BioRxiv
15 September 2020

Computationally Grafting An IgE Epitope Onto A Scaffold

Sari Sabban[1]*

Abstract
Due to the increased hygienic life style of the developed world allergy is an increasing disease. Once allergy develops, sufferers are permanently trapped in a hyper immune response that makes them sensitive to innocuous substances. This paper discusses the strategy and protocol employed which designed proteins displaying a human IgE motif very close in proximity to the IgE's FcεRI receptor binding site. The motif of interest was the FG motif and it was excised and grafted onto the protein scaffold 1YN3. The new structure (scaffold + motif) was fixed-backbone sequence designed around the motif to find an amino acid sequence that would fold to the designed structure correctly. Ten computationally designed proteins showed successful folding when simulated using the AbinitioRelax folding simulation and the IgE epitope was clearly displayed in its native three dimensional structure in all of them. Such a designed protein has the potential to be used as a pan anti-allergy vaccine by guiding the immune system towards developing antibodies against this strategic location on the body's own IgE molecule, thus neutralising it and presumably permanently shutting down a major aspect of the Th_2 immune pathway.

Keywords
Protein Design — Epitope Grafting — Vaccine Design — Computational Structural Biology — Allergy — Type I Hypersensitivity

[1] *Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Makka, Kingdom of Saudi Arabia*
***Corresponding author**: sari.sabban@gmail.com

Introduction

Allergy was first defined by Clemens von Pirquet in 1906 when he discovered that second injections of horse serum caused a severe inflammatory reaction in some, but not all, individuals. He termed this condition allergy, from the Greek words allos "other" and ergon "works" and, therefore, the allergy causing agent an "allergen" [21]. In the 1960s Kimishige Ishizaka and Teruko Ishizaka demonstrated that allergic reactions are mediated by a new class of antibodies that they discovered and called immunoglobulin E [6] [11].

Humans have five antibodies; immunoglobulin G (IgG) is the most abundant type since it targets viral and bacterial pathogens. Immunoglobulin E (IgE) on the other hand, is concerned with parasitic immunity. Since parasites are eukaryotes and closer to other eukaryotes phylogenically, compared to bacteria and viruses, this pathway can target innocuous substances, that look like parasites but are not usually harmful, leading to a type of inflammatory reaction termed an allergic reaction, or known medically as type I hypersensitivity. IgE binds onto its high-affinity Fc receptor (FcεRI) that is found on mast and basophil cells.

Thus IgE antibodies are best known for their role as mediators of the allergic response, which in its most serious manifestations, causes asthma or an anaphylactic shock. Reports of an increase in the number of individuals suffering from allergic manifestations began in the second half of the last century and the incidence of allergy has now reached pandemic proportions [20]. IgE mediated allergic responses have diverse manifestations, which range from mild to severe and can be life threatening. Mammals including humans, dogs, and horses are known to suffer the clinical symptoms of IgE-mediated type I hypersensitivity responses. In spite of extensive worldwide research efforts, no effective active therapeutic intervention strategies are currently available.

One of the perceived reasons for the continual increase in allergy incidence, especially in the developed world, is a hypothesis termed the Hygiene Hypothesis, originally formulated by Strachan [28] [29] [19], it states that a lack of exposure to infectious pathogens in early childhood, i.e. living in too clean of an environment, can lead to inadequate immune system development, i.e. a shift from the Th_1 immune response (bacteria, viruses) to that of the Th_2 immune response (parasite, allergy), resulting in an increase in susceptibility to develop allergy. Further studies in this immunological pathway has shed light into the viability of this hypothesis, and showed a correlation between tuberculosis infections in childhood and lack of allergy in adulthood [30].

The most widely used therapy against allergy is pharmacotherapy, which is a passive immunotherapeutic intervention strategy, employing the use of anti-histamines, corticosteroids, or epinephrine, all of which alleviate the symptoms of al-

lergy without curing its underlying cause. The quest to treat allergy is not a new concept, it was first attempted in 1911 [18] in which subcutaneous injections of an allergen extract were administered in an effort to desensitise atopic patients to certain allergens (Homeopathy). This was successful to treat certain allergies such as anaphylaxis and allergic rhinitis, but it was unsuccessful in treating asthma [16]. This protocol has remained controversial as it has the potential to sensitise patients even more, thus worsening their condition [17] A novel immunotherapy called sublingual immunotherapy is currently being researched where allergen extracts are given to patients under their tongue [9]. The efficacy of these therapies varies greatly between individuals since doctors do not have a standard protocol to follow, they usually develop their own protocols according to their own observations and individual successes.

Since allergy incidence have been on the rise globally, a new form of therapy is under development. Though still a passive immunetherapeutic strategy, it employs non-anaphylactogenic antibodies which have demonstrated their capacity to treat type I hypersensitivity responses. These humanised mouse monoclonal antibodies (mAbs), of which Omalizumab [22] is best characterised, are now successful but have been shown to be associated with a number of drawbacks: 1) poor effectiveness in obese patients, 2) logistics and cost, 3) treatment only reduces symptoms temporarily, hence it is a passive immunetherapeutic strategy.

These draw backs logically lead to the potential to develop a new active form of immunetherapeutic strategies, where a vaccine primes the immune system against its own IgE antibody, at which point the IgE is neutralised and the allergy disease is terminated. Even though current mainstream research is concentrating on the passive immunisation approach, it is believed that active immunisation is a viable form of treatment against this disease (figure 1).

This paper proposes a strategy by which a pan-anti-allergy vaccine can be computationally designed by excising the motif of interest from the IgE structure and grafting in onto a scaffold protein structure, thus displaying only the motif of interest in its original three dimensional form without any of the surrounding native structure, allowing the immune system to target that particular motif only.

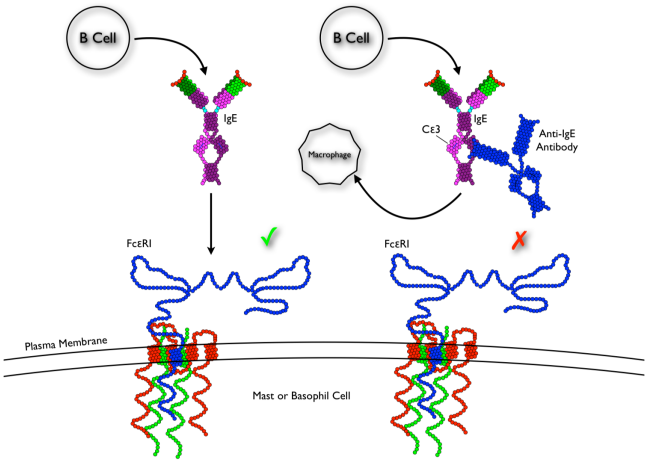


Figure 1. Summery of the pan-anti-allergy vaccine therapy concept. Administering a vaccine that is capable of producing antibodies against the body's own IgE molecule. This will neutralise the IgE molecule, thus disrupting the entire allergy pathway, potentially curing the disease [24].

Methods

The following steps were used to generate a database of scaffold structures as well as isolate the IgE motif, graft it onto a scaffold, then design the scaffold to fold onto the designed structure.

Motif determination and excision

The FG motif (2Y7Q chain B 420-429 VTHPHLPRAL) was chosen due to its very close proximity to the receptor binding site (2Y7Q 331-338 chain B SNPRGVSA). The motif has a ridged structure (since the motif looks like a heart shape with an anchoring 425 lysine pointing into the core fixing its shape). Since the receptor binding site (named R) was not anchored and had a higher degree of movement its surrounding area was not well modelled in the crystal structure, thus it was not chosen. The FG motif was isolated along with the receptor (2Y7Q chain A) as separate files in preparation for grafting.

Scaffold database generation

The scaffold database was generated by downloading the entire PDB database, then isolating only protein structures and separating each chain into separate .pdb files. Each structure was cleaned (removed of any none-peptide atoms) then passed through (scored) by the Rosetta modelling software [15] to make sure each structure will not crash the software. Structures that were not satisfactory were discarded. A script was developed that would generate a better, smaller, and more targeted database, but was not used here, nonetheless is available.

Motif grafting

The desired motif between positions 420 and 429 in the 2Y7Q chain B protein was isolated along with the receptor in chain A then a grafting search was performed that matched the backbone of the motif to backbones within the database, if there was a match within an RMSD value of 1.0 88 Å or less the motif was grafted onto the scaffold structure (replacing the original backbone) and measured for its clash with the receptor (i.e: to make sure the backbone was not grafted inward or was buried within the structure). This protocol was developed by [2] [1].

Selective fixed-backbone sequence design

The final structure was tested for folding, which failed as predicted, thus some human guided mutations were employed to push the structure to fold onto its designed structure. After many failed attempts the fixed-backbone design protocol was

employed [12] [7] [14] [10] [13], where the side chain (sequence) of the structure is stochastically mutated and packed using a rotated library to find the lowest energy structure that would fold into the designed structure. In this protocol we changed that REF2015 energy function weights to include aa_rep 1.0, aspartimid_penalty 1.0, buried_unsatisfied_penalty 1.0, and approximate_buried_unsat_penalt 5.0 which assisted in designing an adequate sequence that fits the backbone structure and increasing the energy gap between the desired structure and any other possible undesired fold.

Folding simulation

To get insight into whether the design process was successful, the structures were simulated for their folding using the AbinitioRelax protocol [23] [5] [3] [4] [26] [27], where the sequence is folded using first principals and some statistical weights through the REF2015 scoring function [31] which uses the following equation (details of which are explained in the original paper):

$$\Delta E_{total} = \sum_i w_i E_i(\Theta_i, aa_i)$$

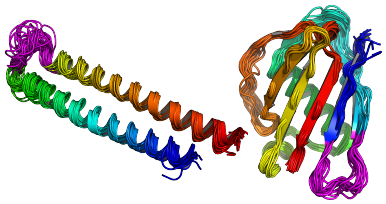
To reduce the folding space and speed up the search for the global minima fragment were developed from the FASTA sequence, were backbone torsion angles are statistically analysed and inserted to help the algorithm fold the structure.

Results

Analysis of the motif position revealed that the R loop and the FG loop from the human IgE (2Y7Q) are the best candidates for a targeted vaccine due to their proximity to the binding site on the α chain of the FcεRI receptor (figure 2). After several attempts at grafting and designing the R loop on 3Q4H replacing the sequence QGDTGMTY at positions 44-51, the FG loop seemed to be the better choice, this was due to the FG loop having an inward pointing leucine resulting in a ridged loop structure, compared with the R loop that had a high degree of angle freedom which resulted in a wide range of different structures when grafted, see figure 3.

The scaffold search algorithm resulted in the FG loop motif being grafted onto the 1YN3 [8] structure as well as several other structures (figure 4). The 1YN3 structure was chosen since it had a backbone that was easily simulated by forward folding using the AbinitioRelax protocol (figure 5). Another reason was that the 1YN3 protein is an EAP domains from *Staphylococcus aureus* which was expressed in *Escherichia coli* when it was crystallised, thus it is predicted to easily crystallise for final structural evaluation. The motif was grafted between positions 164 and 173 on the 1YN3 structure replacing the sequence ITVNGTSQNI with VTHPHLPRAL (figure 6). As predicted the freshly grafted structure failed a forward fold using AbinitioRelax, this was because the addition of the motif backbone and side chains severely disrupted the stability of the entire structure. To overcome this the entire structure

Figure 2. The structure of the human IgE bound to its FcεRIα receptor (2Y7Q). The colours show the different loops that are closet in proximity or forms hydrogen bonds with the receptor when bound. Purple for the FG loop, green for the R loop, orange for the BC loop, and yellow for the DE loop.



(a) (b)

Figure 3. Comparison of grafting the R motif to grafting the FG motif. A: One of the structures (3Q4H) that successfully grafted the R loop motif (in purple), showing the large variability of the motif backbone since it lacked an anchor (average RMSD = 1.29 88 Å to the natives motif). **B:** One of the structures (1YN3) that successfully grafted the FG loop motif (in purple) showing better motif stability (average RMSD = 0.62 88 Å to the natives motif). Structures rendered through PyMOL [25]

was sequence designed by changing and optimising the side chains (except for the motif) while fixing the backbone to stabilise the structure and accommodate the new motif backbone and side chains. Since there exists a failure rate between a successful forward fold and a successful crystal structure, the sequence design step was repeated ten times, this resulted in ten structures all of which had a successful forward fold, see figure 7, this should increase that probability of synthesising a correctly folded vaccine structure since only one of these structures must pass a crystallography evaluation to be tested as a potential vaccine.

All structures were predicted to fold within a sub angstrom level of the designed structure, giving high confidence that these will be the structures of the proteins when synthesised biologically. Each structure must be crystallised to definitively confirm the correct fold of the protein and the motif before they are tested on animals.

Conclusion

This paper communicates the protocol for computationally designing proteins that correctly display the three dimensional structure of the FG strategic motif of the IgE molecule, where the motif was grafted onto the 1YN3 scaffold protein, then the scaffold/motif was sequence designed resulting in ten struc-

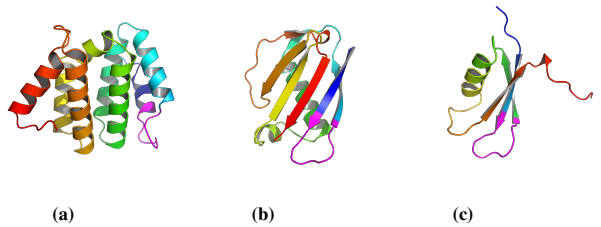


Figure 4. Grafting the FG motif onto three different scaffolds. This figure is showing three structures that successfully grafted the FG loop **A:** 3LDZ chain A where the ATSEMNTAED sequence at positions 16-25 was replaced by the FG motif. **B:** 1YN3 chain A where the sequence ITVNGTSQNI at positions 164-173 was replaced by the FG sequence. **C:** 3EGR chain B where the VRSKQGLEHK sequence at positions 13-22 was replaced by the FG motif. 1YN3 was chosen since the native structure was easily forward folded using AbinitioRelax (figure 5). Thus it was be easier to redesign this structure and it did result is a structure with a large energy gap between the desired structure and any other potential structure.

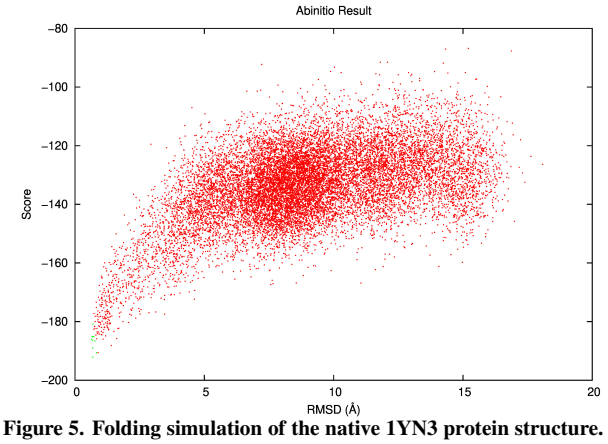


Figure 5. Folding simulation of the native 1YN3 protein structure. AbinitioRelax result of the native 1YN3 protein showing a successful simulation, a funnel shaped plot with the lowest simulated energy close to the predicted energy and RMSD of the structure.

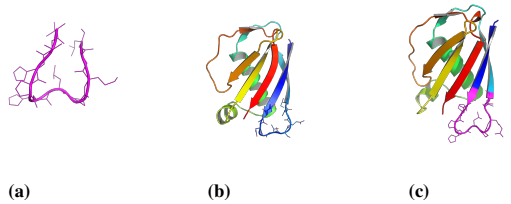


Figure 6. Stages of the grafting protocol. A: The structure of the motif. **B:** The structure of 1YN3 before grafting the motif and **C:** after grafting the motif.

tures, opening the possibility of using such protein structures as a vaccine against self-IgE and permanently shutting down the allergy pathway regardless of the offending allergen (a pan-anti-allergy vaccine). The resulting structures showed agreement in their final folds when simulated with the Rosetta AbinitioRelax folding algorithm, yet the only definitive way to determine their realistic folds is to solve their structures through X-Ray crystallography. Furthermore, the efficacy of the proteins in pushing the immune system into developing antibodies against self-IgE at a higher binding affinity than

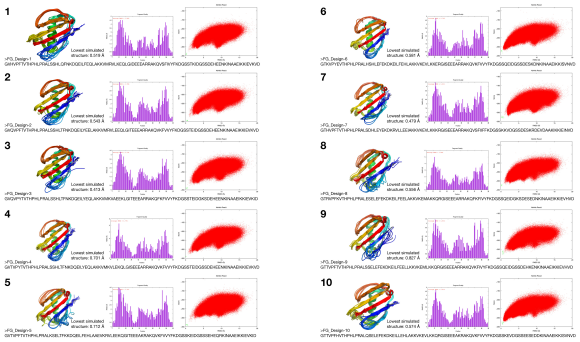


Figure 7. Final designed structures. Ten designed structures that display the FG loop in its native three dimensional structure. The figure shows each designed structure (cartoon) superimposed onto the lowest energy and RMSD structures from the AbinitioRelax simulation (wire) and the corresponding lowest RMSD value of the simulation, thus all structures were predicted to fold within a sub angstrom level of the designed structure. Also showing are the FASTA sequences of each structure, the fragment quality used in each AbinitioRelax simulation, and the AbinitioRelax plot showing a successful funnel shaped plot for all structures. The green points in each folding simulation are the REF2015 (Rosetta Energy Function 2015) energy score values of the corresponding computationally designed structure after being relaxed thus indicating the lowest possible energy score for each structure.

IgE/FcεRI receptor's binding affinity could not be computationally simulated, and thus must be tested on animals to reach a definitive answer. The script that was used to design these proteins is available at this GitHub repository which includes an extensive README file and a video that explains how to use it.

Grant information

The authors declared that no grants were involved in supporting this work.

Competing interests

The author has used these results to apply for a patent.

Acknowledgements

The corresponding author would like to thank the High Performance Computing Center at King Abdulaziz University for making available the Aziz high performance computer where the corresponding author was able to perform the Epitope Grafting search and the AbinitioRelax folding simulations.

References

[1] Mihai L. Azoitei, Yih-En Andrew Ban, Jean-Philippe Julien, Steve Bryson, Alexandria Schroeter, Oleksandr Kalyuzhniy, Justin R. Porter, Yumiko Adachi, David Baker, Emil F. Pai, and William R. Schief. Computational design of high-affinity epitope scaffolds by backbone grafting of a linear epitope. *Journal of Molecular Biology*, 415(1):175 – 192, 2012.

[2] Mihai L. Azoitei, Bruno E. Correia, Yih-En Andrew Ban, Chris Carrico, Oleksandr Kalyuzhniy, Lei Chen, Alexandria Schroeter, Po-Ssu Huang, Jason S. McLellan, Peter D. Kwong, David Baker, Roland K. Strong,

and William R. Schief. Computation-guided backbone grafting of a discontinuous motif onto a protein scaffold. *Science*, 334(6054):373–376, 2011.

[3] Richard Bonneau, Charlie E.M Strauss, Carol A Rohl, Dylan Chivian, Phillip Bradley, Lars Malmström, Tim Robertson, and David Baker. De novo prediction of three-dimensional structures for major protein families. *Journal of Molecular Biology*, 322(1):65 – 78, 2002.

[4] Richard Bonneau, Jerry Tsai, Ingo Ruczinski, Dylan Chivian, Carol Rohl, Charlie E. M. Strauss, and David Baker. Rosetta in casp4: Progress in ab initio protein structure prediction. *Proteins: Structure, Function, and Bioinformatics*, 45(S5):119–126, 2001.

[5] Philip Bradley, Kira M. S. Misura, and David Baker. Toward high-resolution de novo structure prediction for small proteins. *Science*, 309(5742):1868–1871, 2005.

[6] Martin D Chapman. Allergens. *Elsevier*, 1998.

[7] Gautam Dantas, Brian Kuhlman, David Callender, Michelle Wong, and David Baker. A large scale test of computational protein design: Folding and stability of nine completely redesigned globular proteins. *Journal of Molecular Biology*, 332(2):449 – 460, 2003.

[8] Brian V Geisbrecht, Brent Y Hamaoka, Benjamin Perman, Adam Zemla, and Daniel J Leahy. The crystal structures of eap domains from staphylococcus aureus reveal an unexpected homology to bacterial superantigens. *Journal of Biological Chemistry*, 280(17):17243–17250, 2005.

[9] G. B. Gidaro, F. Marcucci, L. Sensi, C. Incorvaia, F. Frati, and G. Ciprandi. The safety of sublingual-swallow immunotherapy: an analysis of published studies. *Clinical & Experimental Allergy*, 35(5):565–571, 2005.

[10] Xiaozhen Hu, Huanchen Wang, Hengming Ke, and Brian Kuhlman. High-resolution design of a protein loop. *Proceedings of the National Academy of Sciences*, 104(45):17668–17673, 2007.

[11] Kimishige Ishizaka, Teruko Ishizaka, and Margaret M. Hornbrook. Physico-chemical properties of human reaginic antibody. *The Journal of Immunology*, 97(1):75–85, 1966.

[12] Brian Kuhlman, Gautam Dantas, Gregory C. Ireton, Gabriele Varani, Barry L. Stoddard, and David Baker. Design of a novel globular protein fold with atomic-level accuracy. *Science*, 302(5649):1364–1368, 2003.

[13] A. Leaver-fay, B. Kuhlman, J. Snoeyink, Andrew Leaver-fay, Brian Kuhlman, and Jack Snoeyink. An adaptive dynamic programming algorithm for the side chain placement problem. In *In Pacific Symposium on Biocomputing*, pages 17–28. World Scientific, 2005.

[14] Andrew Leaver-Fay, Brian Kuhlman, and Jack Snoeyink. Rotamer-pair energy calculations using a trie data structure. In Rita Casadio and Gene Myers, editors, *Algorithms in Bioinformatics*, pages 389–400, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[15] Andrew Leaver-Fay, Michael Tyka, Steven M. Lewis, Oliver F. Lange, James Thompson, Ron Jacak, Kristian W. Kaufman, P. Douglas Renfrew, Colin A. Smith, Will Sheffler, Ian W. Davis, Seth Cooper, Adrien Treuille, Daniel J. Mandell, Florian Richter, Yih-En Andrew Ban, Sarel J. Fleishman, Jacob E. Corn, David E. Kim, Sergey Lyskov, Monica Berrondo, Stuart Mentzer, Zoran Popović, James J. Havranek, John Karanicolas, Rhiju Das, Jens Meiler, Tanja Kortemme, Jeffrey J. Gray, Brian Kuhlman, David Baker, and Philip Bradley. Chapter nineteen - rosetta3: An object-oriented software suite for the simulation and design of macromolecules. In Michael L. Johnson and Ludwig Brand, editors, *Computer Methods, Part C*, volume 487 of *Methods in Enzymology*, pages 545 – 574. Academic Press, 2011.

[16] David B Lewis. Allergy immunotherapy and inhibition of th2 immune responses: a sufficient strategy? *Current Opinion in Immunology*, 14(5):644 – 651, 2002.

[17] R. Movérare, L. Elfman, E. Vesterinen, T. Metso, and T. Haahtela. Development of new ige specificities to allergenic components in birch pollen extract during specific immunotherapy studied with immunoblotting and pharmacia cap system™. *Allergy*, 57(5):423–430, 2002.

[18] L. Noon. Prophylactic inoculation against hay fever. *Lancet*, 177:1572–1573, 1911.

[19] H. Okada, C. Kuhn, H. Feillet, and J.-F. Bach. The 'hygiene hypothesis' for autoimmune and allergic diseases: an update. *Clinical & Experimental Immunology*, 160(1):1–9, 2010.

[20] Ruby Pawankar, GW Canonica, ST Holgate, RF Lockey, and M Blaiss. World allergy organization (wao) white book on allergy. *Wisconsin: World Allergy Organisation*, 2011.

[21] C. Von Pirquet. MÜnchener medizinische wochenschrift. *Allergie*, 1909.

[22] Leonard G Presta, SJ Lahr, RL Shields, JP Porter, CM Gorman, BM Fendly, and PM Jardieu. Humanization of an antibody directed against ige. *The Journal of Immunology*, 151(5):2623–2632, 1993.

[23] Srivatsan Raman, Robert Vernon, James Thompson, Michael Tyka, Ruslan Sadreyev, Jimin Pei, David Kim, Elizabeth Kellogg, Frank DiMaio, Oliver Lange, et al. Structure prediction for casp8 with all-atom refinement using rosetta. *Proteins: Structure, Function, and Bioinformatics*, 77(S9):89–99, 2009.

[24] Sari Sabban. *Development of an in vitro model system for studying the interaction of Equus caballus IgE with its high-affinity FcεRI receptor*. PhD thesis, University of Sheffield, 2011.

[25] LLC Schrödinger. The pymol molecular graphics system, version 1.8. *Schrödinger, LLC New York*, 2015.

[26] Kim T Simons, Charles Kooperberg, Enoch Huang, and David Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *Journal of molecular biology*, 268(1):209–225, 1997.

[27] Kim T Simons, Ingo Ruczinski, Charles Kooperberg, Brian A Fox, Chris Bystroff, and David Baker. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins: Structure, Function, and Bioinformatics*, 34(1):82–95, 1999.

[28] David P Strachan. Hay fever, hygiene, and household size. *BMJ: British Medical Journal*, 299(6710):1259, 1989.

[29] David P Strachan. Family size, infection and atopy: the first decade of the'hygiene hypothesis'. *Thorax*, 55(Suppl 1):S2, 2000.

[30] Leena von Hertzen, T Klaukka, H Mattila, and T Haahtela. Mycobacterium tuberculosis infection and the subsequent development of asthma and allergic conditions. *Journal of allergy and clinical immunology*, 104(6):1211–1214, 1999.

[31] Alford, Rebecca F. and Leaver-Fay, Andrew and Jeliazkov, Jeliazko R. and O'Meara, Matthew J. and DiMaio, Frank P. and Park, Hahnbeom and Shapovalov, Maxim V. and Renfrew, P. Douglas and Mulligan, Vikram K. and Kappel, Kalli and Labonte, Jason W. and Pacella, Michael S. and Bonneau, Richard and Bradley, Philip and Dunbrack, Roland L. and Das, Rhiju and Baker, David and Kuhlman, Brian and Kortemme, Tanja and Gray, Jeffrey J. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design *Journal of Chemical Theory and Computation*, 13(6):3031-3048, 2017