger volumes of intersection suggest greater electrostatic complementarity and smaller volumes suggest the opposite.

To evaluate the electrostatic effect of a modeled change on subunit A of complex AB, we first modify A to produce A'. We then replace A with A' to produce the complex AB. The methods in Section 2.2 describe what modifications are made and how. Once modification is complete, we

evaluate C, the electrostatic complementarity of the complex AB, and subtract it from C, for each member of the range R of electrostatic thresholds. The difference $C - C$ represents the change in complementarity induced by the modification of A to A. If $C - C$ is close to zero, then we assume that at this isopotential threshold, the effect of mutation on binding does little to affect affinity. When $C - C$ is negative, we assume that the changes in A increase electrostatic complementarity and affinity with B, relative to A. Likewise, when $C - C$ is positive, we surmise that the changes in A decrease affinity.

While changes in complementarity, measured in this way, can be predictors of amino acids that have an electrostatic influence on specificity Chen (2014), the magnitude of the difference in complementarity does not necessarily correlate to differences in binding affinity Chong et al. (1998). Generally speaking, we find that after evaluating a range of electrostatic thresholds, one positive and one negative threshold are generally most informative, with other thresholds reiterating similar results.

When examining the contribution of individual amino acids, at a given isopotential threshold, we predict that an amino acid has an electrostatic influence on affinity when the difference in complementarity between the wild type complex and the modified complex is greater than one half the maximum positive or the maximum negative difference observed for all amino acids at this isopotential. We refer to half the maximum positive difference or half maximum negative difference as the postiive and negative prediction thresholds, respectively. This approach was applied in earlier methods, where such differences in electrostatic isopotential were effective for distinguishing influential amino acids Chen (2014).

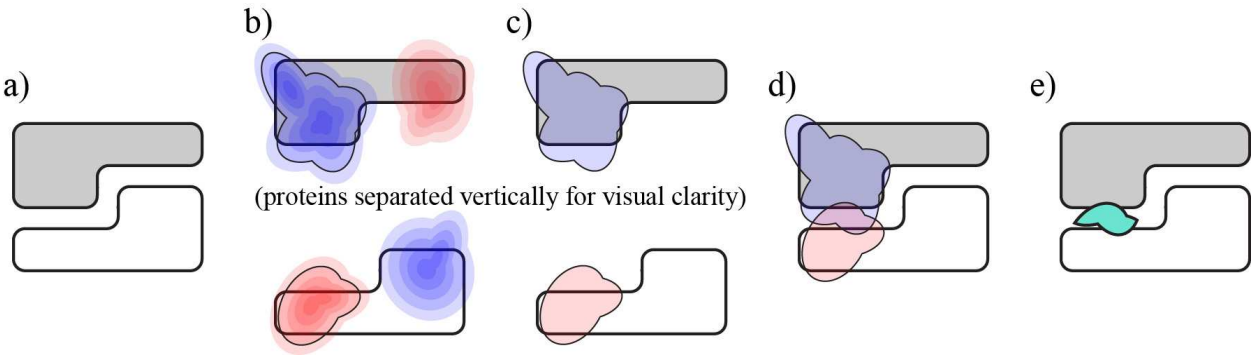

Fig. 2: Computing Electrostatic Complementarity with Boolean Intersections. a) Molecular surfaces of two proteins in complex are shown in dark and light grey with solid outlines. b) Electrostatic isopotentials at increasingly positive and negative isopotential thresholds are shown in increasingly opaque shades of red and blue. The analysis is performed at a user defined isopotential threshold, shown in a thin black line. c) User defined isopotentials, shown in transparent red and blue, are determined for each protein in isolation. One protein uses the negative of the user defined isopotential. d) The spatial relationship between isopotentials. e) A Boolean intersection of the two selected isopotentials.

2.2. Creating Electrostatic Variations at the Protein Protein Interface

To evaluate techniques for examining the contribution of an amino acid on electrostatic complementarity, we considered three three methods for generating electrostatic variations in the protein-protein interface. The first two methods, glycine scanning and nullification, eliminate the electrostatic contribution of individual amino acids. By detecting the removal of charged amino acids that add or detract from electrostatic complementarity, they yield insights into the effect of mutating charged amino acids into uncharged amino acids. The third method, using SCWRL for sidechain structure prediction, both adds and removes charged amino acids. This approach adds the further

capability of evaluating the effect of mutating uncharged amino acids into charged ones.

2.2.1. Glycine Scanning. Glycine scanning modifies the input complex by removing the side chain from one amino acid in one protein of in the input complex. No compensation for the absence of this sidechain is made: the changes in sidechain packing and backbone conformation that inevitably occur are not modeled in this approach. After a sidechain is removed, the complex is examined in the manner described in Section 2.1. This process is repeated for every amino acid in the oligomer. Finally, the difference in complementarity between the original complex and the modified complex is plotted for every amino acid.

2.2.2. Electrostatic Nullification. Earlier, we developed Electrostatic Nullification as one technique to evaluate the influence of one amino acid in molecular interactions Chen (2014). This technique uses Delphi Rocchia et al. (2001) to eliminate the electrostatic contribution of one amino acid in one of the proteins in the complex. While the electrostatic contribution is eliminated, the atomic geometry of the amino acid remains fixed, thereby decoupling changes to the electrostatic field and the structure. The effect is that a region of space that was once occupied by the amino acid remains inaccessible to the solvent, creating a low-dielectric zone that would have reverted to the high dielectric of the solvent if the amino acid were removed completely. Once the electrostatic contribution of the amino acid is eliminated and the electrostatic potential field is recomputed, we evaluate the change in electrostatic complementarity. This process is repeated for every amino acid in both proteins in the complex.

2.2.3. Sidechain Structure Prediction. We use SCWRL Krivov et al. (2009) to modify the input complex by replacing the side chain of one amino acid with the sidechains of the remaining 19

standard amino acids. SCWRL selects the conformation of the side chain to avoid steric hindrance with the existing structure, but no compensation is made for changes at the protein-protein interface that may cause steric hindrance with the protein binding partner. For each substituted amino acid, the entire complex is examined in the manner described in Section 2.1. Only single mutants are considered at a user-selected range of amino acids, to limit the computational cost of examining the huge of all possible single mutants. Once the electrostatic isopotentials are created, we plot the difference in complementarity between the original complex and the sidechain-substituted complex.

2.3. SMAD2-SMAD4 Trimer

Mothers against decapentaplegic homolog 4 (SMAD4) is involved in cell signaling by modulating members of the Transforming Growth Factor B (TGF-B) protein superfamily Chacko et al. (2004). SMAD4 forms heteromeric complexes with other SMAD homologs, such as SMAD2 and SMAD3 Lagna et al. (1996). These complexes accumulate in the nucleus of a cell and function to regulate the transcription of many target genes by acting as transcription factors Massagué et al. (2005). These intracellular mediators are important for transcriptional and anti-proliferative responses to TGF-B. The signals that propagate through the SMAD/TGF-B pathway relate to cell growth, adhesion, migration, cell-fate determination and differentiation, and apoptosis. Since these signals are vital to the life of a cell, malfunctions in the SMAD/TGF-B pathway have been discovered in many human diseases, such as pancreatic cancer, head and neck cancers, colorectal cancer, as well as juvenile polyposis and hereditary hemorrhagic telangiectasia Gallione et al. (2004). More specifically, the SMAD4 protein has been found to be inactive in nearly half of pancreatic carcinomas Hahn et al. (1996). Due to the severity of these diseases, it is desirable to determine the source

of these malfunctions.

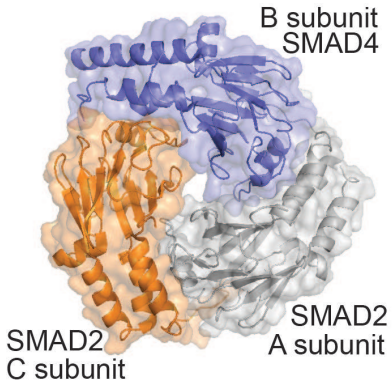

Fig. 3: The SMAD2/SMAD4 trimer. The B subunit is shown in blue. SMAD2 subunits, A and C, are shown in orange and gray.

Previous studies reveal that SMAD2 and SMAD4 form a trimer through a conserved protein-protein interface, and the majority of tumor-derived missense mutations map to this interface Chacko et al. (2004). These mutations disrupt the formation of the trimer, indicating that the assembly of the SMAD4 is crucial for cell signaling and is disturbed by mutations that are ultimately tumorigenic. The crystal structure of the SMAD2/SMAD4 trimer (pdb: 1u7v), described earlier by Chako and coworkers Chacko et al. (2004), is illustrated in Figure Figure 3. SMAD4, hereon referred to as the B subunit, forms two separate interfaces with the SMAD2 subunits, A and C, which share a separate interface. Together, the three subunits have no common interface, leaving an empty region in the center of the trimer. For this reason, we examined SMAD4 mutations on each interface independently from the third subunit.

The Department of Pathology at the University of Utah offers a manually curated database, hereon referred to as the ARUP database, of many known clinically significant mutations in SMAD4 Wooderchak et al. (2010). The mutations identified are associated with JPS and with HHT. To supplement the detailed literature search that populates the SMAD4 database, we performed our

own literature search to identify mutations not mentioned. Altogether, these data form a dataset of experimentally verified mutations that are associated with JPS and HHT. Using this data and other published literature as a gold standard, our computational results evaluate how accurately the three methods considered here can be used to predict a mutation that corresponds to one of these two conditions.

3. RESULTS

We used the techniques described above to create structural variations of the SMAD2/SMAD4 trimer. Glycine scanning and nullification were used to predict charged residues that have an electrostatic influence on binding, by examining the effect of removing them. Sidechain structure prediction was used to predict influential mutations at specific residue positions. All methods were evaluated against findings established in the experimental literature. We demonstrate the validity of our findings by explicitly citing the literature that supports our predictions. Results were therefore summarized in paragraph rather than table form.

3.1. Glycine scanning and Nullification on the AB interface

Glycine scanning, performed on the AB interface of the SMAD4 trimer, identified three negatively charged and two positively charged amino acids that surpassed the prediction threshold (Figure 4a). These residues were E330, E337, R372, E377 and R380. All those identified play an experimentally established role in binding. Nullification detected the same amino acids (Figure 4b).

Mutations of GLU330 have been associated with JP in studies that observed the mutations E330G and E330K in patients and tumors Sayed et al. (2002). GLU337 is known to form a charge-charge

interaction with R243 on Smad3 Schiro et al. (2011). ARG372 and ARG380 are totally conserved in SMAD proteins and they form four and three hydrogen bonds, respectively, that stabilize the loop helix binding region of SMAD4 Shi et al. (1997). GLU 377 is also totally conserved in SMAD Shi et al. (1997). The ARUP database associates several loss of charge mutations with R361, but glycine scanning and nullification of R361 did not create differences in electrostatic complementarity that were sufficient for detection.

3.2. Glycine scanning and Nullification on the BC interface

Glycine scanning (Figure 4c) and nullification (Figure 4d) identified seven influential amino acids at the BC interface. These residues were D493, R496, R502, K519, E526, D537 and D547. All those identified play an experimentally established role in binding: D493 is frequently mutated to histidine in cancer Prokova et al. (2007); Hahn et al. (1996). Mutations of R496 to histidine likely disturb the the formation of homo and heteromeric SMAD complexes Lazzereschi et al. (2005); Chacko et al. (2001). Mutations of R502 disrupt the the capacity for SMAD4 to activate the transcriptional response, thus disrupting its antitumor activity Chacko et al. (2001). Loss of charge in K519 prevents monoubiquitylation of K519 and prevents complex formation with SMAD2 Dupont et al. (2012). E526 forms intramolecular hydrogen bond networks that stabilize the binding interfaces of SMAD4 Shi et al. (1997). D537 is mutated in colon cancer Shi et al. (1997); Eppert et al. (1996), though K537 was below the threshold in glycine scanning, but only slightly above threshold in nullification. Established evidence of mutations of D547 destabilizing the SMAD2/SMAD4 trimer could not be found.

3.3. Substitutions modeled with SCWRL

Using SCWRL, we modeled all 20 possible mutations of every cancer-related missense mutation identified by the ARUP database as well as the amino acids identified via glycine scanning and nullification. These results are illustrated in heat maps in Figure 5.

The electrostatic effect of SMAD4 mutations modeled with SCWRL generally influenced electrostatic complementarity at only one interface. Two mutations, E390 and G510, did not have a substantial influence on electrostatic complementarity. Both amino acids lie distant from either SMAD2 subunit and lie buried in the SMAD4 core. The remaining mutations tested exhibited larger variations in electrostatic complementarity.

Loss of charge mutations caused notable changes in complementarity. Mutations to glycine, which partially duplicated the measurements from glycine scanning, affected complementarity less than opposite charge mutations, which had the greatest effect on complementarity in all cases. The ARUP database indicates that R361C, R361G, R361H, R361L, and R361S are all associated with hereditary disease, and all exhibited a similar disruption of electrostatic complementarity. Indeed, quite a few other mutations of R361, as predicted with SCWRL, appear to achieve the same outcome. Other mutations of R361 appear to achieve similar changes in electrostatic complementarity and may thus also degrade antitumor performance.

Gain-of-charge mutations exhibited changes in complementarity that corresponded to the identity and charge of the substituted side chain. Mutations to argenine, lysine, glutamate, aspartate, and histidine, performed on C363, Y353, G352, L364, G386, C324, W509, G491, L533, W524, all had notable effects on complementarity. Electrostatic complementarity correctly predicted all gain of charge mutations associated by ARUP with hereditary disease.

We also evaluated mutations from uncharged amino acids to uncharged amino acids that are associated by ARUP with hereditary disease. Unsurprisingly, these amino acids were never predicted to have an electrostatic influence on affinity. Among these mutations Y353S, L364V, L364W, G491V, L533V, L533P, W524L, no individual mutation altered electrostatic complementarity substantially from the wild type.

4. DISCUSSION

We have performed a direct comparison between glycine scanning, nullification, and sidechain structure prediction as tools for modeling the influence of mutation on affinity on the SMAD2/SMAD4 trimer. Overall, predictions made with glycine scanning and nullification were similar, both identifying several amino acids that have experimentally established roles in binding. While small variations in electrostatic complementarity may have resulted in the prediction of slightly different amino acids, it is clear that both methods generally identify similar amino acids.

Differences in electrostatic complementarity in glycine scanning models generally tend towards zero when an uncharged residue is changes to glycine. In contrast, differences in electrostatic complementarity in nullification models often result in regions of electrostatic focusing, such as in the SMAD4 residues between 361 and 369. These amino acids have no net charge, but because their electrostatic contribution is totally eliminated. Because they continue to prevent the polar solvent from occupying the same space as the original amino acid, the effect is an enhancement in electrostatic potentials nearby. Thus, the effect of nullifying uncharged amino acids nearby a charged one still changes electrostatic complementarity. This effect was not large enough in SMAD4 to create substantial inaccuracies. Glycine scanning does not exhibit this effect, and should generally

achieve superior prediction specificity, while nullification achieves greater prediction sensitivity.

Unlike glycine scanning and nullification, which only evaluate the loss of charged residues, sidechain structure prediction can identify gain-of-charge mutations that alter electrostatic complementarity. In every case where experimental data is available, gain-of-charge mutations were accurately predicted using SCWRL. Loss-of-charge and opposite-charge mutations were identified as accurately as glycine scanning and nullification. Finally, given the exclusively electrostatic nature of the analysis performed, influential mutations of uncharged sidechains to other uncharged sidechains could not be distinguished from other mutations, as expected. Overall, these data indicate that glycine scanning has considerable potential for the rapid identification of potentially influential charged amino acids and that SCWRL provides an excellent tool for evaluating the effect of any mutation on electrostatic complementarity.

ACKNOWLEDGMENTS

This work was supported in part by National Science Foundation Grant 1320137 to Brian Chen and Katya Scheinberg.

AUTHORS' CONTRIBUTIONS

B.E.N. and E.F. found the dataset, implemented and automated the software, and performed the experiments. B.Y.C. concieved of the experiment and wrote the article.

AUTHOR DISCLOSURE STATEMENT

The authors declare that no competing financial interests exist.

References

Blumenthal, S., Tang, Y., Yang, W., and Chen, B. Y. 2013. Isolating influential regions of electrostatic focusing in protein and dna structure. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on*, 10(5):1188–1198.

Chacko, B. M., Qin, B., Correia, J. J., Lam, S. S., de Caestecker, M. P., and Lin, K. 2001. The l3 loop and c-terminal phosphorylation jointly define smad protein trimerization. *Nature Structural & Molecular Biology*, 8(3):248–253.

Chacko, B. M., Qin, B. Y., Tiwari, A., Shi, G., Lam, S., Hayward, L. J., De Caestecker, M., and Lin, K. 2004. Structural basis of heteromeric smad protein assembly in tgf-β signaling. *Molecular cell*, 15(5):813–823.

Chen, B. Y. 2014. Vasp-e: Specificity annotation with a volumetric analysis of electrostatic isopotentials. *PLoS Comput Biol*, 10(8).

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. 2009. Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 66(1):12–21.

Chong, L. T., Dempster, S. E., Hendsch, Z. S., Lee, L.-P., and Tidor, B. 1998. Computation of electrostatic complements to proteins: A case of charge stabilized binding. *Protein science*, 7(1):206–210.

Ding, Z., Wu, C.-J., Chu, G. C., Xiao, Y., Ho, D., Zhang, J., Perry, S. R., Labrot, E. S., Wu, X., Lis, R., et al. 2011. Smad4-dependent barrier constrains prostate cancer growth and metastatic progression. *Nature*, 470(7333):269–273.

Dupont, S., Inui, M., and Newfeld, S. J. 2012. Regulation of tgf-β signal transduction by mono-and deubiquitylation of smads. *FEBS letters*, 586(14):1913–1920.

Eppert, K., Scherer, S. W., Ozcelik, H., Pirone, R., Hoodless, P., Kim, H., Tsui, L.-C., Bapat, B., Gallinger, S., Andrulis, I. L., et al. 1996. Madr2 maps to 18q21 and encodes a tgfβ–regulated mad–related protein that is functionally mutated in colorectal carcinoma. *Cell*, 86(4):543–552.

Gallione, C. J., Repetto, G. M., Legius, E., Rustgi, A. K., Schelley, S. L., Tejpar, S., Mitchell, G., Drouin, É., Westermann, C. J., and Marchuk, D. A. 2004. A combined syndrome of juvenile polyposis and hereditary haemorrhagic telangiectasia associated with mutations in madh4 (smad4). *The Lancet*, 363(9412):852–859.

Gallione, C. J., Richards, J. A., Letteboer, T., Rushlow, D., Prigoda, N. L., Leedom, T. P., Ganguly, A., Castells, A., van Amstel, J. P., Westermann, C., et al. 2006. Smad4 mutations found in unselected hht patients. *Journal of medical genetics*, 43(10):793–797.

Hahn, S. A., Schutte, M., Hoque, A. S., Moskaluk, C. A., da Costa, L. T., Rozenblum, E., Weinstein, C. L., Fischer, A., Yeo, C. J., Hruban, R. H., et al. 1996. Dpc4, a candidate tumor suppressor gene at human chromosome 18q21. 1. *science*, 271(5247):350–353.

Hardwick, J. C., Kodach, L. L., Offerhaus, G. J., and Van den Brink, G. R. 2008. Bone morphogenetic protein signalling in colorectal cancer. *Nature Reviews Cancer*, 8(10):806–812.

Harvey, S. 1989. Treatment of Electrostatic Effects in Marcomolecular Modeling. *Proteins*, 5:78–92.

Hendsch, Z. and Tidor, B. 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci*, 3(2):211–26.

Hernanda, P. Y., Chen, K., Das, A., Sideras, K., Wang, W., Li, J., Cao, W., Bots, S., Kodach, L., de Man, R., et al. 2015. Smad4 exerts a tumor-promoting role in hepatocellular carcinoma. *Oncogene*, 34(39):5055–5068.

Honig, B., Hubbell, W., and Flewelling, R. 1986. Electrostatic Interactions in Membranes and Proteins. *Annu Rev Biophys Biophys Chem*, 15:163–193.

Honig, B. and Nicholls, A. 1995. Classical electrostatics in biology and chemistry. *Science*, 268(5214):1144–1149.

Howe, J. R., Roth, S., Ringold, J. C., Summers, R. W., Järvinen, H. J., Sistonen, P., Tomlinson, I. P., Houlston, R. S., Bevan, S., Mitros, F. A., et al. 1998. Mutations in the smad4/dpc4 gene in juvenile polyposis. *Science*, 280(5366):1086–1088.

Itatani, Y., Kawada, K., Fujishita, T., Kakizaki, F., Hirai, H., Matsumoto, T., Iwamoto, M., Inamoto, S., Hatano, E., Hasegawa, S., et al. 2013. Loss of smad4 from colorectal cancer cells promotes ccl15 expression to recruit ccr1+ myeloid cells and facilitate liver metastasis. *Gastroenterology*, 145(5):1064–1075.

Kangas, E. and Tidor, B. 2001. Electrostatic Complementarity at Ligand Binding Sites: Application to Chorismate Mutase. *J Phys Chem B*, 105(4):880–888.

Kinoshita, K. and Nakamura, H. 2003. Protein informatics towards function identification. *Curr Opin Struct Biol*, 13(3):396–400.

Klapper, I., Hagstrom, R., Fine, R., Sharp, K., and Honig, B. 1986. 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*, 1(1):47–59.

Krivov, G. G., Shapovalov, M. V., and Dunbrack, R. L. 2009. Improved prediction of protein side-chain conformations with scwrl4. *Proteins: Structure, Function, and Bioinformatics*, 77(4):778–795.

Lagna, G., Hata, A., Hemmati-Brivanlou, A., and Massagué, J. 1996. Partnership between dpc4 and smad proteins in tgf-β signalling pathways.

Lazzereschi, D., Nardi, F., Turco, A., Ottini, L., D'Amico, C., Mariani-Costantini, R., Gulino, A., and Coppa, A. 2005. A complex pattern of mutations and abnormal splicing of smad4 is present in thyroid tumours. *Oncogene*, 24(34):5344–5354.

Lee, L.-P. and Tidor, B. 2001. Optimization of binding electrostatics: Charge complementarity in the barnase-barstar protein complex. *Protein Science*, 10(2):362–377.

Massagué, J., Seoane, J., and Wotton, D. 2005. Smad transcription factors. *Genes & development*, 19(23):2783–2810.

Matthew, J. 1985. Electrostatic effects in proteins. *Annu Rev Biophys Biophys Chem*, 14:387–417.

McLaughlin, S. 1989. The electrostatic properties of membranes. *Annu Rev Biophys Biophys Chem*, 18:113–36.

Nakamura, H. 1996. Roles of electrostatic interaction in proteins. *Q Rev Biophys*, 29:1–90.

Nassar, N., Horn, G., Herrmann, C., Block, C., Janknecht, R., and Wittinghofer, A. 1996. Ras/rap effector specificity determined by charge reversal. *Nature Structural & Molecular Biology*, 3(8):723–729.

O'Shea, E., Rutkowski, R., and Kim, P. 1992. Mechanism of specificity in the Fos-Jun oncoprotein heterodimer. *Cell*, 68(4):699–708.

Polticelli, F., Honig, B., Ascenzi, P., and Bolognesi, M. 1999. Structural determinants of trypsin affinity and specificity for cationic inhibitors. *Protein Sci*, 8(12):2621–2629.

Prokova, V., Mavridou, S., Papakosta, P., Petratos, K., and Kardassis, D. 2007. Novel mutations in smad proteins that inhibit signaling by the transforming growth factor β in mammalian cells. *Biochemistry*, 46(48):13775–13786.

Record, M., Anderson, C., and Lohman, T. 1978. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q Rev Biophys*, 11(2):103–178.

Rocchia, W., Alexov, E., and Honig, B. 2001. Extending the applicability of the nonlinear poisson-boltzmann equation: Multiple dielectric constants and multivalent ions. *J Phys Chem B*, 105(28):6507–6514.

Rogers, N. 1986. The modelling of electrostatic interactions in the function of globular proteins. *Prog Biophys Mol Biol*, 48(1):37–66.

Rohs, R., West, S. M., Sosinsky, A., Liu, P., Mann, R. S., and Honig, B. 2009. The role of dna shape in protein–dna recognition. *Nature*, 461(7268):1248–1253.

Sayed, M., Ahmed, A., Ringold, J., Anderson, M., Bair, J., Mitros, F., Lynch, H., Tinley, S., Petersen, G., Giardiello, F., et al. 2002. Germlinesmad4 orbmpria mutations and phenotype of juvenile polyposis. *Annals of Surgical Oncology*, 9(9):901–906.

Schiro, M. M., Stauber, S. E., Peterson, T. L., Krueger, C., Darnell, S. J., Satyshur, K. A., Drinkwater, N. R., Newton, M. A., and Hoffmann, F. M. 2011. Mutations in protein-binding hot-spots on the hub protein smad3 differentially affect its protein interactions and smad3-regulated gene expression. *PloS one*, 6(9):e25021.

Schreiber, G. and Fersht, A. R. 1993. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry*, 32(19):5145–5150.

Schwenter, F., Faughnan, M. E., Gradinger, A. B., Berk, T., Gryfe, R., Pollett, A., Cohen, Z., Gallinger, S., and Durno, C. 2012. Juvenile polyposis, hereditary hemorrhagic telangiectasia, and early onset colorectal cancer in patients with smad4 mutation. *Journal of gastroenterology*, 47(7):795–804.

Sharp, K. and Honig, B. 1990. Electrostatic Interactions in Macromolecules: Theory and Applications. *Annu Rev Biophys Biophys Chem*, 19:301–332.

Shi, Y., Hata, A., Lo, R. S., Massagué, J., and Pavletich, N. P. 1997. A structural basis for mutational inactivation of the tumour suppressor smad4. *Nature*, 388(6637):87–93.

Sindelar, C., Hendsch, Z., and Tidor, B. 1998. Effects of salt bridges on protein structure and design. *Protein Sci*, 7(9):1898–1914.

Warshel, A. and Russell, S. 1984. Calculations of electrostatic interactions in biological systems and in solutions. *Q Rev Biophys*, 17(3):283–422.

Wooderchak, W., Spencer, Z., Crockett, D., McDonald, J., and Bayrak-Toydemir, P. 2010. Repository of smad4 mutations: Reference for genotype/phenotype correlation. j data mining in genom proteomics 1: 101 doi: 10.4172/2153-0602.1000101.

Xu, X., Kobayashi, S., Qiao, W., Li, C., Xiao, C., Radaeva, S., Stiles, B., Wang, R.-H., Ohara, N., Yoshino, T., et al. 2006. Induction of intrahepatic cholangiocellular carcinoma by liver-specific disruption of smad4 and pten in mice. *The Journal of clinical investigation*, 116(7):1843–1852.

Zhang, B., Halder, S. K., Kashikar, N. D., Cho, Y.-J., Datta, A., Gorden, D. L., and Datta, P. K. 2010. Antimetastatic role of smad4 signaling in colorectal cancer. *Gastroenterology*, 138(3):969–980.

Zhang, X., Bajaj, C. L., Kwon, B., Dolinsky, T. J., Nielsen, J. E., and Baker, N. A. 2006. Application of new multiresolution methods for the comparison of biomolecular electrostatic properties in the absence of global structural similarity. *Multiscale Modeling & Simulation*, 5(4):1196–1213.

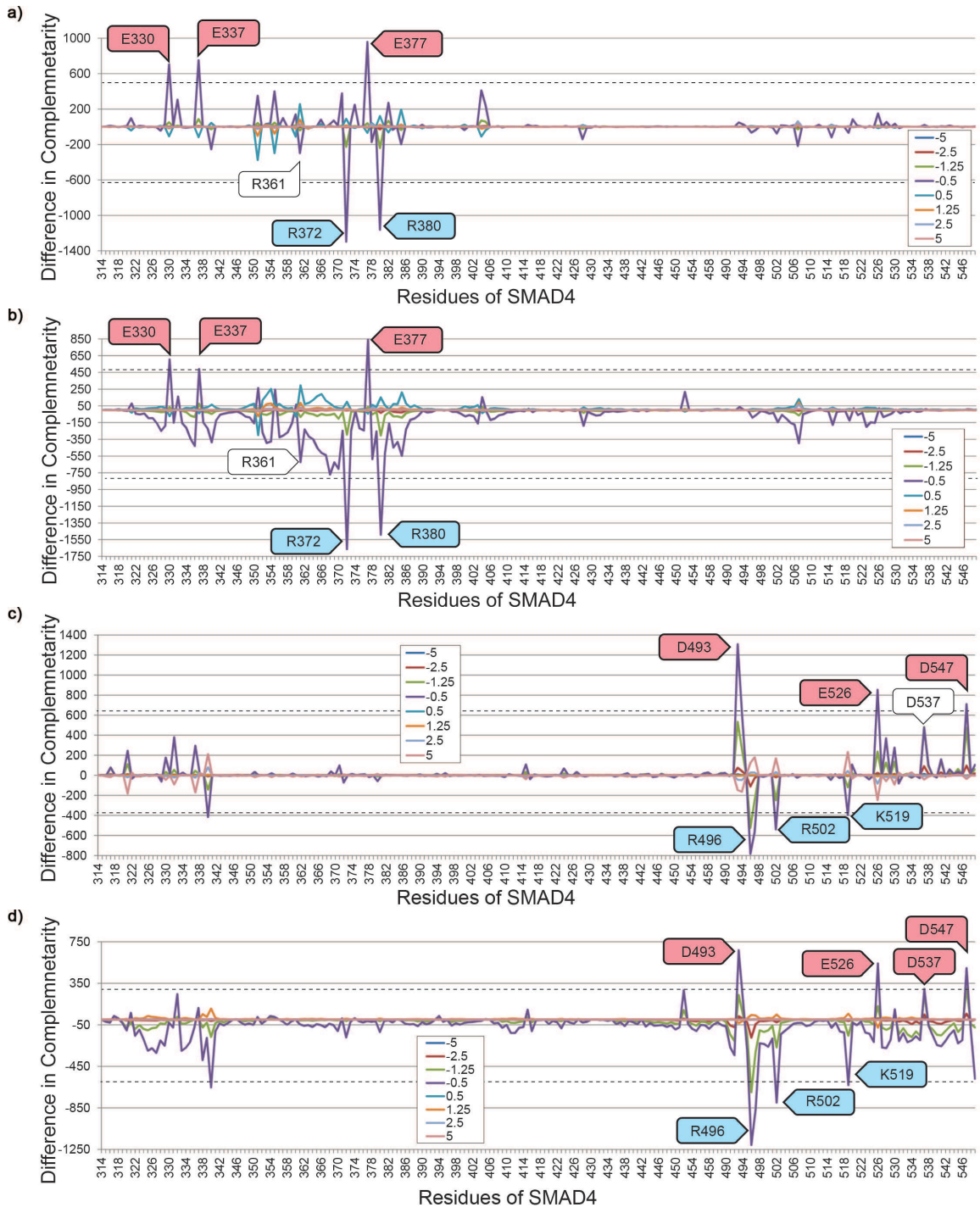

Fig. 4: Changes in electrostatic complementarity, in the BA (a,b) and BC (c,d) interfaces, caused by B subunit modifications. Horizontal axes indicate which residues of B were changed to glycine (a,c) or nullified (b,d). Vertical axes indicate the electrostatic complementarity of the unmodified complex minus electrostatic complementarity in the modified complex. Line colors correspond to potential thresholds. Dotted lines are positive and negative prediction thresholds for 0.5 kt/e and -0.5 kt/e potentials. Residues surpassing the positive threshold are labeled red, those surpassing

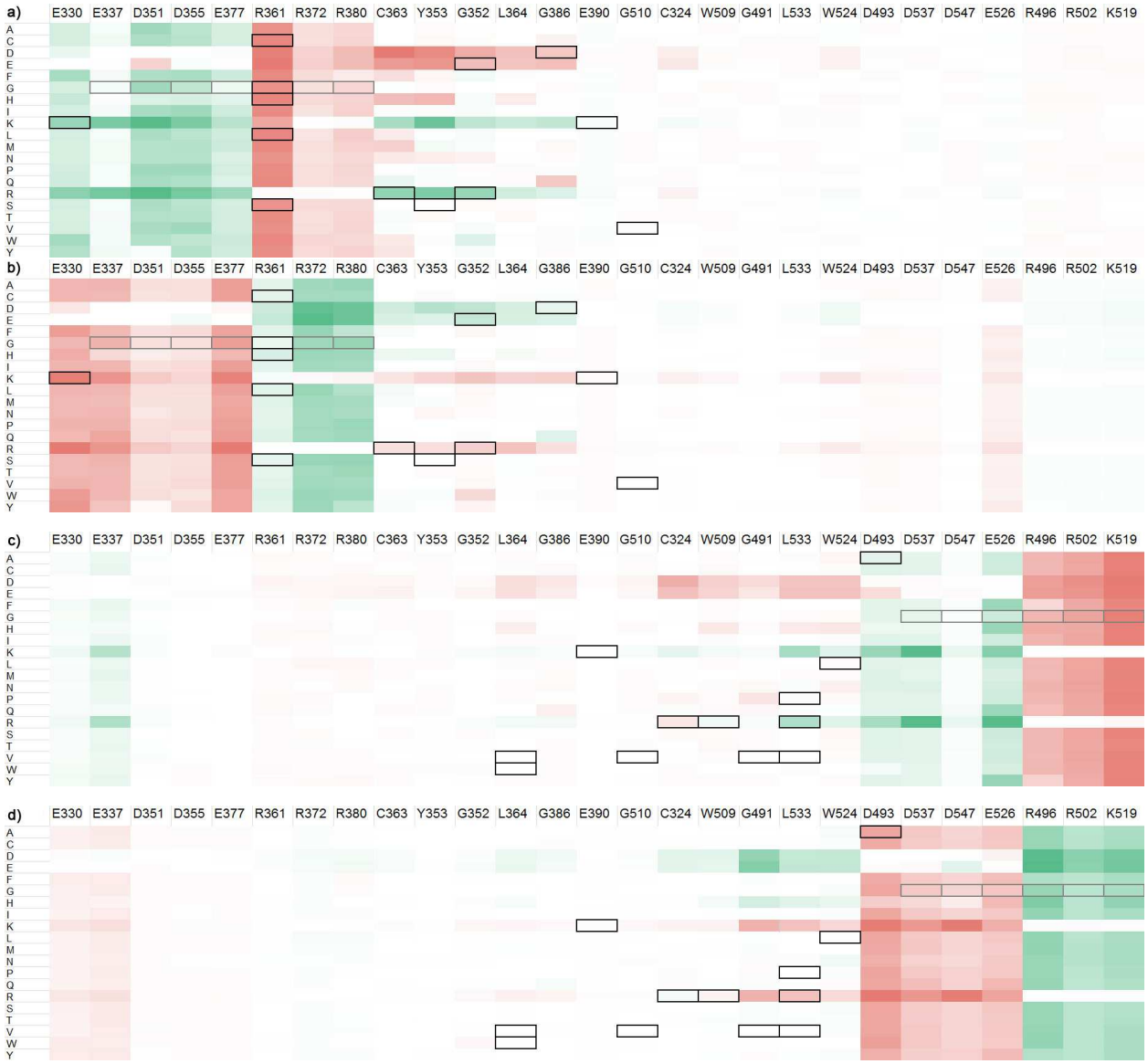

Fig. 5: Changes in electrostatic complementarity, in the BA (a,b) and BC (c,d) interfaces, caused by sidechain replacement in B. Horizontal axes indicate which residue was mutated. Vertical axes indicate what each residue was replaced with. Rectangles in each heatmap plot the degree of change in electrostatic complementarity from most negative (red) to zero (white) to most positive (green). Electrostatic complementarity was evaluated at 0.5 kt/e (a,c) and -0.5 kt/e (b,d). Black rectangles outline mutations identified by the ARUP database as associated with disease. Grey boxes outline mutations to glycine that are not specified by the ARUP database.