

1 **Can charge-reversal be considered as a strategy for**
2 **attaining thermal stability in proteins?**

3 **Authors:** Suman Hait*, Sudipto Basu and Sudip Kundu*

4 **Author affiliations:**

5 Department of Biophysics, Molecular Biology and Bioinformatics, 92, Acharya Prafulla Chandra
6 Road, University of Calcutta, Kolkata 700009, India.

7 *Correspondence. skbmbg@caluniv.ac.in (S.K.), shbmbb_rs@caluniv.ac.in (S.H.)

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

1 **ABSTRACT**

2 Do charge reversal mutations (CRM) naturally occur in mesophilic-thermophilic/hyperthermophilic (M-
3 T/HT) orthologous proteins? Do they contribute to thermal stability by altering charge-charge
4 interactions? A careful investigation on 1550 M-T/HT orthologous protein pairs with remarkable
5 structural and topological similarity extracts the role of buried and partially exposed CRMs in enhancing
6 thermal stability. Our findings could assist in engineering thermo-stable variants of proteins.

7 **SIGNIFICANCE**

8 Protein engineering is one of the hot topics for decades specifically for its applications in different fields
9 like de-novo protein design, directed evolution, making highly stable variants for food and drug industry
10 etc. Proteins from organisms living in extreme environments are therefore a matter of common interest for
11 scientists from different disciplines. Over three decades of study has already found several sequence and
12 structural adaptations related to thermal stability, while charge reversal study remains ignored to a large
13 extent. Influenced by nature's strategy, our study provides a systemic understanding of how proper
14 designing of few partially exposed and buried CRMs significantly contributes to thermal stability by
15 altering the short distance electrostatic interactions.

16 **Keywords:** protein stability, thermal adaptation, electrostatic interaction, charge reversal, solvent
17 interaction

18

19

20

21

22

23

24

25

26

27

28

29

30

1 INTRODUCTION

2 Some proteins are considered useful in various research and industrial sectors because of their efficient
3 and cost-effective catalytic properties. But naturally occurring proteins are not considered potential
4 candidates for industrial applications because of their inadequate stability in various working conditions
5 (1). Condition like extreme temperature demands special designing of proteins in thermophile and
6 hyperthermophiles to function and survive. As proteins in thermophilic and hyperthermophilic organisms
7 are more repellent to proteolysis and denaturation, engineering thermostable variants of biocatalysts on the
8 same mechanism that the nature uses (2–4), is a topic of immense interest for decades (5, 6). Among
9 underlying molecular adaptations behind thermostability, number of charged amino acids is one of the
10 few factors those exhibit a consistent incremental tendency from mesophilic to thermophilic and
11 hyperthermophilic organisms (6). This motivates to look for the possible role of electrostatic interaction
12 in thermostability. Previous studies show that mutations of amino acids in the positions involved in
13 attractive (favourable for stability) electrostatic interactions can destabilize the structure (7), while
14 removal of repulsive (unfavourable for stability) interactions can enhance stability (8). Computational 3D
15 designing and rational optimization of charge-charge interactions for designing thermostable proteins
16 have also been reported in several literature (9–13). Apart from these studies, the instances of charge
17 reversal studies are rare, despite the fact that a single CRM can change local interaction pattern to a large
18 extent (**Fig 1A,B**).

19 Charge reversal is a specific type of mutation where charged amino acid in one sequence is substituted by
20 an amino acid with opposite charge in another sequence of a pair-wise sequence alignment (**Fig 1C**).
21 Previous efforts of charge-reversal didn't find any stabilizing effect. In their study in 1982, Hollecker and
22 Creighton converted all the amino groups to acids groups of a few proteins (14) that can results in the
23 disruption of the salt-bridges, and creation of some new repulsive interactions with the previously present
24 acid groups, hence, found no stabilizing effect. In another study in 2003, Schwehm et. al reversed all the
25 surface charges of a protein and found a destabilizing effect (15), because reversing all the charges at an
26 instance is not a feasible option for gaining thermal stability, as it would disrupt many native interactions
27 and may interfere with the protein's function. So, in the current study we assess whether the CRMs
28 naturally occur in putative M-T/HT orthologous proteins. If yes, what are the possible roles of CRMs in
29 thermal adaptation? To address this, we have used a carefully curated dataset from our previous study of
30 1550 M-T/HT naturally occurring putative orthologous protein pairs that (i) share common evolutionary
31 ancestry and exhibit similar (ii) lengths, (iii) domain content and (iv) 3D topologies. Statistical
32 comparison of the effect of putative CRMs further exhibits that charge reversal of a few buried and
33 moderately exposed amino acids may contribute substantially in enhancing stability, which contradicts
34 with the outcome of previous studies (14, 15). Moreover, there are other stabilizing factors for the surface
35 CRMs, like- enhancement of solvent interaction, which weren't considered in previous studies. This work
36 in its limited scope could be a step forward to understand the broader role of charge reversal in
37 thermostability and provide a new perspective for protein engineers.

38 MATERIALS AND METHODS

39 Data Collection

1 Evolutionary expansion of protein families involves several innovations that directly or indirectly affect
2 their domain architectures, functions and global topologies. So statistical comparison of thermophilic and
3 mesophilic putative orthologous proteins to find out thermal adaptation signature often lead to
4 inconsistent results because it is difficult to assess whether the observed differences between orthologs are
5 associated with thermal adaptations or evolutionary innovations. Taking this into account, in our previous
6 study (6), starting from collecting a set of 3346 mesophilic, 988 thermophilic and 407 hyperthermophilic
7 x-ray crystallographic structures (≤ 3.0 Å resolution, and $\geq 80\%$ protein sequence coverage) from Protein
8 Data Bank (PDB) (16), we constructed a carefully curated datasets of M-T/HT orthologous proteins that
9 not only share common evolutionary ancestry (reciprocal BLAST (17) search with $\geq 50\%$ query sequence
10 coverage and 1^{-10} expected threshold), also exhibit similar lengths (L), 3D topologies and domain
11 architectures. In the current study, we have used a dataset of 1550 M-T/HT orthologous protein pairs
12 (**Data S1**) that exhibit similar length and ACO (18, 19) (ACO, a widely used topological parameter, is
13 calculated as the average separation of contacting amino acids in the primary chain).

14 **Sequence and structural analysis**

15 **Pairwise Sequence Alignment of putative orthologs**

16 Pairwise local alignment on 1550 putative orthologous pairs was performed using Water program (20),
17 implemented in EMBOSS Package (21) with a gap-opening and gap-extension penalty of 10.0 and 0.5
18 respectively. These alignments are used to extract mesophilic-to-thermophilic/hyperthermophilic charge
19 reversal point-mutations within the aligned regions (**Data S1**).

20 **Solvent accessible surface area**

21 Solvent accessible surface area (*SASA*) of charged amino acids in the protein structures is computed using
22 Surface Racer tool (22) using a probe radius of 1.4 Å roughly similar to that of a water molecule.

23 **Charge-charge and Cation- π interactions**

24 The interaction energy (G) for an ion pair is calculated as

$$25 \quad G = 16.583 \frac{q_1 q_2}{r}$$

26 In the above expression, charges (q_1 and q_2) are given in units of electron/proton charge at *pH* 7 (23),
27 distances (r) are given in angstroms (Å), and interaction energies (G) are obtained in kilocalories per
28 mole (kcal/mol); the numerical factor 16.583 is obtained by the vacuum permittivity, proteins' average
29 dielectric constant of 20 (24), electron/proton charges in units of coulomb and the conversion factor for
30 energy from joules to kcal/mol. To calculate the total energy, associated with charge-reversal mutations
31 (CRMs) in a protein, all the putative CRMs are first detected from its alignment and then the algebraic
32 sum of G s of CRMs with their nearby charges are taken as shown in **Fig 1A-C**.

33 Cation- π interaction occurs between positively charged side chains of amino acids, Lysine and Arginine
34 and the aromatic rings of Phenylalanine, Tryptophan and Tyrosine (that are placed approximately 4 Å

1 apart), is believed to be a key contributor to proteins folding and thermal stability (25). Our own in-house
2 python scripts are used to identify these interactions.

3 **Statistical analysis and Figure Preparation**

4 PAST (PAleontological STatistics) software (26) and our own in-house python scripts are used for the
5 statistical analyses. Test used, number of data points and respective p -values are provided in the plots. All
6 the images are produced using OriginPro and Seaborn package (Michael Waskom and the Seaborn
7 development team) of Python 2.7.

8 **RESULTS AND DISCUSSION**

9 In order to function in the extreme environments, thermophilic proteins have been designed differently
10 than mesophilic proteins (1). This designing principle includes changes in the sequence level that is
11 further associated with several secondary and tertiary structural factors of thermostability (6).
12 Thermophilic proteins feature a higher number of charge-charge interactions than their mesophilic
13 orthologs (6, 27–30), which may be attributed to the fact that thermophilic sequences are generally more
14 enriched with charged residues (6, 31, 32). Moreover, each amino acid can participate in several attractive
15 and repulsive interactions simultaneously. Henceforth, the (attractive – repulsive) interactions per charged
16 residue (charge metric (33), CM) may provide a more appropriate signature than counting charge-charge
17 interactions. Here, we have calculated the CM for thermo (TCM) and meso (MCM) for varying distance
18 cut-offs from 4 Å to 15 Å, and we observe that TCM and MCM significantly differ up to a distance of 12
19 Å (Two sample t-test p value ≤ 0.01) (**Fig 1D**). Since interaction energy of any charge amino acid pair at
20 11 and 12 Å becomes negligibly smaller (**Supplementary Fig 1**), distance cut-offs up to 10 Å are
21 considered for further study.

22 **CRMs possess more stabilizing energy in thermophilic proteins at short range**

23 We started with extracting all the meso to thermo charge-reversal point mutations for all 1550
24 orthologous pairs from their alignments (**Fig 1C**). Net interaction energy (IE , where $IE = \Sigma G$) for
25 amino acids in mesophiles are distributed in two groups: R_M for reversed and NR_M for not reversed,
26 similarly, R_T and NR_T for thermophiles. Mann Whitney U (MW) test is performed for R_M, R_T and
27 NR_M, NR_T for all distance cutoffs from 4 Å to 10 Å. At every cutoff, the MW p value of comparison
28 in R_M and R_T is more significant than that of NR_M and NR_T (**Fig 1E, Supplementary Fig 2A-D**).
29 Thermophilic proteins are enriched with charged residues (31, 32) whose positions in the 3D structure are
30 such that in turns they create more stabilizing charge-charge interactions (27–30) than mesophilic
31 proteins. Among all the charged residues in thermo, putative CRMs contribute even more in stabilizing
32 energy gain than the rest, and this trend is stronger at short distance cutoffs (**Fig 1E, Supplementary Fig**
33 **2A-D**).

34 A comparison of IEs of the common CRMs in different distance bins of 0-4, 4-6, 6-8, 8-10 Å shows that
35 the average interaction energy gain ($IE_T - IE_M$) per CRM (-1.54 kcal/mol, -1.36 kcal/mol, -0.39
36 kcal/mol, and -0.14 kcal/mol at the bins respectively) significantly differ from one another (Kruskal
37 Wallis, (KW) p value = $5.07e^{-21}$) (**Fig 1F**). This indicates that CRMs do not contribute equally at every
38 distance and the overall energy gain arises mainly from the interactions at a shorter distance (≤ 6 Å).

1 Beyond that distance the attractive and repulsive interaction energies associated to the CRMs nearly
2 counter balance each other in both thermophilic and mesophilic proteins. Since, more than one CRM may
3 occur in a meso-thermo orthologous protein pair, how they collectively contribute in the stability of the
4 protein will give more realistic picture of their role in thermostability.

5 To understand how the CRMs contribute in thermostability at protein level we estimated two parameters
6 for each meso-thermo orthologous pair, $E_M = \sum IE_M^j$ and $E_T = \sum IE_T^j$ that represent the total interaction
7 energy for meso and thermo, where IE_M^j and IE_T^j stand for the energy associated with j^{th} CRM in meso
8 and thermo respectively. In these comparisons, our null hypothesis H_0 is $E_T = E_M$ and the alternative
9 hypothesis H_1 is $E_T > E_M$ for the statistical test at each distance bins. The difference is more significant
10 for the 0-4 Å (MW p value = $1.44e^{-73}$) and 4-6 Å (MW p value = $3.06e^{-71}$) bins compared to 6-8 Å (MW
11 p value = $4.32e^{-52}$) and 8-10 Å (MW p value = $7.04e^{-15}$) bins (**Fig 2A, Supplementary Fig 3A-D**). This
12 result indicates that the contribution of CRMs is limited to a larger extent on the short-range interaction
13 compared to long-range interactions. The difference of E_T and E_M ($E_T - E_M$) plotted for the distance
14 bins and their statistical comparison (Kruskal Wallis (KW) p value = $3.62e^{-74}$) further supports the
15 previous result (**Fig 2B**).

16 **Majority of the total stabilizing energy gain appears from the short range interaction by** 17 **buried and moderately exposed CRMs**

18 Charged amino acids prefer to occur near the surface in both mesophilic and thermophilic. However, this
19 tendency is stronger in thermophiles as it contributes by increasing solvent interactions (34–36). On the
20 other hand, burial of charged residue comes with desolvation entropic penalty which may be reduced in
21 thermophilic by proper networking of charged residue (37). To investigate where in the structure the
22 CRMs occur, we distributed all the CRMs in three groups based on their solvent accessibility i) buried
23 (Charged solvent accessible surface area, $C_SASA \leq 35\%$, M_0-35 and T_0-35), ii) moderately exposed
24 ($35\% < C_SASA \leq 70\%$, M_35-70 and T_35-70), and iii) exposed ($70\% < C_SASA \leq 100\%$, M_70-100
25 and T_70-100), where M and T stand for meso and thermo respectively. Number of mutations occur in
26 the moderately exposed group account for ~55% of the total mutation in both meso and thermo (**Fig 2C**).
27 However, CRMs may fall in other groups in meso and thermo. Still the amino acids that remain
28 moderately exposed in both (36.2%) account for more than two fold than the next populated group, where
29 they are buried in both (16.6%) (**Fig 2C**). Total energy comparison of these sets of mutations for
30 orthologous meso-thermo pairs, using the same methodology described above, demonstrates that
31 optimization of the short-range interactions occurs predominantly at the moderately exposed and buried
32 segments of the protein structure (**Fig 2D**). Here, we hypothesise that rational optimization of moderately
33 exposed short range electrostatic interaction by charge-reversal may be used for engineering thermostable
34 variant of mesophilic proteins.

35 **Additional factors**

36 However, only 75.3%, 72.5%, 66.3% and 57.6% of the total orthologous pairs exhibit energy gain
37 associated with CRMs at 0-4, 4-6, 6-8, and 8-10 Å bins (**Supplementary Fig 4**) respectively. For those
38 who possess energy loss we looked for other counter balancing factors. Here, we limit our discussion to

1 the effect of CRMs only. Comparison of cation- π contacts for energy loss associated M-T/HT orthologs
2 exhibit no significance difference. Another possibility is the interaction of the charged amino acids with
3 its solvent environment. Exposure of charged groups on protein surface facilitates thermal stability by
4 blocking penetration of the solvent within the protein. This is achieved by a highly connected network of
5 water-water hydrogen bonds, attached to the protein surface by amino acids-water hydrogen bonds (34–
6 36). Irrespective of energy gain or loss cumulative solvent accessibility of charge residues in thermophilic
7 proteins exhibit more C_{SASA} compared to those in their mesophilic orthologs (**Fig 3A**). However, when
8 considered individually, energetically destabilizing CRMs (i.e. those associated with interaction energy
9 loss) in thermophilic proteins exhibit more exposed surface than stabilizing CRMs, indicating a counter
10 balancing role of C_{SASA} locally (**Fig 3B**).

11 In our study, working on a large, carefully curated dataset, we aim to understand, in general, how CRMs
12 contribute in thermostability. However, when we calculate the charge-charge interaction energy, all the
13 interactions are treated equally, although several molecular dynamics and experimental studies show that
14 the interaction energy depends on several other factors like location of the charges on the ions, positions
15 of the amino acids in the structure, dielectric constant, folded and denatured state of the protein etc (33,
16 38, 39). Moreover, dielectric constant varies from core to surface and also with temperature (40).
17 However the inclusion of the effect of the variation of dielectric constant and temperature will not alter
18 the conclusion of our study.

19 **Conclusion**

20 While strategies for attaining thermal stability and its application in protein engineering have been studied
21 for decades, CRMs remain unpracticed. In our study, statistical comparison of sequence and structural
22 features and semi-empirical energetics of 1550 M-T/HT orthologous protein pairs with remarkable
23 structural and topological similarity reveals the following key points that could be useful for protein
24 engineers as a future reference.

- 25 1. CRMs prevalently occur in naturally occurring putative orthologous pairs. Not only that, they
26 contribute subsequently to protein stability by altering short distance charge-charge interactions.
27 Hence, CRM acts as one of the key mechanisms for gaining thermal stability.
- 28 2. Positional comparison of interaction energy of the CRMs exhibits higher energy gain for buried
29 and partially exposed CRMs compared to those who are exposed to the solvent environment.
- 30 3. CRMs associated with energy loss are more solvent-exposed than the rests, indicating a higher
31 role of solvent interaction in the former.

32 **Acknowledgement**

33 The authors would like to acknowledge Raktim Maity and Soumya Kundu for their valuable comments
34 during the discussions.

35 **Author Contribution**

36 S.H. and S.K. designed research; S.H. implemented computational methodologies, S.H. and S.B.
37 performed research; S.H., S.B. and S.K. analyzed data and wrote the paper.

38 **Funding**

1 S.H. is supported by DBT-BINC SRF fellowship (Fellow Number: DBT-BINC/2017/CU/12).

2 **References**

- 3 1. Rigoldi, F., S. Donini, A. Redaelli, E. Parisini, and A. Gautieri. 2018. Review: Engineering of
4 thermostable enzymes for industrial applications. *APL Bioeng.* 2.
- 5 2. Scandurra, R., V. Consalvi, R. Chiaraluce, L. Politi, and P.C. Engel. 1998. Protein thermostability
6 in extremophiles. *Biochimie.* 80:933–941.
- 7 3. Van Den Burg, B., G. Vriend, O.R. Veltman, G. Venema, and V.G.H. Eijssink. 1998. Engineering
8 an enzyme to resist boiling. *Proc. Natl. Acad. Sci. U. S. A.* 95:2056–2060.
- 9 4. Van den Burg, B. 2003. Extremophiles as a source for novel enzymes. *Curr. Opin. Microbiol.*
10 6:213–218.
- 11 5. Pucci, F., and M. Rooman. 2017. Physical and molecular bases of protein thermal stability and
12 cold adaptation. *Curr. Opin. Struct. Biol.* 42:117–128.
- 13 6. Hait, S., S. Mallik, S. Basu, and S. Kundu. 2020. Finding the generalized molecular principles of
14 protein thermal stability. *Proteins Struct. Funct. Bioinforma.* 88:788–808.
- 15 7. Tissot, A.C., S. Vuilleumier, and A.R. Fersht. 1996. Importance of two buried salt bridges in the
16 stability and folding pathway of barnase. *Biochemistry.* 35:6786–6794.
- 17 8. Koide, A., M.R. Jordan, S.R. Horner, V. Batori, and S. Koide. 2001. Stabilization of a fibronectin
18 type III domain by the removal of unfavorable electrostatic interactions on the protein surface.
19 *Biochemistry.* 40:10326–10333.
- 20 9. Spassov, V.Z., A.D. Karshikoff, and R. Ladenstein. 1994. Optimization of the electrostatic
21 interactions in proteins of different functional and folding type. *Protein Sci.* 3:1556–1569.
- 22 10. Loladze, V. V., B. Ibarra-Molero, J.M. Sanchez-Ruiz, and G.I. Makhatadze. 1999. Engineering a
23 thermostable protein via optimization of charge-charge interactions on the protein surface.
24 *Biochemistry.* 38:16419–16423.
- 25 11. Strickler, S.S., A. V. Gribenko, A. V. Gribenko, T.R. Keiffer, J. Tomlinson, T. Reihle, V. V.
26 Loladze, and G.I. Makhatadze. 2006. Protein stability and surface electrostatics: A charged
27 relationship. *Biochemistry.* 45:2761–2766.
- 28 12. Gribenko, A. V., and G.I. Makhatadze. 2007. Role of the Charge-Charge Interactions in Defining
29 Stability and Halophilicity of the CspB Proteins. *J. Mol. Biol.* 366:842–856.
- 30 13. Gribenko, A. V., M.M. Patel, J. Liu, S.A. McCallum, C. Wang, and G.I. Makhatadze. 2009.
31 Rational stabilization of enzymes by computational redesign of surface charge-charge interactions.
32 *Proc. Natl. Acad. Sci. U. S. A.* 106:2601–2606.
- 33 14. Hollecker, M., and T.E. Creighton. 1982. Effect on protein stability of reversing the charge on
34 amino groups. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* 701:395–404.
- 35 15. Schwehm, J.M., C.A. Fitch, B.N. Dang, E. Bertrand García-Moreno, and W.E. Stites. 2003.
36 Changes in stability upon charge reversal and neutralization substitution in staphylococcal
37 nuclease are dominated by favorable electrostatic effects. *Biochemistry.* 42:1118–1128.

- 1 16. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and
2 P.E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
- 3 17. Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T.L. Madden.
4 2009. BLAST+: Architecture and applications. *BMC Bioinformatics.* 10.
- 5 18. Ivankov, D.N., S.O. Garbuzynskiy, E. Alm, K.W. Plaxco, D. Baker, and A. V. Finkelstein. 2003.
6 Contact order revisited: Influence of protein size on the folding rate. *Protein Sci.* 12:2057–2062.
- 7 19. Mallik, S., and S. Kundu. 2018. Topology and Oligomerization of Mono- and Oligomeric Proteins
8 Regulate Their Half-Lives in the Cell. *Structure.* 26:869-878.e3.
- 9 20. Smith, T.F., and M.S. Waterman. 1981. Identification of common molecular subsequences. *J. Mol.*
10 *Biol.* 147:195–197.
- 11 21. Rice, P., L. Longden, and A. Bleasby. 2000. EMBOSS: The European Molecular Biology Open
12 Software Suite. *Trends Genet.* 16:276–277.
- 13 22. Tsodikov, O. V., M. Thomas Record, and Y. V. Sergeev. 2002. Novel computer program for fast
14 exact calculation of accessible and molecular surface areas and average surface curvature. *J.*
15 *Comput. Chem.* 23:600–609.
- 16 23. Cameselle, J., J. Meireles Ribeiro, and A. Sillero. 1986. Derivation and use of a formula to
17 calculate the net charge of acid-base compounds. Its application to amino acids, proteins and
18 nucleotides. *Biochem. Educ.* 14.
- 19 24. Williams, J.C., A.L.M. Haffa, J.L. McCulley, N.W. Woodbury, and J.P. Allen. 2001. Electrostatic
20 interactions between charged amino acid residues and the bacteriochlorophyll dimer in reaction
21 centers from rhodobacter sphaeroides. *Biochemistry.* 40:15403–15407.
- 22 25. Dougherty, D.A. 1996. Cation- π interactions in chemistry and biology: A new view of benzene,
23 Phe, Tyr, and Trp. *Science (80-).* 271:163–168.
- 24 26. Hammer, D.A.T., P.D. Ryan, Ø. Hammer, and D.A.T. Harper. 2001. Past: Paleontological
25 Statistics Software Package for Education and Data Analysis. .
- 26 27. Folch, B., Y. Dehouck, and M. Rooman. 2010. Thermo- and mesostabilizing protein interactions
27 identified by temperature-dependent statistical potentials. *Biophys. J.* 98:667–677.
- 28 28. Kumar, S., C.J. Tsai, and R. Nussinov. 2000. Factors enhancing protein thermostability. *Protein*
29 *Eng.* 13:179–191.
- 30 29. Elcock, A.H. 1998. The stability of salt bridges at high temperatures: Implications for
31 hyperthermophilic proteins. *J. Mol. Biol.* 284:489–502.
- 32 30. Dominy, B.N., H. Minoux, and C.L. Brooks. 2004. An electrostatic basis for the stability of
33 thermophilic proteins. *Proteins Struct. Funct. Genet.* 57:128–141.
- 34 31. Zhou, X.X., Y.B. Wang, Y.J. Pan, and W.F. Li. 2008. Differences in amino acids composition and
35 coupling patterns between mesophilic and thermophilic proteins. *Amino Acids.* 34:25–33.
- 36 32. Pack, S.P., and Y.J. Yoo. 2004. Protein thermostability: Structure-based difference of amino acid
37 between thermophilic and mesophilic proteins. *J. Biotechnol.* 111:269–277.

- 1 33. Sawle, L., J. Huihui, and K. Ghosh. 2017. All-Atom Simulations Reveal Protein Charge
2 Decoration in the Folded and Unfolded Ensemble Is Key in Thermophilic Adaptation. *J. Chem.*
3 *Theory Comput.* 13:5065–5075.
- 4 34. Sterpone, F., C. Bertonati, G. Briganti, and S. Melchionna. 2009. Key role of proximal water in
5 regulating thermostable proteins. *J. Phys. Chem. B.* 113:131–137.
- 6 35. Mou, Z., Y. Ding, X. Wang, and Y. Cai. 2014. Comparison of Protein-water Interactions in
7 Psychrophilic, Mesophilic, and Thermophilic Fe-SOD. *Protein Pept. Lett.* 21:578–583.
- 8 36. Sterpone, F., C. Bertonati, G. Briganti, and S. Melchionna. 2010. Water around thermophilic
9 proteins: The role of charged and apolar atoms. *J. Phys. Condens. Matter.* 22.
- 10 37. Zhou, H.X. 2002. Toward the physical basis of thermophilic proteins: Linking of enriched polar
11 interactions and reduced heat capacity of unfolding. *Biophys. J.* 83:3126–3133.
- 12 38. Grimsley, G.R., K.L. Shaw, L.R. Fee, R.W. Alston, B.M.P. Huyghues-Despointes, R.L. Thurlkill,
13 J.M. Scholtz, and C.N. Pace. 1999. Increasing protein stability by altering long-range coulombic
14 interactions. *Protein Sci.* 8:1843–1849.
- 15 39. Nick Pace, C., R.W. Alston, and K.L. Shaw. 2000. Charge-charge interactions influence the
16 denatured state ensemble and contribute to protein stability. *Protein Sci.* 9:1395–1398.
- 17 40. Pitera, J.W., M. Falta, and W.F. Van Gunsteren. 2001. Dielectric properties of proteins from
18 simulation: The effects of solvent, ligands, pH, and temperature. *Biophys. J.* 80:2546–2555.

19

20

21

22

23

24

25

26

27

28

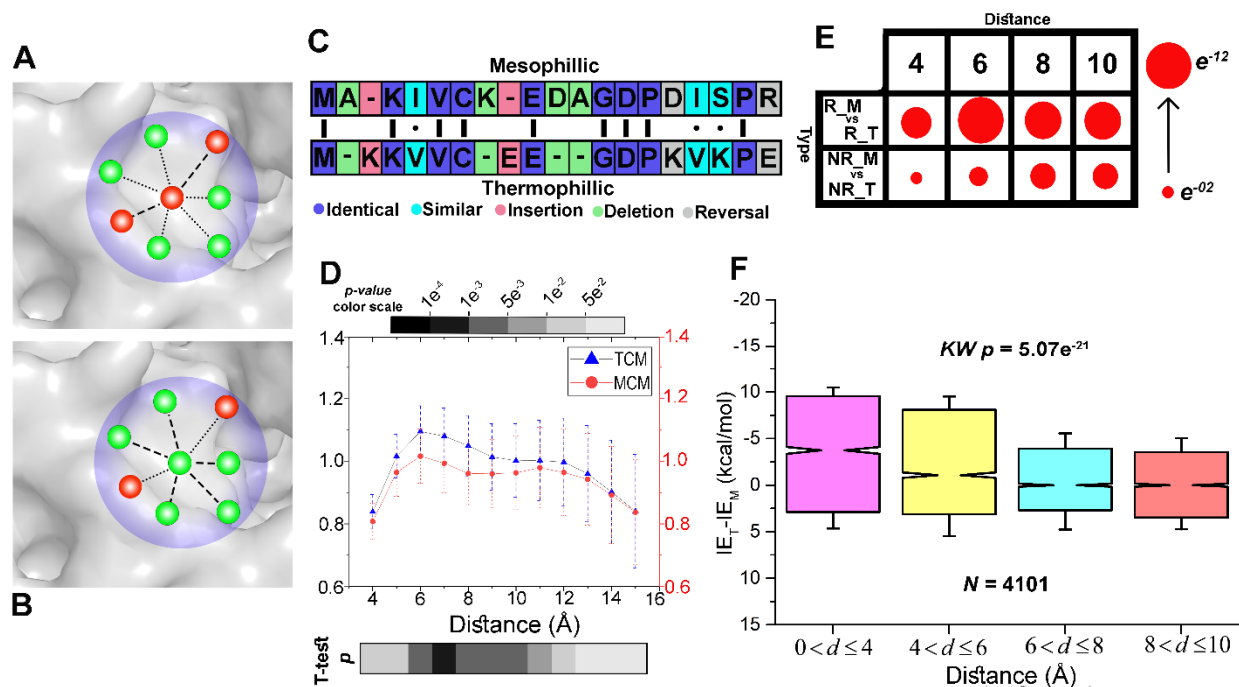
29

30

31

1 **FIGURES AND FIGURE LEGENDS:**

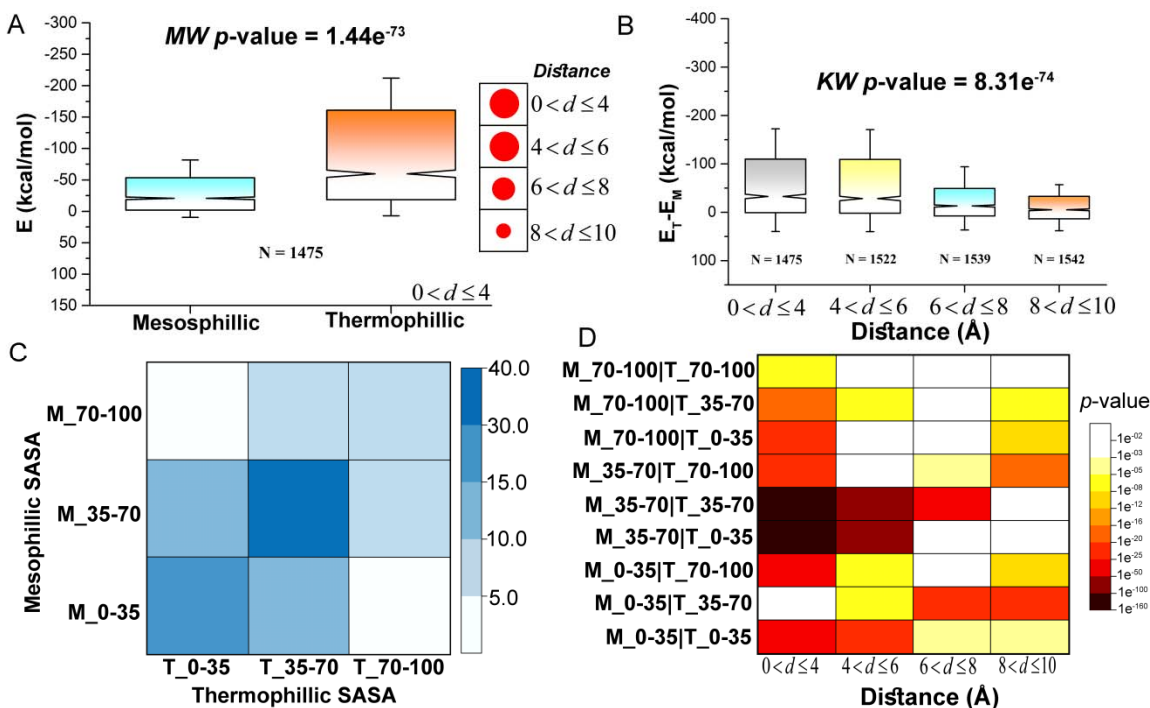
2 **Figure 1:**



3
 4 **Fig 1. (A-B)** A schematic diagram showing how a single charge reversal can completely change the
 5 existing interaction pattern. Here, the positive and negative charges are represented as red and green
 6 spheres respectively. The attractive and the repulsive interactions are represented as dotted and dashed
 7 lines respectively. **(C)** It shows how putative charge reversal mutations are found from the pairwise
 8 alignment of mesophilic and thermophilic orthologous proteins. **(D)** Comparison of charge-metrics at
 9 different distance cut-offs in M-T/HT orthologous proteins. The t-test p value for each distance cut-off is
 10 represented by a grey scale gradient shown below. **(E)** Interaction energies of charged reversed and non-
 11 reversed amino acids in thermophilic proteins have been statistically compared with those in mesophilic
 12 proteins using Mann-Whitney U (MW) test. MW p values are represented as red circles, where size of the
 13 circles represent approximately the level of significance (respective box plots and their p values are
 14 provided in **Supplementary Fig 2A-D**). **(F)** Thermo-meso energy difference ($IE_T - IE_M$) associated
 15 with the same set of CRMs at different distance bins (d) are statistically compared using Kruskal Wallis
 16 test.

17
 18
 19
 20
 21
 22
 23

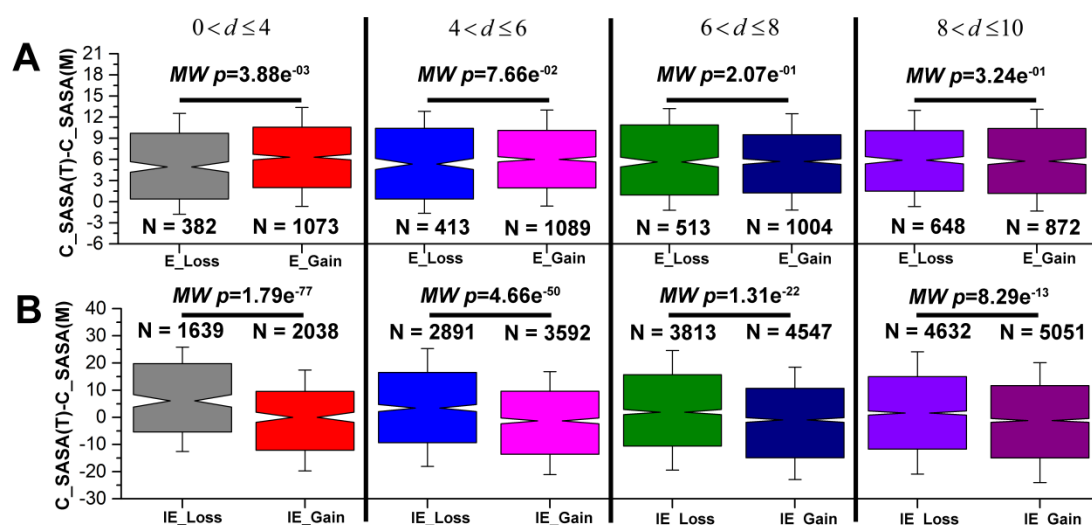
1 **Figure 2:**



2
3 **Fig 2. (A)** Statistical comparison (MW test) and its p value of total energy associated with CRMs in
4 mesophilic proteins and their respective thermophilic orthologs for the distance bin 0-4 Å is provided. For
5 the rest of the distance bins the p values (represented as red circles, where there sizes roughly represent
6 the significance level) of MW test are provided alongside. Box plots for all the distance bins are given in
7 **Supplementary Fig 3A-D.** **(B)** Stabilizing energy gain associated with CRMs for the M-T/HT
8 orthologous proteins at different distance bins are compared using Kruskal Wallis test. **(C)** Occurrence of
9 CRMs in different Solvent Accessible Surface Area (SASA) ranges in M-T/HT orthologous proteins are
10 represented as heatmap. A darker shade of blue represents higher percentage of occurrence. **(D)** MW p
11 values of comparisons of interaction energies associated to CRMs for M-T/HT orthologous pairs at
12 different SASA ranges at different distance bins are represented as heatmap. Level of significance
13 increases from white to dark-brown.

14
15
16
17
18
19
20
21
22

1 **Figure3:**



2
 3 **Fig 3. (A)** Difference of the percentage of SASA occupied by charged amino acids in thermophilic and
 4 mesophilic orthologous proteins are plotted for the pairs that exhibit a gain and loss in interaction energy
 5 due to CRMs at different distance bins. MW test is performed for each set and their *p* values are provided.
 6 **(B)** Difference of the percentage of area of CRMs, exposed to the surface in thermo and meso are
 7 compared where a gain and loss in interaction energy is found due to CRMs at different distance bins.
 8 MW test is performed for each set and their *p* values are provided.