Energy landscapes and heat capacity signatures for peptides correlate with phase separation propensity

Nicy,[1] Rosana Collepardo-Guevara,[1,2,3] Jerelle A. Joseph*,[4,a)] and David J. Wales*[1,b)]

[1)] *Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK.*

[2)] *Department of Physics, University of Cambridge, Cambridge, UK.*

[3)] *Department of Genetics, University of Cambridge, Cambridge, UK.*

[4)] *Department of Chemical and Biological Engineering, Princeton University, USA.*

(Dated: 13 July 2023)

[a)]Electronic mail: jerellejoseph@princeton.edu

[b)]Electronic mail: dw34@cam.ac.uk

SUPPLEMENTARY INFORMATION

Effect of capping peptides: The methyl group at the C-terminal cap may interact with aromatic residues (GFGGFG, and GYGGYG). The polar contacts established between caps may constrain the structure, and the loss of these polar contacts may help the stickers interact with each other better (GRGGFG). When the cap interacts with a sticker, the distance between the stickers may be larger to minimise the repulsive interaction between the cap and the other sticker. In contrast, when the cap is not interacting with the sticker, it may be possible for the stickers to establish better contact by reducing the distance between them (GFGGFG, GYGGYG, GKGGFG).

Weak interactions in polymorphism Several weak interactions play an important role in driving phase separation, such as quadrupole-quadrupole (Burley and Petsko, 1985), charge-quadrupole interactions (Burley and Petsko, 1986), and hydrogen-bonding (Burley and Petsko, 1988). These weaker interactions are anisotropic, and for the same pair of residues these can be alternative favourable conformations, which may then pack into different arrangements (Berkovitch-Yellin and Leiserowitz, 1984), possibly leading to the existence of different phases. This suggestion is consistent with the conformational polymorphism (Nangia, 2008) seen in organic crystals. While the distinction between competing intramolecular and intermolecular interactions is beyond the scope of the current study, the smaller and larger distances between the interacting stickers within the monomers may be useful in visualising the various possible ways in which the interactions may occur in oligomeric systems. Polymorphism is also known to occur in amino acids (Purohit and Venugopalan, 2009), and protein complexes (Tompa and Fuxreiter, 2008), and H-bonding ability is important in polymorphic molecules (Cruz-Cabeza et al., 2015; Kumar and Nangia, 2014). The stability of non-linear H-bonds may not differ significantly from the collinear H-bonds (Baker and Hubbard, 1984). Metastable conformers may be stabilised by favourable packing arrangements (Babu et al., 2010), and some polymorphs are known to have different strengths for $\pi-\pi$ interactions (Fan et al., 2009).

Hexapeptide	Peak/Inflection(*) temperatures			
GFGGFG	0.158*	0.323*	0.493	-
GYGGYG	0.029	0.143	0.349	-
GRGGYG	0.028	0.053	0.495	-
GKGGYG	0.045	0.203	0.373	-
GRGGFG	0.022	0.450	-	-
GKGGFG	0.232	0.542	-	-
FGGGGF	0.098	0.415	0.600	-
YGGGGY	0.061*	0.194*	0.565	
RGGGGY	0.120*	0.474	0.676	-
KGGGGY	0.011	0.107	0.322	0.712
RGGGGF	0.128*	0.462	-	-
KGGGGF	0.105	0.252	0.625	-
FFGGFF	0.363	0.622	-	-
YYGGYY	0.012	0.133	0.464*	0.630
FLGGLF	0.131*	0.348*	0.736	-
RYGGYR	0.023^{i}	0.096	0.273	0.836
KYGGYK	0.073	0.380	0.734	-
YKGGKY	0.057	0.333	0.862	-
REGGER	0.325	0.690	-	-
KEGGEK	0.378	0.962	-	-
AAGGAA	0.227	0.354	-	-

TABLE I: $k_{\mathrm{B}}T$ in kcal mol^{-1} at which peaks and/or distinct inflection points are observed for selected peptides. The inflection points are marked with an asterisk (*) and one very small peak is marked with 'i' (^{i}).

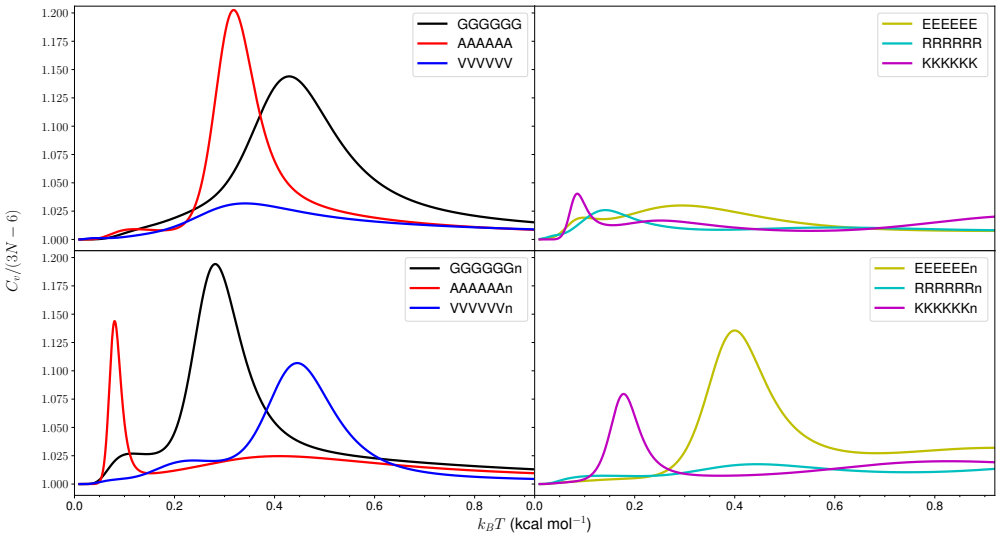

FIG. 1: Heat capacity versus $k_{\mathrm{B}}T$ for various hexapeptides. The uncapped peptide sequences end with a lowercase 'n'.

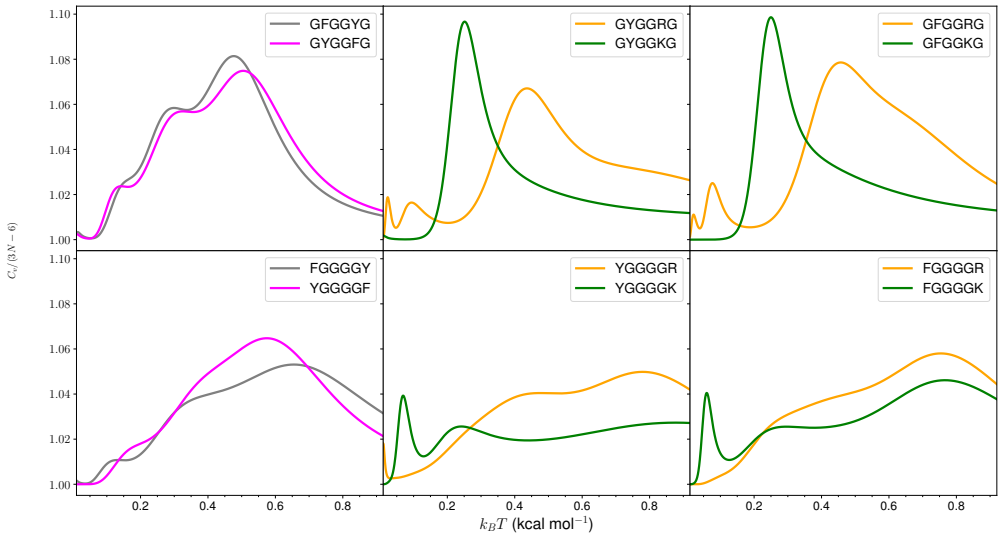

FIG. 2: Heat capacity versus $k_{\mathrm{B}}T$ for various hexapeptides.

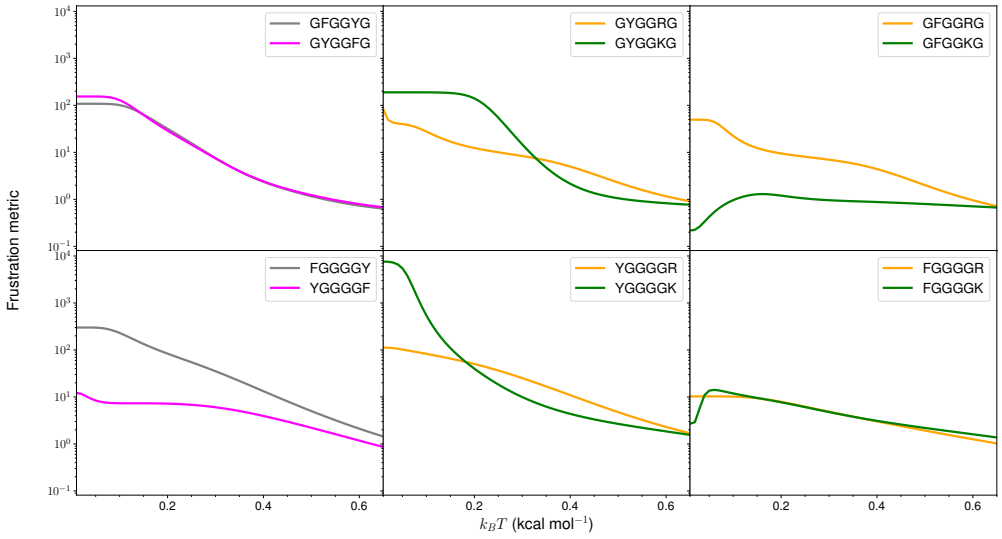

FIG. 3: Frustration metric $[\widetilde{f}(T)]$ versus $k_{\mathrm{B}}T$ for various hexapeptides.

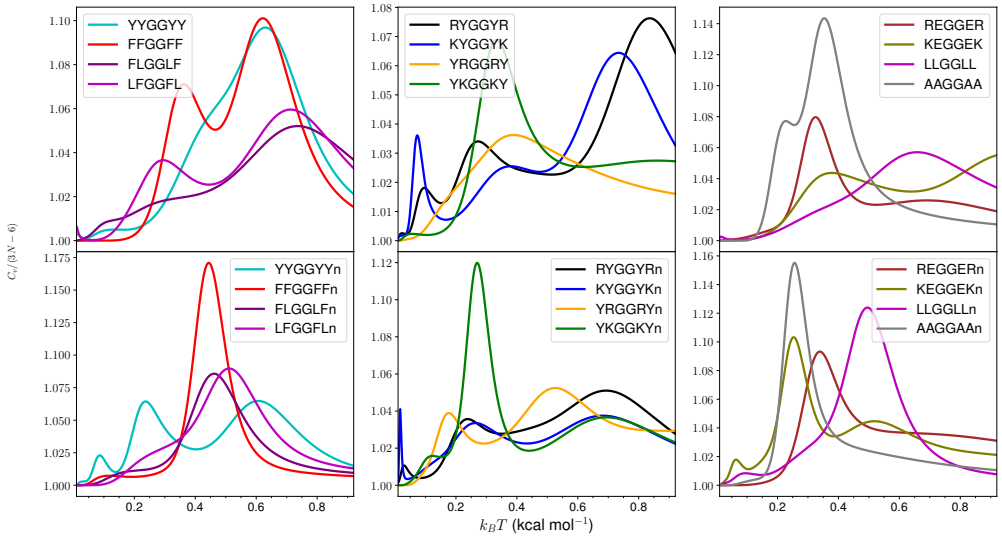

FIG. 4: Heat capacity versus $k_{\mathrm{B}}T$ for various hexapeptides. The uncapped peptide sequences end with a lowercase 'n'.

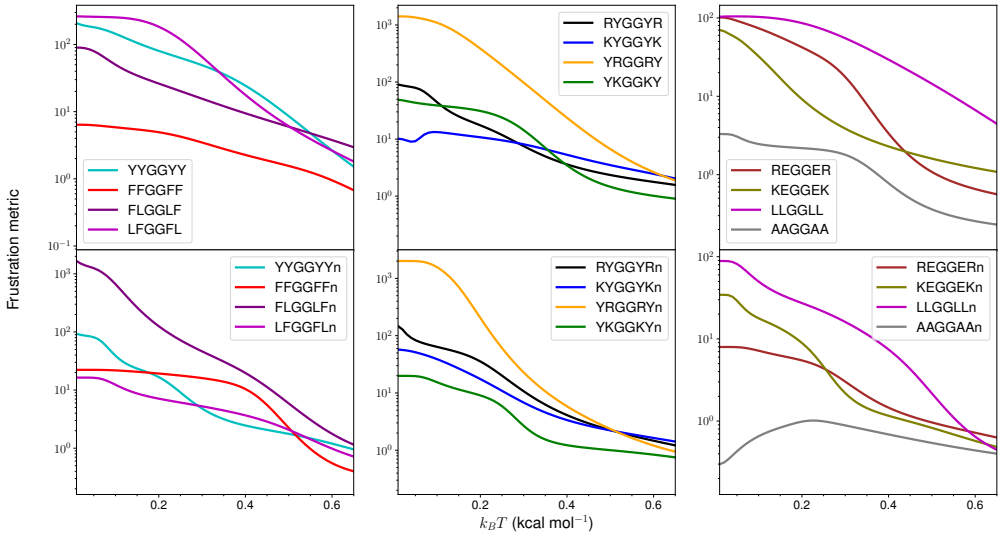

FIG. 5: Frustration metric $[\widetilde{f}(T)]$ versus $k_{\mathrm{B}}T$ for various hexapeptides. The uncapped peptide sequences end with a lowercase 'n'.

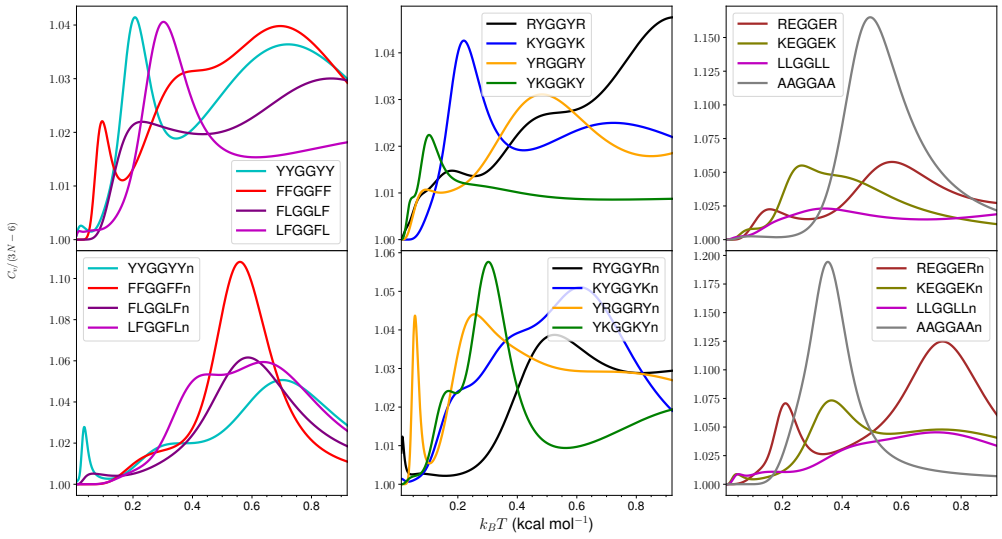

FIG. 6: Heat capacity versus $k_{\mathrm{B}}T$ for various hexapeptides modelled using FF19SB force field in AMBER20. The uncapped peptide sequences end with a lowercase 'n'.

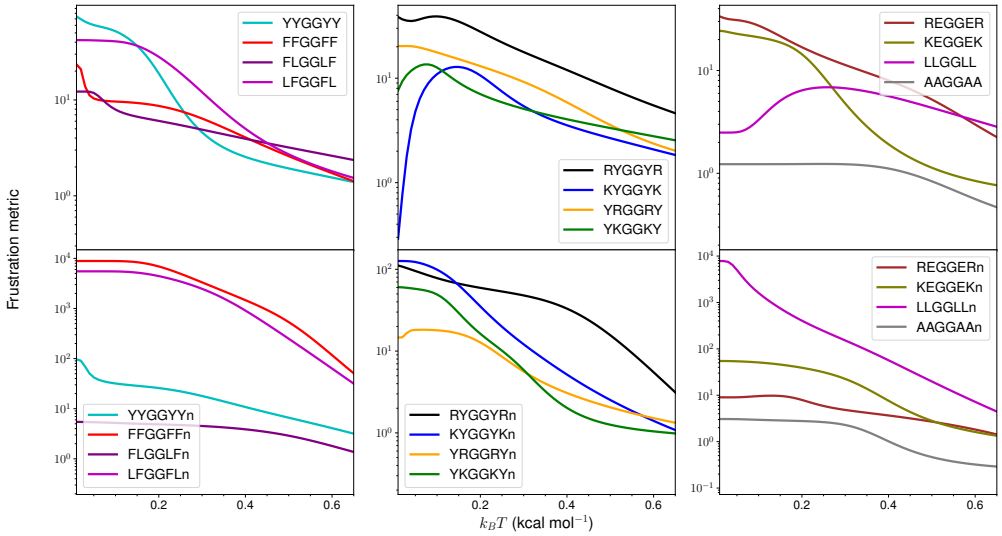

FIG. 7: Frustration metric $[\widetilde{f}(T)]$ versus $k_{\mathrm{B}}T$ for various hexapeptides modelled using FF19SB force field in AMBER20. The uncapped peptide sequences end with a lowercase 'n'.

Hexapeptide	Peak/Inflection(*) temperatures			
GFGGYG	0.014	0.165*	0.299	0.477
GYGGFG	0.012	0.142	0.325	0.505
GYGGRG	0.023	0.094	0.436	-
GYGGKG	0.250	-	-	-
GFGGRG	0.020	0.076	0.457	-
GFGGKG	0.250	-	-	-
FGGGGY	0.132	0.430*	0.622	-
YGGGGF	0.197*	0.574	-	-
YGGGGR	0.475	0.779	-	-
YGGGGK	0.069	0.242	0.743	-
FGGGGR	0.468*	0.754	-	-
FGGGGK	0.060	0.296	0.766	-
LFGGFL	0.293	0.703	-	-
YRGGRY	0.386	-	-	-
LLGGLL	0.015^{i}	0.355*	0.659	-
FFGGFFn	0.117	0.444	-	-
YYGGYYn	0.033*	0.086	0.237	0.607
LFGGFLn	0.287*	0.511	-	-
FLGGLFn	0.208*	0.461	-	-
RYGGYRn	0.030	0.239	0.693	-
KYGGYKn	0.017	0.264	0.682	-
YRGGRYn	0.175	0.518	0.876*	0.918
YKGGKYn	0.123	0.270	0.694	-
REGGERn	0.339	-	-	-
KEGGEKn	0.064	0.253	0.518	-
LLGGLLn	0.095	0.470	-	-
AAGGAAn	0.256	-	-	-

TABLE II: $k_{\mathrm{B}}T$ at which peaks and/or distinct inflection points are observed for selected hexapeptides. The inflection points are marked with an asterisk (*) and a very small peak is marked with 'i' (^{i}). The uncapped peptide sequences end with a lowercase 'n'.

Hexapeptide	Peak/Inflection(*) temperatures			
YYGGYY	0.025	0.207	0.721	-
FFGGFF	0.096	0.413*	0.695	-
FLGGLF	0.230	0.866	-	-
LFGGFL	0.020^{i}	0.304	0.807	-
RYGGYR	0.088*	0.180	0.538*	0.928
KYGGYK	0.220	0.724	-	-
YRGGRY	0.094	0.445	-	-
YKGGKY	0.131	1.078	-	-
REGGER	0.157	0.568	-	-
KEGGEK	0.104*	0.263	0.372*	-
LLGGLL	0.059*	0.235	0.535*	0.885
AAGGAA	0.494	-	-	-
YYGGYYn	0.036	0.368*	0.702	-
FFGGFFn	0.290*	0.536	-	-
FLGGLFn	0.067	0.587	-	-
LFGGFLn	0.448*	0.599	-	-
RYGGYRn	0.014	0.076^{i}	0.525*	0.670
KYGGYKn	0.012	0.222*	0.417*	0.611
YRGGRYn	0.056	0.256	0.674*	-
YKGGKYn	0.056*	0.169	0.299	-
REGGERn	0.209	0.737	-	-
KEGGEKn	0.046	0.364	0.734	-
LLGGLLn	0.052	0.171*	0.438*	0.656
AAGGAAn	0.352	-	-	-

TABLE III: $k_{\mathrm{B}}T$ at which peaks and/or distinct inflection points are observed for selected hexapeptides using the FF19SB force field in AMBER20. The inflection points are marked with an asterisk (*) and two very small peaks are marked with 'i' (^{i}). The uncapped peptide sequences end with a lowercase 'n'.

Hexapeptide	Peak temperatures		
GGGGGG	0.429	-	-
AAAAAA	0.121	0.318	-
VVVVVV	0.042	0.341	-
RRRRRR	0.141	0.596	-
KKKKKK	0.085	0.251	1.021
EEEEEE	0.098	0.294	-
GGGGGGn	0.116	0.281	-
AAAAAAn	0.08	0.409	-
VVVVVVn	0.238	0.445	-
RRRRRRn	0.137	0.445	-
KKKKKKn	0.178	0.828	-
EEEEEEn	0.56	0.923	-

TABLE IV: $k_{\mathrm{B}}T$ at which peaks and/or distinct inflection points are observed for selected hexapeptides using the FF99IDPS force field. The uncapped peptide sequences end with a lowercase 'n'.

Computational protocol

The AMBER input files can be created by following the method given at https://wikis.ch.cam.ac.uk/ro-walesdocs/wiki/index.php/Preparing_input_files_for_a_peptide_using_AMBER. The steps to clone the softwarewales repository are given at https://wikis.ch.cam.ac.uk/ro-walesdocs/wiki/index.php/Git_Workflow. The steps to obtain the A12GMIN, A12OPTIM, and PATHSAMPLE executables are given at https://wikis.ch.cam.ac.uk/ro-walesdocs/wiki/index.php/Compiling_Wales_Group_codes_using_cmake. The example input files for running basin-hopping parallel tempering and discrete path sampling are given in the next section. The explanation of all the keywords can be found at https://www-wales.ch.cam.ac.uk/GMIN.doc/node7.html for the data file, https://www-wales.ch.cam.ac.uk/OPTIM.doc/node4.html for the odata.connect and odata.addminxyz files, and https://www-wales.ch.cam.ac.uk/PATHSAMPLE.2.1.doc/node6.html for the pathdata file. The explanation of how to add the minima obtained using A12GMIN to a PATHSAMPLE database is given at https://wikis.ch.cam.ac.uk/ro-walesdocs/wiki/index.php/

Adding_several_minima_obtained_using_GMIN_(maybe_using_BHPT)_to_min.data. The creation of input files and executables required for heat capacity analysis is given at https://wikis.ch.cam.ac.uk/ro-walesdocs/wiki/index.php/Decoding_heat_capacity_curves. The disconnectivity graph can be constructed using the disconnectionDPS program documented at https://www-wales.ch.cam.ac.uk/disconnectionDPS.doc/. Example input files used for one of the peptides are given below.

Input files

```
1. data file for BHPT
MPI
! 300 K to 575 K
! 16 replicas determined by the number of nodes
BHPT 0.5962 1.142 10
STEPS 100000 1.0
! Step size can vary based on the peptide.
STEP 0.8 0.0
EDIFF 0.01D0
GROUPROTATION
RADIUS 1000.0
SLOPPYCONV 1.0D-4
TIGHTCONV 1.0D-7
MAXERISE 1.0D-4
DUMPINT 10000
UPDATES 1500
MAXIT 3000 5000
MAXBFGS 0.2D0
SAVE 400
DUMPSTRUCTURES
FIXBOTH
AMBER12
```

2. odata.addminxyz file for adding minima found using BHPT to a PATHSAMPLE database

```
BFGSMIN 1.0D-6
NOCISTRANS
LPERMDIST 11 0.5 5.0 0.06
NOHESS
ENDNUMHESS
GEOMDIFFTOL 0.3
EDIFFTOL 1.0D-4
DUMPDATA
AMBER12 start
```

3. odata.connect file for running discrete path sampling using the PATHSAMPLE program

```
REOPTIMISEENDPOINTS
BFGSMIN 1.0D-6
NOCISTRANS
LPERMDIST 11 0.5 5.0 0.06
NEBK 10.0
ADJUSTK 5 20.0 1.03D 0.125 1.0D2
NEWNEB 25 400 0.01
MAXBFGS 0.2
BFGSSTEPS 20000
MAXERISE 1.0D-4 1.0D-2
UPDATES 175
NOHESS
ENDNUMHESS
TRAD 0.2
MAXMAX 0.3
```

```
STEPS 4000
BFGSTS 1000 20 200 0.01
MAXTSENERGY 10.0D0
GEOMDIFFTOL 0.3
EDIFFTOL 1.0D-4
MAXSTEP 0.31
PUSHOPT 0.2 0.001 1000
DIJKSTRA EXP
NEWCONNECT 100 5 6.0 50.0 31 2.0 0.001
DUMPALLPATHS
AMBER12 start
```

4. Extra keywords present in odata.connect to perform QCI calculation

```
! 110 is the number of atoms in QCI and QCIMAXACTIVE keywords
CONCUTABS 0.2 0.05
QCIRESET 2000
QCI 1.D-3 1.0 1.0 9.0 0.0 110 0 300000 2000 400000 1.0 1.5
KINT 3.0
CHECKREP 100 1.25
QCIIMAGE 0.0 1.5 25 25 1 12 501
QCIADJUSTK 5 12.0 1.01 1.0 3.0
INTSPRINGACTIVE
QCILPERMDIST 2501 1.2D0
QCIMAXACTIVE 110
INTFREEZE 0.02D0 5
MAXCON 10
QCICONCUT 50.0
QCICYCLES 5.0 200
```

```
QCIAMBER
AMBER12 start
```

5. pathdata file for running PATHSAMPLE. Some of the keywords are commented out (using !) as they are used during separate PATHSAMPLE runs for the initial setup.

```
NATOMS 110
SLURM
! ADDMINXYZ alllowest.xyz
! CYCLES 6400
! CONNECTPAIRS connectfile
UNTRAP 2.0 4.0 3.0
CYCLES 1000
! CV 0.01 1.2 0.001
! SHANNON 0.01 1.2 0.01 0.5 -1
TEMPERATURE 0.592
PLANCK 9.536D-14
EXEC /sharedscratch/nn320/softwarewales/OPTIM/builds/gfortran/A12OPTIM
COPYFILES perm.allow min.in coords.prmtop coords.inpcrd
COPYOPTIM
DIRECTION AB
CONNECTIONS 1
PERMDIST
ETOL 1.0D-4
GEOMDIFFTOL 3.0D-1
ITOL 1.0D0
AMBER12
```

```
6. dinfo file to construct disconnectivity graphs
first -74.8
maxtsenergy -74.8 ! 12 kcal/mol higher than global minimum
levels 26
delta 0.5
minima min.data
ts ts.data
monotonic
orderbyenergy
lowest 2610
scalebar 2
```

REFERENCES

Babu NJ, Cherukuvada S, Thakuria R and Nangia A (2010) Conformational and synthon polymorphism in furosemide (Lasix). Crystal Growth & Design **10**, 1979–1989.

Baker EN and Hubbard RE (1984) Hydrogen bonding in globular proteins. Progress in Biophysics and Molecular Biology **44**, 97–179.

Berkovitch-Yellin Z and Leiserowitz L (1984) The role played by C–H$\cdots$O and C–H$\cdots$ N interactions in determining molecular packing and conformation. Acta Crystallographica Section B: Structural Science **40**, 159–165.

Burley SK and Petsko GA (1985) Aromatic-aromatic interaction: A mechanism of protein structure stabilization. Science **229**, 23–28.

Burley SK and Petsko GA (1986) Amino-aromatic interactions in proteins. FEBS Letters **203**, 139–143.

Burley SK and Petsko GA (1988) Weakly polar interactions in proteins. Advances in Protein Chemistry and Structural Biology **39**, 125–189.

Cruz-Cabeza AJ, Reutzel-Edens SM and Bernstein J (2015) Facts and fictions about polymorphism. Chemical Society Reviews **44**, 8619–8635.

Fan Y, Zhao Y, Ye L, Li B, Yang G and Wang Y (2009) Polymorphs and pseudopolymorphs of N, N-di (n-butyl) quinacridone: Structures and solid-state luminescence properties. Crystal Growth & Design **9**, 1421–1430.

Kumar SS and Nangia A (2014) A solubility comparison of neutral and zwitterionic polymorphs. Crystal Growth & Design **14**, 1865–1881.

Nangia A (2008) Conformational polymorphism in organic crystals. Accounts of Chemical Research **41**, 595–604.

Purohit R and Venugopalan P (2009) Polymorphism: An overview. Resonance **14**, 882–893.

Tompa P and Fuxreiter M (2008) Fuzzy complexes: Polymorphism and structural disorder in protein–protein interactions. Trends in Biochemical Sciences **33**, 2–8.