Amino acids that are essential for the structure and function of a protein are likely to be conserved through evolution. However, if a mutation in one of those amino acids is rescued by a complementary mutation in a different amino acid, the two mutations together also may be conserved. A recently published algorithm called evo-fold uses this evolutionary covariance in the amino acid sequence to predict 3-dimentional structures of cytosolic and membrane proteins to sub-4 Å accuracy [(Marks et al. 2012;](http://wizfolio.com/?citation=1&ver=3&ItemID=651&UserID=14332&AccessCode=4870F096D12A4BFAB925C026D374371B&CitationSuffix=) [Hopf et al. 2012)](http://wizfolio.com/?citation=1&ver=3&ItemID=650&UserID=14332&AccessCode=AD6AE867FBE7486DBBD02CFAFDD92694&CitationSuffix=). With over 30 million non-redundant protein sequences now publically available [(Suzek et al. 2007)](http://wizfolio.com/?citation=1&ver=3&ItemID=712&UserID=14332&AccessCode=0384242697DC4FFFB59B5A48AD43C3EA&CitationSuffix=), the algorithm used by evo-fold seems to lend itself to a wide range of applications. My project involves integrating the evolutionary information that can be obtained through evo-fold with the previously-developed protein design and protein-protein interface prediction algorithms, in order to improve their performance.

One application of this “meta-design” algorithm would be to improve the thermal stability of protein-based pharmaceuticals. Most proteins have a closely-matched energy profile of folding and unfolding, and this allows them to function optimally under physiological conditions [(Korkegian et al. 2005)](http://wizfolio.com/?citation=1&ver=3&ItemID=699&UserID=14332&AccessCode=5143F69ADCCD4D8CA3CE3274F7B4DF06&CitationSuffix=). However this also makes those proteins vulnerable to degradation and limits their shelf-life and stability in pharmaceutical preparations. Many approaches have been used previously to improve the thermal stability of proteins, including structure-guided protein redesign with RosettaDesign [(Korkegian et al. 2005)](http://wizfolio.com/?citation=1&ver=3&ItemID=699&UserID=14332&AccessCode=5143F69ADCCD4D8CA3CE3274F7B4DF06&CitationSuffix=), optimisation of the protein amino acid sequence with machine learning algorithms [(Li and Fang 2012)](http://wizfolio.com/?citation=1&ver=3&ItemID=702&UserID=14332&AccessCode=80CEB0ABA76845A09DBBD60CE64DFF9D&CitationSuffix=), mutation of destabilising residues selected with high-temperature molecular dynamics simulations [(Xiaoqin et al. 2011)](http://wizfolio.com/?citation=1&ver=3&ItemID=700&UserID=14332&AccessCode=CD5ABC574E4D4C7CA5E10DDF61732CCD&CitationSuffix=), and a combination of computational techniques with directed evolution and high-throughput screening [(Narasimhan et al. 2012)](http://wizfolio.com/?citation=1&ver=3&ItemID=708&UserID=14332&AccessCode=065DA8F354E440569A1767EA8D053E00&CitationSuffix=). In all of those cases, specific residues that were deemed necessary for the stability and function of the protein were left unchanged, but the selection of those residues was manual and was confined only to the well-characterised regions such as the catalytic pocket. Using the information obtained through evo-fold, we could improve on those techniques by confining the mutations only to those residues that do not show evidence of evolutionary constraint. This would both reduce the search space for potential mutations and decrease the likelihood that a mutation will disrupt the basic structure of the protein.

Another application would be in locating and re-engineering specific protein-protein interaction interfaces. Since evo-fold is able to locate intermolecular contacts for a dimeric protein [(Hopf et al. 2012)](http://wizfolio.com/?citation=1&ver=3&ItemID=650&UserID=14332&AccessCode=AD6AE867FBE7486DBBD02CFAFDD92694&CitationSuffix=), it seems likely that it also would be able to locate intermolecular contacts between different proteins [(de Juan et al. 2013)](http://wizfolio.com/?citation=1&ver=3&ItemID=686&UserID=14332&AccessCode=0C3781303EB44971B64D9E9B23D81FB2&CitationSuffix=). In the case of putative partners in a signalling cascade, this information could be used to predict the interface of interaction, to characterise the strength of interaction and to model the structure of the multi-protein complex. Protein-based pharmaceuticals then could be designed to stabilise the complex or to prevent its formation. While experimental techniques such as phage display still outperform computational approaches in producing high affinity binding, they usually are not specific to a particular region on the target. We could use evo-fold to find the evolutionary constraints between the region of interest on the target and proteins that bind to that region, and we could include those constraints in the site-directed diversification libraries to be used in phage display. This would allow us to maintain the specificity to a particular region on the target while experimentally optimising the affinity of binding.

[**References**](https://wizfolio.com/?style=1&ver=3&UserID=14332&StyleName=Medline-AuthorYear%20(Alexey%20mod%20Final))

de Juan D, Pazos F, Valencia A. Emerging methods in protein co-evolution. *Nat Rev Genet* 2013;14(4):249-261.

Hopf TA, Colwell LJ, Sheridan R, Rost B, Sander C, Marks DS. Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* 2012;149(7):1607-1621.

Korkegian A, Black ME, Baker D, Stoddard BL. Computational thermostabilization of an enzyme. *Science* 2005;308(5723):857-860.

Li Y, Fang J. PROTS-RF: a robust model for predicting mutation-induced protein stability changes. *PloS one* 2012;7(10):e47247.

Marks DS, Hopf TA, Sander C. Protein structure prediction from sequence variation. *Nat Biotechnol* 2012;30(11):1072-1080.

Narasimhan D, Woods JH, Sunahara RK. Bacterial cocaine esterase: a protein-based therapy for cocaine overdose and addiction. *Future medicinal chemistry* 2012;4(2):137-150.

Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 2007;23(10):1282-1288.

Xiaoqin H, Daquan G, Chang-Guo Z. Computational design of a thermostable mutant of cocaine esterase via molecular dynamics simulations. *Organic & Biomolecular Chemistry* 2011;9(11)