Communication

Dominant Chemical Interactions Governing the Folding Mechanism of Oligopeptides

Michele Larocca [1,*], Giuseppe Floresta [2], Daniele Verderese [3] and Agostino Cilibrizzi [4,5,*]

1 Istituto di Metodologie per l'Analisi Ambientale–Consiglio Nazionale delle Ricerche (CNR-IMAA), Contrada, Santa. Loja, 85050 Potenza, Italy
2 Department of Drug and Health Sciences, University of Catania, Viale A. Doria 6, 95125 Catania, Italy
3 Dipartimento di Scienze Economiche e Statistiche, Università di Salerno, via Giovanni Paolo II, 132, 84084 Salerno, Italy
4 Institute of Pharmaceutical Science, King's College London, Stamford Street, London SE1 9NH, UK
5 Centre for Therapeutic Innovation, University of Bath, Bath BA2 7AY, UK
* Correspondence: michele.larocca@studio.unibo.it (M.L.); agostino.cilibrizzi@kcl.ac.uk (A.C.)

Abstract: The hydrophobic effect is the main factor that drives the folding of polypeptide chains. In this study, we have examined the influence of the hydrophobic effect in the context of the main mechanical forces approach, mainly in relation to the establishment of specific interplays, such as hydrophobic and CH–π cloud interactions. By adopting three oligopeptides as model systems to assess folding features, we demonstrate herein that these finely tuned interactions dominate over electrostatic interactions, including H-bonds and electrostatic attractions/repulsions. The folding mechanism analysed here demonstrates cooperation at the single-residue level, for which we propose the terminology of "single residues cooperative folding". Overall, hydrophobic and CH–π cloud interactions produce the main output of the hydrophobic effect and govern the folding mechanism, as demonstrated in this study with small polypeptide chains, which in turn represent the main secondary structures in proteins.

Keywords: folding mechanisms; main mechanical forces; dominant chemical interactions; dihedral angle calculations

1. Introduction

The folding mechanism of polypeptide chains has been recently described as the output of main mechanical forces (MMFs), which develop as the result of highly specific chemical interactions established amongst side chains of amino acid residues in close spatial proximity within a sequence [1–4]. Earlier MMFs methodology studies support the assumption that there is a finely tuned mechanism regulating how each single amino acid in a polypeptide chain sequence 'encodes' its three-dimensional (3D) native structure. In agreement with previously reported studies [5], this proposed mechanism specifically relies on the cooperativity amongst amino acid side chain residues, namely the formation of suitable inter-residue chemical interactions amongst the amino acid side chains, such as hydrophobic, H-bond, electrostatic attraction/repulsion, CH–π cloud and π–π (pi-stacking) interactions [1–4].

Although advancements continue to be reported in the literature, focusing on improved methods of analysis and prediction [6], the folding process is still far from a full understanding. The main factor that drives the folding of polypeptide chains is well-known as the hydrophobic effect [7–10]. In this study, this effect was considered determinant to maximize the chemical interactions mentioned above, by simultaneous effect of the solvent, through the formation of hydrophobic cores. In this context, the work presented herein aims to provide further insights on the analysis of protein folding by describing the dominant chemical interactions that rule the mechanism itself. By applying the MMFs method and

calculations [1–4], three small polypeptide chains, namely chignolin (PDB code: 1uao) [11], Leu-enkephalin [12] and CCR5 ECL2 (PDB code: 2mzx) [13], were used as model systems for this investigation. In particular, the dihedral angles in these reference polypeptide chains were calculated using the recently reported MMFs approach [4]. Subsequently, the structures obtained via MMFs analysis were modelled using the Molefacture plug-in [14] in VMD [15] and compared to experimentally determined structures by superimposition (see details in Supplementary Materials File S1: paragraph 1. Determination of the internal contacts). Briefly, the MMF approach focuses on interpreting, 'numerically', the interactions amongst the side chains of neighbour residues that, based on the length of the side chain, may establish interactions up to a certain distance [1–4]. This space length was defined by us as "threshold distance" r_i (see calculation spreadsheets for details: Supplementary Materials Files S2–S4), i.e., the initial distance that is used for the calculations [2,4]. Once r_i is obtained, it is possible to calculate the Coulombian force acting through the interacting atoms in the side chains and, in turn, define the equation that reports their respective potential energy [1,2]. The potential energy equation contains a partial dihedral angle to be calculated, where for a partial dihedral angle we take into consideration the rotation obtained after each single interaction of the side chain of a residue with its neighbour interacting residues. Specifically, given the threshold distance, it is possible to decipher the number of interactions that a residue might establish with the closest neighbour residues, enabling the calculation of all the allowed partial dihedral angles, which can then be added up to obtain the total dihedral angle. This latter is then compared to the experimental value. Furthermore, the MMFs approach allows to infer the equation of the torque moment, which also contains the partial dihedral angle to be calculated. Thus, equalizing potential energy and torque moment, it is possible to calculate the desired partial dihedral angle for a specific interaction by using the arctangent of the angle [4] (see details on calculations, reported in the calculations spreadsheets: Supplementary Materials Files S2–S4).

Overall, by applying the MMFs approach, we achieved a high match between the predicted/experimental structures examined in this work, which was further improved through energy minimization using the web server Chiron [16]. By data analysis, it was found that hydrophobic interactions, as well as CH–π clouds, dominate over all other types of possible interactions. Based on the MMFs approach, these two types of interplays may be considered as direct outputs of the hydrophobic effect, being the most represented for the determination of partial dihedral angles, which are then used for the calculations of total dihedral angles. For each of the polypeptides adopted in this study, we have considered all the possible interactions that can occur in the residues of the side chains on the basis of both threshold distances and evident chemical interactions [4]. Furthermore, the interactions that were useful to calculate the dihedral angles (highlighted in black, in the Supporting Information spreadsheets, Supplementary Materials File S2; Supplementary Materials File S3; Supplementary Materials File S4) showed that the best match was obtained when the largest partial rotations (or partial dihedral angles) were considered.

2. Results

Primarily, the abovementioned evidence led to discovering that the most useful interactions are primarily the hydrophobic and CH–π cloud interactions, these being also the strongest interactions, as compared to H-bonds and electrostatic attractions (Figure 1), in agreement with the MMFs approach [1–4]. In contrast, pi stacking possessed an intermediate behaviour, while electrostatic repulsions were not represented much (Figure 1), even if these latter are the strongest of all the other types of interactions.
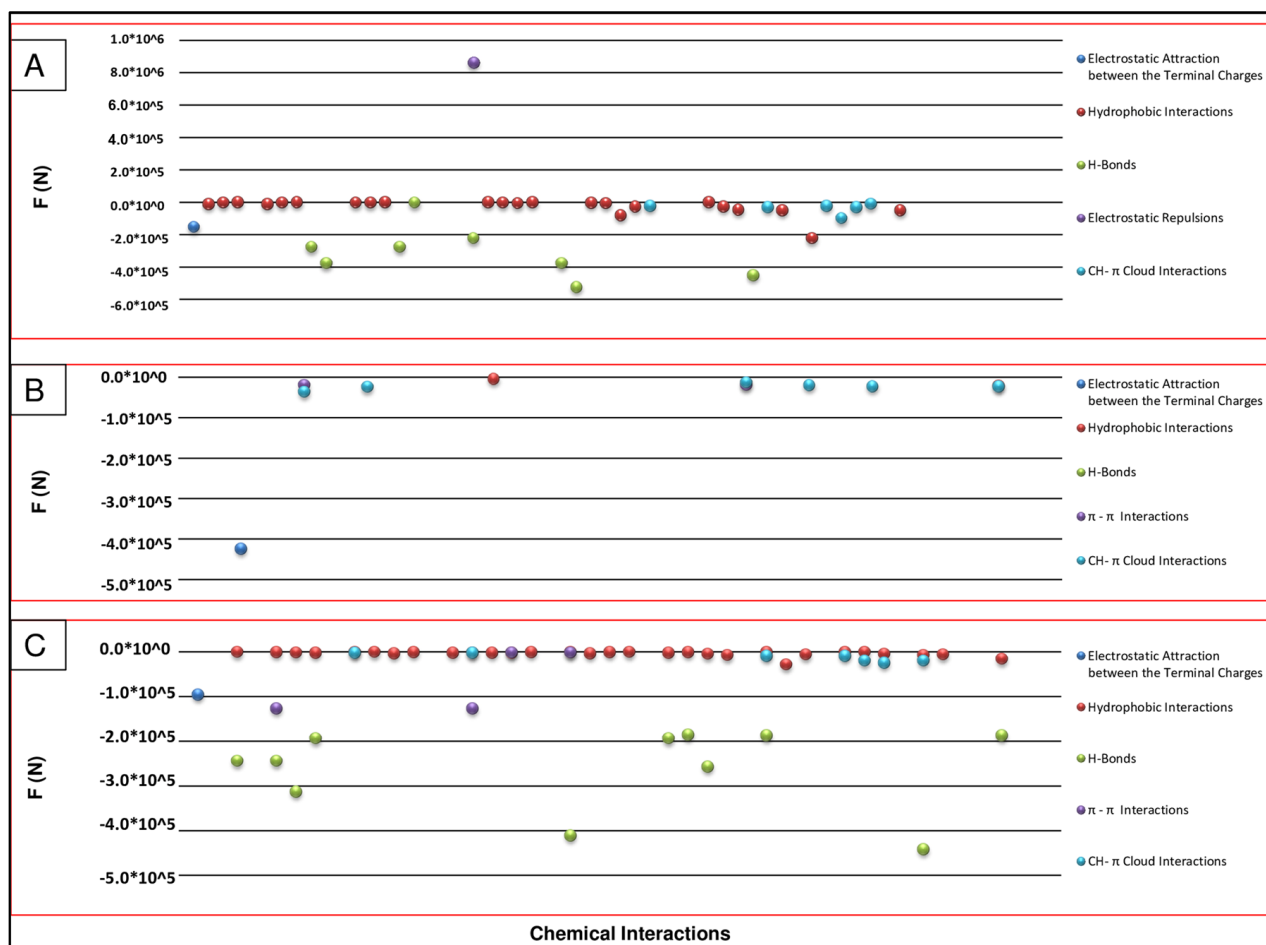
Figure 1. Mechanical forces related to (**A**) chemical interactions of the chignolin (1uao) folding process; (**B**) chemical interactions of the Leu–enkephalin folding process; and (**C**) chemical interactions of the CCR5 ECL2 (2mzx) folding process.

A single residue belonging to 2mzx (triptophan190 or TRP190) was also analysed in more detail, with the aim to further confirm these findings. We found that the folding mechanism of TRP190 relies on the establishment of several chemical interactions due to the chemical nature of the neighbour side chain residues. Specifically, TRP190 results involved hydrophobic, π–π, H-bond and electrostatic interactions between the two terminal charges, the latter being maintained constant for all the residues, in order to calculate accurately the related dihedral angles. As reported in Figure 2, there is the need to consider the largest values of the calculated partial dihedral angles for the calculation of the TRP190 dihedral angle φ. This confirms that the larger the rotations (partial dihedral angles), the higher the match with the experimental values. This feature implies that the larger rotation is provided by the longest threshold distance for the occurring chemical interaction. For all the possible interplays with each neighbour residue, the partial dihedral angle with the higher value of potential energy (and related mechanical force) is the one to take into consideration for the calculations, as shown in Figure 2 (i.e., black arrows) for the TRP190 interactions to achieve the calculation of the dihedral angle φ.
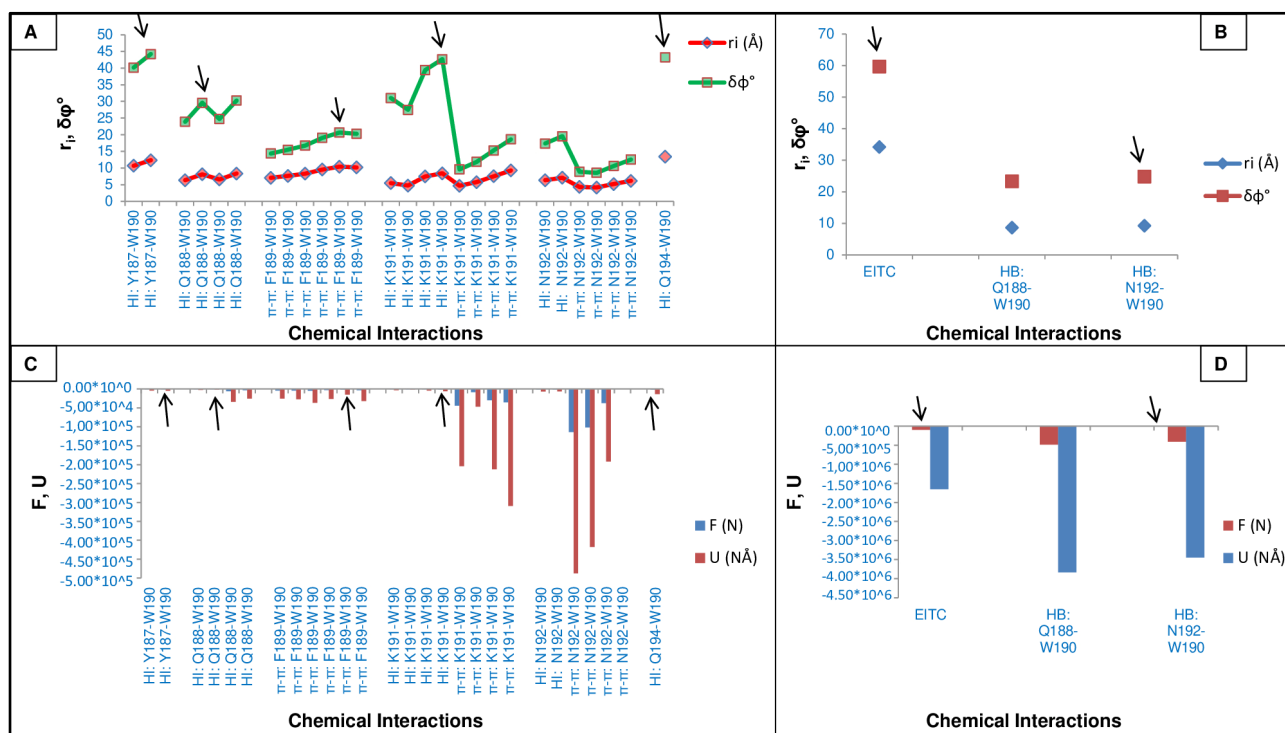
Figure 2. Physical measures that rule the folding mechanism at the single-residue level of TRP190. (**A**,**B**) Comparison between threshold distances and partial dihedral angles. (**C**,**D**) Comparison between developed mechanical forces and the related potential energy. Abbreviations: ri = threshold distance, $\delta\varphi°$ = partial dihedral angle, F = mechanical force, U = potential energy, HI = hydrophobic interaction, HB = H-bond, EITC = electrostatic interaction between the terminal charges, π–π = π–π interaction or pi stacking, W = Tryptophan, Y = Tyrosine, Q = Glutamine, F = Phenylalanine, K = Lysine, N = Asparagine.

The approach applied to TRP190 was also used for all the amino acid residues in the three oligopeptides considered in this work. Notably, glycines express a variable behaviour within this reproducible workflow of calculations, which is normally applied throughout the peptide chain. That is, depending on the specific position of glycines in the chain, in many cases the conventional calculations are not suited to obtaining reliable values of dihedral angles, the estimation of which requires to adopt an alternative equation, specifically tailored to fulfil the specific case. For instance, the conventional procedure used for all residues can be applied to glycine if the amino acid is located in positions 1 or 2 of a peptide sequence. In contrast, if the glycine appears in a different position within the chain, the newly designed Equation (1) is employed (for GLY in position n, with $n \neq 0, 1$ and the sign $\pm$ in agreement with the sign of $\varphi°_{n-1}$):

$$\delta\varphi°_{GLY} = (\pm 3) \cdot (\varphi°_{n-1}) \cdot (\delta\varphi°_{tc}/180°) \quad n \neq 1, 2 \tag{1}$$

In Equation (1), $\varphi°_{n-1}$ is the total dihedral angle at position $n-1$ with respect to a GLY in position n, whereas $\delta\varphi°_{tc}$ is the partial dihedral angle arising from the postulated and constant electrostatic interaction between the two terminal charges (i.e., charges on the first and last residue).

This equation does not apply when GLY is found at position $n = 1$ or 2, since in the case of $n = 1$ the dihedral angle to calculate is ψ_1 (being $\varphi_1 = 0°$). For this angle, the same rules considered as for the calculation of the dihedral angles φ [1–4] are valid and apply consistently. In contrast, if the GLY residue is located in position $n = 2$, it is clearly not possible to apply Equation (1), due to $\varphi_{n-1} = 0°$. Additionally, the methodology of the calculation of dihedral angles ψ has required the introduction of suitable modifications

concerning Equation (2), as in "Equation (9)" and as established previously [4], normally applied to obtain ψ values demonstrated not applicable for some residues.

$$\psi_n = \pm\, f \cdot [(\varphi_n \cdot \varphi_{n+1})/(\varphi_n + \varphi_{n+1})] \pm (\times\, 180°)\ n \neq 1 \tag{2}$$

In particular, we have determined that the value of "f" (the constant required in the prior Equation (9) to calculate the dihedral angles ψ) needs to be changed in specific cases in order to achieve a better match with the experimentally determined dihedral angle values. This may be indicative of differences between small and relatively longer peptide chains that need to be taken into consideration, in order to obtain robust outputs from the prediction analysis. Details concerning the changes about the values of the constant "f" (adjusted values of f concerning chignolin are f = −1/2 for THR6 and f = 1/5 for GLY7, whereas Leu-enkephalin concerns GLY2, GLY3 and PHE4, which have, respectively, f = 3/4, −1/10 and −1/10. In the case of CCR5 ECL2, f = 3/2 for PHE193 and 1/2 for GLN194) are reported in the calculation spreadsheets within the Supporting Information.

By adopting these modifications and applying the workflow previously developed [4] for the calculation of dihedral angles in peptides, we have analysed herein chignolin, Leu-enkephalin and CCR5 ECL2. Each dihedral angle in these polypeptides was calculated. The final structures were modelled in VMD [14,15] and compared by superimpositions to those experimentally available. Overall, a robust match in terms of RMSD was obtained throughout the analysis. By submitting the calculated structures to energy minimization using the web server Chiron [16], an RMSD improvement was achieved, as evident from the decrease of all RMSD values, which in each case are below 1.70 Å. Details of these results are shown in Figure 3, where the superimposed structures A–C are those without energy minimization, whereas D–F are those after energy minimization.
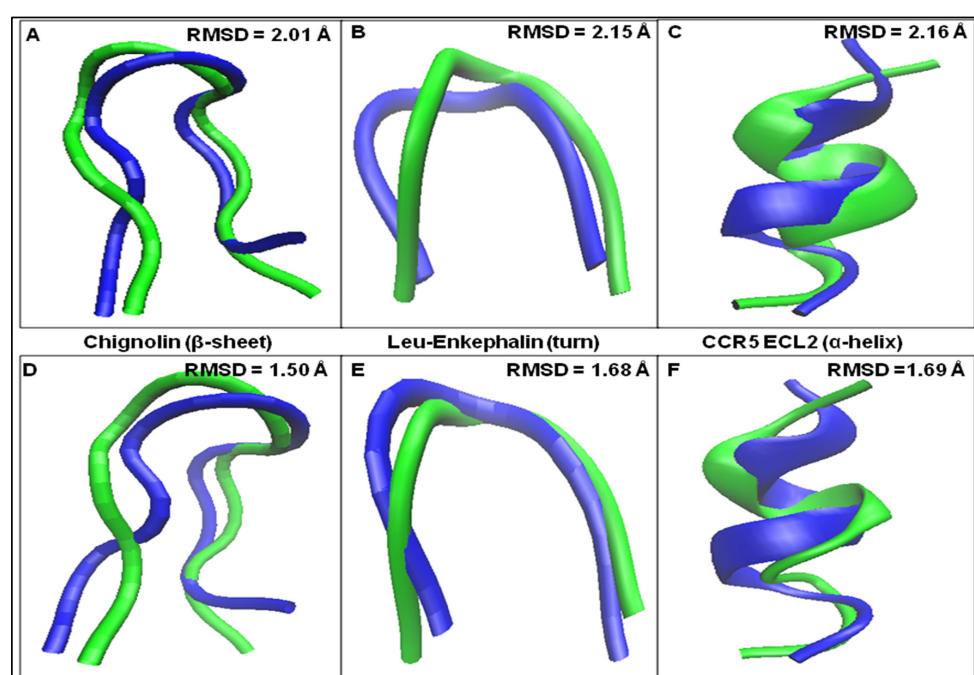

Figure 3. (**A**) Superimposition of the calculated without energy minimization (green) and the experimental structure (blue) of chignolin; (**B**) superimposition of the calculated without energy minimization (green) and the experimental structure (blue) of Leu-enkephalin; (**C**) superimposition of the calculated without energy minimization (green) and the experimental structure (blue) of CCR5 ECL2; (**D**) superimposition of the calculated after energy minimization (green) and the experimental structure (blue) of chignolin; (**E**) superimposition of the calculated after energy minimization (green) and the experimental structure (blue) of Leu-enkephalin; and (**F**) superimposition of the calculated after energy minimization (green) and the experimental structure (blue) of CCR5 ECL2.

Moreover, we have submitted the three oligopeptides to molecular dynamics (MD) simulations. The MD simulations of the complexes were performed with YASARA. A periodic simulation cell extending 10 Å from the protein surfaces were employed. The cell was filled with water, with a maximum sum of all water bumps of 1.0 Å and a density of 0.997 g/mL. The setup included optimizing the hydrogen bonding network [17] to increase the solute stability and a pKa prediction to fine-tune the protonation states of protein residues at the chosen pH of 7.4 [18]. With an excess of either Na or Cl to neutralize the cell, NaCl ions were supplied at a physiological concentration of 0.9%. The cutoff was 10 Å for Van der Waals forces (the default used by AMBER) [19], and no cutoff was applied to electrostatic forces (using the Particle Mesh Ewald algorithm) [20]. The equations of motions were integrated with multiple time steps of 2.5 fs for bonded interactions and 5.0 fs for nonbonded interactions at a temperature of 298 K and a pressure of 1 atm. Initially, a short MD simulation was run on the solvent only to remove clashes. The entire system was then energy minimized using a steepest descent minimization to remove conformational stress, followed by a simulated annealing minimization until convergence (<0.01 kcal/mol Å). Finally, 100 ns MD simulation without any restrictions was conducted and the conformations were recorded every 200 ps (details on the MD simulations can be accessed via the Supporting Information file, Figures S1–S18). We have subsequently compared the structures obtained by calculations with the corresponding ones having the lower RMSD from MD simulations in water (1.81 Å for chignolin, 1.43 Å for Leu-Enkephalin and 1.2 Å for CCR5 ECL2). The comparison was performed to show how the internal contacts of each residue are determined (method reported in Supporting Information, Supplementary Materials File S1). Substantially, we found that between the calculated oligopeptides and the corresponding structures with the lowest values of RMSD in MD simulations, the number of internal contacts is comparable, except for GLY10 in chignolin (Figure 4). We explain this output as a confirmation of the fact that the hydrophobic effect arising after solubilisation in water, proceeds by maximizing the internal contacts, allowing to establish the interactions of interest that drive the folding process of all the three oligopeptides investigated in this work.
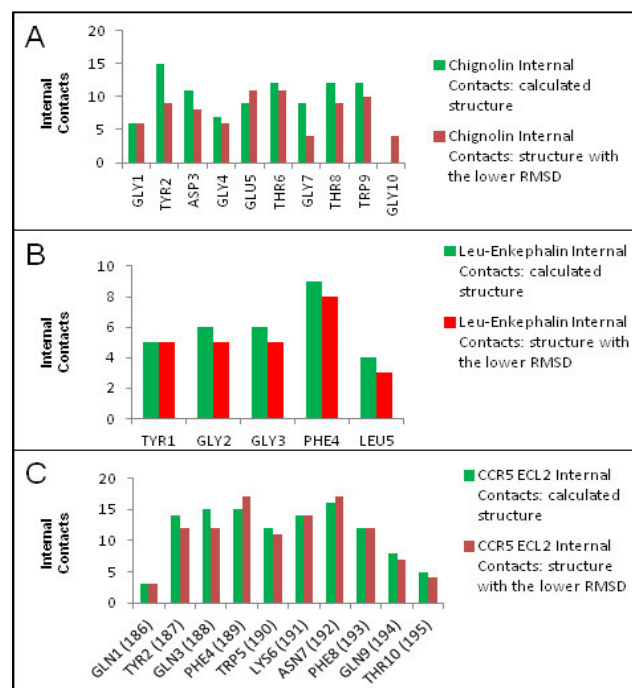

Figure 4. Per-residue internal contacts between the calculated/predicted structures (green) and the structures with the lowest RMSD during MD simulations (red). (**A**) Chignolin, (**B**) Leu-enkephalin and (**C**) CCR5 ECL2.

3. Discussion and Conclusions

To conclude, the MMFs approach demonstrated a valid tool to analyse the folding mechanism of the oligopeptides selected for this study, enabling us to obtain information at the single-residue level. To this end, we propose the adoption of "single residues cooperative folding" as a suitable term to define the folding mechanism under the circumstances assessed in this work. Clearly, we have established that certain type of interactions, namely hydrophobic and CH–π cloud, dominate over others, such as H-bonds and electrostatic attraction/repulsion. Hydrophobic and CH–π cloud interactions were thus considered as the main output of the hydrophobic effect, while H-bonds and inter-residue electrostatic interplays are likely playing a role in stabilizing the peptide structures. It is noteworthy that hydrophobic and CH–π cloud interactions confer the largest rotations in terms of partial dihedral angles and result in the most frequent types of interplays that provide a better match with the experimental data. The outcomes of this study support widely earlier findings [5,21], indicating that the folding is the result of a "cooperative activity" within amino acid residues. The process might be influenced and, primarily, proceed through the interplay amongst apolar or "hydrophobic" amino acids, although larger and more polar side chains (e.g., methionine or, alternatively, charged groups as in lysine and arginine) may also play a role in particular circumstances concerning the dominant interactions pointed out in this study.

References

1. Larocca, M.; Foglia, F.; Cilibrizzi, A. Dihedral Angle Calculations to Elucidate the Folding of Peptides through its Main Mechanical Forces. *Biochemistry* **2019**, *58*, 1032–1037. [CrossRef] [PubMed]
2. Larocca, M.; Floresta, G.; Cilibrizzi, A. Principles that Rule the Calculation of Dihedral Angles in Secondary Structures: The cases of an α-helix and a β-sheet. *J. Mol. Struct.* **2021**, *1229*, 129802. [CrossRef]
3. Larocca, M. Polypeptides folding: Rules for the calculation of the backbone dihedral angles φ starting from the amino acid sequence. *ChemRxiv* **2020**, Preprint. [CrossRef]
4. Larocca, M.; Floresta, G.; Verderese, D.; Cilibrizzi, A. 'Main Mechanical Forces-Chemical Interactions' Interplay as a Tool to Elucidate Folding Mechanisms. *Pept. Sci.* 2024, in press. [CrossRef]
5. Gromihaa, M.M.; Selvaraj, S. Inter-residue interactions in protein folding and stability. *Prog. Biophys. Mol. Biol.* **2004**, *86*, 235–277. [CrossRef] [PubMed]
6. Chen, S.-J.; Hassan, M.; Jernigan, R.L.; Jia, K.; Kihara, D.; Kloczkowski, A.; Kotelnikov, S.; Kozakov, D.; Liang, J.; Liwo, A.; et al. Protein folds vs. protein folding: Differing questions, different challenges. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2214423119. [CrossRef] [PubMed]
7. Kauzmann, W. The Three Dimensional Structures of Proteins. *Biophys. J.* **1964**, *4*, 43–54. [CrossRef] [PubMed]
8. Camilloni, C.; Bonetti, D.; Morrone, A.; Giri, R.; Dobson, C.M.; Brunori, M.; Gianni, S.; Vendruscolo, M. Towards a structural biology of the hydrophobic effect in protein folding. *Sci. Rep.* **2016**, *6*, 28285. [CrossRef] [PubMed]

9. Baldwin, R.L. Dynamic Hydration Shell Restores Kauzmann's 1959 Explanation of How the Hydrophobic Factor Drives Protein Folding. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13052–13056. [CrossRef] [PubMed]

10. Perunov, N.; England, J.L. Quantitative Theory of Hydrophobic Effect as a Driving Force of Protein Structure. *Protein Sci.* **2014**, *23*, 387–399. [CrossRef] [PubMed]

11. Honda, S.; Yamasaki, K.; Sawada, Y.; Morii, H. 10 Residue Folded Peptide Designed by Segment Statistics. *Structure* **2004**, *12*, 1507–1518. [CrossRef] [PubMed]

12. Ishida, T.; Kenmotsu, M.; Mino, Y.; Inoue, M.; Fujiwara, T.; Tomita, K.; Kimura, T.; Sakakibara, S. X-ray diffraction studies of enkephalins. Crystal structure of [(4′-bromo) Phe4,Leu5]-enkephalin. *Biochem. J.* **1984**, *218*, 677–689. [CrossRef] [PubMed]

13. Abayev, M.; Moseri, A.; Tchaicheeyan, O.; Kessler, N.; Arshava, B.; Naider, F.; Scherf, T.; Anglister, J. An extended CCR5 ECL2 peptide forms a helix that binds HIV-1 gp120 through non-specific hydrophobic interactions. *FEBS J.* **2015**, *282*, 1906–1921. [CrossRef] [PubMed]

14. Mayne, C.G.; Saam, J.; Schulten, K.; Tajkhorshid, E.; Gumbart, J.C. Rapid parameterization of small molecules using the Force Field Toolkit. *J. Comput. Chem.* **2013**, *34*, 2757–2770. [CrossRef] [PubMed]

15. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef] [PubMed]

16. Ramachandran, S.; Kota, P.; Ding, F.; Dokholyan, N.V. Automated minimization of steric clashes in protein structures. *Proteins Struct. Funct. Bioinform.* **2011**, *79*, 261–270. [CrossRef] [PubMed]

17. Krieger, E.; Dunbrack, R.L.; Hooft, R.W.; Krieger, B. Assignment of protonation states in proteins and ligands: Combining pKa prediction with hydrogen bonding network optimization. In *Computational Drug Discovery and Design*; Springer: New York, NY, USA, 2012; pp. 405–421.

18. Krieger, E.; Nielsen, J.E.; Spronk, C.A.; Vriend, G. Fast empirical pKa prediction by Ewald summation. *J. Mol. Graph. Model.* **2006**, *25*, 481–486. [CrossRef] [PubMed]

19. 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. Bioinform.* **2006**, *65*, 712–725. [CrossRef] [PubMed]

20. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [CrossRef]

21. Dyson, H.J.; Wright, P.E.; Scheraga, H.A. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13057–13061. [CrossRef] [PubMed]