



Lang, E. J. M., Baker, E. G., Woolfson, D. N., & Mulholland, A. J. (2021). *Implicit solvent models fail to reproduce secondary structures of de novo designed peptides*. (ChemRxiv). ChemRxiv. Advance online publication. <https://doi.org/10.26434/chemrxiv-2021-1x2zl>

Early version, also known as pre-print

License (if available):  
CC BY-NC-ND

Link to published version (if available):  
[10.26434/chemrxiv-2021-1x2zl](https://doi.org/10.26434/chemrxiv-2021-1x2zl)

[Link to publication record on the Bristol Research Portal](#)  
PDF-document

This is the submitted manuscript (SM). It first appeared online via ChemRxiv at [10.26434/chemrxiv-2021-1x2zl](https://doi.org/10.26434/chemrxiv-2021-1x2zl). Please refer to any applicable terms of use of the publisher.

## University of Bristol – Bristol Research Portal

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

# Implicit solvent models fail to reproduce secondary structures of *de novo* designed peptides

Eric J. M. Lang<sup>1,2,\*</sup>, Emily G. Baker<sup>1,2,3</sup>, Derek N. Woolfson<sup>1,2,3</sup>, Adrian J. Mulholland<sup>1,2\*</sup>

<sup>1</sup>School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK.

<sup>2</sup>BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ,

UK. <sup>3</sup>School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol, BS8 1TD, UK.

**KEYWORDS:** computational protein design; *de novo* peptides; generalized Born model; implicit solvent; molecular dynamics simulations.

**ABSTRACT:**

We test a range of standard implicit solvent models and protein forcefields for a set of 5 experimentally characterized, designed  $\alpha$ -helical peptides. 65 combinations of forcefield and implicit solvent models are evaluated in >800  $\mu$ s of molecular dynamics simulations. The data show that implicit solvent models generally fail to reproduce the experimentally observed secondary structure content, and none performs well for all 5 peptides. The results show that these models are not usefully predictive.

## MAIN TEXT

The use of molecular dynamics (MD) simulations in aqueous media is established in structural biology.<sup>1</sup> There is a growing need for the application of such methods in the rapidly advancing fields of protein structure prediction and *de novo* protein design. However, in both cases, many model structures have to be evaluated, which is computationally expensive. For example, in protein design, MD simulations are often currently limited to short backbone-restrained simulations to improve side-chain packing.<sup>2,3</sup> Computational protein design would benefit from incorporating longer unrestrained MD simulations routinely to test the stabilities of designed structures and to improve understanding of sequence-to-structure/function principles.<sup>4-6</sup> An attractive, practical solution is to use implicit solvent models. These have the benefit of including dielectric shielding and other effects of aqueous solvation whilst avoiding the computational cost of explicit representations of large numbers of water molecules.<sup>7,8</sup>

Although explicit solvent simulations are accepted as being more accurate, currently they are too expensive computationally for routine use in computational protein design pipelines that generate many constructs. Simulations using implicit solvent—such as those based on generalized Born (GB) models—are faster and easier to set up and analyze. Moreover, because protein dynamics are not damped by solvent viscosity, conformational space sampling is accelerated.<sup>9,10</sup> Balanced against this, many studies have highlighted limitations of implicit solvent models: e.g., different GB model–forcefield combinations can lead to very different results, with many combinations unable to reproduce native folds.<sup>11-13</sup> Nonetheless, recent work suggests that the latest implicit solvent models and forcefields have improved accuracy and reproduce the observed structures for a test set of small proteins.<sup>14</sup>

Systematically designed *de novo* peptides provide well-defined and comparable test set for examining the performance of solvent models in MD simulations. Previously, we designed and characterized a series of *de novo* peptides to explore local and macroscopic electrostatic interactions in single  $\alpha$  helices.<sup>15</sup> These comprise blocks of charged glutamic acid (Glu, E) and lysine (Lys, K) residues in sequences of the type  $(E_xK_x)_n$  or  $(K_xE_x)_n$ , Table 1. Changes in peptide length and the order and sizes of the blocks result in different experimental  $\alpha$ -helical content. *N.B.* sequences with  $(E_4K_4)_n$  and similar repeats occur in natural proteins and form single  $\alpha$ -helical (SAH) domains.<sup>16</sup> Thus, interest in such sequences is not limited to *de novo* design. Here, we simulate 5 *de novo* sequences (Table 1) using 5 different GB models and 13 different forcefields, all included in the AMBER16 biomolecular simulation program<sup>17</sup>.

**Table 1. *De novo* peptides<sup>a</sup> used for testing different forcefield-GB model combinations.**

Peptide	Sequence <sup>b</sup>	Length (residues)	Experimental $\alpha$ helicity (%) <sup>c</sup>
A <sub>4</sub> (K <sub>4</sub> E <sub>4</sub> ) <sub>1</sub> A <sub>4</sub> (K <sub>4</sub> E <sub>4</sub> ) <sub>1</sub> A <sub>4</sub>	Ac-GAAAAK <b>KKK</b> E <b>EEEE</b> AAAAK <b>KKK</b> E <b>EEEE</b> AAAAGW-NH <sub>2</sub>	31	97
(E <sub>4</sub> K <sub>4</sub> ) <sub>2</sub>	Ac-G <b>EEEE</b> K <b>KKK</b> E <b>EEEE</b> K <b>KKK</b> GW-NH <sub>2</sub>	19	65
(K <sub>4</sub> E <sub>4</sub> ) <sub>2</sub>	Ac-G <b>KKK</b> K <b>EEEE</b> E <b>KKK</b> K <b>EEEE</b> GW-NH <sub>2</sub>	19	22
(E <sub>4</sub> K <sub>4</sub> ) <sub>3</sub>	Ac-G <b>EEEE</b> K <b>KKK</b> E <b>EEEE</b> K <b>KKK</b> E <b>EEEE</b> K <b>KKK</b> GW-NH <sub>2</sub>	27	74
(K <sub>4</sub> E <sub>4</sub> ) <sub>3</sub>	Ac-G <b>KKK</b> K <b>EEEE</b> E <b>KKK</b> K <b>EEEE</b> E <b>KKK</b> E <b>EEEE</b> GW-NH <sub>2</sub>	27	62

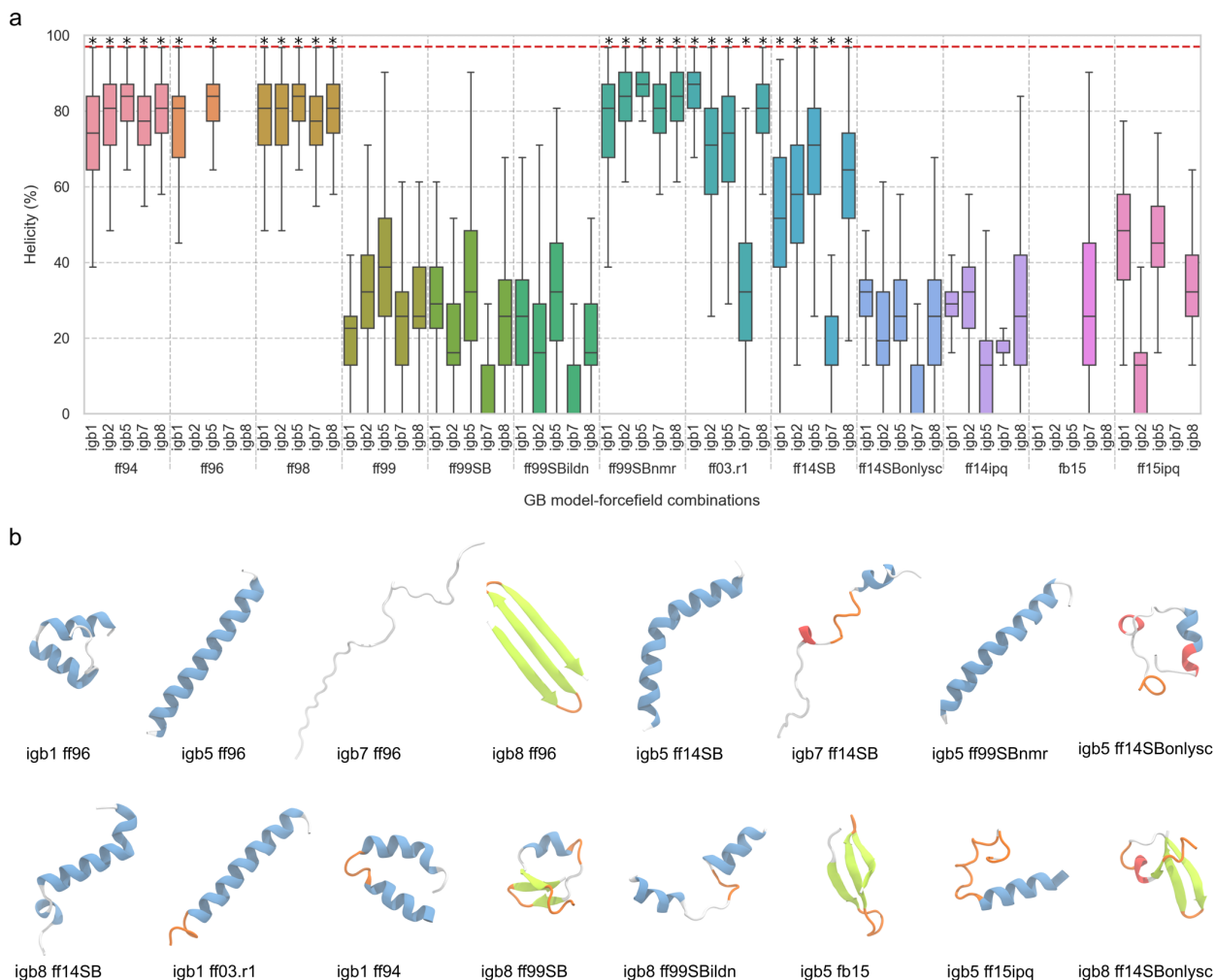
<sup>a</sup>Previously characterized experimentally by Baker *et al.*<sup>15</sup> The same naming convention is used here. <sup>b</sup>Peptides were capped with acetyl (Ac) and amide (NH<sub>2</sub>) groups. Lys residues are colored blue and Glu red. <sup>c</sup>Experimentally determined  $\alpha$  helicities from circular dichroism (CD) spectroscopy measurements at 5 °C in phosphate-buffered saline, pH 7.4.<sup>15</sup>

To test implicit solvent models in MD simulations, we applied the 5 GB models available in AMBER: the Hawkins, Cramer, Truhlar model<sup>18</sup> (igb1); the Onufriev, Bashford, Case model<sup>19,20</sup> (igb2) and its modified version<sup>20</sup> (igb5); and the GBn model by Mongan, Simmerling, McCammon, Case and Onufriev<sup>21</sup> (igb7) and its modified version by Nguyen, Roe and Simmerling<sup>22</sup> (igb8). In

combination with these, thirteen AMBER forcefields were tested with these GB models: ff94,<sup>23</sup> ff96,<sup>24</sup> ff98,<sup>25</sup> ff99,<sup>26</sup> ff99SB,<sup>27</sup> ff99SBildn,<sup>28</sup> ff99SBnmr,<sup>29</sup> ff03.r1,<sup>30</sup> ff14SB,<sup>31</sup> ff14SBonlysc,<sup>31</sup> ff14ipq,<sup>32</sup> fb15,<sup>33</sup> and ff15ipq.<sup>34</sup>

The choice of GB models and the AMBER suite was made for the following reasons. Firstly, with the exception of CHARMM,<sup>35</sup> one or more of these GB models have been implemented in all of the major biomolecular MD packages: all 5 in AMBER and OpenMM; igb1, igb2 and igb5 in GROMACS; and igb2 in NAMD. Secondly, the most recent GB model, igb8, used in combination with the ff14SBonlysc forcefield was developed in AMBER. Finally, the GPU MD engine pmemd.cuda<sup>36</sup> within AMBER is fast and ideal for rapid testing of multiple GB model–forcefield combinations.

Conditions from the experimental study<sup>15</sup> were replicated as closely as possible: the peptides were *N*-terminally acetylated and *C*-terminally amidated; simulations were run at 278.15 K; the ionic concentration was set to 0.137 mol/L; and Glu and Lys side chains fully ionized, *i.e.* negatively and positively charged, respectively, yielding neutral peptides. Initially, single 6  $\mu$ s simulations were run for each GB model–forcefield combination, with some systems repeated four times to test reproducibility (see below). All simulations used the same minimization, heating and equilibration protocol (see Methods in Supporting Information), starting from a fully helical structure of each peptide created using Chimera.<sup>37</sup> The first 250 ns were discarded as further equilibration, yielding 5.75  $\mu$ s of production MD for each. A total of 5750 frames (saved every 1 ns) were analyzed.



**Figure 1.** Predicted  $\alpha$  helicities and final structures from MD simulations of  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$  using 65 GB model–forcefield combinations. This peptide has been shown experimentally to be 97%  $\alpha$  helical.<sup>15</sup> **(a)** Percentage helicity of  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$  calculated with DSSP<sup>38</sup> for the MD trajectories generated for all 65 combinations. The results are presented as boxplots with the boxes indicating the first quartile, the median and the third quartile of the sample. The whiskers indicate 1.5 times the interquartile range. Gaps correspond to combinations for which the percentage  $\alpha$  helicity is negligible. For each combination, 5750 frames were analyzed. Each forcefield is represented with a different color. The red dotted line represents the experimental helicity. Asterisks indicate the combinations selected for further

testing. **(b)** Backbone structures of  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$  from selected GB model–forcefield combinations after 6  $\mu$ s of MD. The peptides are colored by structure:  $\alpha$  helix, blue; extended  $\beta$ -strand and  $\beta$  bridge, green;  $\pi$  helix, red;  $3_{10}$  helix, purple; turn, orange; coil, white.

We began our study with peptide  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$  because this is the most helical experimentally (97%, Table 1). The helicities and structures from the MD simulations of this peptide for each GB model–forcefield combination are shown in Figure 1.

None of the GB model–forcefield combinations gave fully helical structures throughout the trajectories (Figure 1a). Moreover, a disconcerting array of conformations was observed, Figure 1b, with some structures being completely unfolded, and others that did not maintain a stable secondary structure for more than a few nanoseconds or were completely reconfigured into  $\beta$  sheets (Movies S1&S2). There is no experimental evidence that the  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$  peptide accesses these states.<sup>15</sup>

The best performing combination was igb5 with ff99SBnmr (Movie S1) with a median helicity of  $\approx 87\%$  and an interquartile range of less than 10% over the course of the simulation. The helicity was due to the unfolding of the terminal residues on both ends. Irrespective of the implicit solvent model, ff99SBnmr performed reasonably well for  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$ , in contrast to the related ff99SB ff99SBildn and ff99 forcefields (median  $\alpha$  helicity  $< 50\%$ ). The older forcefields ff94 and ff98 yielded a median helicity above 75% and an interquartile range below 20%, regardless of the GB model used and, were among the best forcefields at capturing the helical structure of this peptide.

The most recent GB model–forcefield combination, igb8 with ff14SBonlysc, which is recommended by the AMBER developers,<sup>14,22</sup> did not maintain the starting  $\alpha$  helix (Figure 1b,

bottom right) or even a stable overall secondary structure (Movie S1). Changing the GB model did not significantly modify the outcome with this forcefield. Perhaps surprisingly, the ff14SB forcefield, which was parameterized for use with the explicit TIP3P water model, led to a higher simulated  $\alpha$  helicity than ff14SBonlysc for each respective GB model, Figure 1 & Movie S1. Perhaps unsurprisingly, forcefields ff14ipq, ff15ipq and fb15, which were not designed to be compatible with GB models, were consistently poor at maintaining the peptide's starting conformation (median  $\alpha$  helicity < 50%) regardless of the GB model, Figure 1b&S2.

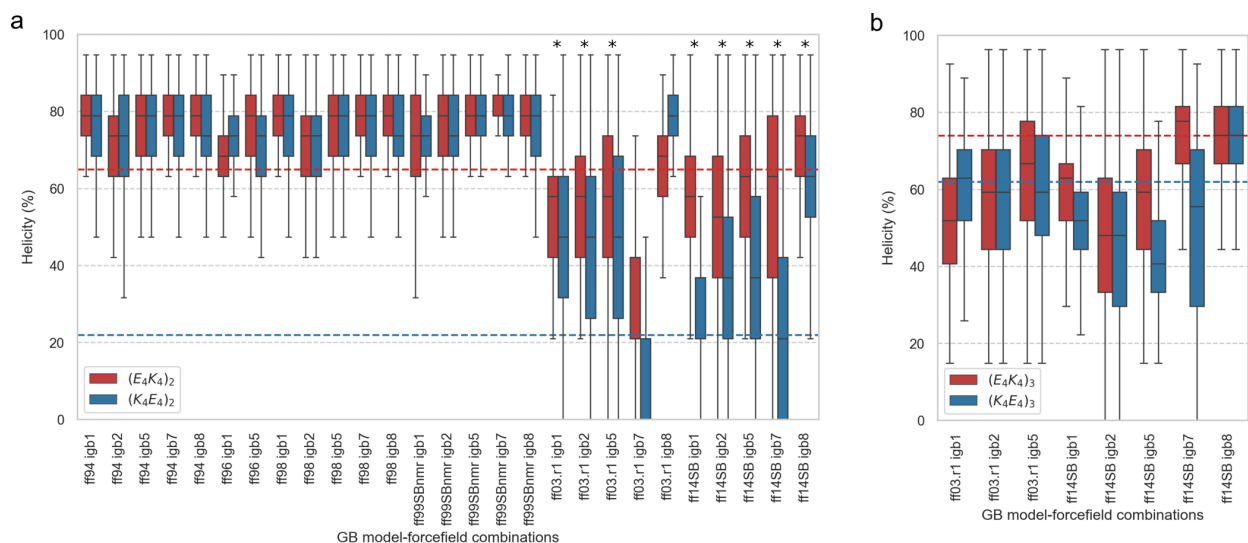
The simulations were sensitive to both the GB model and to the forcefield used: different predominant secondary structures were predicted by the same forcefield when different GB models were used, and vice versa. Some forcefields, e.g. ff96, ff03.r1 and ff14SB, were much more sensitive to the choice of implicit solvent model, with some GB models leading to a median helicity above 70% and others below 50%. For example, ff96 with igb5 captured the  $\alpha$  helicity relatively well, with a median helicity of  $\approx$ 83%. However, the same forcefield with igb8 predicted the formation of a  $\beta$  hairpin within 1  $\mu$ s, which developed into a three-stranded sheet over the remainder of the simulation. In contrast, using this forcefields with igb7, the A<sub>4</sub>(K<sub>4</sub>E<sub>4</sub>)<sub>1</sub>A<sub>4</sub>(K<sub>4</sub>E<sub>4</sub>)<sub>1</sub>A<sub>4</sub> peptide remained unfolded (Figure 1b&S1 and Movie S2).

To test the reproducibility of these simulations, we ran a total of 4 replicas for each of several GB model–forcefield combinations (Figures S3&S4). These replica simulations gave similar results for predicted structures and the differences between replicas are, in most cases, minimal. Reassuringly, the same major conformational changes were sampled in all four replica runs albeit at different times, such the  $\alpha$  helix-to- $\beta$  structure transitions observed with the igb8–ff96 combination. These results indicate that running a single 6  $\mu$ s simulation sufficiently captures the performance of a GB model–forcefield combination for peptides of this size, but replica



simulations should be run to ensure the reproducibility of the results for the best performing combinations.

Although none of the GB model–forcefield combinations maintained the experimentally observed  $\alpha$  helicity of  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$ , several gave a high degree of helicity (median helicity  $>75\%$ ). Therefore, we tested these combinations, along with some others for comparison, on other peptides with lower experimental  $\alpha$ -helical content; *e.g.*, the pair  $(E_4K_4)_2$  (65%  $\alpha$  helix) and  $(K_4E_4)_2$  (22%  $\alpha$  helix), Table 1, which differ only in the order of the Glu<sub>4</sub> and Lys<sub>4</sub> blocks. 27 implicit solvent–forcefield combinations were tested, starting from fully helical conformations of both peptides, and using the same protocol as for  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$ .



**Figure 2.** DSSP<sup>38</sup> calculated  $\alpha$  helicities of  $(E_4K_4)_n$  and  $(K_4E_4)_n$  peptides, where  $n=2,3$ . (a) Percentage  $\alpha$  helicities of  $(E_4K_4)_2$  (red) and  $(K_4E_4)_2$  (blue) for the MD trajectories generated from 27 GB model–forcefield combinations. Asterisks indicates the combinations selected for the MD simulations of  $(E_4K_4)_3$  and  $(K_4E_4)_3$  (b) Percentage  $\alpha$  helicities of  $(E_4K_4)_3$  (red) and  $(K_4E_4)_3$  (blue) for the MD trajectories from 8 GB model–forcefield combinations. The results are

presented as boxplots as in Figure 1. The dotted lines show the experimentally measured helicities for  $(E_4K_4)_n$  in red and  $(K_4E_4)_n$  in blue).

Notably, the older forcefields (ff94, ff96, ff98 and ff99SBnmr) that captured the high  $\alpha$ -helical content of  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$ , consistently predicted high  $\alpha$  helicities for  $(E_4K_4)_2$  and  $(K_4E_4)_2$ , Figure 2a. They did not discriminate between  $(E_4K_4)_2$  and  $(K_4E_4)_2$ , regardless of the implicit solvent model, with median helicities for both peptides of  $\approx 80\%$  compared with the experimental values of 67% and 22%, respectively. This shows that these forcefields systematically favor, and overestimate,  $\alpha$ -helical structures, at least for Glu/Lys-rich peptides. In contrast, the more-recent forcefields ff03.r1 and ff14SB gave lower  $\alpha$  helicities, closer to the experimentally determined values (except in the case of ff03.r1 with igb8), Figure 2a. Moreover, they correctly predicted  $(E_4K_4)_2$  to be more helical than  $(K_4E_4)_2$ ; although the differences in predicted helicities between the two peptides were less than that measured experimentally. ff14SB performed better than ff03.r1. The model combining ff14SB with igb7, which failed to predict the experimentally determined  $\alpha$  helicity of  $A_4(K_4E_4)_1A_4(K_4E_4)_1A_4$  (13% vs. 97%) (Figure 1), gave the best prediction of the difference in helicity between  $(E_4K_4)_2$  and  $(K_4E_4)_2$  at 63% and 21%, respectively. Overall, the results show that even when a particular GB model–forcefield combination describes a particular peptide well, it may fail for a closely related homologue (e.g. differing only in the order of the amino acid residues).

To test this last point further, we took the 8 GB model–forcefield combinations that performed best for  $(E_4K_4)_2$  and  $(K_4E_4)_2$  and modelled another comparable pair of peptides, namely  $(E_4K_4)_3$  and  $(K_4E_4)_3$ , Table 1, which have a smaller difference in experimental  $\alpha$  helicities, *i.e.*, 74% and 62%, respectively. Again, the different methods predicted a wide range of helicities (Figure 2b). Both ff03.r1 with igb5 and ff14SB with igb7 gave results in reasonable agreement with the

experimental helicity for the two peptides, with median  $\alpha$  helicities of 67% and 78% respectively for (E<sub>4</sub>K<sub>4</sub>)<sub>3</sub>, and 59% and 56% respectively for (K<sub>4</sub>E<sub>4</sub>)<sub>3</sub>. ff14SB with igb1 and igb5 also correctly predicted (E<sub>4</sub>K<sub>4</sub>)<sub>3</sub> as the more helical peptide, although the median helicities were 10% to 20% lower than experiment.

Overall, the combination of ff14SB with igb7 stood out, having predicted the percentage helicities for (E<sub>4</sub>K<sub>4</sub>)<sub>2</sub>, (K<sub>4</sub>E<sub>4</sub>)<sub>2</sub>, (E<sub>4</sub>K<sub>4</sub>)<sub>3</sub> and (K<sub>4</sub>E<sub>4</sub>)<sub>3</sub> reasonably well; *i.e.*, all were within 10% of the experimental values (Figure 2). However, it failed to predict the high helicity of A<sub>4</sub>(K<sub>4</sub>E<sub>4</sub>)<sub>1</sub>A<sub>4</sub>(K<sub>4</sub>E<sub>4</sub>)<sub>1</sub>A<sub>4</sub>, Figure 1a, returning random conformations instead (Figure 1b). Moreover, the combinations that best predicted the helicity of A<sub>4</sub>(K<sub>4</sub>E<sub>4</sub>)<sub>1</sub>A<sub>4</sub>(K<sub>4</sub>E<sub>4</sub>)<sub>1</sub>A<sub>4</sub> did not give the correct percentage helicities for the other four peptides, and almost always failed to predict the more helical of the paired designs. Thus, none of the GB-forcefield combinations tested here is a reliable predictor for all of the peptides that we tested. Comparison with other experimental measures of helicity did not improve the overall agreement with experiment and did not change the conclusion that none of the models quantitatively predict secondary structure (Supplementary results and Table S1).

In summary, these simulations show that none of the implicit solvent model–forcefield combinations that we have tested accurately predict the experimentally measured  $\alpha$  helicities of *de novo* designed Glu/Lys-rich peptides. While some combinations systematically favor a high degree of helicity, irrespective of the peptide, others are only predictive for peptides with high or intermediate  $\alpha$  helicities. Furthermore, some combinations predict entirely incorrect conformations, including  $\beta$ -rich structures, for which there is no experimental evidence. Therefore, our simulations serve as a warning not only of the unreliability of implicit solvent models, but also a reminder of the importance of the forcefield: changing the forcefield for a given GB model will

usually lead to changes that are more marked than changing the GB model for a given forcefield. They also highlight the importance, whenever possible, of rigorously testing combinations of GB models and forcefields against experimental data. With this in mind, the systematic set of *de novo* designed peptides that we have studied here provide a usefully sized training set for testing solvent models and protein forcefields for simulating experimentally observed secondary structures.

The present study could be extended in several ways. Replica exchange<sup>39</sup> was not used because the goal was to look at a fast and easy approach that can be applied to computational protein design. Replica exchange might improve convergence, but is not likely to change the overall conclusion that the implicit solvent models are not predictive of secondary structures adopted by short peptides. Also, the Glu and Lys residues were treated as always ionized. However, we note that interactions between such charged residues can lead to changes in effective  $pK_a$  and hence in protonation state.<sup>40</sup> Using constant pH MD<sup>41,42</sup> could shed further light on those phenomena and on their relative importance in secondary structure formation of the peptides. Finally, it would be interesting to explore combinations of forcefields with explicit solvent models on the same Glu and Lys rich peptides.

## **Supporting Information**

Methods, Supplementary Results, Supplementary Figures S1 to S11 and Supplementary Table S1 (PDF), Supplementary Movies S1 (mp4) and S2 (mp4).

## AUTHOR INFORMATION

### **Corresponding Authors**

\*Adrian J. Mulholland, email: [adrian.mulholland@bristol.ac.uk](mailto:adrian.mulholland@bristol.ac.uk) and Eric J. M. Lang, email: [eric.lang@bristol.ac.uk](mailto:eric.lang@bristol.ac.uk). School of Chemistry, University of Bristol, Cantock's Close, Bristol, BS8 1TS, UK.

### Author Contributions

E.J.M.L. conceived the project. E.J.M.L., D.N.W. and A.J.M. designed the research. E.J.M.L. performed the research. E.J.M.L analyzed the data with contributions from all authors. E.G.B collected the experimental CD data. All the authors contributed to writing the manuscript. All authors have given approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

E.J.M.L., E.G.B., A.J.M. and D.N.W. were funded by the BBSRC/EPSRC Synthetic Biology Research Centre, BrisSynBio (BB/L01386X/1). A.J.M. thanks the EPSRC for support (EP/M022609/1). D.N.W. held a Royal Society Wolfson Research Merit Award (WM140008). The authors thank Robert E. Arbon for help with the statistical analysis of the MD data. This work was carried out using the computational facilities (BlueCrystal Phase 4) of the Advanced Computing Research Centre, University of Bristol (<http://www.bris.ac.uk/acrc/>).

### ABBREVIATIONS

GB: generalized Born. MD: molecular dynamics. CD: circular dichroism

### REFERENCES

- (1) Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; Kamp, M. W. van der. Biomolecular Simulations: From Dynamics and Mechanisms to Computational Assays of Biological Activity. *WIREs Comput. Mol. Sci.* **2019**, *9* (3), e1393. <https://doi.org/10.1002/wcms.1393>.

- (2) Park, H.; DiMaio, F.; Baker, D. The Origin of Consistent Protein Structure Refinement from Structural Averaging. *Structure* **2015**, *23* (6), 1123–1128. <https://doi.org/10.1016/j.str.2015.03.022>.
- (3) Park, H.; Lee, G. R.; Kim, D. E.; Anishchenko, I.; Cong, Q.; Baker, D. High-Accuracy Refinement Using Rosetta in CASP13. *Proteins Struct. Funct. Bioinforma.* **2019**, *87* (12), 1276–1282. <https://doi.org/10.1002/prot.25784>.
- (4) Crean, R. M.; Gardner, J. M.; Kamerlin, S. C. L. Harnessing Conformational Plasticity to Generate Designer Enzymes. *J. Am. Chem. Soc.* **2020**, *142* (26), 11324–11342. <https://doi.org/10.1021/jacs.0c04924>.
- (5) Bunzel, H. A.; Anderson, J. L. R.; Mulholland, A. J. Designing Better Enzymes: Insights from Directed Evolution. *Curr. Opin. Struct. Biol.* **2021**, *67*, 212–218. <https://doi.org/10.1016/j.sbi.2020.12.015>.
- (6) Ge, Y.; Kier, B. L.; Andersen, N. H.; Voelz, V. A. Computational and Experimental Evaluation of Designed  $\beta$ -Cap Hairpins Using Molecular Simulations and Kinetic Network Models. *J. Chem. Inf. Model.* **2017**, *57* (7), 1609–1620. <https://doi.org/10.1021/acs.jcim.7b00132>.
- (7) Onufriev, A. V.; Case, D. A. Generalized Born Implicit Solvent Models for Biomolecules. *Annu. Rev. Biophys.* **2019**, *48* (1), 275–296. <https://doi.org/10.1146/annurev-biophys-052118-115325>.
- (8) Kleinjung, J.; Fraternali, F. Design and Application of Implicit Solvent Models in Biomolecular Simulations. *Curr. Opin. Struct. Biol.* **2014**, *25*, 126–134. <https://doi.org/10.1016/j.sbi.2014.04.003>.
- (9) Zagrovic, B.; Pande, V. Solvent Viscosity Dependence of the Folding Rate of a Small Protein: Distributed Computing Study. *J. Comput. Chem.* **2003**, *24* (12), 1432–1436. <https://doi.org/10.1002/jcc.10297>.
- (10) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophys. J.* **2015**, *108* (5), 1153–1164. <https://doi.org/10.1016/j.bpj.2014.12.047>.
- (11) Shell, M. S.; Ritterson, R.; Dill, K. A. A Test on Peptide Stability of AMBER Force Fields with Implicit Solvation. *J. Phys. Chem. B* **2008**, *112* (22), 6878–6886. <https://doi.org/10.1021/jp800282x>.
- (12) Robinson, M. K.; Monroe, J. I.; Shell, M. S. Are AMBER Force Fields and Implicit Solvation Models Additive? A Folding Study with a Balanced Peptide Test Set. *J. Chem. Theory Comput.* **2016**, *12* (11), 5631–5642. <https://doi.org/10.1021/acs.jctc.6b00788>.
- (13) Maffucci, I.; Contini, A. An Updated Test of AMBER Force Fields and Implicit Solvent Models in Predicting the Secondary Structure of Helical,  $\beta$ -Hairpin, and Intrinsically Disordered Peptides. *J. Chem. Theory Comput.* **2016**, *12* (2), 714–727. <https://doi.org/10.1021/acs.jctc.5b01211>.
- (14) Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *J. Am. Chem. Soc.* **2014**, *136* (40), 13959–13962. <https://doi.org/10.1021/ja5032776>.
- (15) Baker, E. G.; Bartlett, G. J.; Crump, M. P.; Sessions, R. B.; Linden, N.; Faul, C. F. J.; Woolfson, D. N. Local and Macroscopic Electrostatic Interactions in Single  $\alpha$ -Helices. *Nat. Chem. Biol.* **2015**, *11* (3), 221–228. <https://doi.org/10.1038/nchembio.1739>.

- (16) Sivaramakrishnan, S.; Spink, B. J.; Sim, A. Y. L.; Doniach, S.; Spudich, J. A. Dynamic Charge Interactions Create Surprising Rigidity in the ER/K  $\alpha$ -Helical Protein Motif. *Proc. Natl. Acad. Sci.* **2008**, *105* (36), 13356–13361. <https://doi.org/10.1073/pnas.0806256105>.
- (17) Case, D. A. et al. *Amber 2016*; 2016.
- (18) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246* (1), 122–129. [https://doi.org/10.1016/0009-2614\(95\)01082-K](https://doi.org/10.1016/0009-2614(95)01082-K).
- (19) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104* (15), 3712–3720. <https://doi.org/10.1021/jp994072s>.
- (20) Onufriev, A.; Bashford, D.; Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins Struct. Funct. Bioinforma.* **2004**, *55* (2), 383–394. <https://doi.org/10.1002/prot.20033>.
- (21) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J. Chem. Theory Comput.* **2007**, *3* (1), 156–169. <https://doi.org/10.1021/ct600085e>.
- (22) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, *9* (4), 2020–2034. <https://doi.org/10.1021/ct3010485>.
- (23) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117* (19), 5179–5197. <https://doi.org/10.1021/ja00124a002>.
- (24) Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. The Development/Application of a ‘Minimalist’ Organic/Biochemical Molecular Mechanic Force Field Using a Combination of Ab Initio Calculations and Experimental Data. In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*; van Gunsteren, W. F., Weiner, P. K., Wilkinson, A. J., Eds.; Computer Simulations of Biomolecular Systems; Springer Netherlands: Dordrecht, 1997; pp 83–96. [https://doi.org/10.1007/978-94-017-1120-3\\_2](https://doi.org/10.1007/978-94-017-1120-3_2).
- (25) III, T. E. C.; Cieplak, P.; Kollman, P. A. A Modified Version of the Cornell et al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **1999**, *16* (4), 845–862. <https://doi.org/10.1080/07391102.1999.10508297>.
- (26) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21* (12), 1049–1074. [https://doi.org/10.1002/1096-987X\(200009\)21:12<1049::AID-JCC3>3.0.CO;2-F](https://doi.org/10.1002/1096-987X(200009)21:12<1049::AID-JCC3>3.0.CO;2-F).
- (27) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65* (3), 712–725. <https://doi.org/10.1002/prot.21123>.
- (28) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (8), 1950–1958. <https://doi.org/10.1002/prot.22711>.

- (29) Li, D.-W.; Brüschweiler, R. NMR-Based Protein Potentials. *Angew. Chem. Int. Ed.* **2010**, *49* (38), 6778–6780. <https://doi.org/10.1002/anie.201001898>.
- (30) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999–2012. <https://doi.org/10.1002/jcc.10349>.
- (31) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>.
- (32) Cerutti, D. S.; Swope, W. C.; Rice, J. E.; Case, D. A. Ff14ipq: A Self-Consistent Force Field for Condensed-Phase Simulations of Proteins. *J. Chem. Theory Comput.* **2014**, *10* (10), 4515–4534. <https://doi.org/10.1021/ct500643c>.
- (33) Wang, L.-P.; McKiernan, K. A.; Gomes, J.; Beauchamp, K. A.; Head-Gordon, T.; Rice, J. E.; Swope, W. C.; Martínez, T. J.; Pande, V. S. Building a More Predictive Protein Force Field: A Systematic and Reproducible Route to AMBER-FB15. *J. Phys. Chem. B* **2017**, *121* (16), 4023–4039. <https://doi.org/10.1021/acs.jpcc.7b02320>.
- (34) Debiec, K. T.; Cerutti, D. S.; Baker, L. R.; Gronenborn, A. M.; Case, D. A.; Chong, L. T. Further along the Road Less Traveled: AMBER Ff15ipq, an Original Protein Force Field Built on a Self-Consistent Physical Model. *J. Chem. Theory Comput.* **2016**, *12* (8), 3926–3947. <https://doi.org/10.1021/acs.jctc.6b00567>.
- (35) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30* (10), 1545–1614. <https://doi.org/10.1002/jcc.21287>.
- (36) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542–1555. <https://doi.org/10.1021/ct200909j>.
- (37) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- (38) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22* (12), 2577–2637. <https://doi.org/10.1002/bip.360221211>.
- (39) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314* (1), 141–151. [https://doi.org/10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9).
- (40) Rhys, G. G.; Wood, C. W.; Lang, E. J. M.; Mulholland, A. J.; Brady, R. L.; Thomson, A. R.; Woolfson, D. N. Maintaining and Breaking Symmetry in Homomeric Coiled-Coil Assemblies. *Nat. Commun.* **2018**, *9* (1), 4132. <https://doi.org/10.1038/s41467-018-06391-y>.



- (41) Mongan, J.; Case, D. A.; McCammon, J. A. Constant PH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25* (16), 2038–2048. <https://doi.org/10.1002/jcc.20139>.
- (42) Swails, J. M.; Roitberg, A. E. Enhancing Conformation and Protonation State Sampling of Hen Egg White Lysozyme Using PH Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (11), 4393–4404. <https://doi.org/10.1021/ct300512h>.