SUPPLEMENTARY INFORMATION doi:10.1038/nature19791

Accurate *de novo* design of hyperstable constrained peptides

Gaurav Bhardwaj*, Vikram K. Mulligan*, Christopher D. Bahl*, Jason M. Gilmore, Peta Harvey, Olivier Cheneval, Garry W. Buchko, Surya V.S.R.K. Pulavarti, Quentin Kaas, Alexander Eletsky, Po-Ssu Huang, William A. Johnsen, Per Greisen, Gabriel J. Rocklin, Yifan Song, Thomas W. Linsky, Andrew Watkins, Stephen A. Rettie, Xianzhong Xu, Lauren P. Carter, Richard Bonneau, James M. Olson, Evangelos Coutsias, Colin E. Correnti, Thomas Szyperski, David J. Craik, and David Baker

*These authors contributed equally to this work.

Supplementary information 1: Computational methods and supplementary results

1.1. Modifications to Rosetta's Scoring Function

Rosetta's scoring function consists of a number of individual score terms that are summed together to produce a final score[1,2]. Each term models different aspects of the energy of a protein or peptide in a given conformation. In the past, peptides composed entirely of D-amino acids were designed in the context of an L-amino acid interaction partner by mirroring the entire system, and using Rosetta's standard design tools to design an L-amino acid peptide in a D-amino acid binding partner context[3]. This ensured that the energy function, optimized for L-amino acid design, would be appropriate for the region being designed. This is not an option for designing peptides of mixed chirality, however. For this reason, the manner in which many of the scoring function terms is calculated had to be modified to permit accurate scoring of peptides containing D-amino acids, or peptides with terminal (N-C) peptide bonds or other noncanonical connections.

First, it was necessary to modify the single-residue torsional potentials. In the *talaris2013* scoring function, these terms are called *rama* (a Ramachandran potential dependent on the mainchain torsion angles phi and psi), *p_aa_pp* (a statistical potential that also yields a score based on the phi and psi torsion angles), *omega* (a potential that penalizes non-planar peptide bond geometry), and *fa_dun* (a potential that penalizes unfavorable sidechain conformations given the backbone)[2,4]. Each of these was modified so that it would score D-amino acid residues by inverting the relevant torsion values and using the score tables or analytical potentials for the corresponding L-amino acid. Derivative calculations, necessary for energy-minimization[4], were also modified so that D-amino acid derivatives would be calculated by inverting relevant torsion values, calculating derivatives as for the equivalent L-amino acid, and then inverting the derivatives to yield the appropriate D-amino acid derivatives.

The *rama*, *omega*, and *p_aa_pp* score terms required additional modification to ensure that mirror-image peptide models scored identically: the potentials for glycine, which were based on statistics from the Protein Data Bank, favored glycine in the region of Ramachandran space favored by D-amino acids. While glycine disproportionately favors such conformations in the context of L-amino acid proteins, in a mixed D/L context, one would expect its conformational preferences to by fully symmetric. We therefore added an option to Rosetta, controlled by an input flag ("-symmetric_gly_tables true"), which permits the user to specify that the scoring tables for *rama* and *p_aa_pp*, and that the functional form of the *omega* potential, be made symmetric. In the case of *rama* and *p_aa_pp*, this is done by averaging the probability table values for (phi, psi) and (-phi, -psi), re-normalizing, and converting probabilities to energies. In the case of *omega*, this is done by setting the potential minima, which are normally offset very slightly based on Protein Data Bank statistics, to 0° and 180°.

Of the longer-range interactions, the *fa_atr* (inter-residue attractive part of the van der Waals force), *fa_rep* (inter-residue repulsive part of the van der Waals term) and *fa_sol* (hydrophobic "force" used to model the hydrophobic effect in the absence of explicit solvent) also required minor modifications for cyclic peptides, since the functional form of these terms is altered slightly for residues that are adjacent in linear sequence. We ensured that, rather than assuming that residue N is connected to residues N+1 and N-1 at its C- and N-terminal connection points, respectively, the scoring machinery would check which residues are connected and score them as adjacent residues based on covalent bonds rather than by indices.

Rosetta's *fa_dslf* score term, which holds disulfide-bonded cysteine S_γ residues together and penalizes deviations from ideal disulfide geometry, was updated to score D-Cys, D-Cys disulfide bonds by inverting torsion values; derivatives were similarly updated. The term then required some additional modifications to permit it to score and preserve disulfide geometry in mixed L-Cys, D-Cys disulfide bonds. This score term has energy minima for L-Cys disulfide bonds at values of -86.10° and 92.39° for the $C_{\beta1}$-$S_{\gamma1}$-$S_{\gamma2}$-$C_{\beta2}$ dihedral angle, based on statistics from high-resolution crystal structures of disulfide-containing natural proteins, and the corresponding minima for D-Cys disulfide bonds were set to 86.10° and -92.39°, respectively. Since no such statistics are available for mixed L-Cys, D-Cys disulfide bonds, however, the minima were set to -90° and 90°. Similarly, the well depths for the two minima were set to identical values (the average of the depths of the two wells for L-Cys disulfide bonds).

The *pro_close* score term, which ensures that energy-minimization does not pull open proline ring, was updated to act on both D- and L-proline. A more general term, *ring_close*, has also been added which can be used on any noncanonical residue type that, like proline, contains a ring that could be pulled open by free rotation about single bonds in the absence of a potential holding it closed.

Finally, we altered the amino acid reference energies to ensure that corresponding L- and D-amino acids have the same reference energy values. (The reference energies are a zeroth-order correction factor to compensate for the fact that certain amino acid types can engage in larger numbers of favorable interactions than others, resulting in pathologies during design in which these residue types are disproportionately favored. By assigning a constant bonus or penalty to each type, this pathology is partially suppressed.)

Recently, the default Rosetta scoring function has been updated to *talaris2014*, which re-weights several terms and adds a new term, *yhh_planarity*, which is intended to hold the tyrosine hydroxyl proton in the plane of the tyrosine ring[1]. We ensured that this term also acts on D-tyrosine. A newer, experimental scoring function, currently called *beta_nov15*, has also entered testing, and may replace the current default scoring function at some point in the future. We have ensured that new terms added in *beta_nov15* are also compatible with D-amino acids, are properly differentiable for energy minimization, and are compatible with cyclic geometry, as described above. All scoring function changes have been tested by constructing, scoring, and minimizing mirror-image structures, confirming that the score matches for mirror-image structures, and by constructing and scoring cyclic permutations of cyclic peptides, confirming that the scoring is identical regardless the start and end points of the peptide. Unit tests have been added to ensure that, as the default Rosetta scoring function is replaced in the future, it continues to support D-amino acids and cyclic geometry fully.

1.2. Implementation of the Generalized Kinematic Closure (GenKIC) Algorithm

One of the core challenges in designing peptides with many covalent cross-links is sampling conformations permitted by the covalent geometry. Ideally, one would want an algorithm capable of *only* sampling conformations that yield good cross-link geometry, which would greatly reduce the search space. Kinematic closure approaches, which break the sampling problem into a series of loop closure problems and analytically solve for torsion values that permit loop closure, permit highly efficient constrained sampling[5,6]. In order to apply this to peptides with arbitrary building blocks and staple chemistries, we implemented a generalized form of Rosetta's kinematic closure algorithm, which we call "GenKIC", in which loops can be defined as any covalently-linked chain of atoms, including chains passing through terminal peptide bonds, disulfide bonds, *etc*. A user interface accessible to the RosettaScripts scripting language was also developed to permit precise and versatile control over the sampling.

Internally, the GenKIC algorithm performs the following steps. First, from a user-input list of residue indices, it identifies the covalently-linked chain of atoms that is the loop to be closed, as well as the start and end points of this chain. Given a chain with N degrees of freedom, the requirement that the rigid-body transform from the loop's start point to its end point must be maintained to maintain closure effectively reduces the degrees of freedom of the system by six. In a second step, it is therefore possible to perturb N-6 degrees of freedom in user-specified ways (which the user controls by invoking GenKIC "perturbers"), and then, in a third step, to solve for the values of the remaining six degrees of freedom (the six torsion angles adjacent to three user-defined pivot atoms) needed to preserve the rigid-body transform between the start and end points of the loop. The perturbation and solving steps are performed iteratively to generate a user-defined number of solutions. Since the system of equations solved in the third step can yield anywhere from 0 to 16 solutions from each attempt, the fourth step is to prune undesired solutions (*e.g.* due to clashing geometry, pivot atom torsion values lying outside of desired ranges, *etc.*) from the candidate list by applying optional GenKIC "filters". In a fifth step, the algorithm applies other Rosetta algorithms that modify the structure ("movers"), optionally defined by the user, to every GenKIC solution remaining (allowing things like sequence design, sidechain rotamer optimization, energy minimization, *etc.*) before selecting a single, top solution based on criteria specified by a user-defined GenKIC "selector". The original structure is then updated with the new loop conformation, and can serve as input into subsequent Rosetta modules or can be written to disk. These steps are shown in flowchart form in Extended Data Figure 7.

GenKIC perturbers have been created to permit torsion, bond angle, and bond length degrees of freedom to be set to user-defined values. These perturbers are called "set_dihedral", "set_bondangle", and "set_bondlength", respectively. If a loop starts in a broken or open conformation, these perturbers can be used to define closed geometry at a particular bond, and have been wrapped in a convenient "CloseBond" statement for ease of use from the RosettaScripts user interface. Loop torsion values can also be randomized fully ("randomize_dihedral"), perturbed slightly from a starting value ("perturb_dihedral"), or, in the case of α-amino acid mainchain torsion values, both phi and psi can drawn randomly from the Ramachandran map-biased distribution for a given amino acid type ("randomize_alpha_backbone_by_rama"). The code has been written for versatility and extensibilty, so additional GenKIC perturbers can be added as needed.

Similarly, GenKIC filters have been defined to discard kinematic closure solutions with clashing geometry ("loop_bump_check"), with pivot torsion values in unlikely regions of Ramachandran space ("alpha_aa_rama_check"), or with particular amino acid residues in undesired user-defined regions of Ramachandran space ("backbone_bin"). GenKIC selectors have been implemented to select the lowest-energy solution found ("lowest_energy_selector"), a random solution from the list of solutions found ("random_selector"), or a random solution biased by the energy, with lower-energy solutions weighted more heavily ("boltzmann_energy_selector"). As with GenKIC perturbers, new GenKIC filters and selectors can be implemented easily as needed.

At the level of the Rosetta source code, the GenKIC algorithm is implemented as methods of the `GeneralizedKIC` class, which is defined in the `protocols::generalized_kinematic_closure` namespace. Perturbers, filters, and selectors are defined as helper classes in the sub-namespaces `protocols::generalized_kinematic_closure::perturber`, `protocols::generalized_kinematic_closure::filter`, and `protocols::generalized_kinematic_closure::selector`. Additional perturbers, filters, and selectors can be added by adding methods to the appropriate helper function. Full user documentation for GenKIC is available on the Rosetta help wiki.

1.3. Construction of a Fragment-Free Conformational Sampling Algorithm

Although computational validation of peptide designs containing mixtures of D- and L-amino acids is a particular challenge, those designs with small numbers of isolated D-amino acids can be validated using the classic Rosetta *ab initio* algorithm, with D-amino acid positions mutated to glycine. Classic *ab initio* works by choosing sets of protein fragments from known structures based on sequence alignment, then using the insertion of these fragments as moves in a simulated annealing-based search of conformational space. For a high-quality design, the *ab initio* algorithm reveals an energy landscape with a unique low-energy conformation corresponding to the design conformation. Poor designs either fail to sample conformations close to the design conformation, or have alternative low-energy conformations that they can access that are revealed by the sampling. Unfortunately, peptides with long stretches of D-amino acids cannot be validated in this manner, since there exist too few solved structures of known proteins in the Protein Data Bank that have long stretches of amino acid residues in the region of Ramachandran space uniquely accessed by D-amino acids, which means that suitable fragment lists cannot be generated. With the GenKIC algorithm in hand, it was possible to implement a fragment-free, GenKIC-based conformational sampling tool that could predict lowest-energy peptide structures based on amino acid sequence.

The algorithm works as follows. First, the input sequence is randomly circularly permuted to avoid any possible artifacts that might be introduced by having the cyclization point in a particular place, and a linear peptide with the permuted sequence is constructed. All omega torsion angles are set to 180°. Second, an amino acid residue in the sequence that is not at either of the ends is randomly chosen to be the "anchor" residue. The anchor residue, henceforth indexed as residue M, will be the fixed point lying outside of the chain of residues that will be treated as a loop to be closed by GenKIC. This residue's mainchain phi and psi torsion angles are randomized, biased by the Ramachandran distribution for the residue type.

Third, the GenKIC algorithm is applied to the loop that runs from residue M+1 (immediately past the anchor residue), through the open terminal peptide bond, to residue M-1 (immediately before the anchor residue). Fourth, pivot atoms are selected: C_α atoms of residues M+1 and M-1 are always chosen as pivot atoms, and the third pivot is selected randomly from the C_α atoms in the rest of the loop. Fifth, GenKIC closes the terminal peptide bond with ideal peptide geometry, and randomizes all mainchain torsion values within the loop biased by the Ramachandran distribution for each residue. We found that this random sampling works well for smaller peptides (up to ~15 residues), typically allowing sampling close to the design conformation and across a broad range of alternative conformations. For longer peptides, it is necessary to bias the sampling slightly by setting mainchain torsion values near the middle of secondary structure elements to ideal values for the secondary structure type, then adding a small random perturbation to these values. Loop residues and the ends of secondary structure elements are always sampled fully randomly. Sixth, filters are applied to eliminate solutions with pivot residues in unreasonable regions of Ramachandran space, or solutions with fewer mainchain hydrogen bonds than a user-specified number. In the case of peptides containing disulfide bonds, all disulfide permutations are attempted at this point, and conformations incompatible with any disulfide geometry (*i.e.* yielding *fa_dslf* scores above a given threshold) are also filtered out. Seventh, Each GenKIC solution passing filters is subjected to multiple rounds of the Rosetta FastRelax algorithm[7] which optimizes sidechain rotamers and carries out energy minimization (including optimization of disulfide geometry, if any disulfide bonds are present). Finally, the lowest-energy sample passing filters is circularly de-permuted and written to disk. After many rounds of sampling, the user may then plot the calculated energy of each sample against the RMSD to the design conformation to determine whether the design conformation represents a unique low-energy state. These steps are shown in flowchart form and key steps are illustrated in Extended Data Figure 8.

The peptide structure prediction algorithm has been implemented as a Rosetta protocol. It is a class named `protocols::cyclic_peptide_predict::SimpleCycpepPredictApplication` that can be called from other code. It also exists as a stand-alone application in the Rosetta applications, called `simple_cycpep_predict`. After compiling Rosetta, the `simple_cycpep_predict` application can be invoked from the command-line as shown in the following example (which was used to generate the plot of energy against RMSD from the design state for the NC_cH$_L$H$_R$_D1 design, shown in Figure 6 in the main text).

```
<path_to_Rosetta>/Rosetta/main/source/bin/simple_cycpep_predict.default.linux
gccrelease
      -cyclic_peptide:rand_checkpoint_file rng01.state.gz -
      cyclic_peptide:checkpoint_file check01.txt -out:file:silent
      out01.silent -cyclic_peptide:sequence_file inputs/seq.txt -beta_nov15 -
      symmetric_gly_tables true -score:weights beta_nov15.wts -in:file:native
      inputs/native.pdb -cyclic_peptide:genkic_closure_attempts 50 -
      cyclic_peptide:genkic_min_solution_count 1 -
      cyclic_peptide:require_disulfides true -
      cyclic_peptide:disulf_cutoff_prerelax 2000 -
      cyclic_peptide:min_genkic_hbonds 14 -cyclic_peptide:min_final_hbonds 14
      -cyclic_peptide:fast_relax_rounds 5 -cyclic_peptide:rama_cutoff 2.0 -
      cyclic_peptide:checkpoint_job_identifier check -mute all -unmute
      protocols.cyclic_peptide_predict.SimpleCycpepPredictApplication -
      nstruct 50000 -cyclic_peptide:user_set_alpha_dihedrals 3 -61 -41 180 4
      -61 -41 180 5 -61 -41 180 6 -61 -41 180 7 -61 -41 180 8 -61 -41 180 9 -
      61 -41 180 16 61 41 180 17 61 41 180 18 61 41 180 19 61 41 180 20 61 41
```

```
180 21 61 41 180 22 61 41 180 23 61 41 180 -
cyclic_peptide:user_set_alpha_dihedral_perturbation 5.0
```

The full details of the flags shown above are available on the Rosetta help wiki. A few details are worth noting: this example uses symmetric glycine Ramachandran and p_aa_pp tables (`-symmetric_gly_tables true`). Solutions with fewer than 14 mainchain hydrogen bonds (`cyclic_peptide:min_final_hbonds 14`) or *rama* energy term scores greater than 2.0 for pivot residues (`-cyclic_peptide:rama_cutoff 2.0`) will be filtered out, as will solutions with pre-minimization *fa_dslf* scores greater than 2000 (`-cyclic_peptide:disulf_cutoff_prerelax 2000`).

1.4. Sequence Design

We created a Rosetta protocol called “FastDesign” for design of amino acid sequences for a given backbone. Rosetta designs sequences using a simulated-annealing-based approach called “packing,” where random substitutions are made using the sidechain rotamers found in the Dunbrack library[8], in an attempt to find the sequence with lowest possible energy for each backbone. FastDesign was created as the sequence design analog to the FastRelax protocol, which is used in structure prediction. FastRelax attempts to find an optimal pose conformation with minimal energy via both small backbone movement and sidechain rotamer packing, but does not alter the existing sequence. Briefly, each repeat of FastDesign consists of four design and minimization steps. The first is done with the Lennard-Jones repulsive term down-weighted to 0.088. This allows the sidechains to clash slightly as they search for the most optimal interactions. The repulsive term is increased in the following steps, until the final step when it is at full strength (0.42). As the repulsive term is increased, the most optimal interactions will stay in place as other interactions are broken to account for the increasing repulsive term. By default, three repeats of FastDesign were performed on each backbone. The resulting structures have improved total energy and sidechain packing (as measured by the Rosetta packstat filter) over an equivalent number of packing/minimization steps without alteration to the repulsive term.

1.5. Supplementary Tables

Supplementary Table S1-1: Sequences of Computationally Designed Peptides

Design Name	# of residues	Disulfide(s)	Sequence*
gHH_44	28	C4-C26	AEDCERIRKELEKNPNDEIKKKLEKCQA
gHHH_06	43	C2-C26, C18-C41	PCEDLKERLKKLGMSEECRQRLEKMCKEGTSEDAERMARNCES
gEHE_06	35	C1-C27, C14-C33	CKQRRRYRGSEEECRKYAEELSRRTGCEVEVECET
gEEH_04	38	C2-C17, C9-C36	QCYTFRSECTNKEFTVCRPNPEEVEKEARRTKEEECRK
gHEEE_02	41	C8-C22, C18-C33	SQETRKKCTEMKKKFKNCEVRCDESNHCVEVRCSDTKYTLC

		C28-C41	
gEHEE_06	45	C8-C38, C19-C41, C28-C45	EERRYKRCGQDEERVRRECKERGERQNCQYQIRKEGNCYVCEIRC
gEEHE_02	36	C2-C35, C4-C19, C23-C31	PCECDVNGETYTVSSSEECERLCRKLGVTNCRVHCG
gEEEH_04	41	C1-C41, C3-C34, C9-C23	CRCHITSSCVRVEGDNGEEYRYCSSDEEDLRRFCKEMQKQC
gEEEEEE_02	47	C2-C15, C11-C42, C33-C46	TCEIRVTDTHCKVHCGTQEYKVPPGRTLKVGNCRFTYHDTTCTVECR
NC_cHHH_D1	22	C5-C18	NPEDCRQDPEANKSPEECKKLK
NC_cHH_D1	26	C9-C22	HDPEKRKECEKKYTDPKKREECKRKA
NC_cEE_D1	18	C5-C14	PVTWCVRIpPTVRCTVRp
NC_cH$_L$H$_R$_D1	26	C8-C21	NPELQRKCKELdTRpeaerkcreeSD
NC_EHE_D1	26	C1-C21, C12-C24	CQTWRrVSPEECRKYKEEYnCVRCTE
NC_HEE_D1	27	C4-C18, C14-C27	NDKCKELKKRYPNCEVRCDpPRYEVHC
NC_EEH_D2	26	C2-C11, C5-C26	TCVECapVKVCRPDPEEARREAEERC

*D-amino acids in the sequence are denoted by lower-case letters and coloured in red.

Supplementary Table S1-2: Most significant sequence alignments from PDB database

Design$^{\#}$	Best e-value*	PDB ID	PDB DESCRIPTION
gHH_44	0.001	4R4L	Chain A, Crystal Structure Of Wt Cgmp Dependent Protein Kinase I Alpha (pkgi Alpha) Leucine Zipper
gHHH_06	0.53	5JPQ	Chain A, Cryo-em Structure Of The 90s Pre-ribosome
gEHE_06	2.2	2E9Y	Chain A, Crystal Structure Of Project Ape1968 From Aeropyrum Pernix K1

gEEH_04	0.015	3CNY	Chain A, Crystal Structure Of A Putative Inositol Catabolism Protein Iole (Iole, Lp_3607) From Lactobacillus Plantarum Wcfs1
gHEEE_02	10	2JX3	Chain A, Nmr Solution Structure Of The N-Terminal Domain Of Dek
gEHEE_06	0.017	4BBY	Chain A, Mammalian Wild-Type Alkyldihydroxyacetonephosphate Synthase
gEEHE_02	2.2	1DCF	Chain A, Crystal Structure Of The Receiver Domain Of The Ethylene Receptor Of Arabidopsis Thaliana
gEEEH_04	0.23	1WEH	Chain A, Crystal Structure Of The Conserved Hypothetical Protein Tt1887 From Thermus Thermophilus Hb8
gEEEEEE_02	7.3	4E6F	Chain A, Crystal Structure Of A Hypothetical Protein (Bacova_04320) From Bacteroides Ovatus Atcc 8483
NC_cHH_D1	1.1	3LW5	Chain 1, Improved Model Of Plant Photosystem I
NC_cHHH_D1	1.9	4L6T	Chain A, Gm1 Bound Form Of The Ecx Ab5 Holotoxin
NC_cEE_D1	18	1MC4	Chain A, Crystal Structure Of Aspartate-Semialdehyde Dehydrogenase From Vibrio Cholerae El Tor
NC_EEH_D2	0.099	1V6S	Chain A, Crystal Structure Of Phosphoglycerate Kinase From Thermus Thermophilus Hb8
NC_EHE_D1	2.00E-04	2MA5	Chain A, Solution Nmr Structure Of Phd Type Zinc Finger Domain Of Lysine- Specific Demethylase 5b (plu-1/jarid1b) From Homo Sapiens, Northeast Structural Genomics Consortium (nesg) Target Hr7375c
NC_HEE_D1	2.9	2K35	Chain A, Hydramacin-1: Structure And Antibacterial Activity Of A Peptide From The Basal Metazoan Hydra

*For PSI-BLAST searches, D-amino acids in sequences were replaced with Glycine. [#]NC_cH$_L$H$_R$_D1 could not be searched effectively using PSI-BLAST due to large stretches of D-amino acids. Replacing all D-amino acids with Glycine in NC_cH$_L$H$_R$_D1 did not show any significant alignment.

Supplementary Table S1-3: Most significant sequence alignments from NCBI Non-Redundant database

Design[#]	e-value*	NCBI ID	DESCRIPTION
gHH_44	0.006	WP_010181808.1	glutamyl-tRNA synthetase [Aquimarina agarilytica]

gHHH_06	0.19	XP_013338023.1	tRNA delta(2)-isopentenylpyrophosphate transferase, putative [Eimeria maxima]
gEHE_06	0.068	EWM28578.1	rna recognition motif-containing protein [Nannochloropsis gaditana]
gEEH_04	0.25	KDQ31172.1	hypothetical protein PLEOSDRAFT_1075335 [Pleurotus ostreatus PC15]
gHEEE_02	0.43	EPY17093.1	hypothetical protein AGDE_16431 [Angomonas deanei]
gEHEE_06	0.12	XP_007325846.1	hypothetical protein AGABI1DRAFT_124501 [Agaricus bisporus var. burnettii JB137-S8]
gEEHE_02	0.75	WP_014815876.1	alpha/beta hydrolase [Mycobacterium chubuense]
gEEEH_04	2.0	KCZ72753.1	hypothetical protein ANME2D_01187 [Candidatus Methanoperedens nitroreducens]
gEEEEEE_02	12	XP_014653899.1	mRNA export factor elf1 [Moesziomyces antarcticus]
NC_cHH_D1	1	WP_021876737.1	hypothetical protein [Clostridium chauvoei]
NC_cHHH_D1	0.44	XP_011304041.1	PREDICTED: RNA-binding protein 28 [Fopius arisanus]
NC_cEE_D1	10	WP_056548245.1	primosome assembly protein PriA [Mycobacterium sp. Root135]
NC_EEH_D2	0.76	WP_018525819.1	hypothetical protein [Spirochaeta alkalica]
NC_EHE_D1	0.03	BAC25202.1	unnamed protein product [Mus musculus]
NC_HEE_D1	7.9	XP_014755505.1	PREDICTED: uncharacterized protein LOC104583481 [Brachypodium distachyon]

*For PSI-BLAST searches, D-amino acids in sequences were replaced with ‘X’ (reflecting any amino acid). $^{\#}$NC_cH$_L$H$_R$_D1 could not be searched effectively using PSI-BLAST due to large stretches of D-amino acids. Replacing all D-amino acids with Glycine in NC_cH$_L$H$_R$_D1 did not show any significant alignment.

1.6. Scripts and Inputs to Design Genetically-Encodable Peptides

The command for running the RosettaScripts XML file shown below is as follows:

```
<path_to_Rosetta>/Rosetta/main/source/bin/rosetta_scripts.default.linuxgccrel
ease
```

```
      -in:file:s <arbitrary initial pdb file>
      -parser:protocol <Rosetta Scripts file>
      -out:file:s <output pdb file name>
```

In the above, “linuxgccrelease” should be replaced with the user’s build and compiler (*e.g.* “macosclangrelease” on an Apple Macintosh system using the Clang compiler.)

1.6.1 Example RosettaScripts XML file for designing EHEE topology:

```
<ROSETTASCRIPTS>

    <SCOREFXNS>

    #### centroid score function used for protein backbone design ####
       <SFXN_CENTROID weights="fldsgn_cen">
             <Reweight scoretype="cenpack" weight="1.0" />
             <Reweight scoretype="hbond_sr_bb" weight="1.0" />
             <Reweight scoretype="hbond_lr_bb" weight="1.0" />
             <Reweight scoretype="atom_pair_constraint" weight="1.0" />
             <Reweight scoretype="angle_constraint" weight="1.0" />
             <Reweight scoretype="dihedral_constraint" weight="1.0" />
       </SFXN_CENTROID>

    #### full-atom score function used for amino acid sequence design ####
       <SFXN_FULLATOM weights="talaris2014" />

    </SCOREFXNS>

    <RESIDUE_SELECTORS>
       <Chain name="chain_A" chains="A" />
    </RESIDUE_SELECTORS>

    <TASKOPERATIONS>

    #### restrict residue identity during design by the degree with which the
residue is burried ####
       <LayerDesign name="layer_all"
layer="core_boundary_surface_Nterm_Cterm" verbose="True"
use_sidechain_neighbors="True" >
             <core>
                   <all append="M" />
             </core>
             <boundary>
                   <all append="M" />
             </boundary>
             <surface>
             </surface>
       </LayerDesign>

    #### allow disulfide bonds to repack, but do not mutate ####
       <OperateOnCertainResidues name="no_design_disulf" >
             <RestrictToRepackingRLT />
             <ResidueName3Is name3="CYS" />
       </OperateOnCertainResidues>
```

```
    #### do not allow non-realistic chi angles of aromatic amino acid
sidechains ####
       <LimitAromaChi2 name="limitchi2" include_trp="True" />

    #### restrict amino acid identity of loop regions based on abego profile
####
       <ConsensusLoopDesign name="disallow_nonnative_loop_sequences" />

    #### increase the diversity of rotamers available to the packer ####
       <ExtraRotamersGeneric name="extra_rots" ex1="True" ex2="True" />

       <OperateOnCertainResidues name="no_repack_non-disulf" >
             <PreventRepackingRLT/>
             <ResidueName3Isnt name3="CYS" />
       </OperateOnCertainResidues>

       <LayerDesign name="layer_core_boundary" layer="core_boundary"
verbose="False" use_sidechain_neighbors="True" />

    </TASKOPERATIONS>

    <FILTERS>
       <SheetTopology name="filter_strand_pairing" topology="1-3.A.0;2-3.A.0"
blueprint="./EHEE.blueprint" />
       <CompoundStatement name="compound_toplogy_filter" >
             <AND filter_name="filter_strand_pairing" />
       </CompoundStatement>

       <TaskAwareScoreType name="dslf_quality_check"
task_operations="no_repack_non-disulf" scorefxn="SFXN_FULLATOM"
score_type="dslf_fa13"  mode="individual" threshold="-0.27" confidence="1" />
       <DisulfideEntropy name="entropy" lower_bound="0" tightness="2"
confidence="0"/>

    ############### core assessment ###############
       <SecondaryStructureHasResidue name="ss_contributes_core"
secstruct_fraction_threshold="1.0"
res_check_task_operations="layer_core_boundary" required_restypes="VILMFYW"
nres_required_per_secstruct="1" filter_helix="1" filter_sheet="1"
filter_loop="0" min_helix_length="4" min_sheet_length="3" min_loop_length="1"
confidence="1" />

    ##### verify presence of secondary structure #####
       <SecondaryStructureCount name="count_SS_elements"
filter_helix_sheet="True" num_helix="1" num_sheet="3" num_helix_sheet="4"
min_helix_length="6" min_sheet_length="4" min_loop_length="2" />

       <CompoundStatement name="sequence_quality_compound_filter" >
             <AND filter_name="ss_contributes_core" />
             <AND filter_name="count_SS_elements" />
             <AND filter_name="dslf_quality_check"/>
             <AND filter_name="entropy" />
       </CompoundStatement>
```

```
    </FILTERS>

    <MOVERS>

    #### assess and record the secondary structure ####
       <Dssp name="dssp" />

    #### design the protein mainchain ####
       <SetSecStructEnergies name="assign_secondary_structure_bonus"
scorefxn="SFXN_CENTROID" blueprint="./EHEE.blueprint" />
       <BluePrintBDR name="build_mainchain" scorefxn="SFXN_CENTROID"
use_abego_bias="True" blueprint="./EHEE.blueprint" />

       <ParsedProtocol name="mainchain_building_protocol" >
             <Add mover="build_mainchain" />
             <Add mover="dssp" />
       </ParsedProtocol>

       <LoopOver name="mainchain_building_loop"
mover_name="mainchain_building_protocol"
filter_name="compound_toplogy_filter" iterations="1000" drift="False"
ms_whenfail="FAIL_DO_NOT_RETRY" />

       <Disulfidize name="disulfidizer" set1="chain_A" set2="chain_A"
min_disulfides="2" max_disulfides="3" match_rt_limit="2.0"
score_or_matchrt="true" max_disulf_score="-0.05" min_loop="5"
use_l_cys="true" keep_current_disulfides="false"
include_current_disulfides="false" use_d_cys="false" />

       <FastDesign name="fastdesign"
task_operations="extra_rots,limitchi2,layer_all,no_design_disulf,disallow_non
native_loop_sequences" scorefxn="SFXN_FULLATOM" clear_designable_residues="0"
repeats="3" ramp_down_constraints="0" />

       <ParsedProtocol name="build_mainchain_and_design_sequence" >
             <Add mover_name="assign_secondary_structure_bonus" />
             <Add mover="mainchain_building_loop" />
             <Add mover="dssp" />
             <Add mover_name="disulfidizer" />
             <Add mover_name="fastdesign" />
       </ParsedProtocol>

       <LoopOver name="build_mainchain_and_design_sequence_loop"
mover_name="build_mainchain_and_design_sequence"
filter_name="sequence_quality_compound_filter" iterations="1000"
drift="False" ms_whenfail="FAIL_DO_NOT_RETRY" />

    </MOVERS>

    <PROTOCOLS>
       <Add mover_name="build_mainchain_and_design_sequence_loop" />
    </PROTOCOLS>

</ROSETTASCRIPTS>
```

1.6.2 Blueprint file for designing EHEE topology

```
SSPAIR 1-3.A.0;2-3.A.0
HSSTRIPLET 1,3-1
1 V LE .
2 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V LG R
0 V LB R
0 V LB R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V HA R
0 V LG R
0 V LB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V LE R
0 V LA R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V EB R
0 V LO R
```

1.7. Scripts and Inputs to Design Disulfide-Stapled Peptides

Command line for running the rosetta scripts:

```
<path_to_Rosetta>/Rosetta/main/source/bin/rosetta_scripts.default.linuxgccrel
ease
      -in:file:s <arbitrary initial pdb file>
      -parser:protocol <Rosetta Scripts file>
      -out:file:s <output pdb file name>
      -run:preserve_header
```

1.7.1 Rosetta Scripts Input File

```
<ROSETTASCRIPTS>
<SCOREFXNS>

   ############## Define Score functions ###############
   <SFXN1 weights="fldsgn_cen">
       <Reweight scoretype="cenpack" weight="1.0" />
       <Reweight scoretype="hbond_sr_bb" weight="1.0" />
       <Reweight scoretype="hbond_lr_bb" weight="1.0" />
       <Reweight scoretype="atom_pair_constraint" weight="1.0" />
       <Reweight scoretype="angle_constraint" weight="1.0" />
       <Reweight scoretype="dihedral_constraint" weight="1.0" />
   </SFXN1>


   <SFXN_STD weights= "beta_july15.wts"  />

</SCOREFXNS>
<TASKOPERATIONS>
</TASKOPERATIONS>
<FILTERS>

   <HelixKink name="hk1" blueprint="eeh.blueprint" />
   <SheetTopology name="sf1" blueprint="eeh.blueprint" />
   <SecondaryStructure name="ss1" blueprint="eeh.blueprint" use_abego="1" />
      <CompoundStatement name="cs1">
            <AND filter_name="ss1" />
            <AND filter_name="hk1" />
            <AND filter_name="sf1" />
      </CompoundStatement>

</FILTERS>

<MOVERS>
      <Dssp name="dssp" />
      <SheetCstGenerator name="sheet_new1" cacb_dihedral_tolerance="0.6"
blueprint="eeh.blueprint" />
      <SetSecStructEnergies name="set_ssene1" scorefxn="SFXN1"
blueprint="eeh.blueprint" />
   <BluePrintBDR name="topology_builder" use_abego_bias="1" scorefxn="SFXN1"
constraint_generators="sheet_new1" constraints_NtoC="-1.0"
blueprint="eeh.blueprint" />
      <ParsedProtocol name="build_dssp1" >
            <Add mover_name="topology_builder" />
            <Add mover_name="dssp" />
      </ParsedProtocol>
```

```
   <LoopOver name="lover1" mover_name="build_dssp1" filter_name="cs1"
iterations="10" drift="0" ms_whenfail="FAIL_DO_NOT_RETRY" />

   <ParsedProtocol name="phase1" >
            <Add mover_name="set_ssene1" />
            <Add mover_name="lover1" />
      </ParsedProtocol>

   <ParsedProtocol name="pp1">
            <Add mover_name="phase1" />
      </ParsedProtocol>

   #### Assemble the topology ####
   <LoopOver name="lover2" mover_name="pp1" filter_name="cs1" iterations="10"
drift="0" ms_whenfail="FAIL_DO_NOT_RETRY" />

    #### Add disulfides to the topology ####
   <Disulfidize name="add_disulf" min_disulfides="2" max_disulfides="2"
max_disulf_score="-0.20" match_rt_limit="2" min_loop="5" />

    #### Design and Relax structures with disulfides in place ####
   <MultiplePoseMover name="disulfidizer" >
       <SELECT>
       </SELECT>

       <ROSETTASCRIPTS>

           <SCOREFXNS>
               <SFXN_STD weights= "beta_july15.wts"  />
           </SCOREFXNS>

           <FILTERS>
               <ResidueCount name=cys_count_1 residue_types="CYS"
min_residue_count=4 confidence=1 />
           </FILTERS>
           <TASKOPERATIONS>

               <DisallowIfNonnative name=nocys resnum=0 disallow_aas="C" />

                ############## select CYS residues ###############

               <OperateOnCertainResidues name="no_design_disulf" >
                   <RestrictToRepackingRLT />
                   <ResidueName3Is name3="CYS" />
               </OperateOnCertainResidues>

               ########### layer selection for design ###########

               <LayerDesign name="layer_all"
layer="core_boundary_surface_Nterm_Cterm" verbose="True"
use_sidechain_neighbors="True" >
                   <core>
                       <all append="M" />
                   </core>
```

```
                   <boundary>
                   </boundary>
                   <surface>
                   </surface>
               </LayerDesign>
           </TASKOPERATIONS>

           <MOVERS>
               <FastDesign name=fdesign8 scorefxn=SFXN_STD repeats=8
task_operations=layer_all,no_design_disulf,nocys ramp_down_constraints=true>
                   <MoveMap name=fdesign_mm>
                       <Chain number=1 chi=true bb=true />
                   </MoveMap>
               </FastDesign>
           </MOVERS>

           <PROTOCOLS>
               <Add filter=cys_count_1 />
               <Add mover=fdesign8 />
           </PROTOCOLS>

       </ROSETTASCRIPTS>
   </MultiplePoseMover>
</MOVERS>
<PROTOCOLS>
      <Add mover_name="lover2" />
      <Add mover_name="dssp" />
   <Add mover_name="add_disulf" />
   <Add mover_name="disulfidizer" />


</PROTOCOLS>
</ROSETTASCRIPTS>
```

1.7.2 Blueprint File for designing EEH topology

```
SSPAIR 1-2.A.0
1    V    LX    .
0    V    EB    R
0    V    EB    R
0    V    EB    R
0    V    EB    R
0    V    LG    R
0    V    LG    R
0    V    EB    R
0    V    EB    R
0    V    EB    R
0    V    EB    R
0    V    LB    R
0    V    LA    R
0    V    LB    R
0    V    HA    R
0    V    HA    R
0    V    HA    R
```

```
0    V    HA    R
0    V    HA    R
0    V    HA    R
0    V    HA    R
0    V    HA    R
0    V    HA    R
0    V    HA    R
0    V    HA    R
0    V    LX    R
```

1.8. Scripts and Inputs to Design Peptides With Cyclic Heterochiral Topologies

The command for running the RosettaScripts XML file shown below is as follows:

```
<path_to_Rosetta>/Rosetta/main/source/bin/rosetta_scripts.default.linuxgccrel
ease
      -in:file:fasta <arbitrary initial fasta file>
      -parser:protocol <Rosetta Scripts file>
      -out:file:s <output pdb file name>
```

1.8.1 Example RosettaScripts XML file:

```
<ROSETTASCRIPTS>

<SCOREFXNS>

    <SFXN_STD weights= "beta_july15_cst.wts" />

    <SFXN_hbond_bb weights="empty.wts" symmetric=0>
        <Reweight scoretype=hbond_sr_bb weight=1.17/>
        <Reweight scoretype=hbond_lr_bb weight=1.17/>
    </SFXN_hbond_bb>
</SCOREFXNS>
<TASKOPERATIONS>
</TASKOPERATIONS>
<FILTERS>
</FILTERS>
<MOVERS>

   <PeptideStubMover name=intial_stub reset=true>

       <Append resname="GLY" />

       <Append resname="ALA" />
       <Append resname="ALA" />
       <Append resname="ALA" />
       <Append resname="ALA" />
       <Append resname="ALA" />
       <Append resname="ALA" />
       <Append resname="ALA" />
       <Append resname="ALA" />
       <Append resname="ALA" />
       <Append resname="ALA" />
```

```
       <Append resname="GLY" />
       <Append resname="VAL" />
       <Append resname="VAL" />

       <Append resname="DALA" />
       <Append resname="DALA" />
       <Append resname="DALA" />
       <Append resname="DALA" />
       <Append resname="DALA" />
       <Append resname="DALA" />
       <Append resname="DALA" />
       <Append resname="DALA" />
       <Append resname="DALA" />
       <Append resname="DALA" />

       <Append resname="ALA" />
       <Append resname="GLY" />

   </PeptideStubMover>

   <DeclareBond name=peptide_bond1 res1=1 atom1="N" atom2="C" res2=26
add_termini=true />

   <SetTorsion name=torsion1>
       <Torsion residue=ALL torsion_name=omega angle=180.0 />
       <Torsion residue=1,12,13,14,25,26 torsion_name=rama
angle=rama_biased/>
       <Torsion residue=2,3,4,5,6,7,8,9,10,11 torsion_name=phi angle=-64.8/>
       <Torsion residue=2,3,4,5,6,7,8,9,10,11 torsion_name=psi angle=-41.0/>
       <Torsion residue=15,16,17,18,19,20,21,22,23,24 torsion_name=phi
angle=64.8/>
       <Torsion residue=15,16,17,18,19,20,21,22,23,24 torsion_name=psi
angle=41.0/>
   </SetTorsion>

   <GeneralizedKIC name=genkic1 closure_attempts=1000 name=genkic1
selector="lowest_energy_selector" stop_when_n_solutions_found="50"
stop_if_no_solution=500 selector_scorefunction="SFXN_hbond_bb" >

       <AddResidue res_index=12 />
            <AddResidue res_index=13 />
       <AddResidue res_index=14 />
       <AddResidue res_index=15 />
       <AddResidue res_index=16 />
       <AddResidue res_index=17 />
       <AddResidue res_index=18 />
       <AddResidue res_index=19 />
       <AddResidue res_index=20 />
       <AddResidue res_index=21 />
       <AddResidue res_index=22 />
       <AddResidue res_index=23 />
       <AddResidue res_index=24 />
       <AddResidue res_index=25 />
       <AddResidue res_index=26 />
       <AddResidue res_index=1 />
```

```
       <SetPivots atom1="CA" atom2="CA" atom3="CA" res1=12 res2=26 res3=1 />

       <CloseBond prioratom_res=26 prioratom="CA" res1=26 atom1="C" res2=1
atom2="N" followingatom="CA" followingatom_res=1 angle1=116.199993
angle2=121.69997 bondlength=1.32865 randomize_flanking_torsions=false />

       <AddPerturber effect="set_dihedral">
           <AddAtoms atom1="C" res1=26 res2=1 atom2="N" />
           <AddValue value=180.0 />
       </AddPerturber>

       <AddPerturber effect="randomize_alpha_backbone_by_rama">
       <AddResidue index=12/>
            <AddResidue index=13 />
       <AddResidue index=14 />
       <AddResidue index=25/>
       <AddResidue index=26/>
       <AddResidue index=1/>
       </AddPerturber>

       <AddFilter type="loop_bump_check" />
      <AddFilter type="backbone_bin" bin_params_file="ABBA" residue=12
bin="Bprime" />
      <AddFilter type="backbone_bin" bin_params_file="ABBA" residue=13
bin="A" />
      <AddFilter type="backbone_bin" bin_params_file="ABBA" residue=14
bin="B" />
      <AddFilter type="backbone_bin" bin_params_file="ABBA" residue=25
bin="B" />
      <AddFilter type="backbone_bin" bin_params_file="ABBA" residue=26
bin="A" />
      <AddFilter type="backbone_bin" bin_params_file="ABBA" residue=1 bin="B"
/>
   </GeneralizedKIC>

   <CreateTorsionConstraint name=peptide_torsion_constraint>
       <Add res1=26 res2=26 res3=1 res4=1 atom1="CA" atom2="C" atom3="N"
atom4="CA" cst_func="CIRCULARHARMONIC 3.141592654 0.005" />
       <Add res1=26 res2=26 res3=1 res4=1 atom1="O" atom2="C" atom3="N"
atom4="H" cst_func="CIRCULARHARMONIC 3.141592654 0.005" />
   </CreateTorsionConstraint>

   <CreateAngleConstraint name=peptide_angle_constraints>
       <Add res1=26 atom1="CA" res_center=26 atom_center="C" res2=1 atom2="N"
cst_func="CIRCULARHARMONIC 2.02807247 0.005" />
       <Add res1=26 atom1="C" res_center=1 atom_center="N" res2=1 atom2="CA"
cst_func="CIRCULARHARMONIC 2.12406565 0.005" />
   </CreateAngleConstraint>

   <CreateDistanceConstraint name=N_To_C_dist_cst>
       <Add res1=26 res2=1 atom1="C" atom2="N" cst_func="HARMONIC 1.32865
0.01" />
   </CreateDistanceConstraint>
```

```
   <Disulfidize name="disulf" min_disulfides="1" max_disulfides="1"
max_disulf_score="0.00" match_rt_limit="1" min_loop="3" use_d_cys="1"
use_l_cys="1" />

   <MultiplePoseMover name="disulfidizer" >
       <SELECT>
       </SELECT>
       <ROSETTASCRIPTS>

           <SCOREFXNS>
               <SFXN_STD weights= "beta_july15_cst.wts" />
           </SCOREFXNS>

           <TASKOPERATIONS>
               <ReadResfile name=resfile_daa filename="./resfile1.txt" />
               <ReadResfile name=resfile_laa filename="./resfile2.txt" />
               <DisallowIfNonnative name=nocysgly resnum=0 disallow_aas="CG"
/>
               <DisallowIfNonnative name=nocys resnum=0 disallow_aas="C" />
               <LayerDesign name=laydesign make_pymol_script=0
use_sidechain_neighbors=1 />


               ############## select CYS residues ###############
               <OperateOnCertainResidues name="no_repack_non-disulf" >
                   <PreventRepackingRLT/>
                   <ResidueName3Isnt name3="CYS" />
               </OperateOnCertainResidues>

               <OperateOnCertainResidues name="no_design_disulf" >
                   <RestrictToRepackingRLT />
                   <ResidueName3Is name3="CYS,DCYS" />
               </OperateOnCertainResidues>

               ############ miscellaneous for design ############
               <LimitAromaChi2 name="limitchi2" include_trp="1" />

               ########### layer selection for design ###########

                        ###Design with default layer design settings###

                        <LayerDesign name="layer_all_noALA_Laa"
layer="core_boundary_surface_Nterm_Cterm" verbose="True"
use_sidechain_neighbors="True" pore_radius=2.0 core=4.0 surface=1.8 >
                   <core>
                       <all append="M" exclude="A" />
                   </core>
                   <boundary>
                       <all exclude="A" />
                   </boundary>
                   <surface>
                       <all exclude="A" />
                   </surface>
               </LayerDesign>
```

```
                <LayerDesign name="layer_all_Laa"
layer="core_boundary_surface_Nterm_Cterm" verbose="True"
use_sidechain_neighbors="True" pore_radius=2.0 core=4.5 surface=1.8 >
                    <core>
                        <all append="M" />
                    </core>
                    <boundary>
                        <all />
                    </boundary>
                    <surface>
                        <all />
                    </surface>
                </LayerDesign>


                         ####Design with D-amino acid settings ###

                <LayerDesign name="layer_all_noALA_Daa"
layer="core_boundary_surface_Nterm_Cterm" verbose="True"
use_sidechain_neighbors="True" pore_radius=2.0 core=4.5 surface=1.8 >
                    <core>
                        <all ncaa_append="DPH,DLE,DIL,DPR,DVA,DTR,DTY" />
                    </core>
                    <boundary>
                        <all
ncaa_append="DVA,DTY,DTR,DTH,DSE,DPR,DPH,DLY,DLE,DIL,DGU,DAS,DAN,DAR,DGN" />
                    </boundary>
                    <surface>
                        <all
ncaa_append="DTH,DSE,DPR,DLY,DHI,DGU,DAS,DAN,DAR,DGN" />
                    </surface>
                </LayerDesign>
                <LayerDesign name="layer_all_Daa"
layer="core_boundary_surface_Nterm_Cterm" verbose="True"
use_sidechain_neighbors="True" pore_radius=2.0 core=4.0 surface=1.8 >
                    <core>
                        <all ncaa_append="DPH,DIL,DLE,DPR,DVA,DTR,DTY,DAL" />
                    </core>
                    <boundary>
                        <all
ncaa_append="DVA,DTY,DTR,DTH,DSE,DPR,DPH,DLY,DLE,DIL,DGU,DAS,DAN,DAR,DAL,DGN"
/>
                    </boundary>
                    <surface>
                        <all
ncaa_append="DTH,DSE,DPR,DLY,DHI,DGU,DAS,DAN,DAR,DGN,DAL" />
                    </surface>
                </LayerDesign>
            </TASKOPERATIONS>

            <FILTERS>
                <BuriedUnsatHbonds name=BuriedUnsat scorefxn=SFXN_STD
jump_number=0 cutoff=100 />

            </FILTERS>
```

```
           <MOVERS>
               <CreateTorsionConstraint name=peptide_torsion_constraint>
                   <Add res1=26 res2=26 res3=1 res4=1 atom1="CA" atom2="C"
atom3="N" atom4="CA" cst_func="CIRCULARHARMONIC 3.141592654 0.005" />
                   Add res1=26 res2=26 res3=1 res4=1 atom1="O" atom2="C"
atom3="N" atom4="H" cst_func="CIRCULARHARMONIC 3.141592654 0.005" />
               </CreateTorsionConstraint>

               <CreateAngleConstraint name=peptide_angle_constraints>
                   <Add res1=26 atom1="CA" res_center=26 atom_center="C"
res2=1 atom2="N" cst_func="CIRCULARHARMONIC 2.02807247 0.005" />
                   <Add res1=26 atom1="C" res_center=1 atom_center="N" res2=1
atom2="CA" cst_func="CIRCULARHARMONIC 2.12406565 0.005" />
               </CreateAngleConstraint>

               <CreateDistanceConstraint name=N_To_C_dist_cst>
                   <Add res1=26 res2=1 atom1="C" atom2="N" cst_func="HARMONIC
1.32865 0.01" />
               </CreateDistanceConstraint>


               <FastDesign name=fdesign2 scorefxn=SFXN_STD repeats=2
task_operations=resfile_daa,layer_all_noALA_Daa,resfile_laa,layer_all_noALA_D
aa,nocys,no_design_disulf,limitchi2 ramp_down_constraints=false>

                   <MoveMap name=fdesign_mm>
                       <Chain number=1 chi=true bb=true />

                   </MoveMap>

               </FastDesign>

               <FastDesign name=fdesign6 scorefxn=SFXN_STD repeats=6
task_operations=resfile_daa,layer_all_Daa,resfile_laa,layer_all_Laa,nocys,no_
design_disulf,limitchi2 ramp_down_constraints=false>

                   <MoveMap name=fdesign_mm>
                       <Chain number=1 chi=true bb=true />

                   </MoveMap>

               </FastDesign>

               <DeclareBond name=peptide_bond1 res1=1 atom1="N" atom2="C"
res2=26 add_termini=true />

           </MOVERS>
           <PROTOCOLS>

               <Add mover=peptide_torsion_constraint />
               <Add mover=peptide_angle_constraints />
               <Add mover=N_To_C_dist_cst />
               <Add mover=fdesign2 />
               <Add mover=fdesign6 />
               <Add mover=peptide_bond1 />
```

```
               <Add filter=BuriedUnsat />

           </PROTOCOLS>
       </ROSETTASCRIPTS>
   </MultiplePoseMover>

</MOVERS>
<PROTOCOLS>
      <Add mover=intial_stub />
   <Add mover=torsion1 />
   <Add mover=peptide_bond1 />
   <Add mover=genkic1 />
   <Add mover="disulf" />
   <Add mover_name="disulfidizer" />
</PROTOCOLS>

</ROSETTASCRIPTS>
```

1.8.2 Resfile for designing D-amino acids in the cyclic heterochiral topology

Resfiles control the behaviour of the Rosetta packer, which optimizes sidechain conformations and/or identities given a fixed backbone. Note that, in this case, the following is intended for use with LayerDesign (as shown in script 8.1, above), which will activate D-amino acid design at the “empty” positions.

```
ALLAAwc
EX 1 EX 2
USE_INPUT_SC

start
12 A EMPTY
15 A EMPTY
16 A EMPTY
17 A EMPTY
18 A EMPTY
19 A EMPTY
20 A EMPTY
21 A EMPTY
22 A EMPTY
23 A EMPTY
24 A EMPTY
```

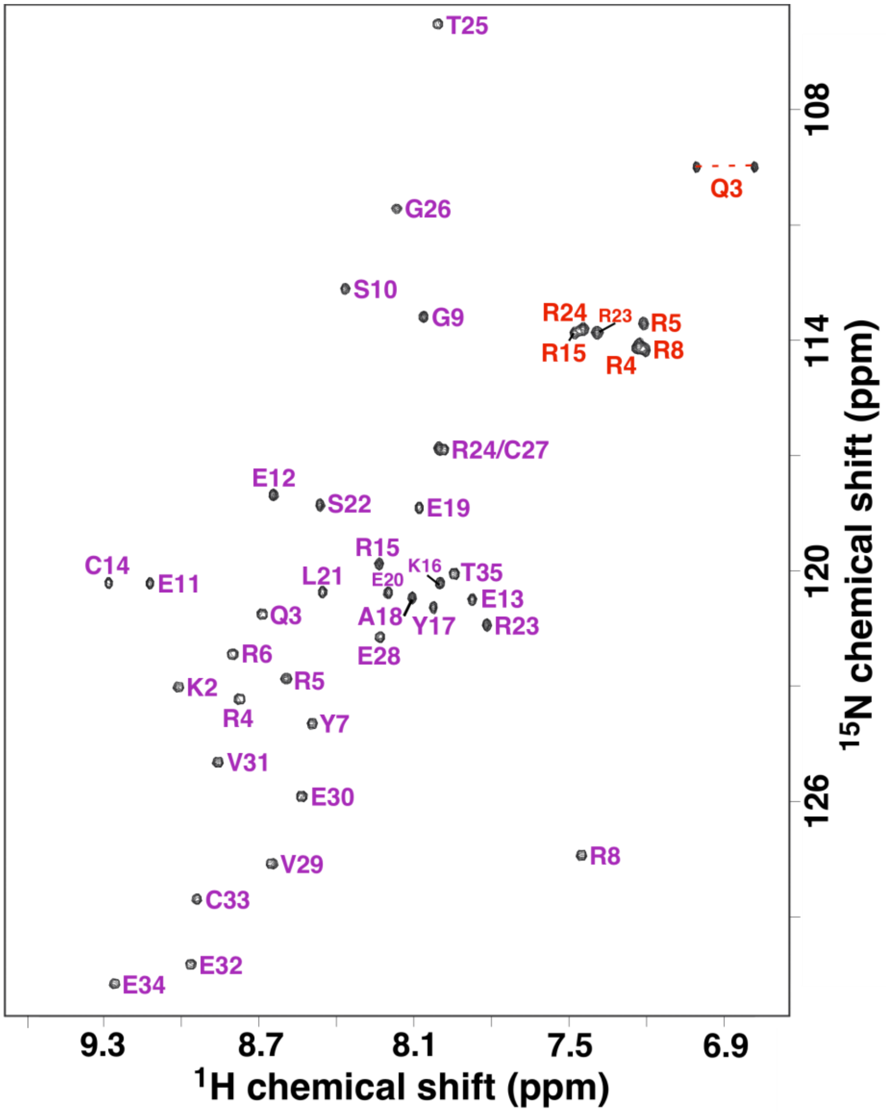
1.8.3 Resfile for designing L-amino acids in the cyclic heterochiral topology

Note that the following is intended for use with LayerDesign (as shown in script 8.1, above); the “RESET” commands are necessary to deactivate D-amino acid design at L-amino acid positions.

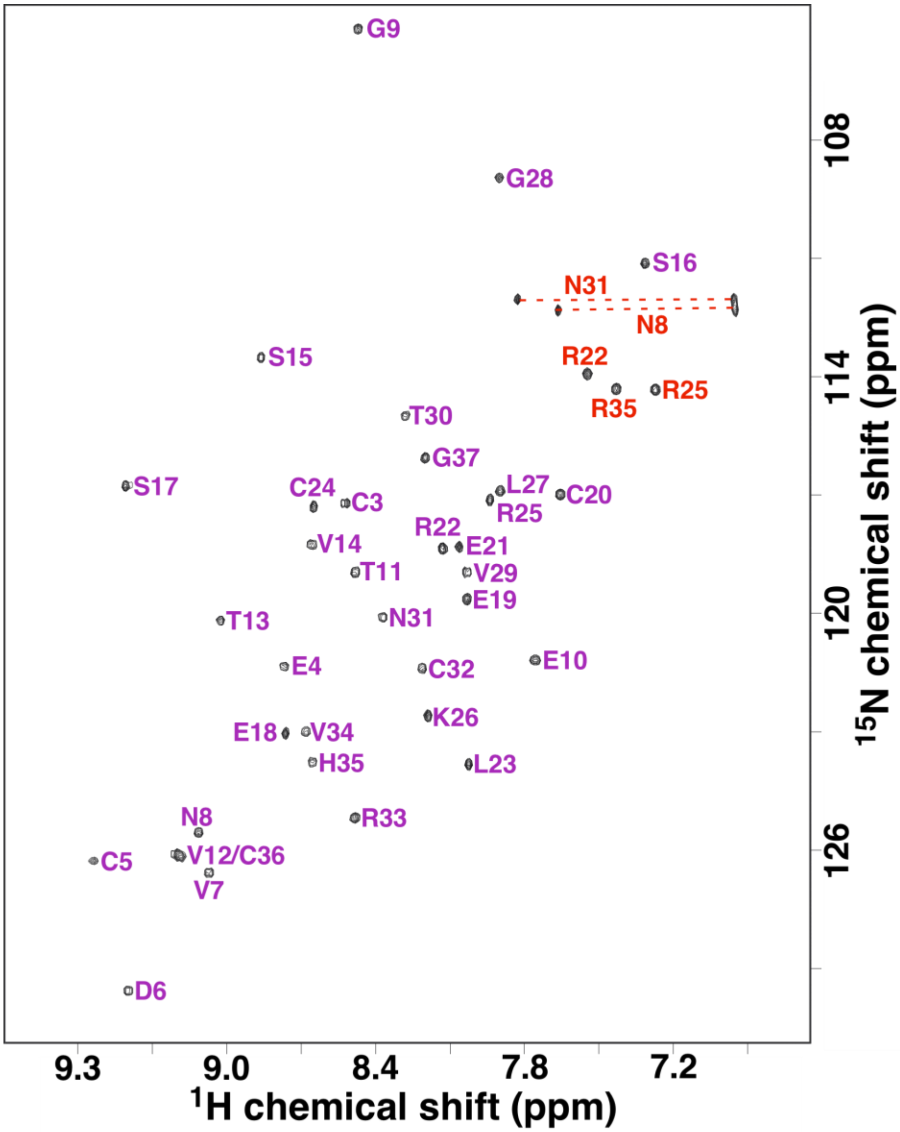
```
start
1 A RESET
2 A RESET
3 A RESET
```

```
4 A RESET
5 A RESET
6 A RESET
7 A RESET
8 A RESET
9 A RESET
10 A RESET
11 A RESET
13 A RESET
14 A RESET
25 A RESET
26 A RESET
```

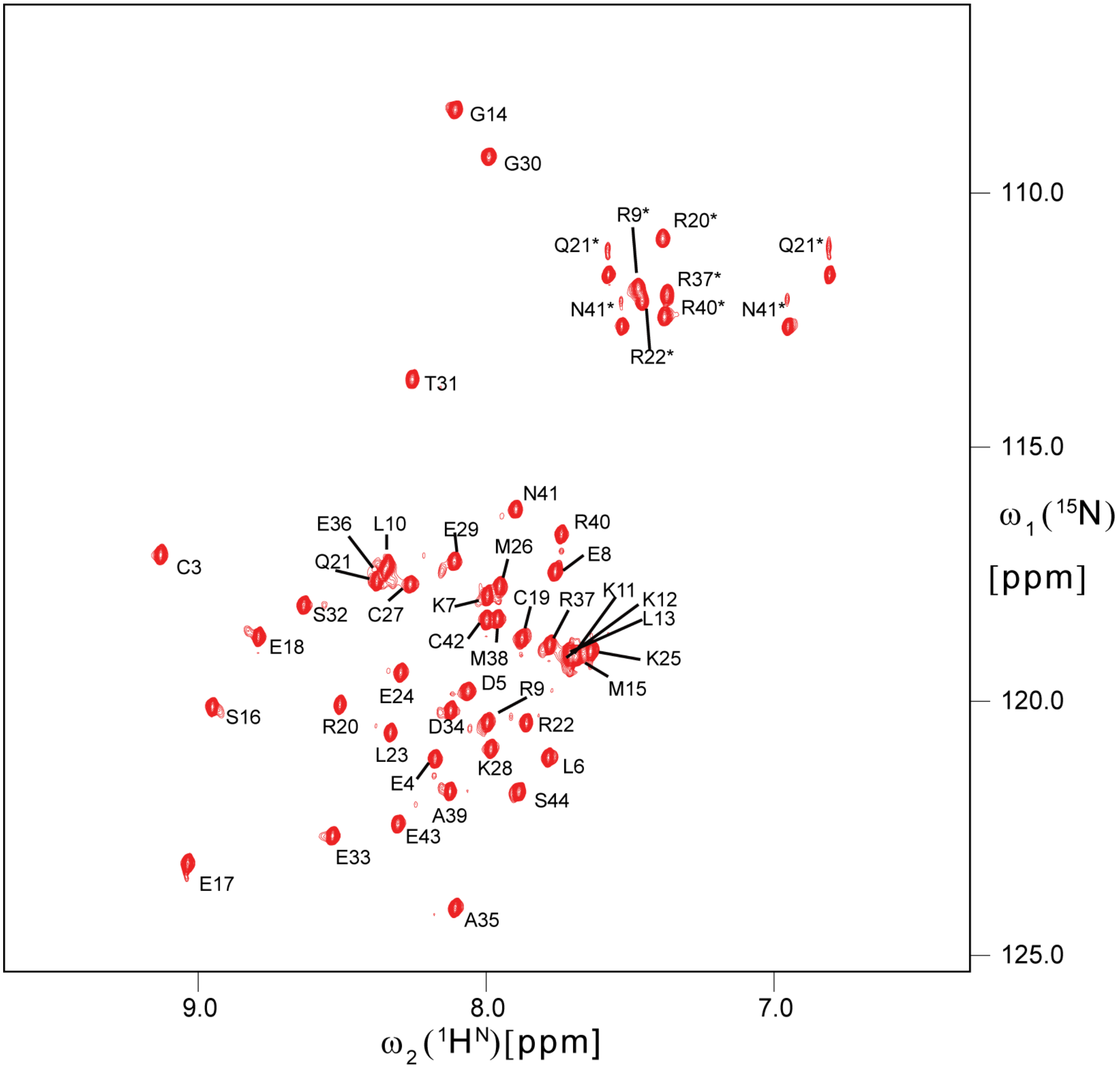
Supplementary information 2: Supplementary structural data



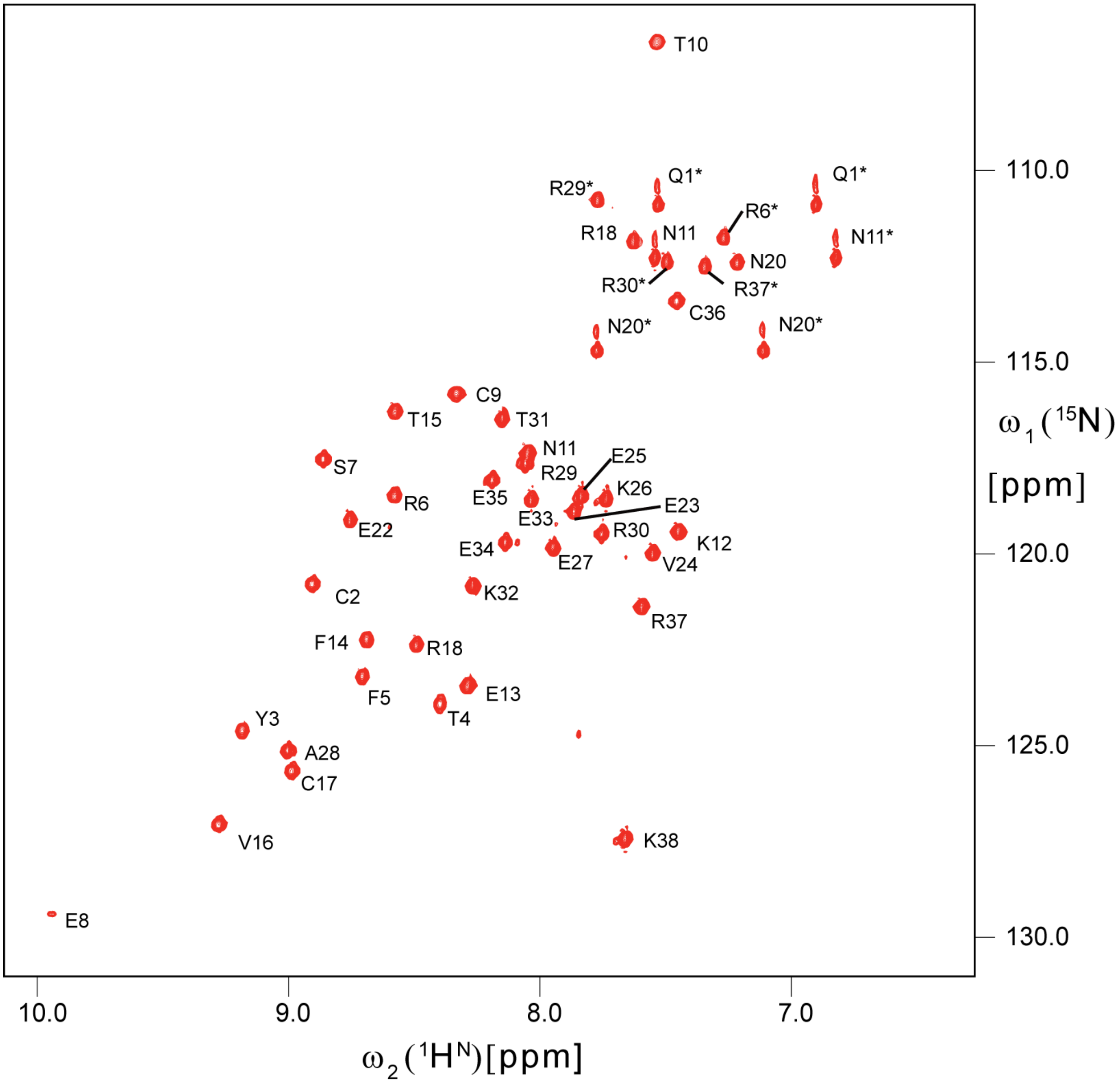
Supplementary Figure S2-1. The ^{1}H-^{15}N HSQC spectrum for gEHE_06 (~1 mM) collected at a proton resonance frequency of 500 MHz, 20 °C, in 50 mM sodium chloride, 25 mM sodium acetate, pH 4.8, with assigned backbone amide resonances and side chain resonances labeled purple and red, respectively. The wide chemical shift dispersion of the amide resonances in the nitrogen and proton dimension is characteristic of a structured protein[9].



Supplementary Figure S2-2. The ^{1}H-^{15}N HSQC spectrum for gEEHE_02 (~0.5 mM) collected at a proton resonance frequency of 500 MHz, 20 °C in 50 mM sodium chloride, 25 mM sodium acetate, pH 4.8, with assigned backbone amide resonances and side chain resonances labeled purple and red, respectively. The wide chemical shift dispersion of the amide resonances in the nitrogen and proton dimension is characteristic of a structured protein[9].



Supplementary Figure S2-3. The ^{1}H-^{15}N HSQC spectrum for gHHH_06 (~1 mM) collected at a proton resonance frequency of 750 MHz, 20 °C, 50 mM sodium phosphate, pH 6.0, 4 uM 4,4-dimethyl-4-silapentane-1-sulfonic acid salt, 0.02% sodium azide with the backbone amide resonances labeled. The side chain Asn, Gln, and Gln resonances are labeled with an asterisk.



Supplementary Figure S2-4. The ^{1}H-^{15}N HSQC spectrum for gEEH_04 (1 mM) collected at a proton resonance frequency of 750 MHz, 20 °C, 50 mM sodium phosphate, pH 6.0, 4 uM 4,4-dimethyl-4-silapentane-1-sulfonic acid, 0.02% sodium azide with the backbone amide resonances labeled. The side chain Asn, Gln, and Gln resonances are labeled with an asterisk.

Supplementary Table S2-1: Summary of the structural statistics^a for gHHH_06, gEHH_4, gEHE_06, and gEEHE_02

Design	*gHHH_06*	*gEEH_04*	*gEHE_06*	*gEEHE_02*
Completeness of ^{1}H resonance assignments [b] **(%)**				
Backbone / Side-chain	100/90	99/70	96/72	97/84
Conformationally-restricting constraints[c]				
Distance constraints				
Total	742	614	317	301
intra-residue (*i* = *j*)	224	135	116	100
sequential (|*i*-*j*| = 1)	220	166	102	96
medium range (1 < |*i* - *j*| < 5)	242	156	43	35
long range (|*i* - *j*| ≥ 5)	56	157	56	70
Dihedral angle constraints	54	44	54	46
Disulfide bond constraints	6	6	6	9
Hydrogen bond constraints	-	-	40	34
No. of constraints per residue	19.0	17.8	11.9	10.5
No. of long range constraints per residue	1.5	4.7	1.6	1.9
Residual constraint violations[c]				
Average no. of distance violations per structure:				
0.1 - 0.2 Å	9.1	5.3	0.4	0.1
0.2 - 0.5 Å	4.75	2.05	0	0
> 0.5 Å	0.7	0	0	0
Average no. of dihedral angle violations per structure:				
1 - 10°	6.6	4.75	0.1	0.35
Model Quality[c]				
RMSD backbone atoms (Å) [c]	0.51 ± 0.10	0.42 ± 0.11	0.55 ± 0.12	0.46 ± 0.09
RMSD heavy atoms (Å) [c]	1.16 ± 0.11	1.12 ± 0.28	1.43 ± 0.11	1.21 ± 0.11
RMSD bond lengths (Å)	0.018	0.021	0.005	0.005
RMSD bond angles (º)	1.2	1.1	0.7	0.6
MolProbity Ramachandran statistics [c]				

Most favored regions (%)	96.9		96.9		97.8		96.5	
Allowed regions (%)	3		2.6		2.2		3.5	
Disallowed regions (%)	0.1		0.4		0.0		0.0	
Global quality scores (Raw / *Z*-score)[c]								
Verify3D	0.34	-1.93	0.22	-3.85	0.35	-1.77	0.42	-0.54
ProsaII	1.38	3.02	0.67	0.88	0.78	0.54	1.14	2.03
Procheck (phi-psi)[c]	0.40	1.89	-0.01	0.28	-0.02	0.24	-0.12	-0.16
Procheck (all)[c]	0.16	0.95	-0.09	-0.53	-0.04	-0.24	-0.19	-1.12
MolProbity clash score	15.6	-1.15	16.8	-1.37	17.34	-1.45	18.5	-1.66
RPF Scores[d]								
Recall / Precision	0.95	0.92	0.92	0.87	0.88	0.91	0.98	0.93
F-measure / DP-score	0.93	0.75	0.89	0.72	0.89	0.55	0.96	0.82
BMRB accession number	26045		26046		30067		30069	
PDB ID	2ND2		2ND3		5JHI		5JI4	

[a] Structural statistics computed for the ensemble of 20 deposited structures.
[b] Computed using AVS software from the expected number of resonances, excluding: highly exchangeable protons (N-terminal, Lys, and Arg amino groups, hydroxyls of Ser, Thr, Tyr), carboxyls of Asp and Glu, and non-protonated aromatic carbons.
[c] Calculated using PSVS 1.5. Average distance violations were calculated using the sum over r^{-6}.
[d] RPF scores reflecting the goodness-of-fit of the final ensemble of structures (including disordered residues) to the NOESY data and resonance assignments.

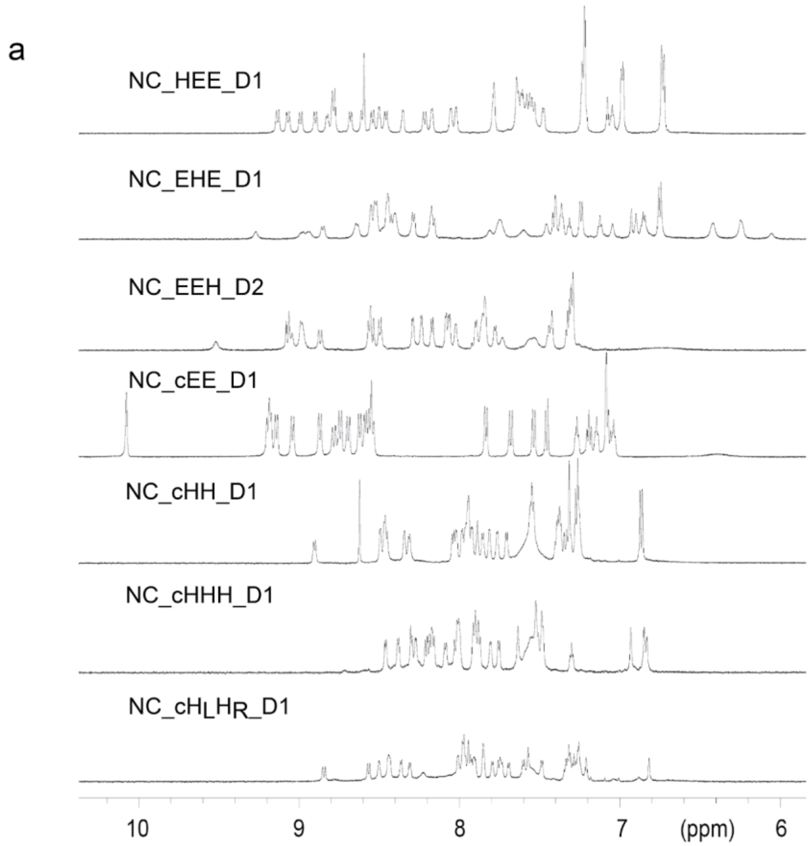
Supplementary Table S2-2. Crystallographic statistics

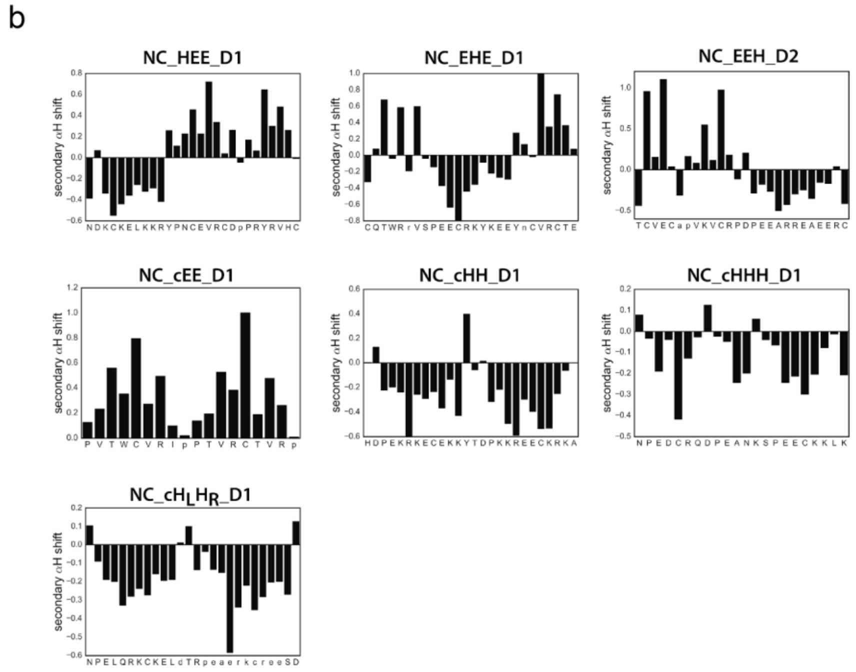
Design	gEHEE_06	EEHE_2.1_02_0008
Data collection		
Space group	$P2_1$	$P2_12_12_1$
a, b, c (Å)	34.9, 45.5, 49.7	68.0, 109.7, 122.7
□, □, □ (°)	90.0, 105.1, 90.0	
Resolution (Å)	50.00-2.09 (2.13-2.09)	50.00-2.90 (2.95-2.90)
Unique reflections	8734	20164
Average redundancy	3.5 (2.8)	3.3 (3.4)
Completeness (%)	96.7 (78.7)	98.7 (99.7)
R_{merge} (%)	11.1 (48.0)	21.1 (56.3)
I/□(I)	14.4 (2.9)	12.0 (3.9)
Refinement statistics		
R_{cryst} (%)	20.0	
R_{free} (%)	24.7	
Number of atoms		
Protein	1226	
Water	75	
R.M.S deviations		
Bond lengths (Å)	0.01	
Bond angles (°)	1.62	
Ramachandran		
Favored (%)	97.8	
Allowed (%)	2.2	
Generously allowed (%)	0	
Disallowed (%)	0	
PDB ID	5JG9	

Highest resolution shell is shown in parenthesis.

Supplementary Table S2-3: Summary of the structural statistic for NC_cHHH_D1, NC_cHH_D1, NC_cEE_D1, NC_HEE_D1, NC_EHE_D1, NC_EEH_D2, NC_cH$_{L}$H$_{R}$_D1

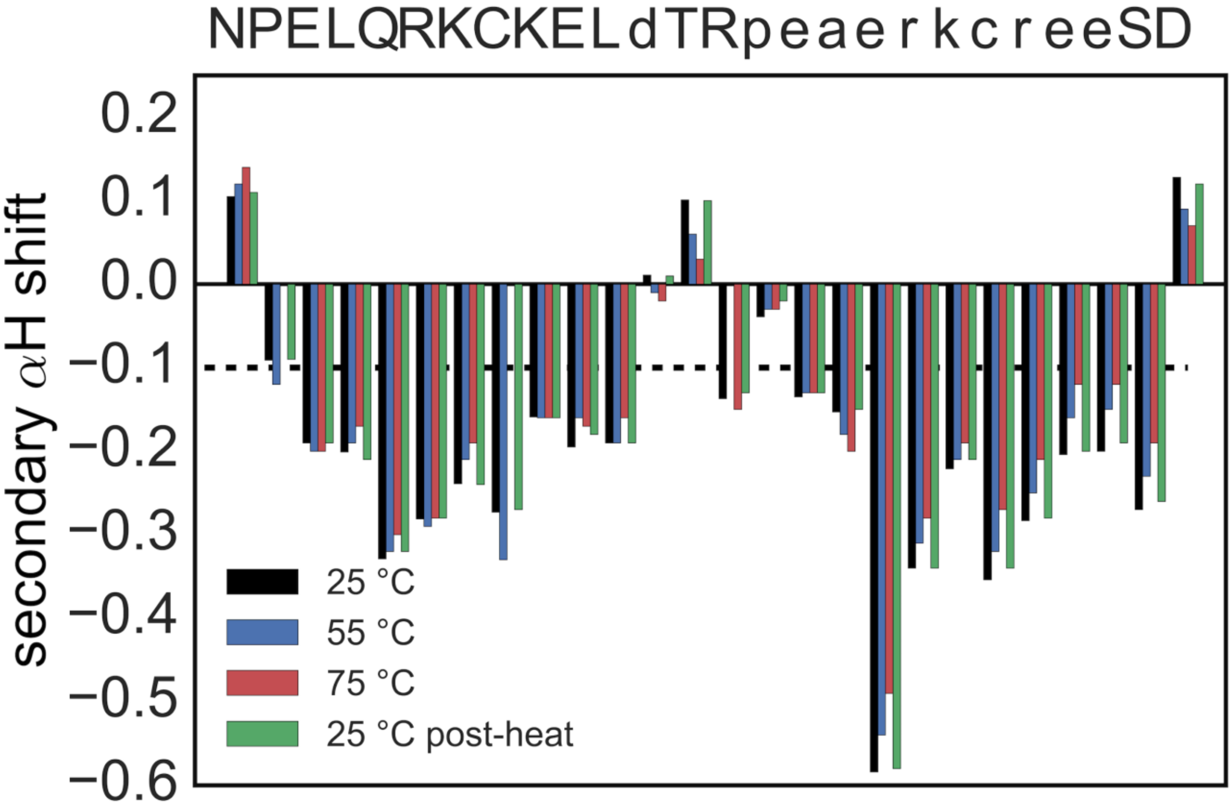
Experimental restraints	NC_cHHH_D1	NC_cHH_D1	NC_cEE_D1	NC_EHE_D1	NC_HEE_D1	NC_EEH_D2	NC_cH$_{L}$H$_{R}$_D1
total no. distance restraints	131	207	119	229	312	220	223
intraresidue	70	84	59	87	100	85	107
sequential	50	74	49	77	108	85	80
medium range, *i-j*<5	7	32	4	36	42	24	31
long range, *i-j*≥5	4	17	7	29	62	26	5
hydrogen bond restraints	6	24	16	18	20	20	16
dihedral angle restraints							
phi	18	21	14	20	21	20	12
psi	17	22	14	18	21	20	9
chi1	7	9	3	8	8	5	5
Deviations from idealized geometry							
bond lengths (Å)	0.008 ± 0.001	0.008 ± 0.000	0.010 ± 0.000	0.010 ± 0.000	0.010 ± 0.001	0.009 ± 0.000	0.008 ± 0.000
bond angles (deg)	0.925 ± 0.064	1.078 ± 0.057	1.029 ± 0.037	1.075 ± 0.033	1.075 ± 0.045	1.077 ± 0.049	1.061 ± 0.048
impropers (deg)	1.32 ± 0.18	1.24 ± 0.15	1.20 ± 0.13	1.21 ± 0.13	1.20 ± 0.14	1.14 ± 0.12	1.23 ± 0.14
NOE (Å)	0.005 ± 0.002	0.010 ± 0.002	0.006 ± 0.003	0.005 ± 0.003	0.011 ± 0.002	0.005 ± 0.003	0.006 ± 0.001
cDih (deg)	0.100 ± 0.090	0.058 ± 0.070	0.092 ± 0.075	0.084 ± 0.084	0.098 ± 0.081	0.091 ± 0.069	0.000 ± 0.000
Mean energies (kcal/mol)							
overall	-796 ± 65	-1154 ± 74	-475 ± 12	-958 ± 68	-1029 ± 57	-985 ± 54	-1049 ± 68
bonds	5.1 ± 0.8	7.2 ± 0.7	7.9 ± 0.7	10.0 ± 1.0	11.2 ± 1.2	8.4 ± 0.7	6.8 ± 0.7
angles	20.0 ± 3.2	31.8 ± 3.8	18.8 ± 1.6	30.9 ± 2.5	31.6 ± 2.8	28.4 ± 3.1	27.9 ± 2.9
improper	9.4 ± 2.1	11.6 ± 2.4	7.8 ± 1.3	11.8 ± 2.1	12.2 ± 2.1	9.6 ± 1.7	11.0± 1.9
van Der Waals	-74.7 ± 5.8	-107.4 ± 4.7	-64.1 ± 2.4	-120.6 ± 6.0	-121.8 ± 5.0	-94.9 ± 6.3	-100.4 ± 5.0
NOE	0.00 ± 0.00	0.02 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.04 ± 0.01	0.01 ± 0.01	0.01 ± 0.00
cDih	0.09 ±0.11	0.05 ±0.08	0.05 ±0.07	0.08 ±0.11	0.10 ±0.14	0.07 ±0.08	0.00 ±0.00
electrostatic	-858 ± 69	-1222 ± 75	-523 ± 10	-1014 ± 71	-1086 ± 59	-1054 ± 58	-1118 ± 70
Violations							
NOE violations exceeding 0.2 Å	0	0	0	0	0	0	0
Dihedral violations exceeding 2.0 Å	0	0	0	0	0	0	0
Rms deviation from mean structure, Å							
backbone atoms	1.14 ± 0.34	0.89 ± 0.31	0.63 ± 0.19	0.93 ± 0.33	1.01 ± 0.32	0.70 ± 0.16	0.70 ± 0.19
all heavy atoms	2.13 ± 0.35	2.06 ± 0.39	1.44 ± 0.26	2.01 ± 0.33	1.96 ± 0.33	1.74 ± 0.30	1.96 ± 0.28
Stereochemical quality[b]							
Residues in most favoured Ramachandran region, %	99.2 ± 1.8	99.8 ± 0.9	92.5 ± 2.5	92.6 ± 2.4	95.4 ± 1.2	95.4 ± 1.2	83.8 ± 4.4
Ramachandran outliers, %	0.0 ± 0.0	0.0 ± 0.0	6.2 ± 0.0	5.7 ± 2.0	4.2 ± 0.0	4.2 ± 0.0	6.9 ± 2.4
Unfavourable sidechain rotamers, %	0.7 ± 2.3	0.4 ± 1.2	0.0 ± 0.0	0.0 ± 0.0	0.2 ± 0.8	0.0 ± 0.0	0.0 ± 0.0
Clashscore, all atoms	7.3 ± 4.0	4.8 ± 2.7	3.7 ± 2.1	6.7 ± 3.2	8.5 ± 3.2	7.4 ± 2.9	5.6 ± 2.6
Overall MolProbity score	1.4 ± 0.2	1.2 ± 0.2	1.5 ± 0.3	1.8 ± 0.2	1.8 ± 0.2	1.7 ± 0.2	1.9 ± 0.2


a
NC_HEE_D1
NC_EHE_D1
NC_EEH_D2
NC_cEE_D1
NC_cHH_D1
NC_cHHH_D1
NC_cH$_L$H$_R$_D1
10
9
8
7
(ppm)
6


b
NC_HEE_D1
NC_EHE_D1
NC_EEH_D2
NC_cEE_D1
NC_cHH_D1
NC_cHHH_D1
NC_cH$_L$H$_R$_D1
secondary αH shift
NDKCKELKKRYPNCEVRCDpPRYRVHC
CQTWRrVSPEECRKYKEEYnCVRCTE
TCVECapVKVCRPDPEEARREAEERC
PVTWCVRIpPTVRCTVRp
HDPEKRKECEKKYTDPKKREECKRKA
NPEDCRQDPEANKSPEECKKLK
NPELQRKCKELdTRpeaerkcreeSD

Supplementary Figure S2-5 : NMR spectroscopy analysis of designed non-canonical peptides

a) Proton NMR spectra for each of the seven designed topologies recorded at a ^{1}H resonance frequency of 600 MHz, 25 °C. Spectra are well-dispersed and sharp, consistent with folded proteins. b) Secondary $^1H_\alpha$ chemical shifts (in ppm) for each of the seven designed topologies.



Supplementary Figure S2-6: Secondary $^1H_\alpha$ chemical shifts at a range of temperatures for peptide NC_cH$_L$H$_R$_D1

NMR spectra were collected at 25 °C (black bars), 55 °C (blue bars), 75 °C (red bars), and again after cooling to 25 °C (green bars). Secondary chemical shifts are largely unchanged during heating, showing clear alpha-helical signatures for residues 2-11 (the designed α_R-helix) and residues 16-25 (the designed α_L-helix), indicating no significant loss of secondary structure resulting from heating. Secondary chemical shifts are identical to the original values after cooling, indicating that the peptide is also not aggregation-prone or otherwise prone to irreversible conformation changes on heating. Overall, these results indicate considerable thermostability.

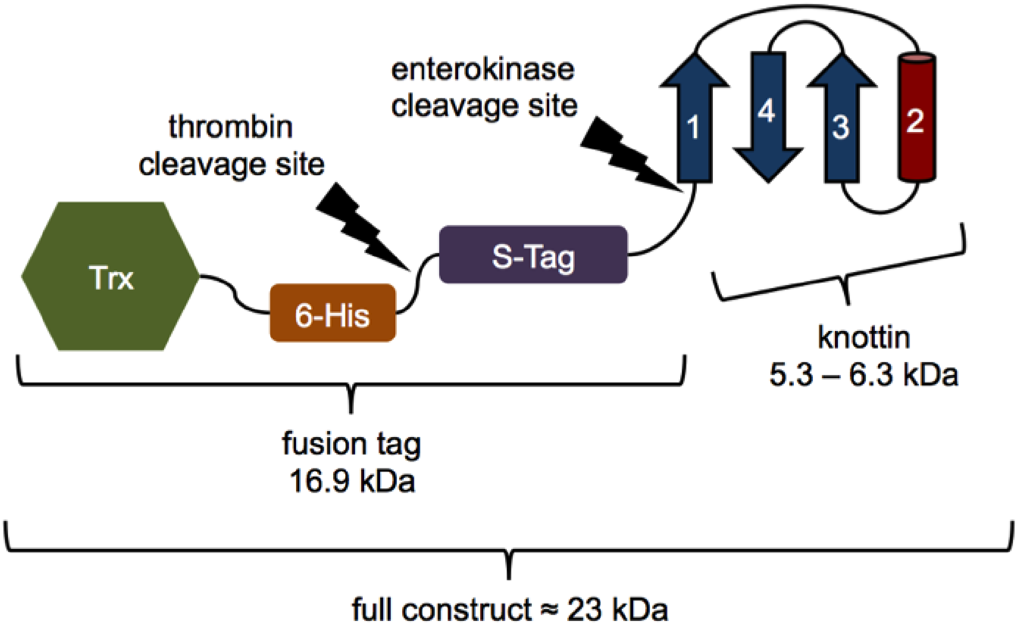
Supplementary information 3: Screening of genetically encodable peptide designs

Experimental screening of genetically encodable disulfide-rich peptides was carried out in successive rounds of gene orders. The amino acid sequence, image of designed model, Rosetta *ab initio* structure prediction results, and where applicable, preliminary circular dichroism and reverse-phase HPLC data are shown for each tested design.

Round 1

This preliminary round of expression were carried out by expressing genes cloned into pET32b (Novagen) and expressed from Rosetta-gami B(DE3) *Escherichia coli* (Novagen).

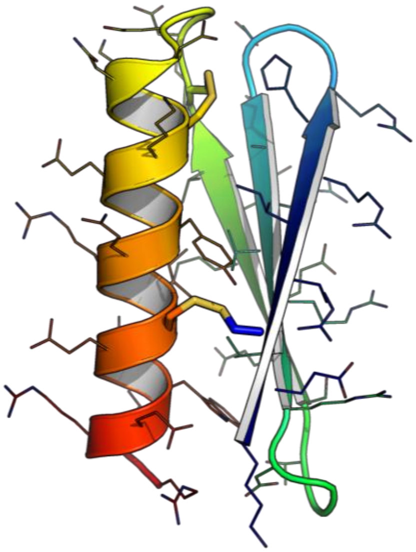
anatomy of the expression constructs

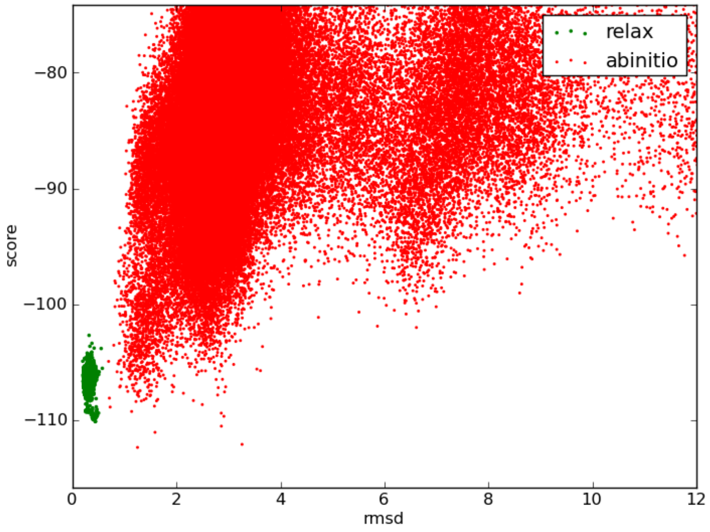


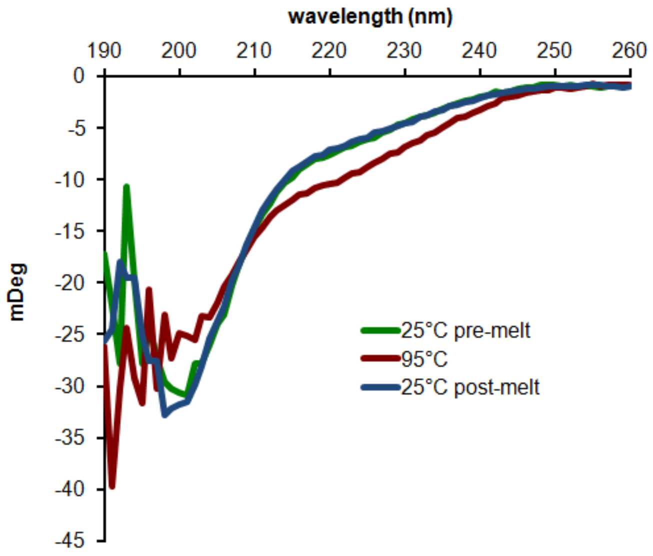
Designs were purified by immobilized metal affinity chromatography (IMAC), and the fusion protein was removed using enterokinase. Protease digestion is inhibited by a proline at the P1' position (first residue of the design). To facilitate removal of the fusion protein, an alanine residue was added to the N-terminus of any design beginning with a proline during generation of the synthetic gene.

EEEH_1.4_03

amino acid sequence:
KECRIRHRGDKARVRVRDGGTSEEREVKCDGDDNKCKEAYQRICEEWERKR

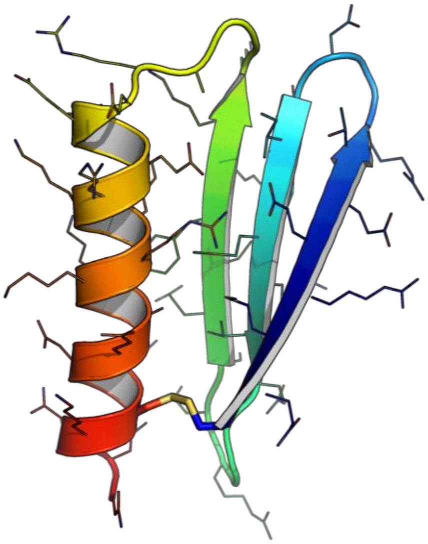


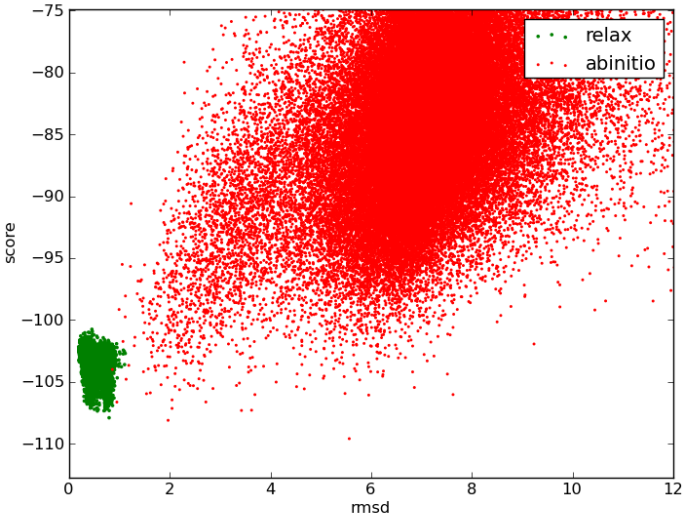




EEEH_1.4_12

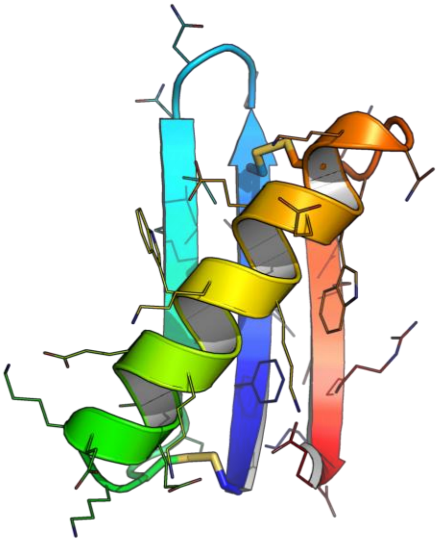
amino acid sequence:
CQMREETRGNTIVMRVQGGRDSEEFRKKGGAREEEERKYRKKAEDKCKNNQ

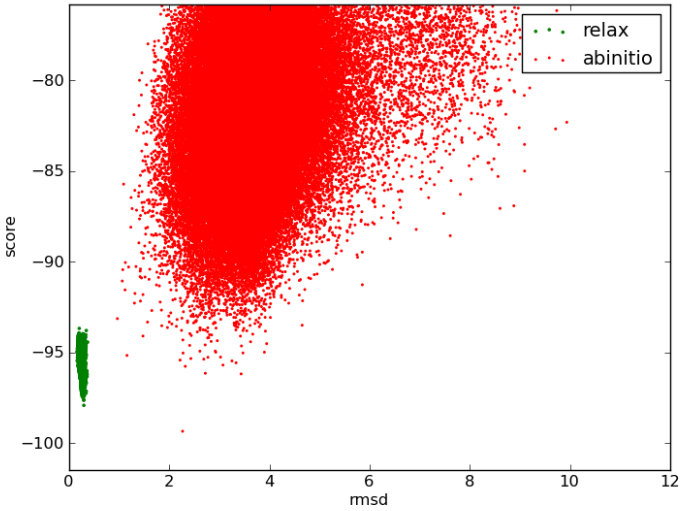


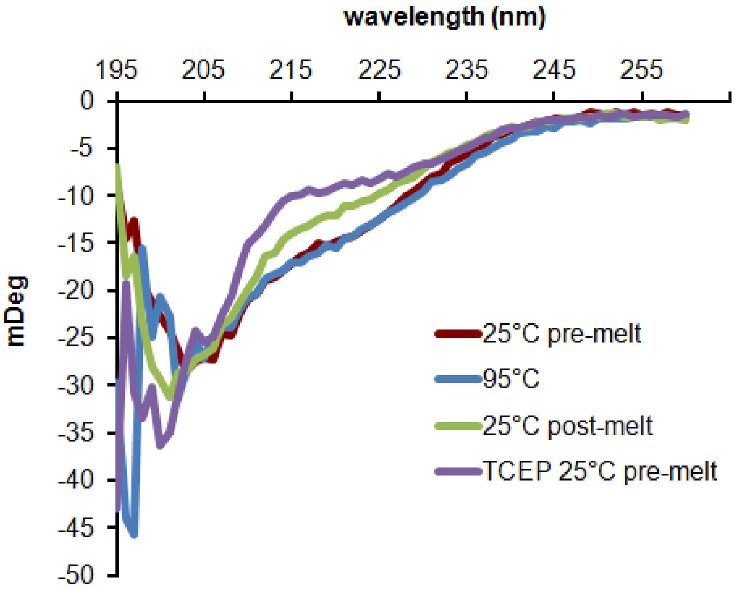


EEHE_1.3_04

amino acid sequence:
CRFRAECQGNNVHVRGDGCKKEEIEKAWKKAEEWCKNGMQSSEREE

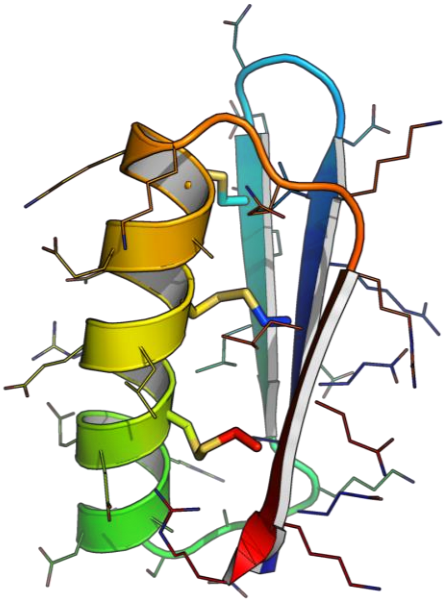


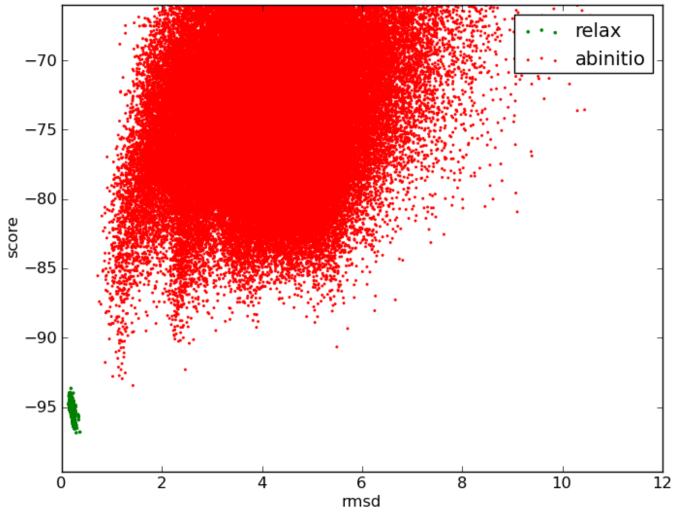




EEHE_1.3_18

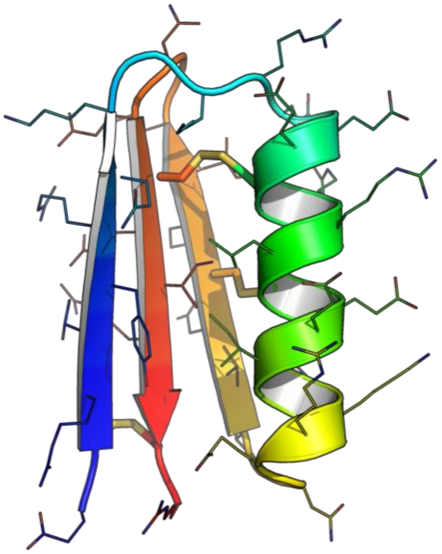
amino acid sequence:
DEAECRIDGNECRLDAKGASDDAREECRELCEEACKKGQKRLQCKR

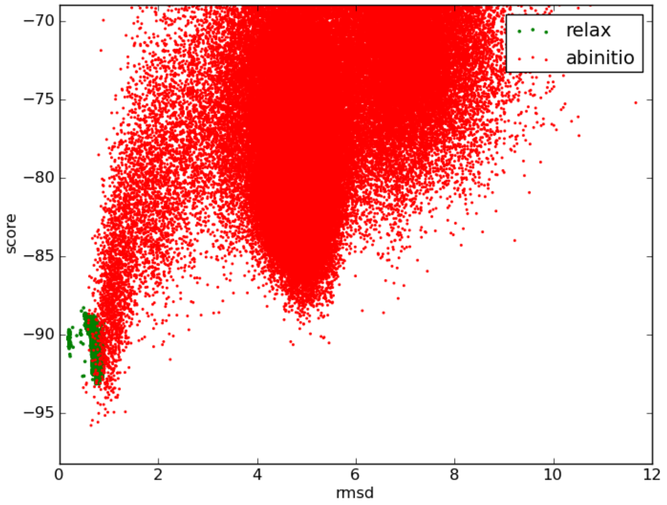


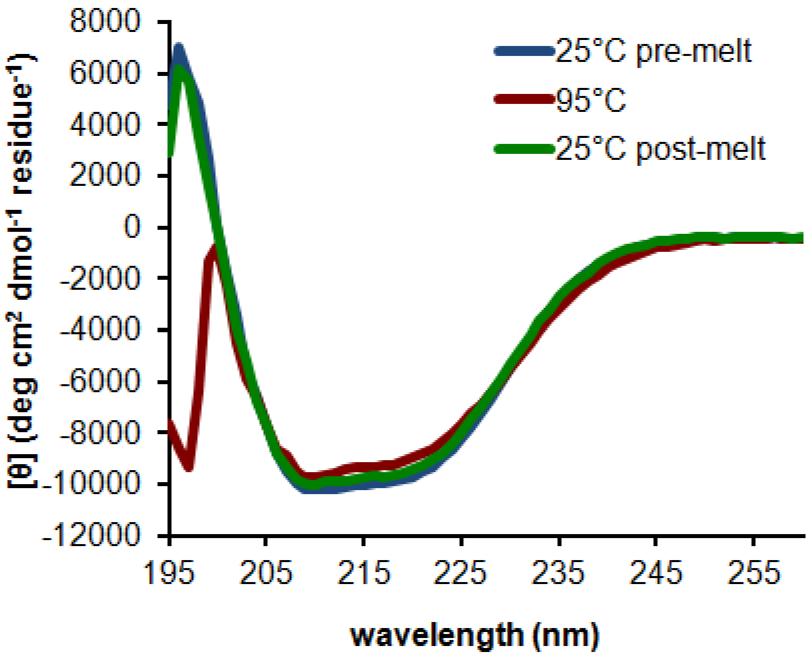


EHEE_1.7_02

amino acid sequence:
ECRTYRQKGKREEECRRLCEEIRKRENGTVDCQIDGNECEIRACR

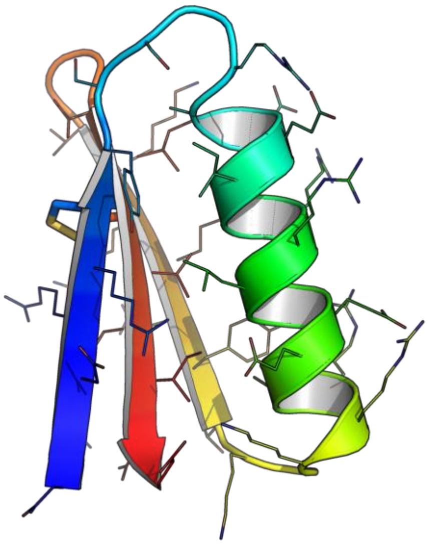


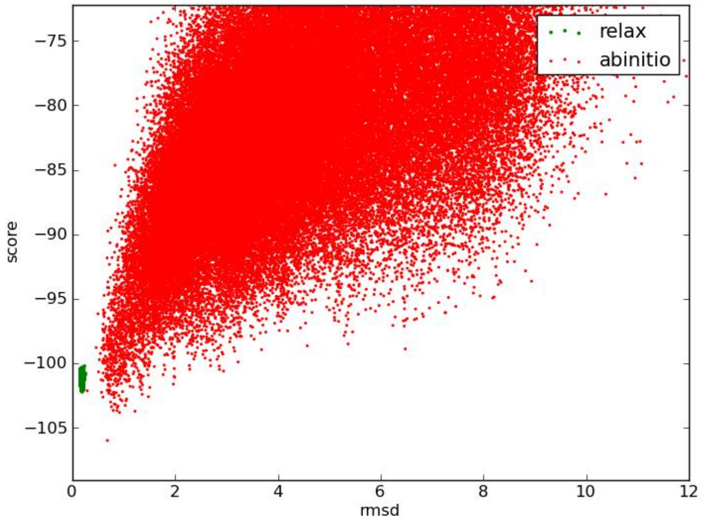


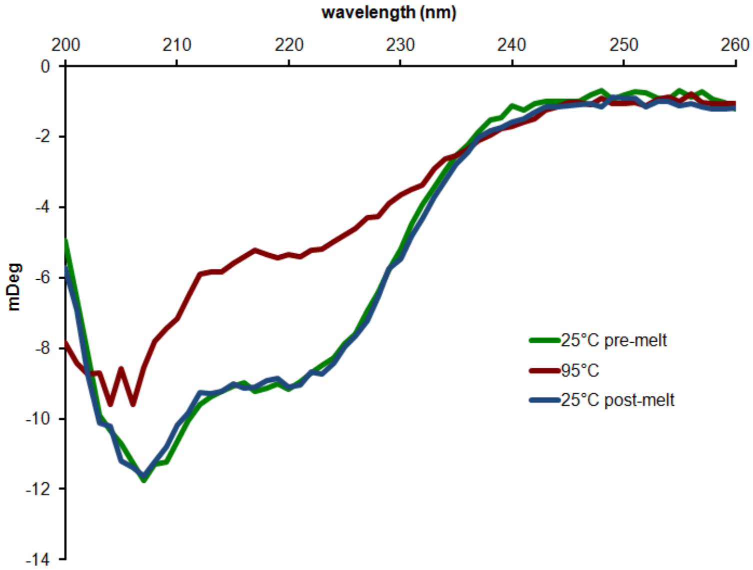


EHEE_1.7_03

amino acid sequence:
NRDRRCYSSGRAEEIARRLAEEARRKGKTYEERKTGGTICVEIDE

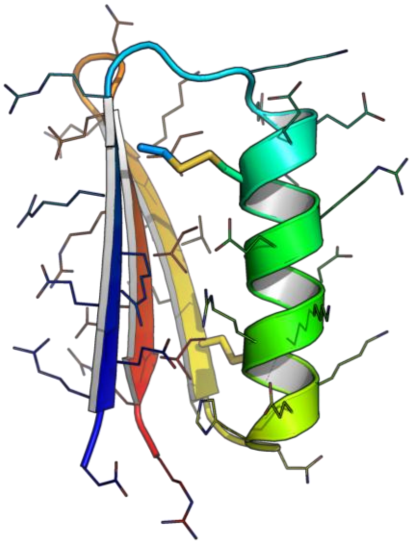


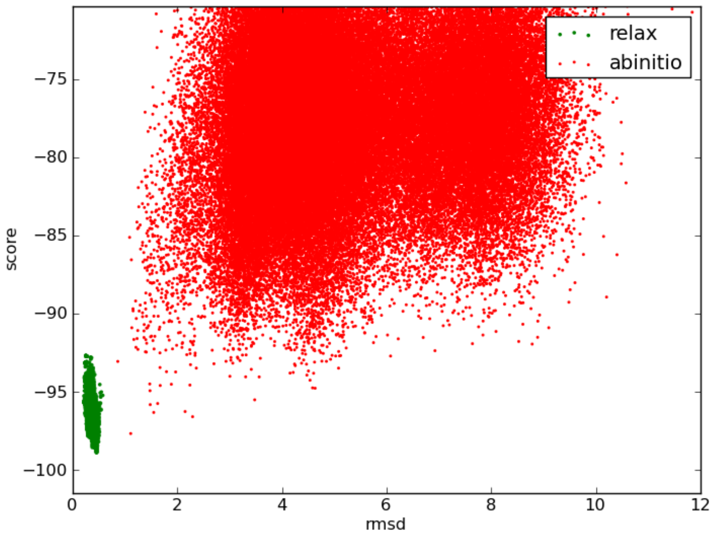


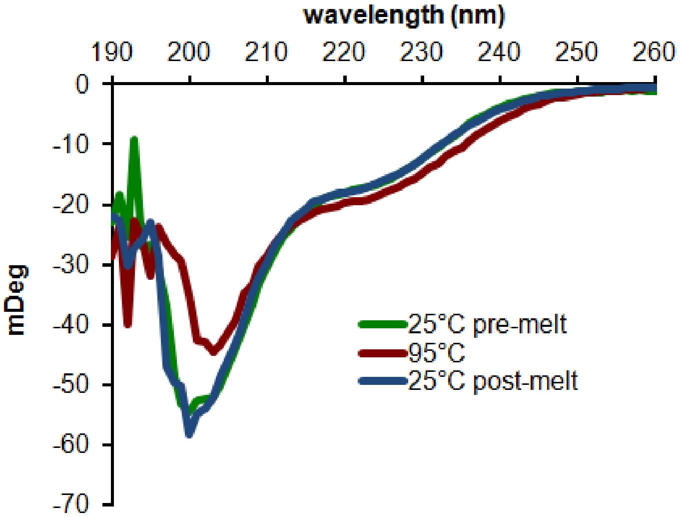


EHEE_1.7_04

amino acid sequence:
DREERRCRGGKEEECRREAEKRCKEHNGTCEVRKQGNEIRIEIRR

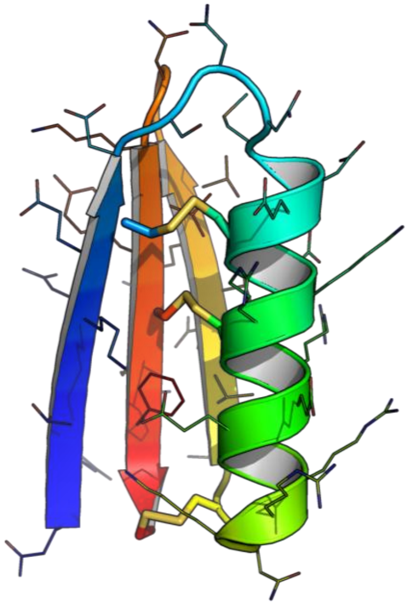


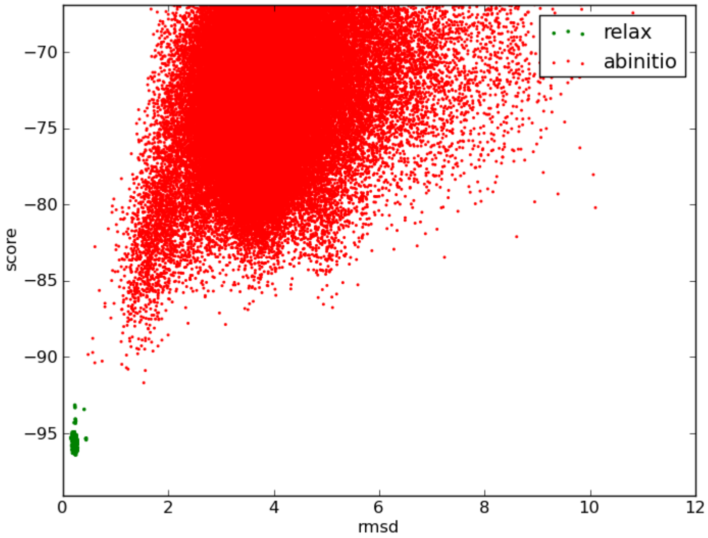


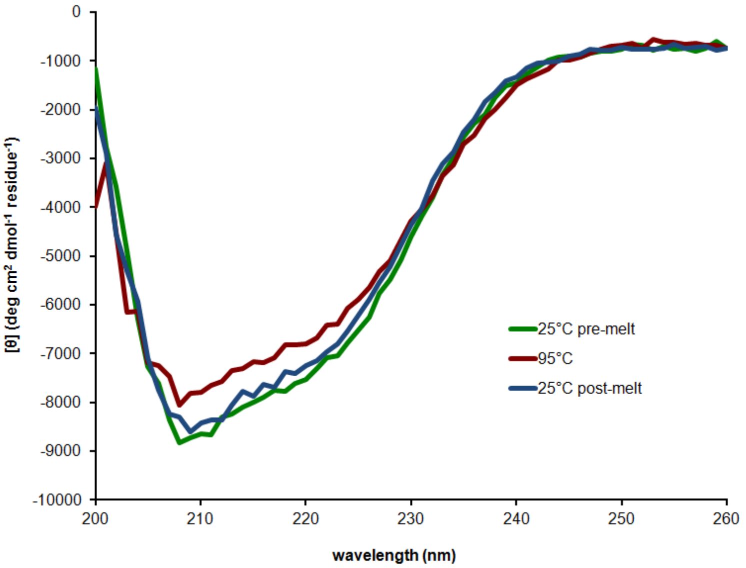


EHEE_1.7_05

amino acid sequence:
QRTRKECDSNNMDECEKRCREEARRKNCRVEIRTRGNKVYCRFEC

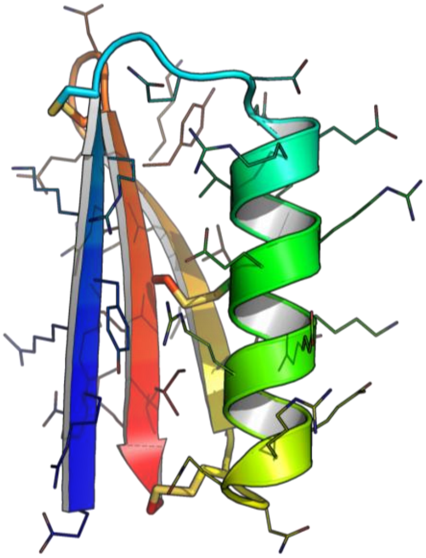


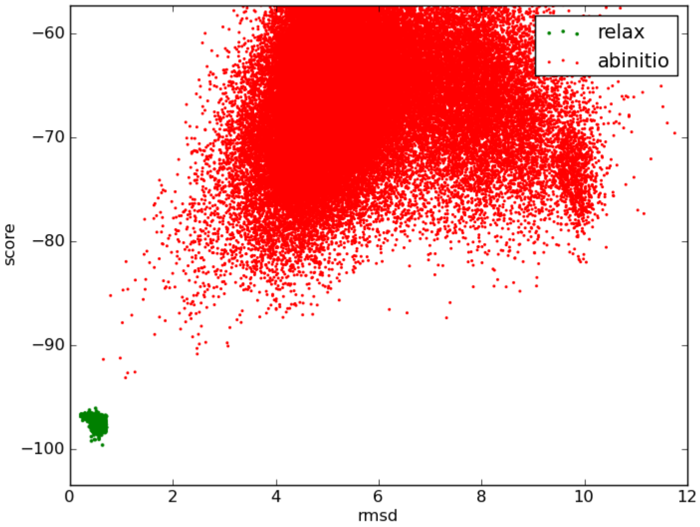


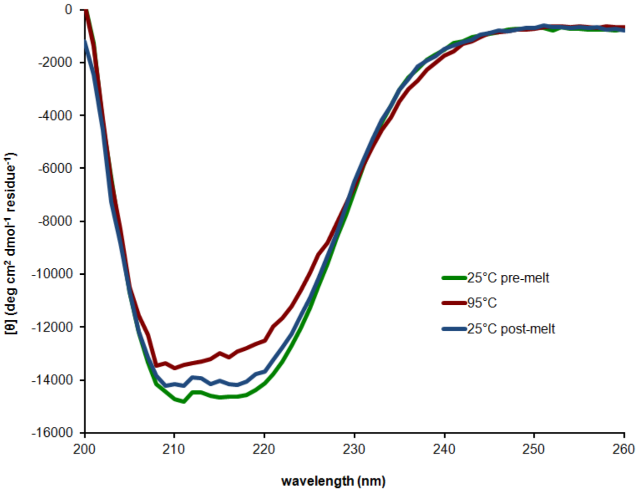


gEHEE_06

also known as: EHEE_1.7_06
amino acid sequence:
EERRYKRCGQDEERVRRECKERGERQNCQYQIRKEGNCYVCEIRC

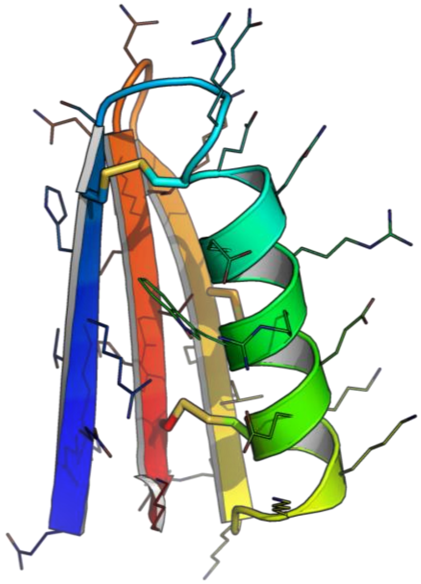


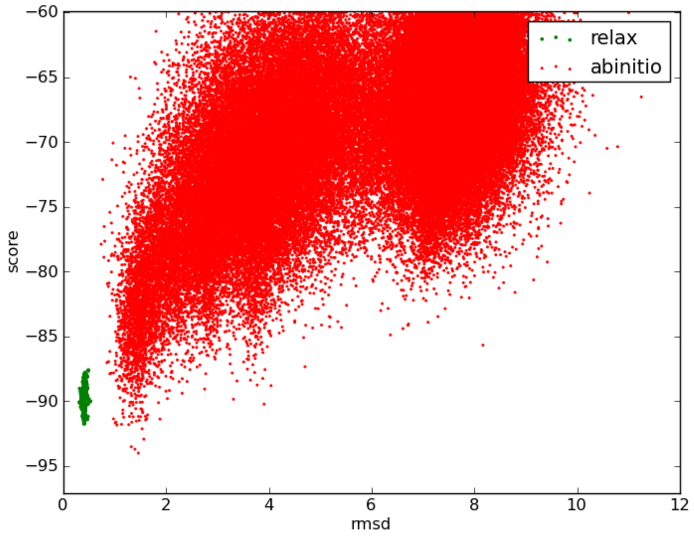




EHEE_1.7_09

amino acid sequence:
QKETRHCSGQRCEQEARRWCEECKKKGKRVRCRKHGNQVEVQCDK





Round 2

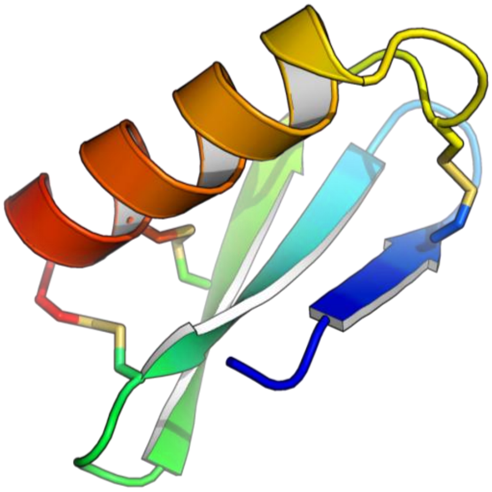
To test designs from this round, synthetic genes coding for designed proteins were cloned into the custom-made expression vector pCDB26, which was built by replacing the multiple cloning site of pET16B (Novagen) and carries an ampicillin resistance cassette, and expressed from BL21*(DE3) *E. coli* (Invitrogen).

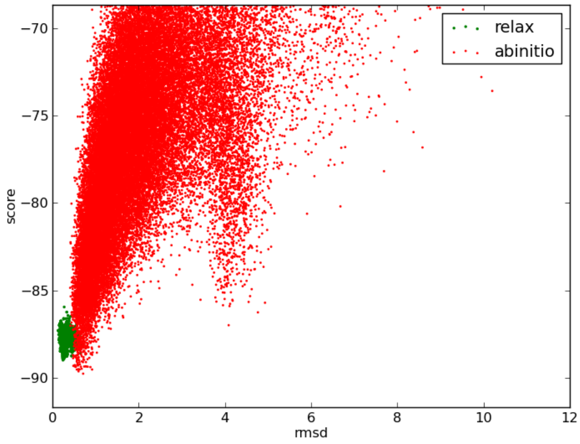
anatomy of the expression constructs

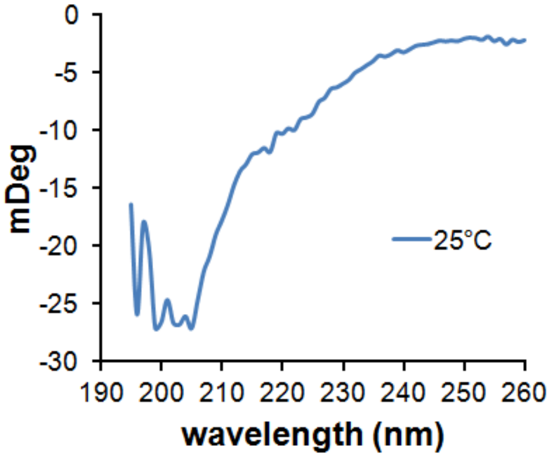
Designs were purified by IMAC from conditioned bacterial medium, and the fusion protein was removed using SUMO protease. Protease digestion is inhibited by a proline at the P1’ position (first residue of the design). To facilitate removal of the fusion protein, an alanine residue was added to the N-terminus of any design beginning with a proline during generation of the synthetic gene.

EEEH_3.0_01

amino acid sequence:
EQYHCHGNYVRYICEDGQDCEYHADCSDEEAEREAKEECERQC

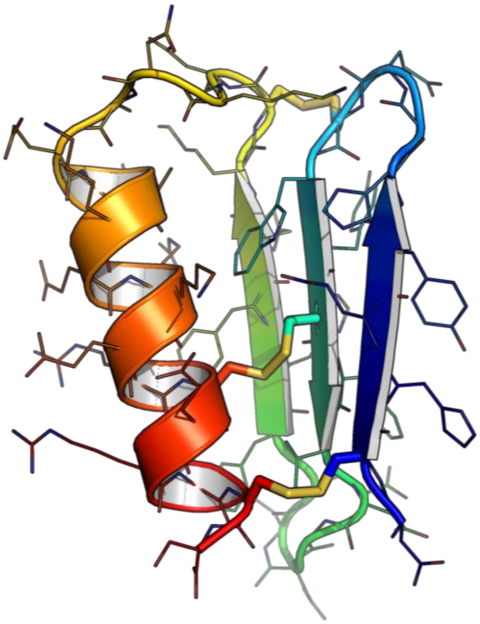


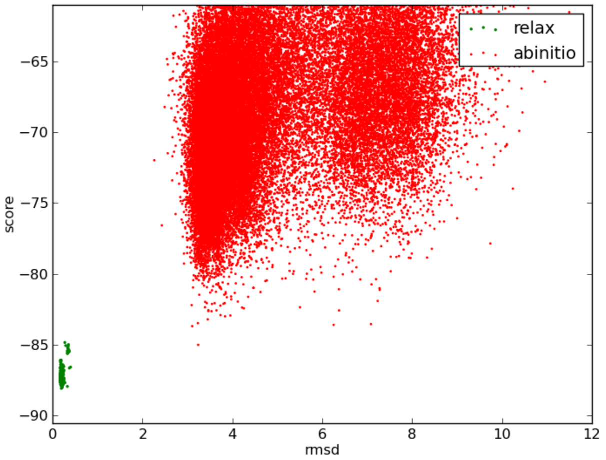


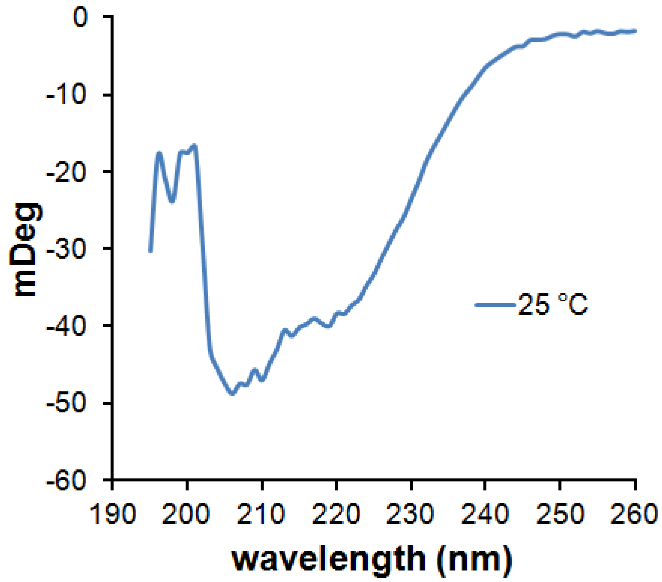


EEEH_3.0_02

amino acid sequence:
NCHEYHGECWYCFVDGDSQFHYHKCDKNAEEAKERKERCERDCS

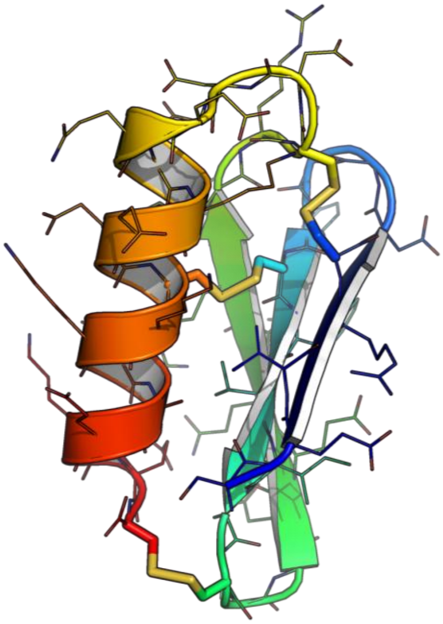


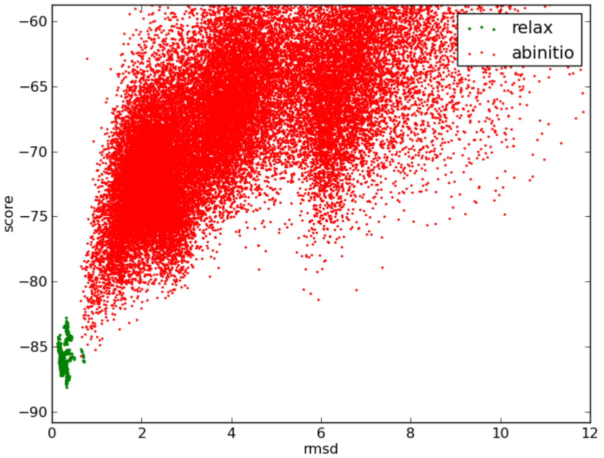


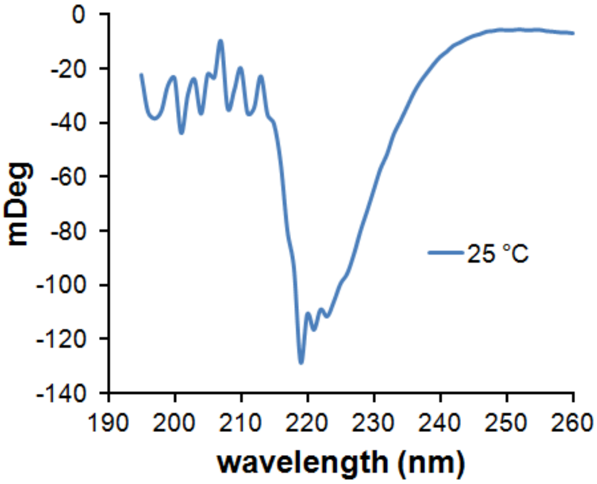


EEEH_3.0_03

amino acid sequence:
SEVRCDGNYCFVIACSGDEQSRDFRCDDEQEKEECKKEAEKEC

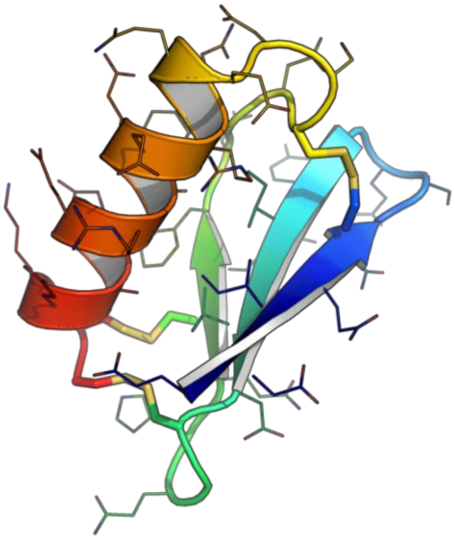


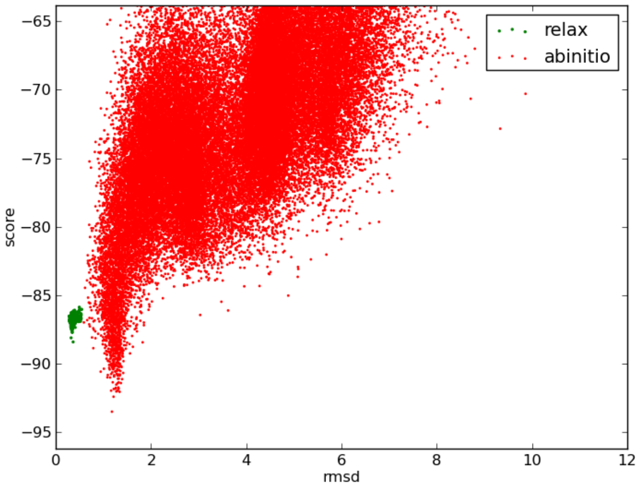


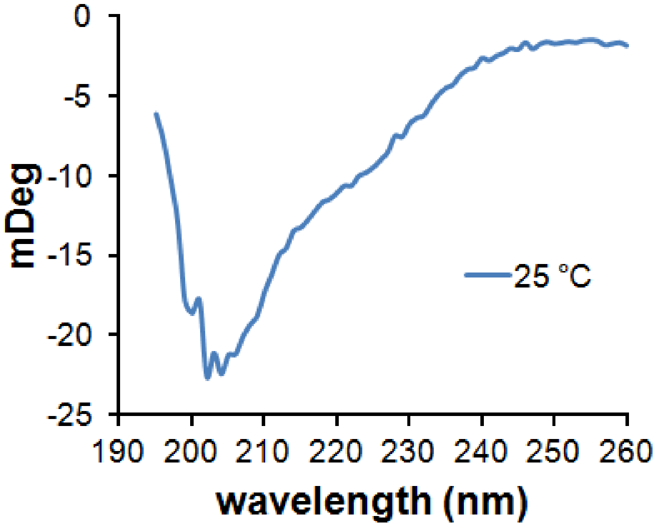


EEEH_3.0_04

amino acid sequence:
EDIQCQSEGYIVVDCGQHQCKFDYDCSDEQQREEAREEAEKCC

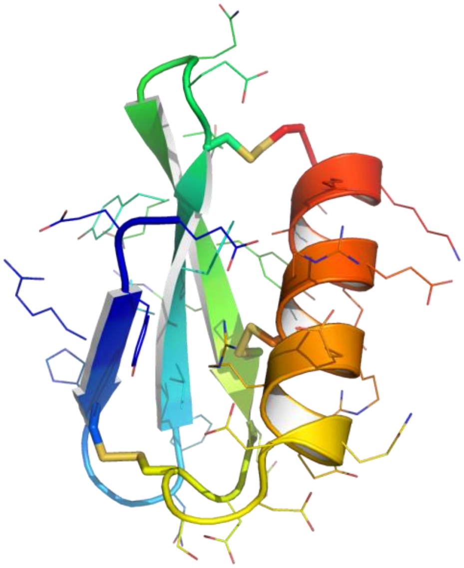


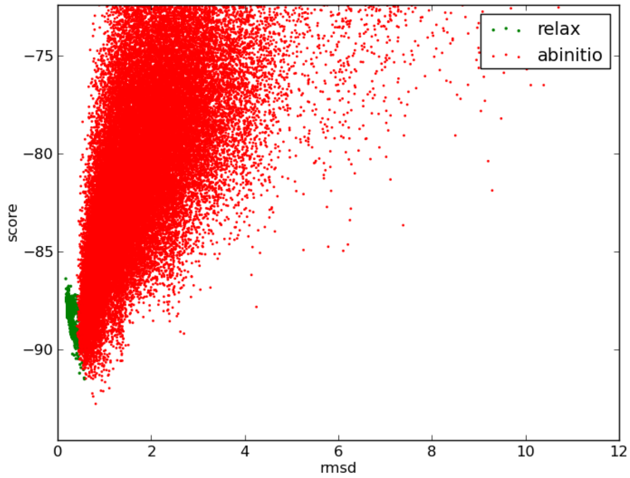


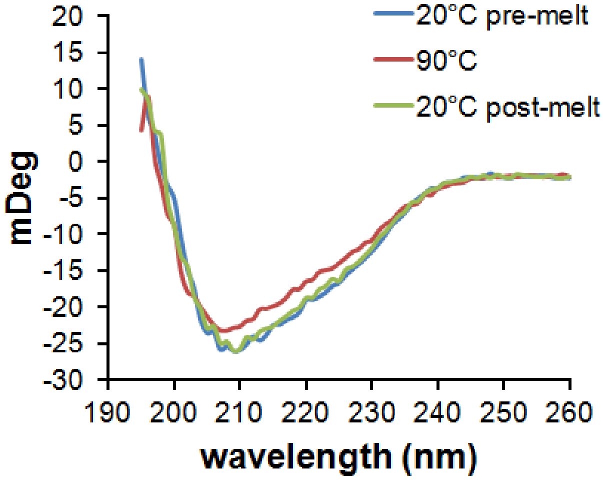


EEEH_3.0_05

amino acid sequence:
EEYRCHGNFVVFYCEQGQEYRCQADCSDEQERERCREEAEKQC

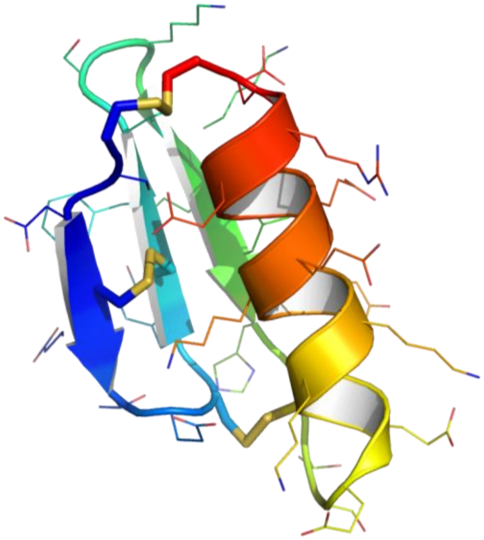


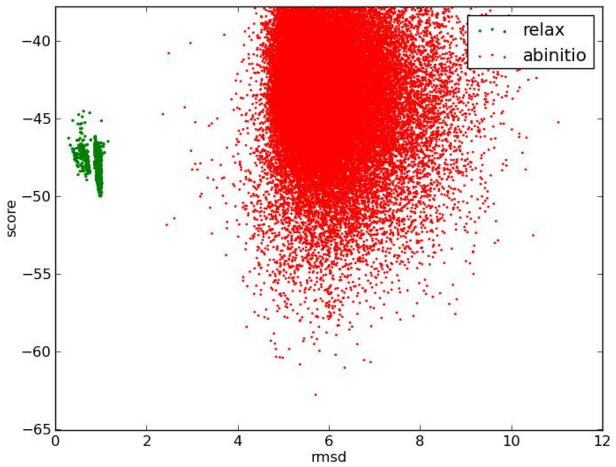


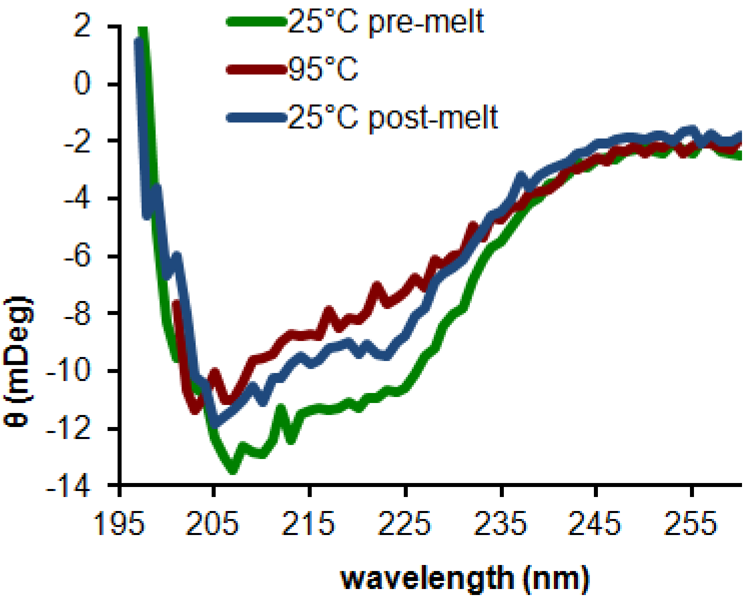


EEEH_3.0_06

amino acid sequence:
CSDCETECYCFVSKGKQWHGTSEECKKYKEEAEREC

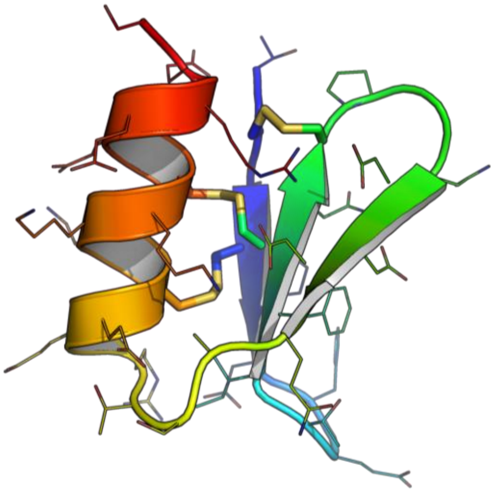


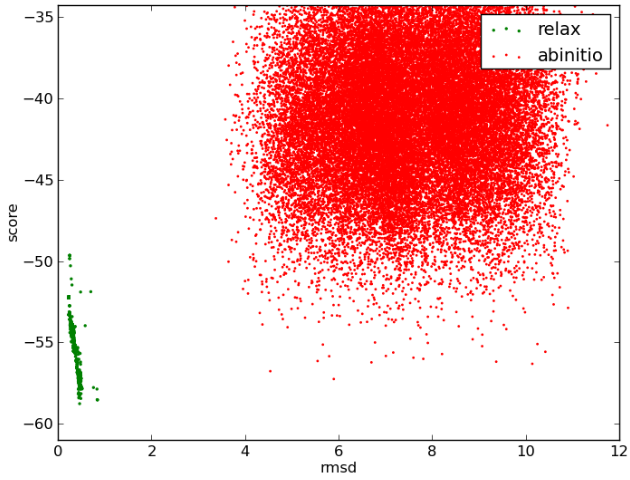


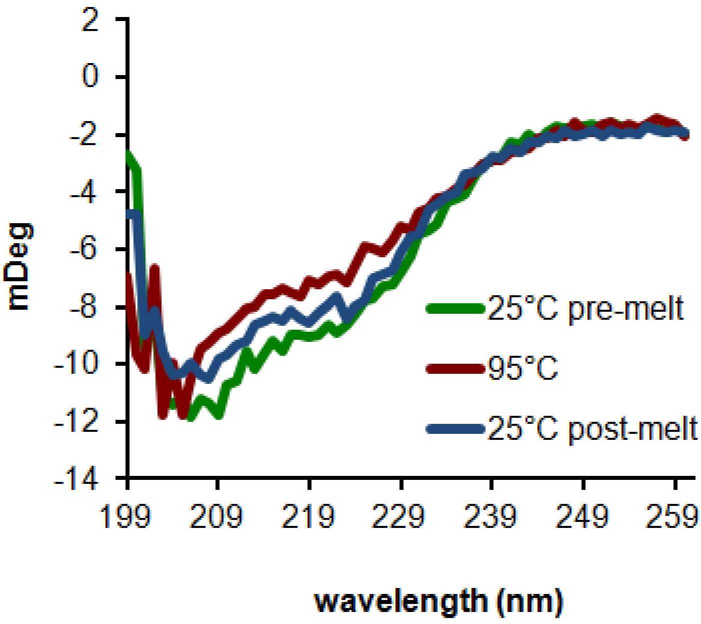


EEEH_3.0_07

amino acid sequence:
TCDCKDHETIFCNCPGNDDDQASTREECKKKCEERES

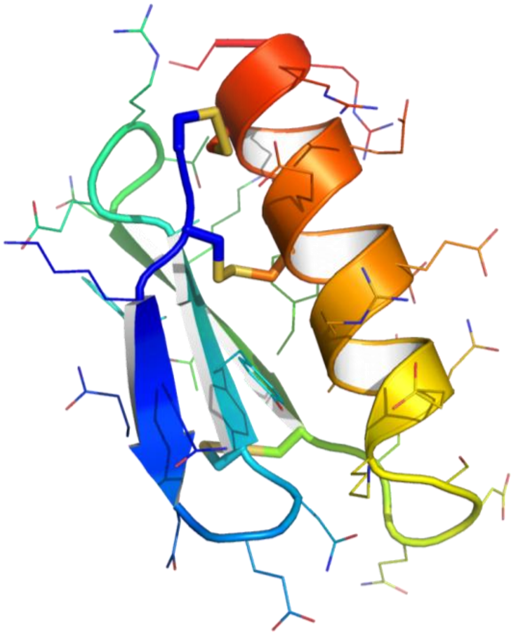


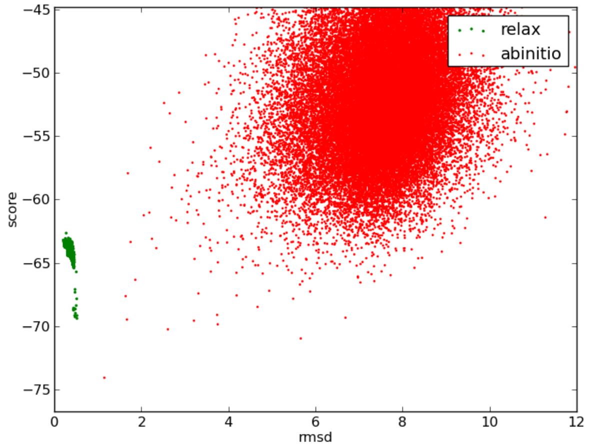


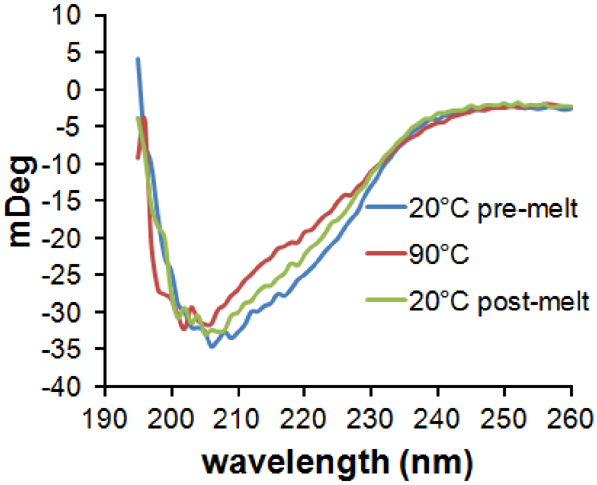


EEEH_3.0_08

amino acid sequence:
CCKQQNENCYFAERTNKTFCYQDSKEQAREDCEEECRRS

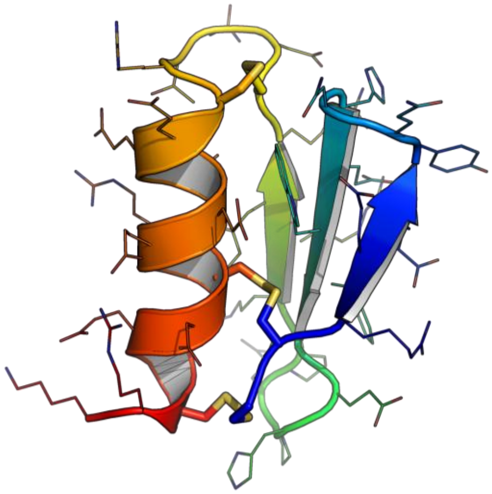


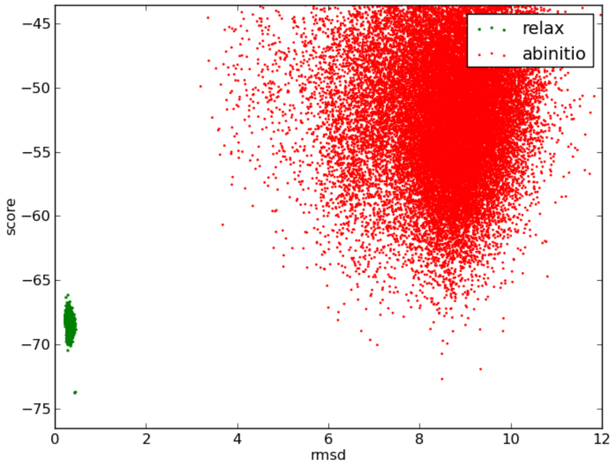




EEEH_3.0_09

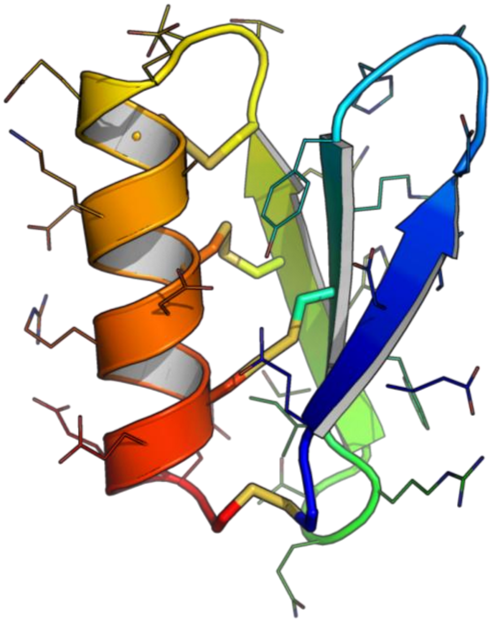
amino acid sequence:
CCREEYQNHEWFVEHPEPRRFRCDNTRCEEAEERCDEECRK

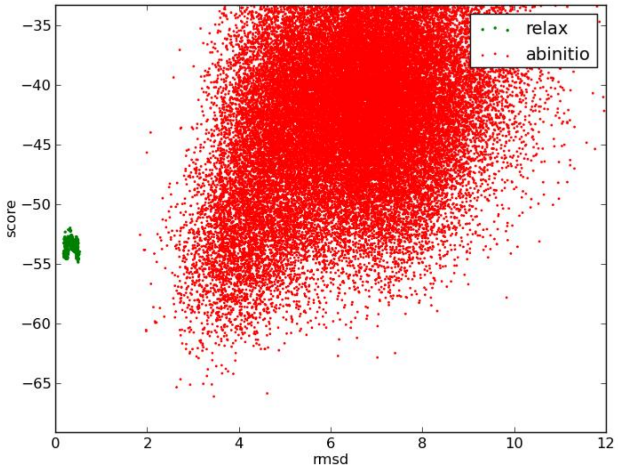


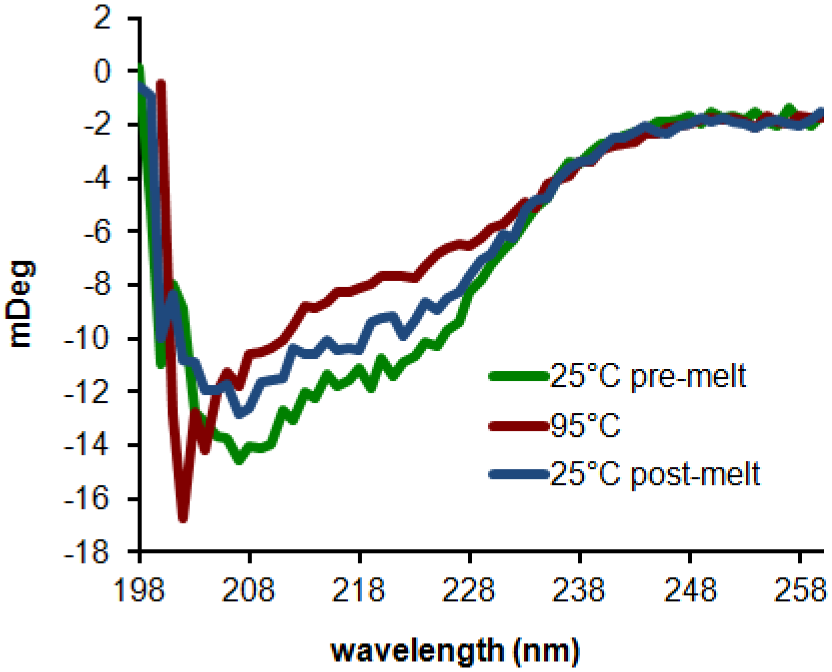


EEEH_3.0_10

amino acid sequence:
CQEDIDGSHYRCFIRQTGSHCQCTTEECAKECDRQCEEEC

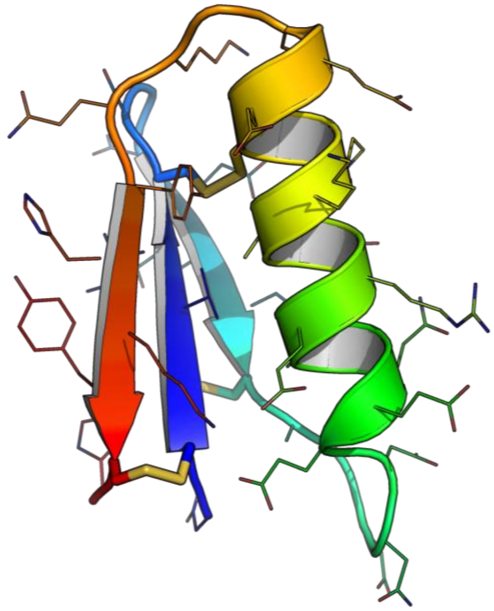


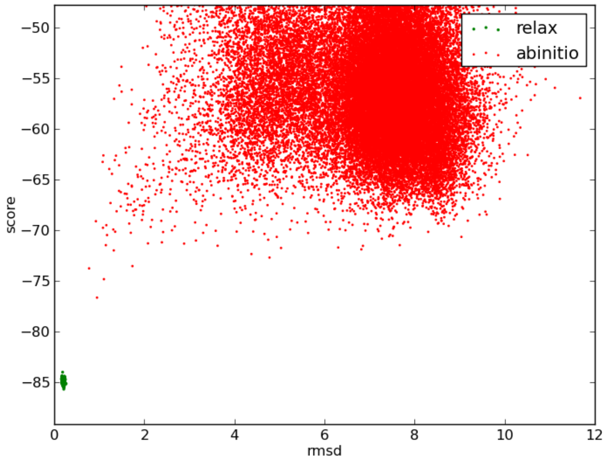




EEHE_2.0_01

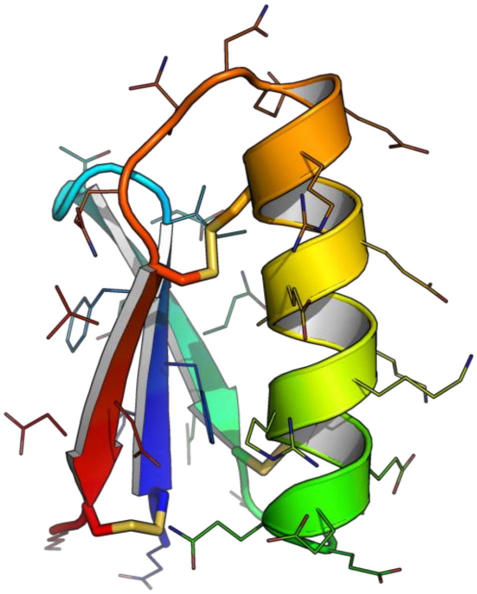
amino acid sequence:
DCCVICSGNDQYCAGDNNEEQAEREAKRCEEEGKQYHKYCH

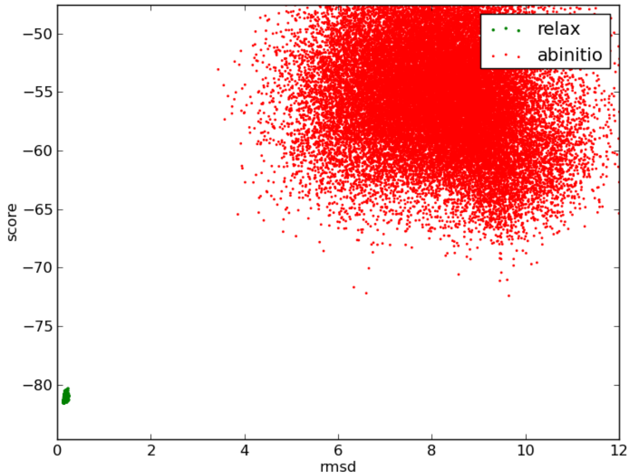




EEHE_2.0_02

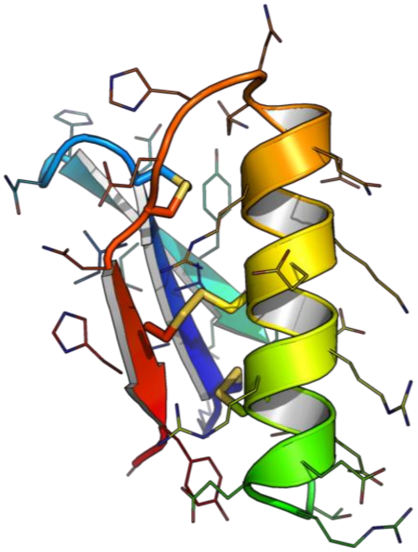
amino acid sequence:
ECYFFIGGTDDQECQSEQEECRKKAEEKCREQNQQCVDDCK

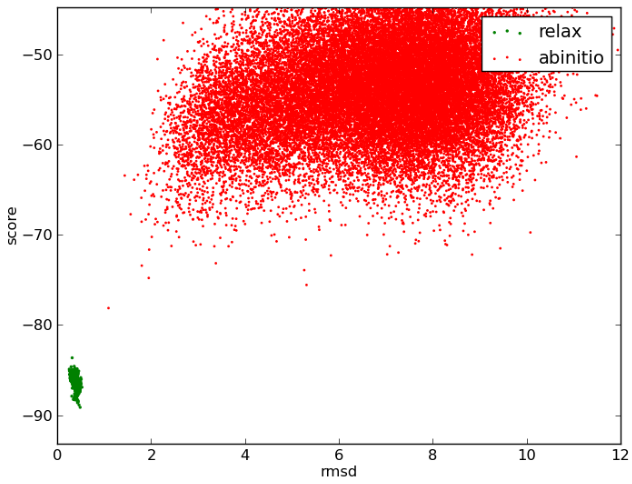


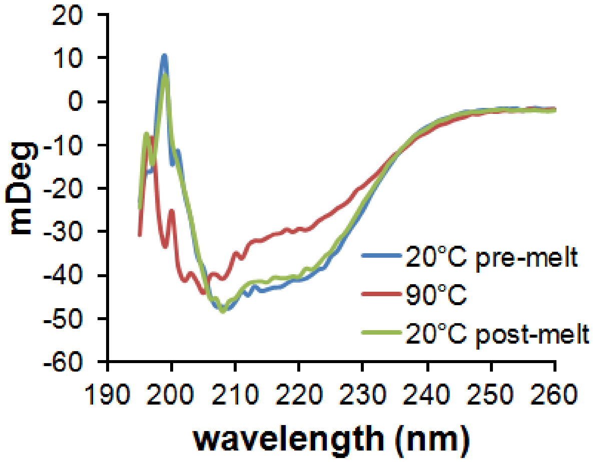


EEHE_2.0_03

amino acid sequence:
KCYVICGNHDDYEFDTTREEECRRECEKARQEQNHECNCHYS

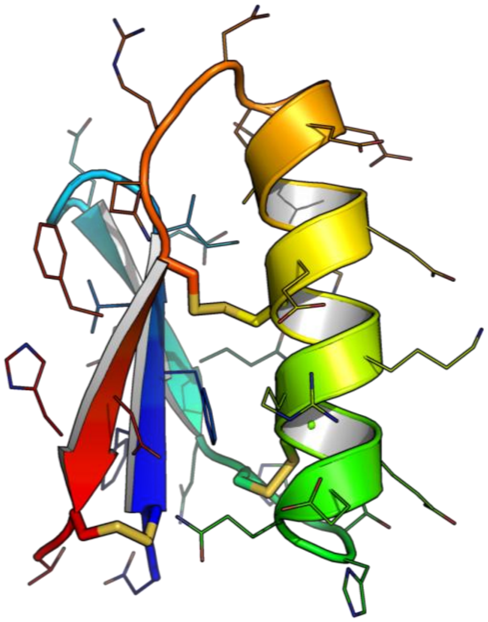


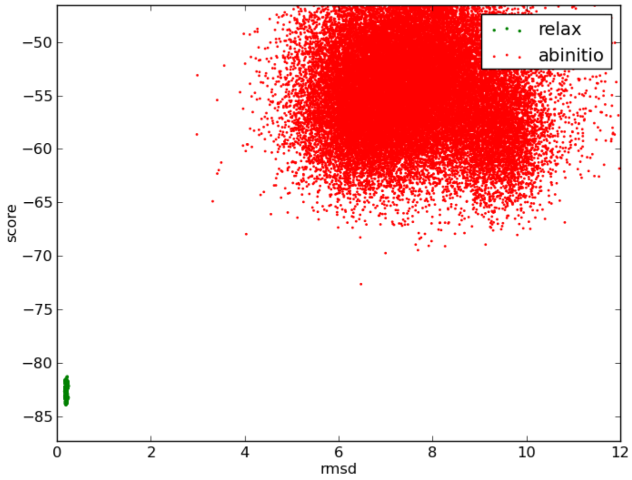




EEHE_2.0_04

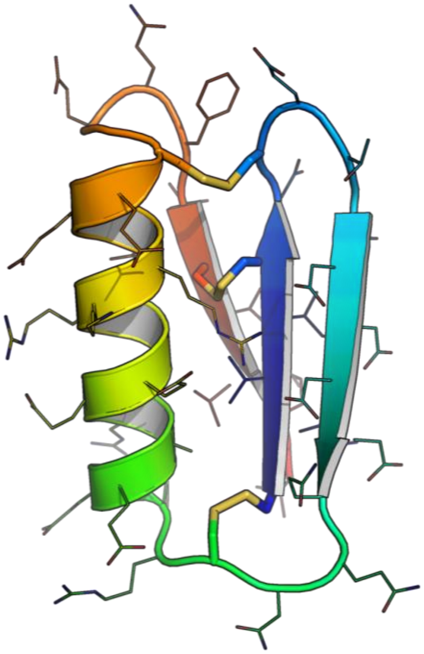
amino acid sequence:
DCFFVIGGQDDQQCHTHQEECRKECEEKAEEQNRQCFDHCT

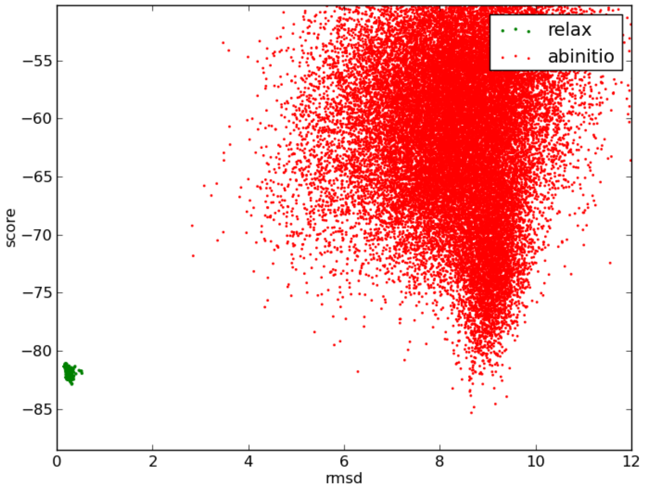




EEHE_2.0_05

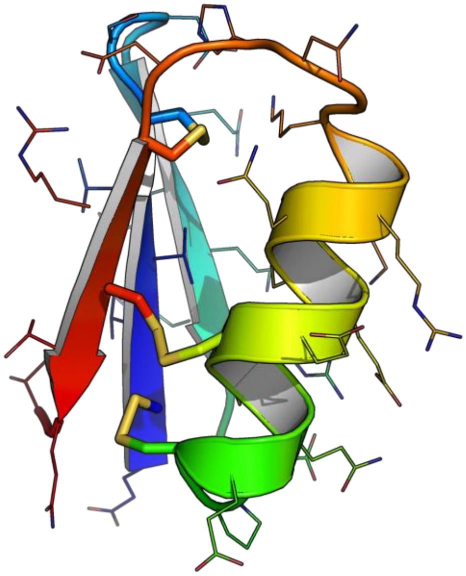
amino acid sequence:
CIVICDCETDDDDDQQNCREEEAREEARKREEECGEQFTCHVQT

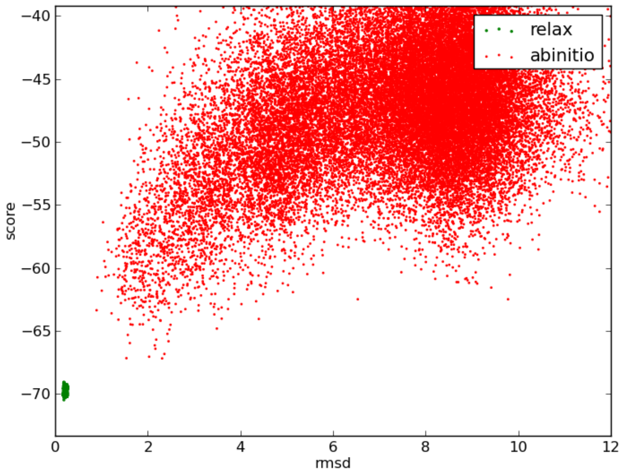




EEHE_2.0_06

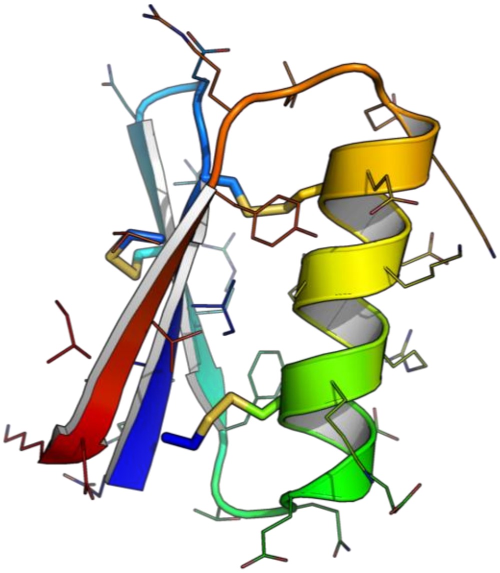
amino acid sequence:
ECVVVCSDGQEQQRQDPCEQVCEEEQRKKGNHDCRCTQT

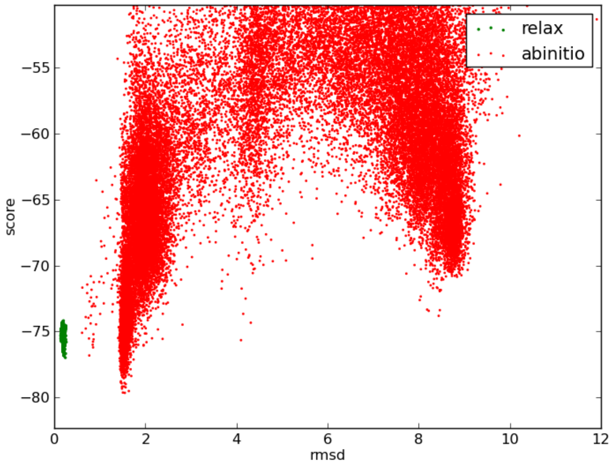




EEHE_2.0_07

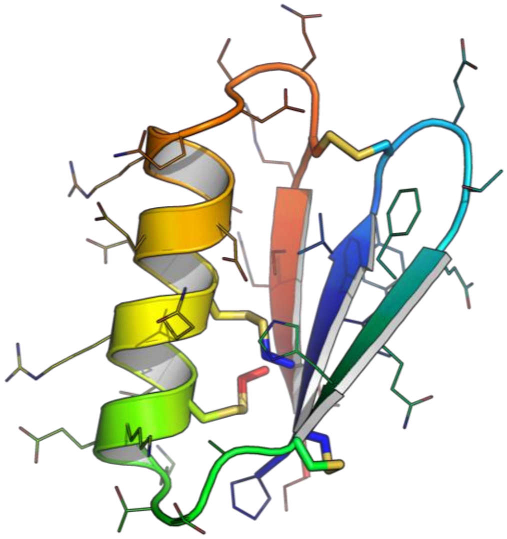
amino acid sequence:
ECIICCEGNQCRKFTQEEECKRQAKECEKQGLRYTTIDK

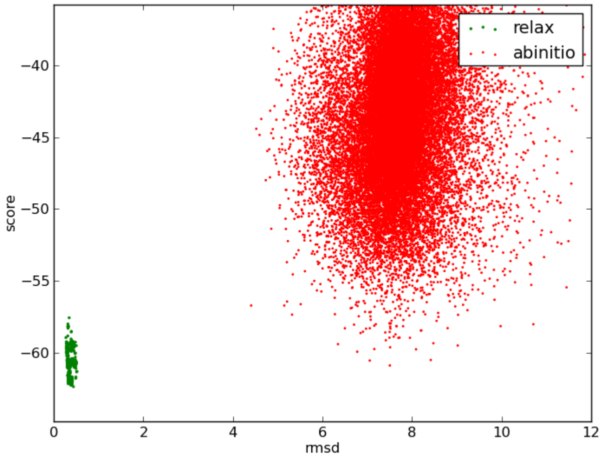




EEHE_2.0_08

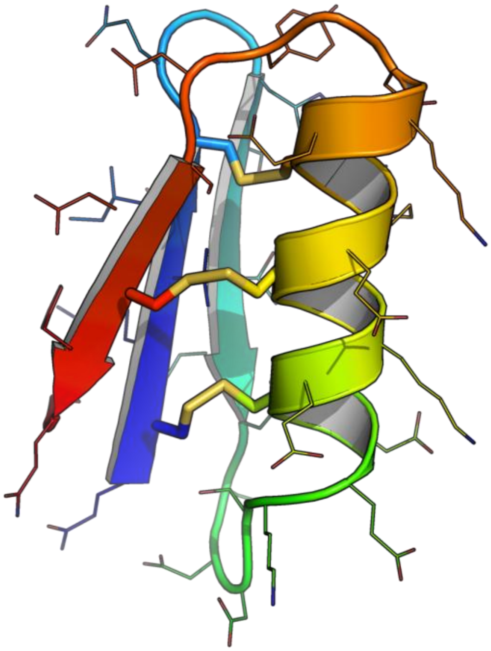
amino acid sequence:
PCCIVYCETQFQHCADTKEKCERQCEEDERQDSQCRSRCTS

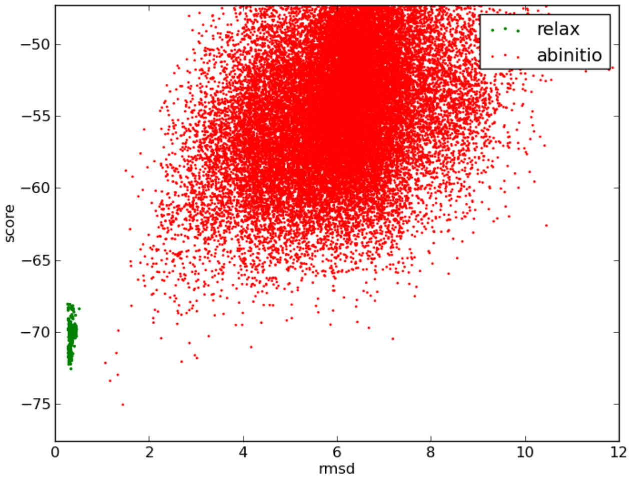




EEHE_2.0_09

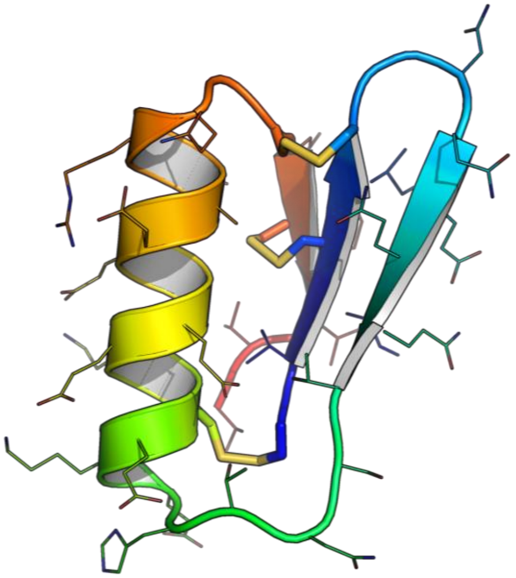
amino acid sequence:
QCRVICQGHSTTEFSDDSKEECEKECERCEKDGYDSDCHQS

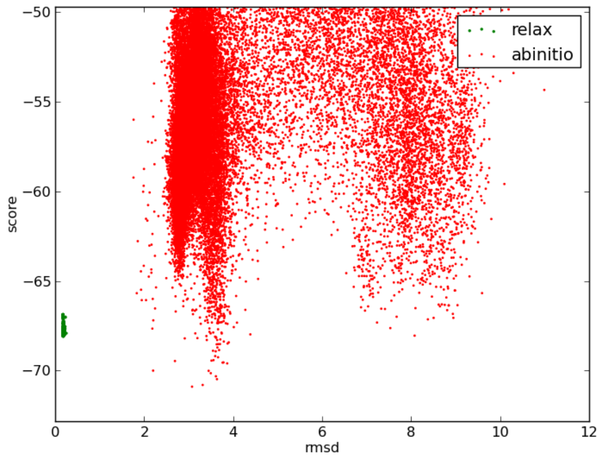




EEHE_2.0_10

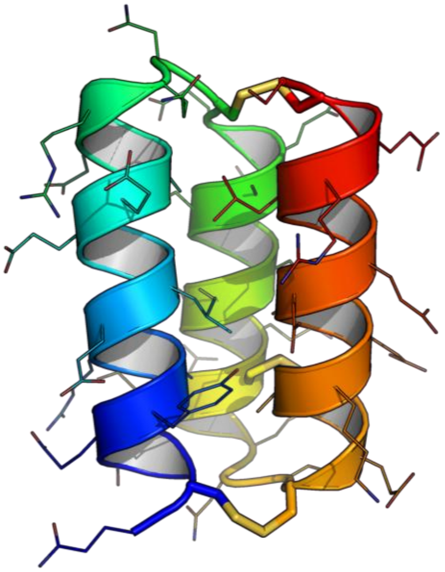
amino acid sequence:
CVICICGNQEQQTSNTHEKECKEEAEEAERQGCDCKVTT

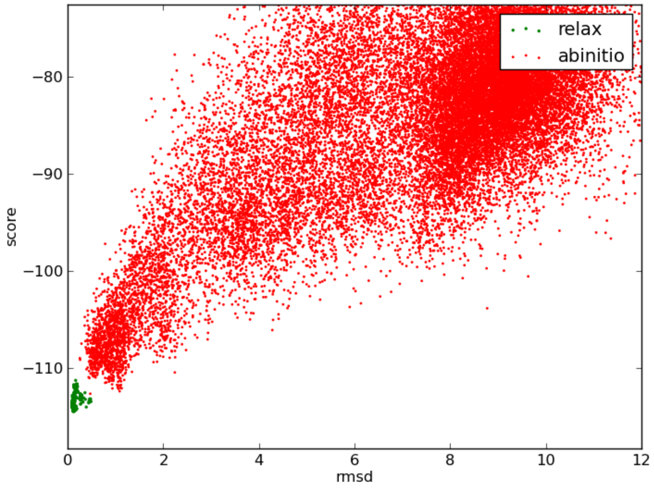




HHH_2.0_01

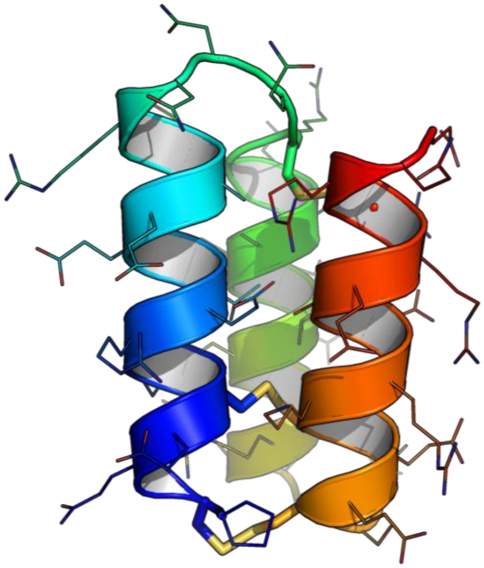
amino acid sequence:
QCEEYARELREEAERQNCEEAREKAEECEEKNDCECAKEAEEKLRECS

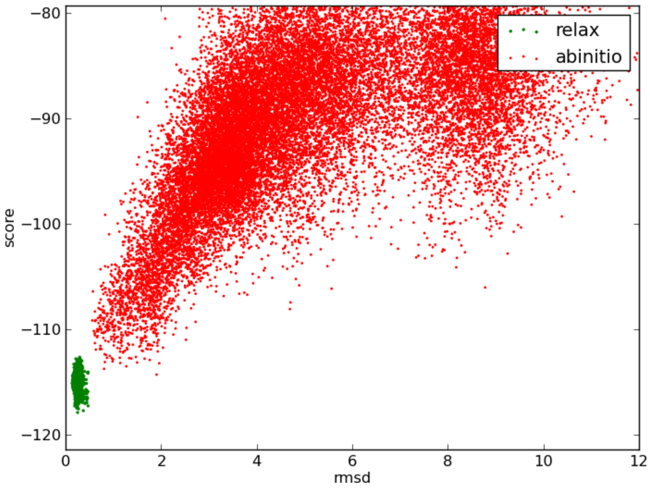


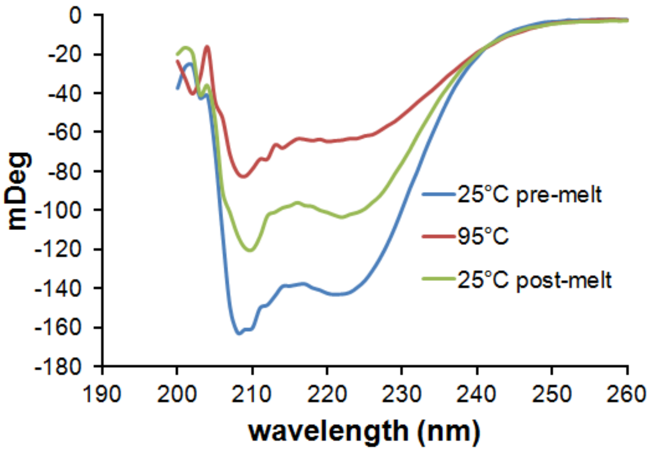


HHH_2.0_02

amino acid sequence:
PCQECERELEEAKRNNQCREERAEEIRREREEGQTSCEECKREAERCRQE

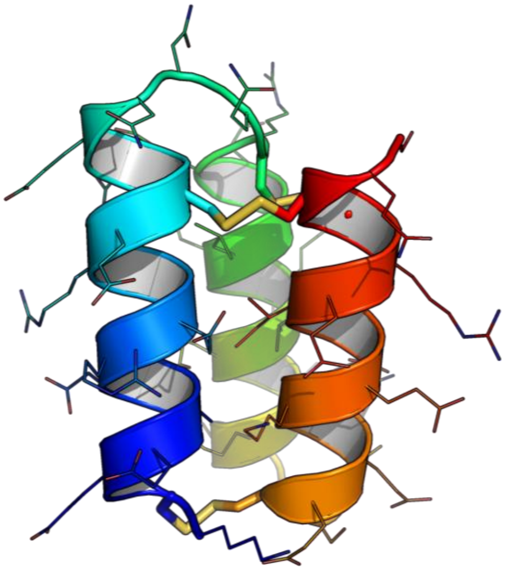


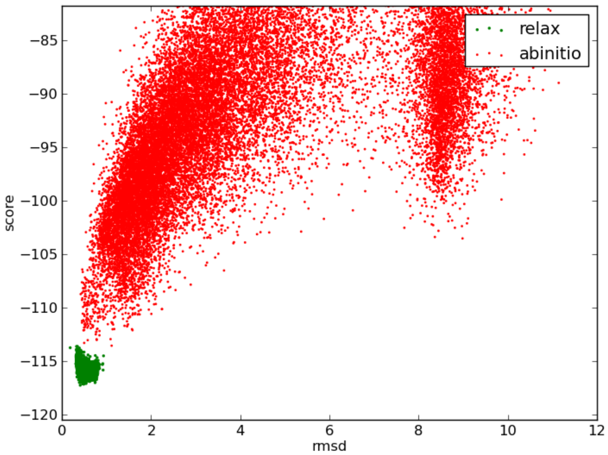




HHH_2.0_03

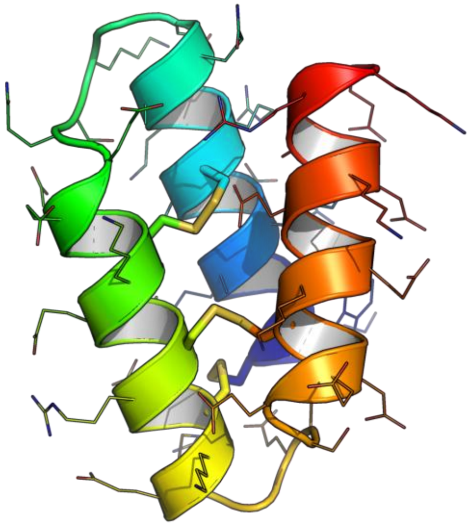
amino acid sequence:
KCEEAEREARECQENNQCREEELEKIEEKREKGETSCEEAKEEIERCCQS

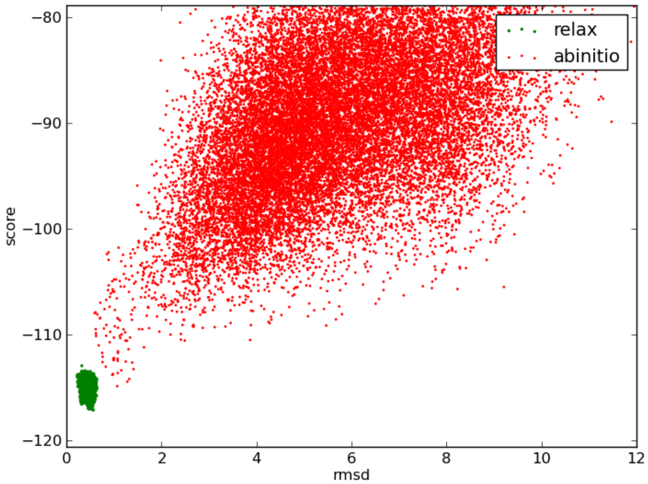




HHH_2.0_04

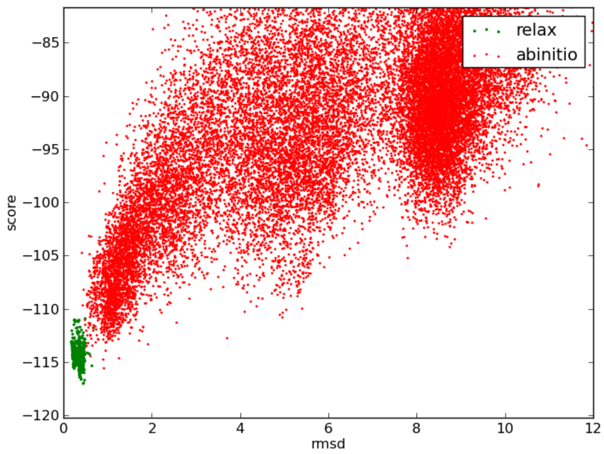
amino acid sequence:
KCWRKAKEECRKAQEGKTQEEECKEACRECKERGESSEEECKEAEKEARKE





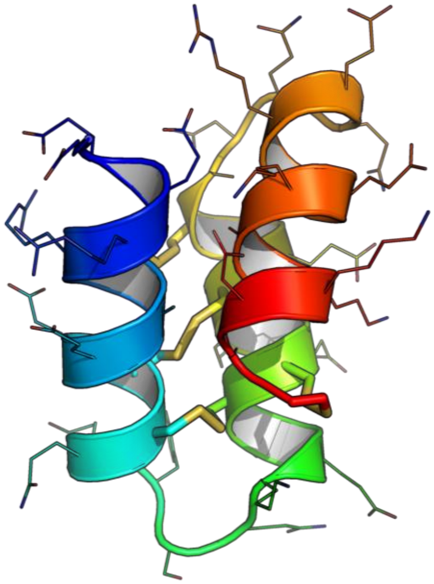
HHH_2.0_05

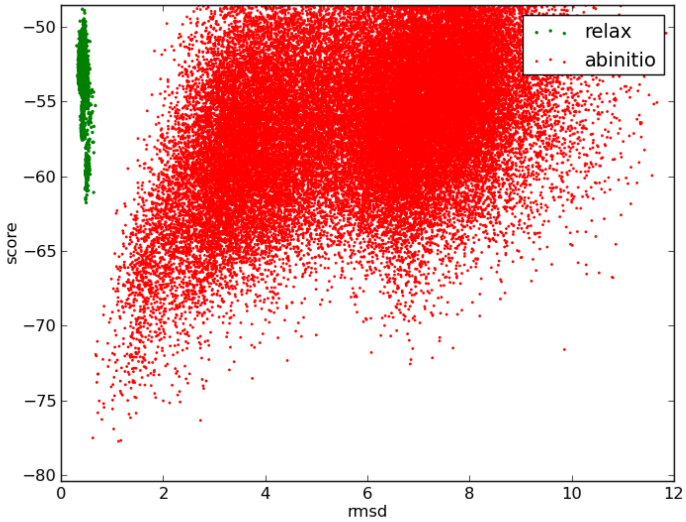
amino acid sequence:
ECEKKAEECKRYAEEQNTSEECAERAEEYARRHCESSEEECREYAEECKKN



HHH_2.0_06

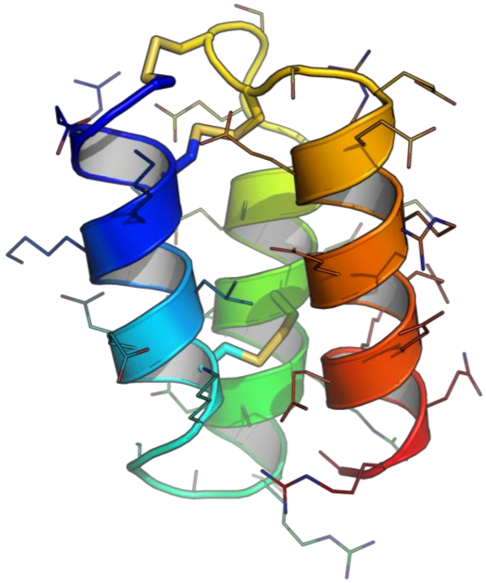
amino acid sequence:
EEERRCAEECCQQFSQKEECCERCEECANQQERAEKAKKDAC

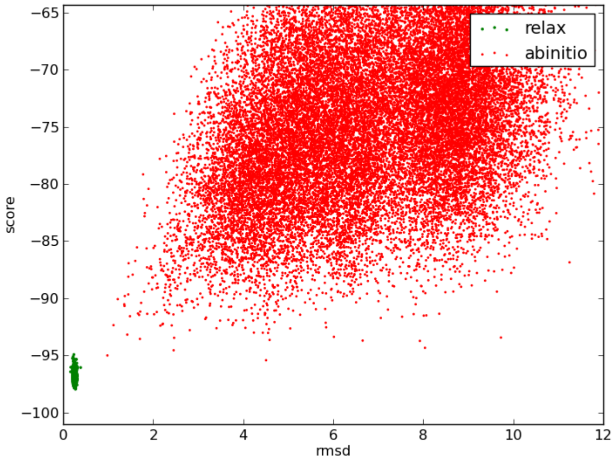




HHH_2.0_07

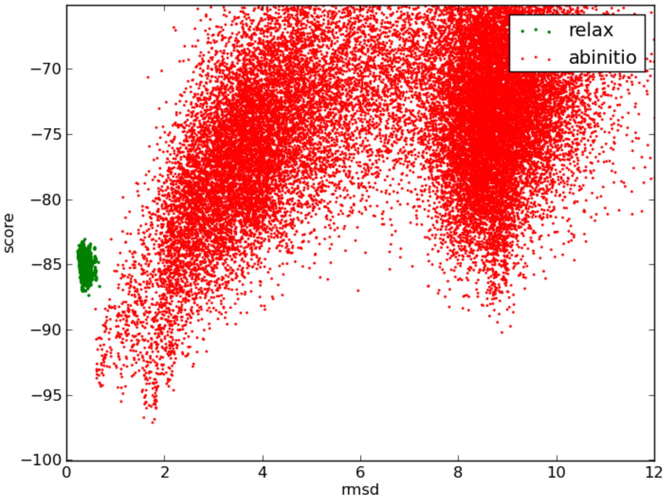
amino acid sequence:
CERCKKKLEECKGSSREDARERCEEAKQESCCSEEERREAEEEKQRA





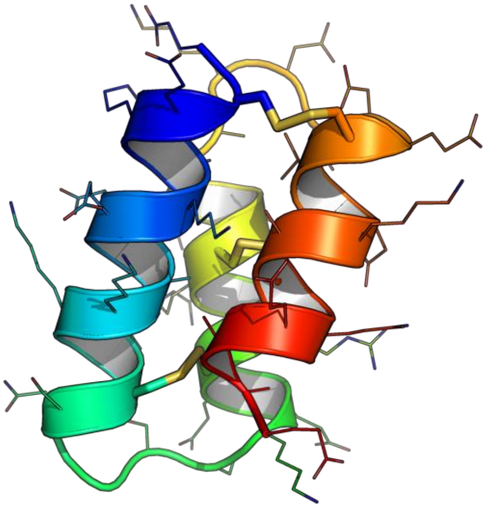
HHH_2.0_08

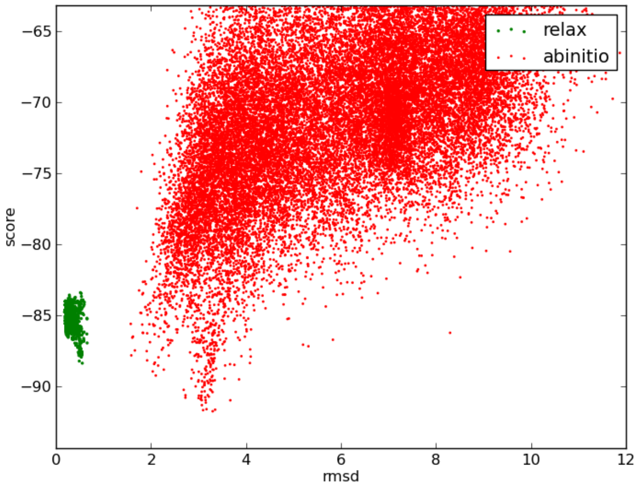
amino acid sequence:
ECYKEYCQEIKECQSTSEEEAEERAREACNTSCEEARKKAEEACQS



HHH_2.0_09

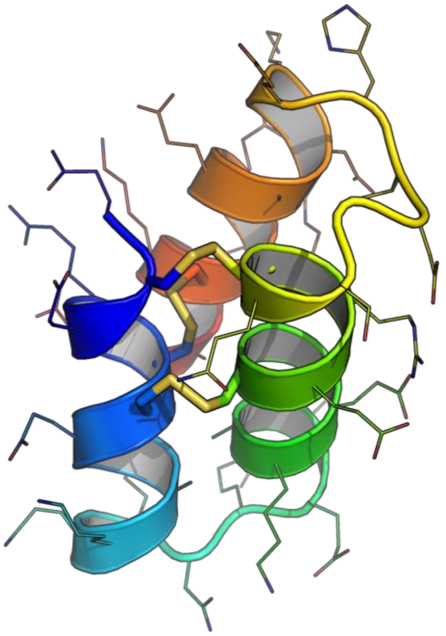
amino acid sequence:
ECERAKEEAKKECSQGSSKEECRERCQEAAKDSDECVEKACQEAAE

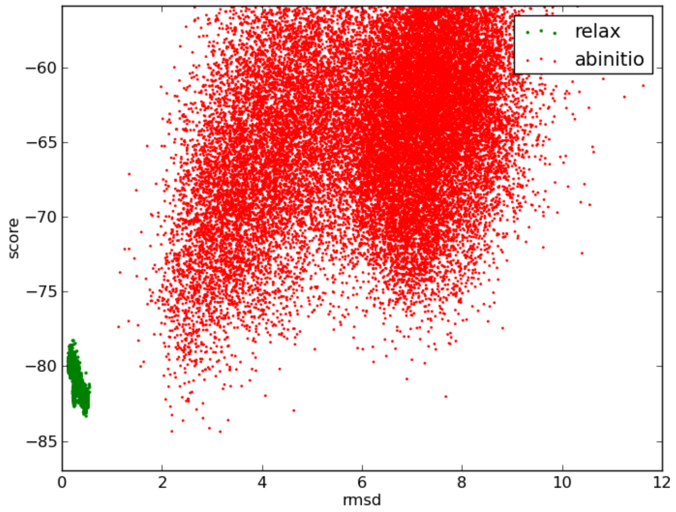




HHH_2.0_10

amino acid sequence:
QCERCCEAAKQKNREEAKEACERCQSGDTHEKDAEERCKEAET



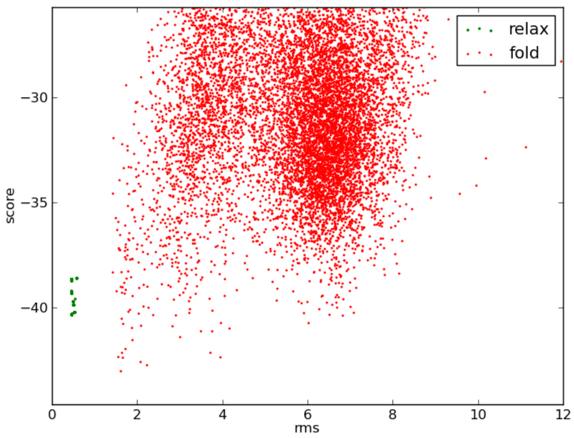


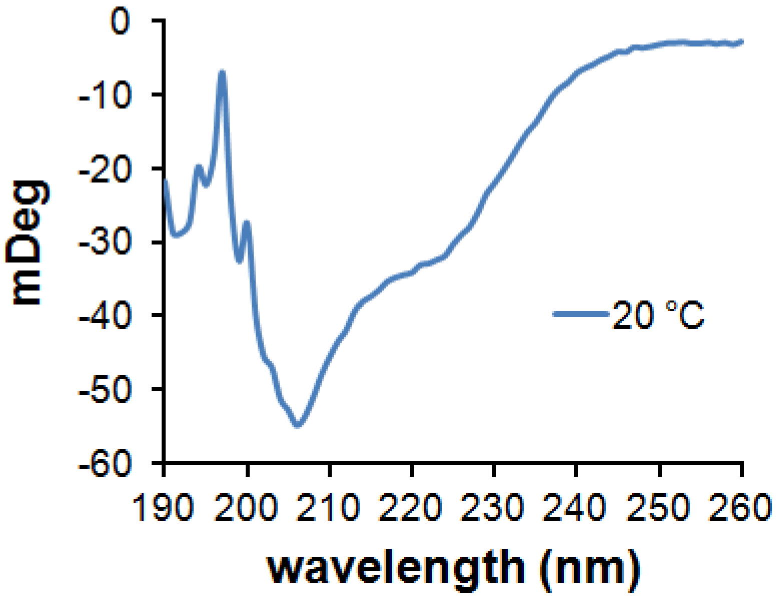
Round 3

To test designs from this round, synthetic genes coding for designed proteins were cloned into pCDB26 and expressed from BL21*(DE3) (Invitrogen) *E. coli*. Designs were purified by IMAC from conditioned bacterial medium, and the fusion protein was removed using SUMO protease. Protease digestion is inhibited by a proline at the P1' position (first residue of the design). To facilitate removal of the fusion protein, an alanine residue was added to the N-terminus of any design beginning with a proline during generation of the synthetic gene.

EEEH_4.0_01

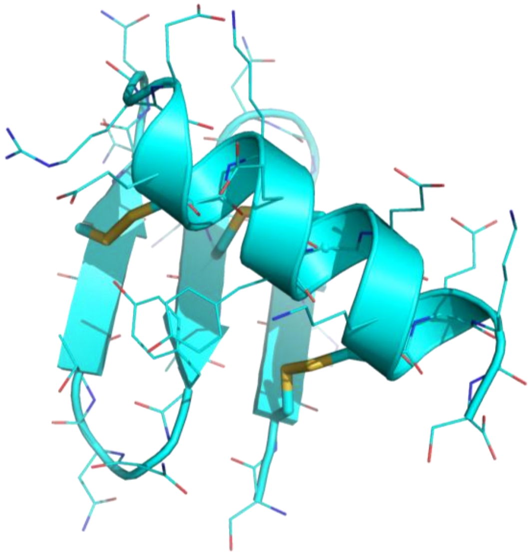
amino acid sequence:
NCQINGDTCQIGNEQCQNQEECKRLCEECEKS

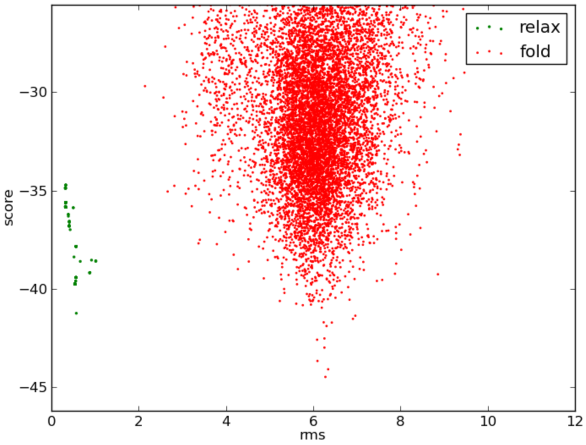


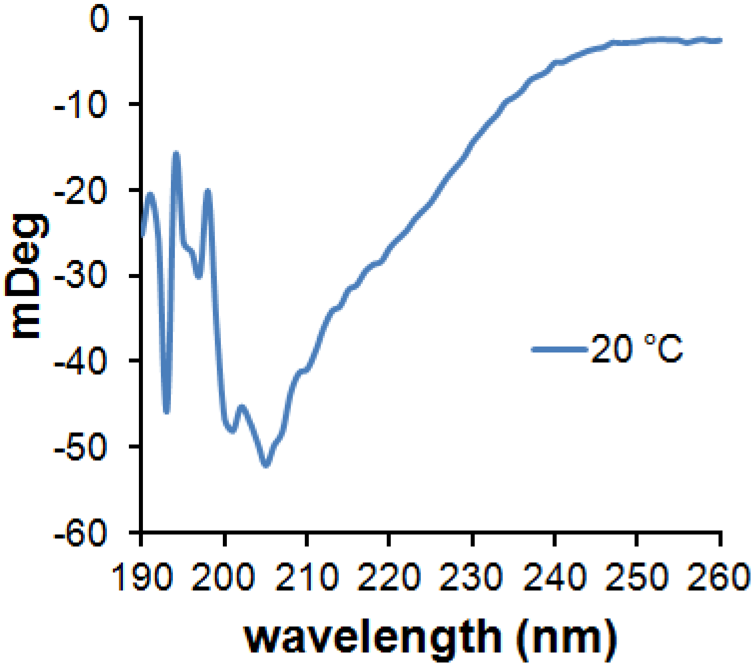


EEEH_4.0_02

amino acid sequence:
SCHIDGNQCTYNNTDCNNREECKEYCEKCEKS

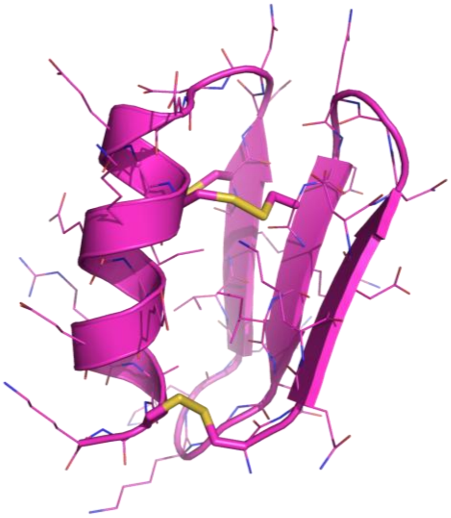


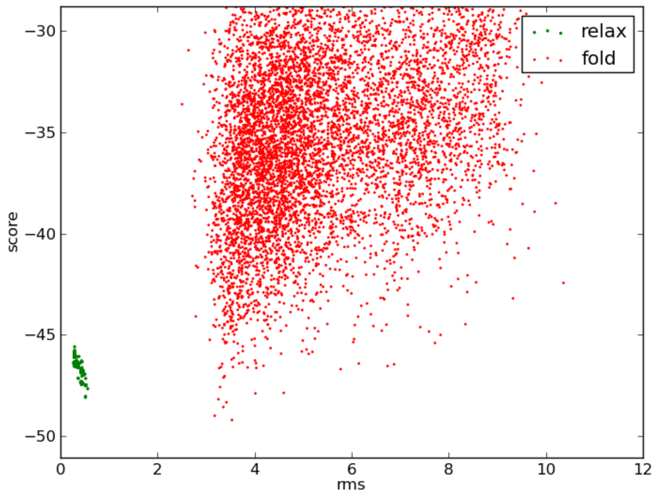


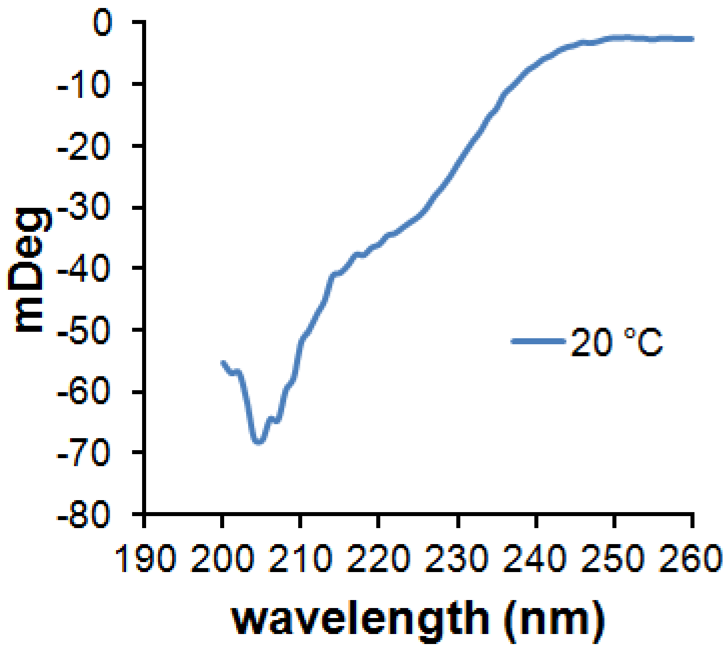


EEEH_4.0_03

amino acid sequence:
CQKDDNGQDCRIDGKHQVECDNDEECCKEIEERACK

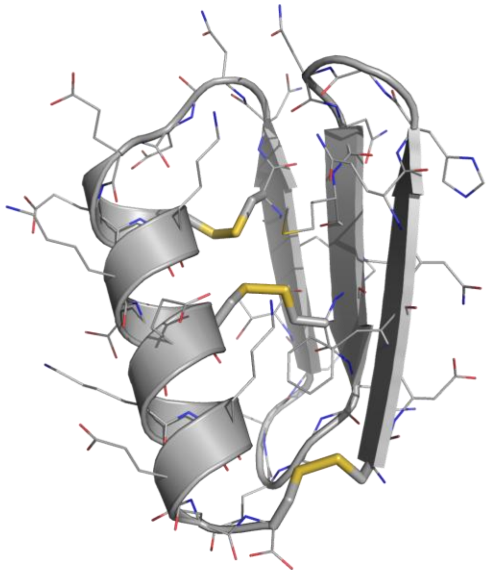


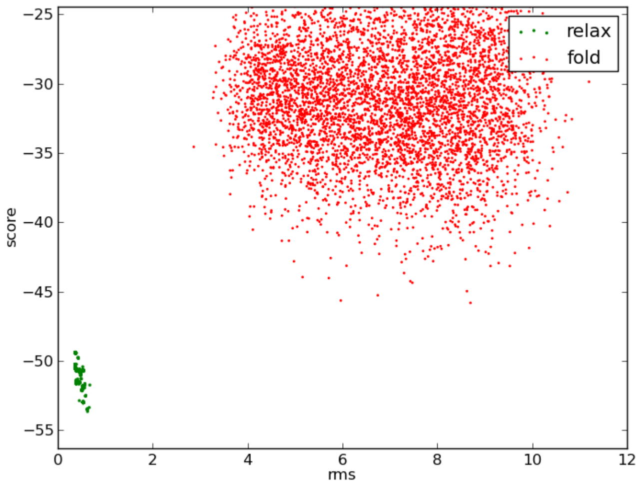


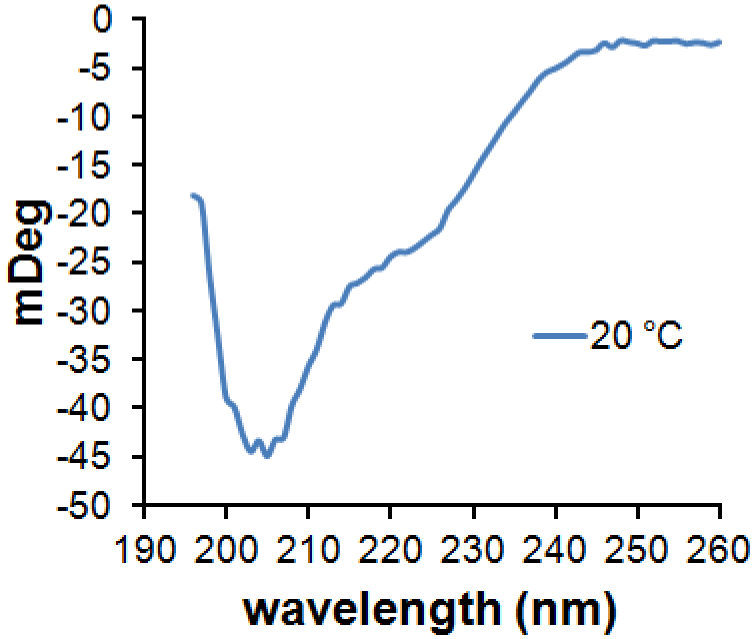


EEEH_4.0_04

amino acid sequence:
CDFNQHGNNMTCNGENDTHCNNDEECKKECEKMKENC

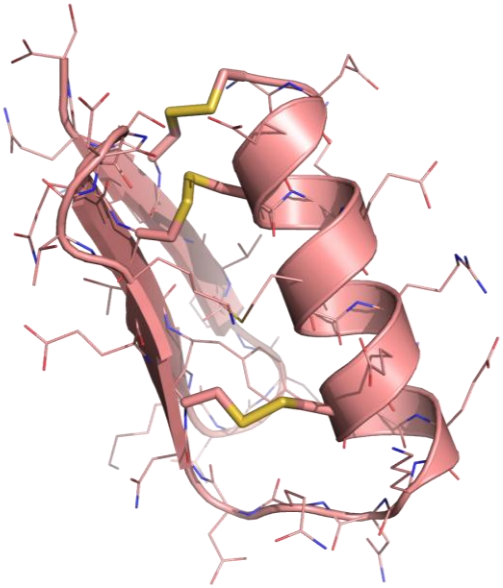


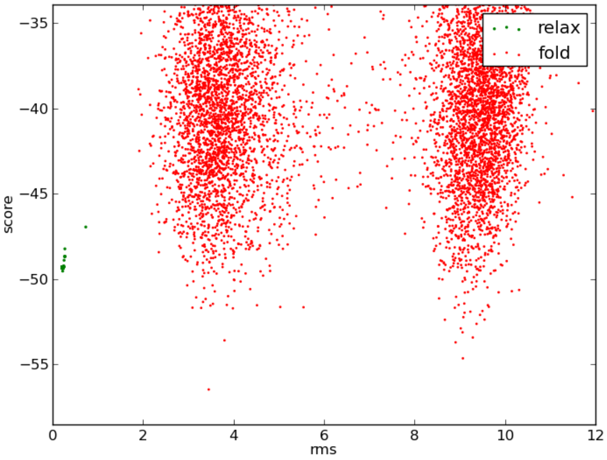


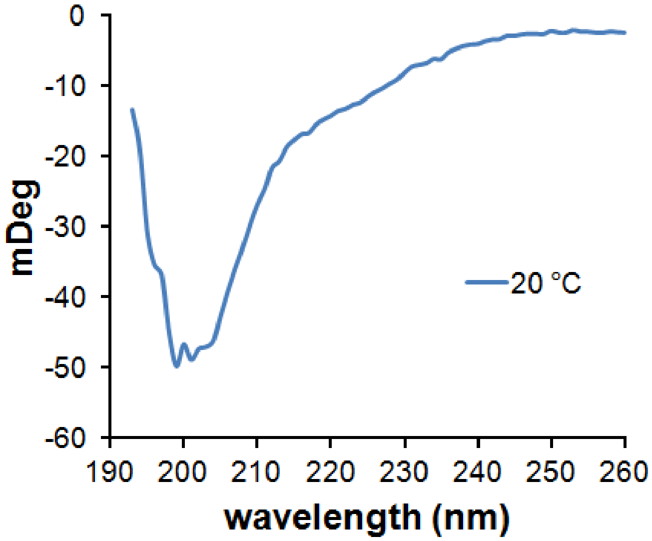


HEEE_3.0_01

amino acid sequence:
CEECKEMARECKEKNQDNCEKTDSQCTYKDNQVKCQS







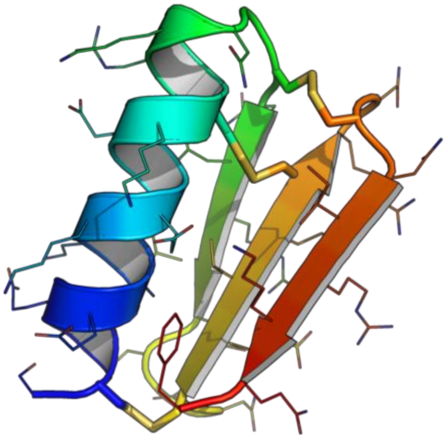
Round 4

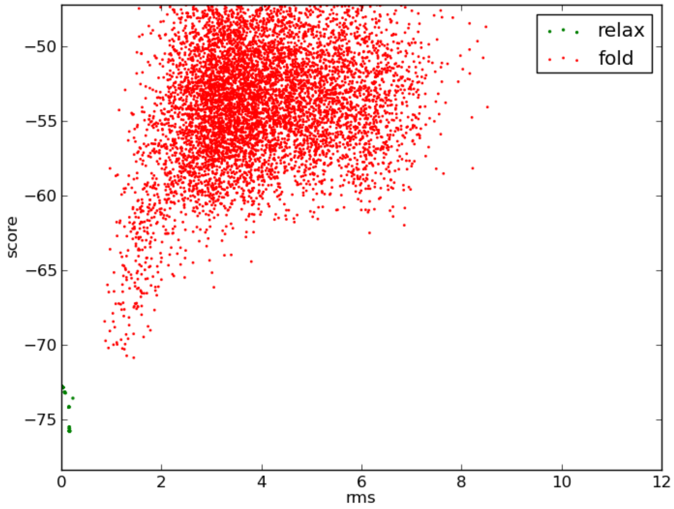
To test designs from this round, synthetic genes coding for designed proteins were cloned into pCDB26 and expressed from BL21*(DE3) (Invitrogen) *E. coli*. Designs were purified by IMAC from conditioned bacterial medium, and the fusion protein was removed using SUMO protease. Protease digestion is inhibited by a proline at the P1' position (first residue of the design). To facilitate removal of the fusion protein, an alanine residue was added to the N-terminus of any design beginning with a proline during generation of the synthetic gene.

Additionally, designed proteins were expressed from the Daedalus system. Designs were purified by IMAC from conditioned cell culture medium, and the fusion protein was removed using TEV protease. Purified designs have a glycine and serine added to the N-terminus.

HEEE_2.1_01

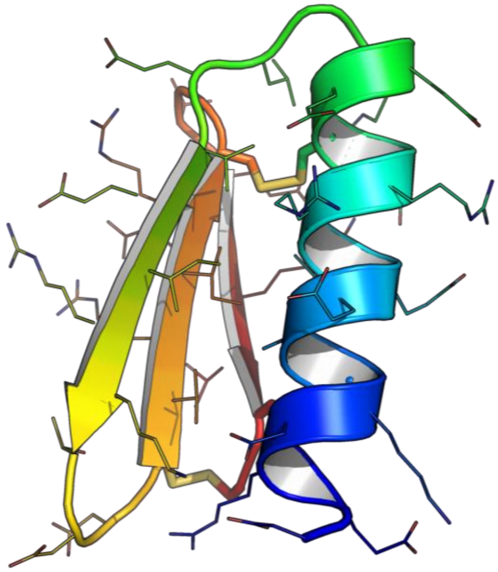
amino acid sequence:
SCEEEAKKEADKCRKNGCQYRVDSDNCEVECRNCNIRKQF

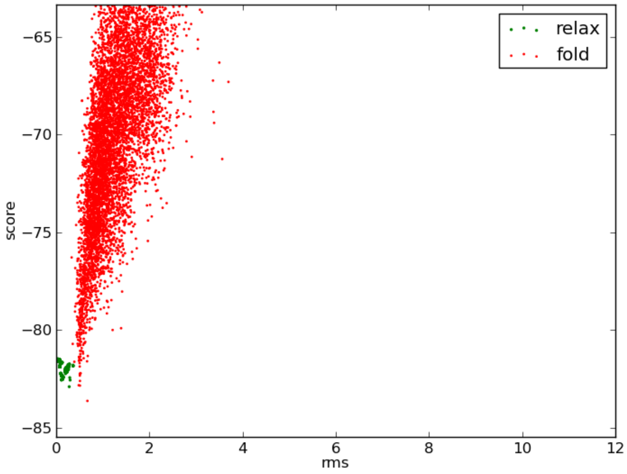




HEEE_2.1_02

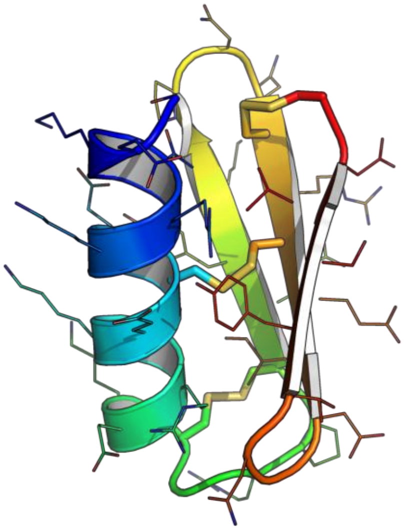
amino acid sequence:
DERDKCAEEIRRECEERGLEVEIRKTDDCVRIRCGTEERTCC

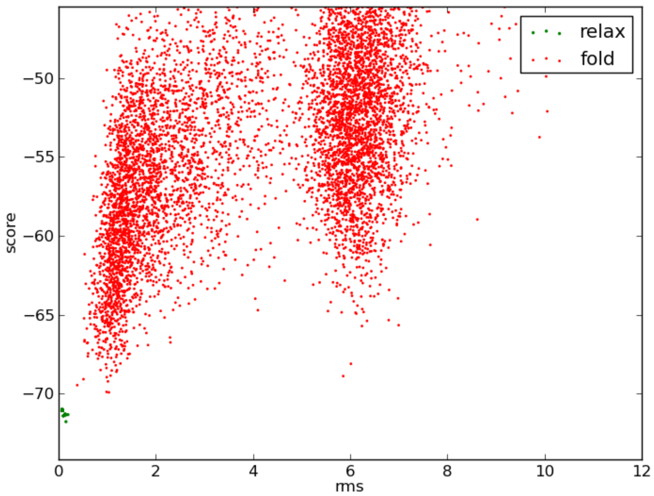


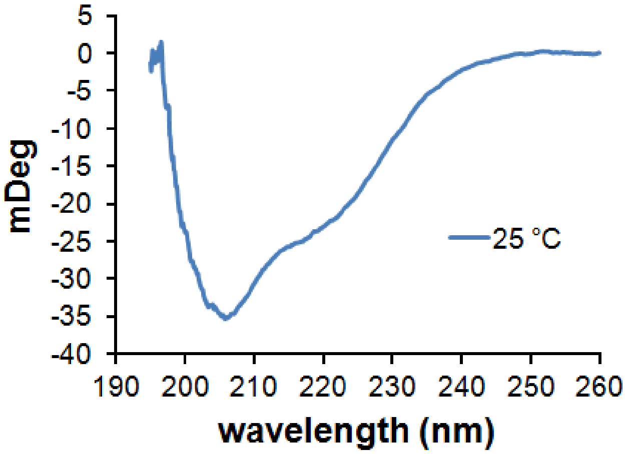


HEEE_2.1_03

amino acid sequence:
SEKTRKECEKQREKCGGRPCEYKGPNNCRCEIDGNTYSVDC

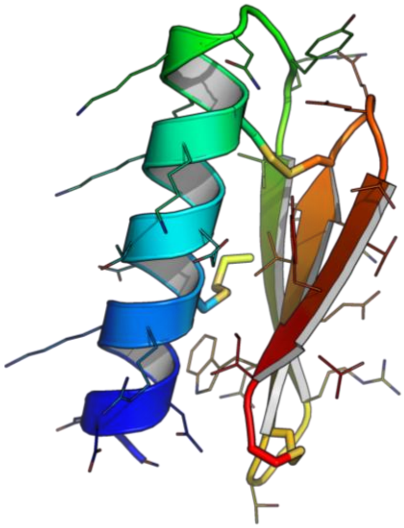


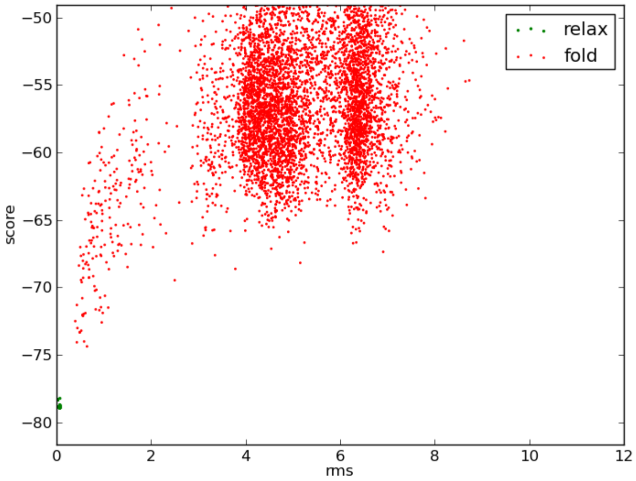


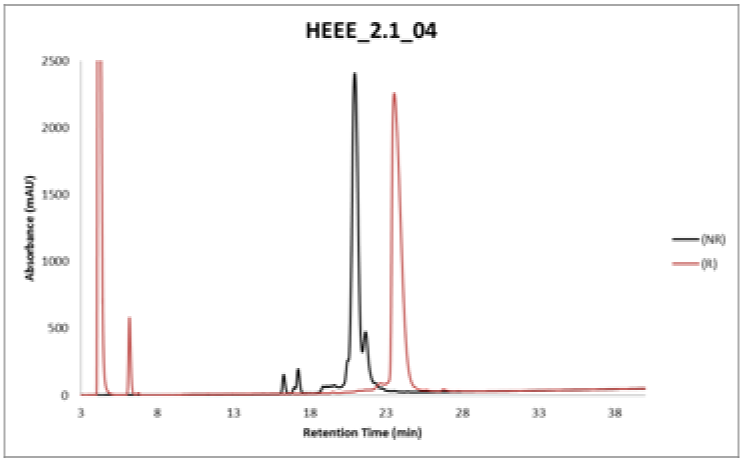


HEEE_2.1_04

amino acid sequence:
SDENKKRCETEAKKCKKNGYRVECRNRGTCWEVDCEETTYTIC

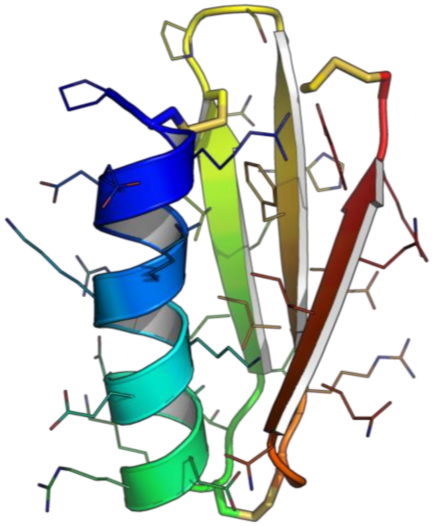


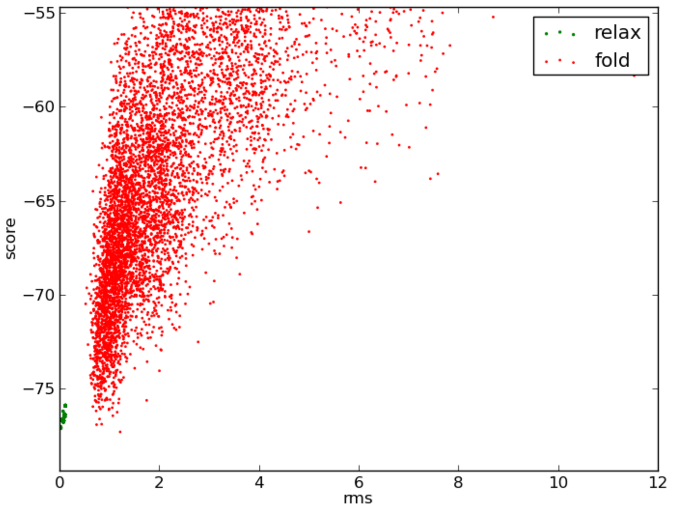


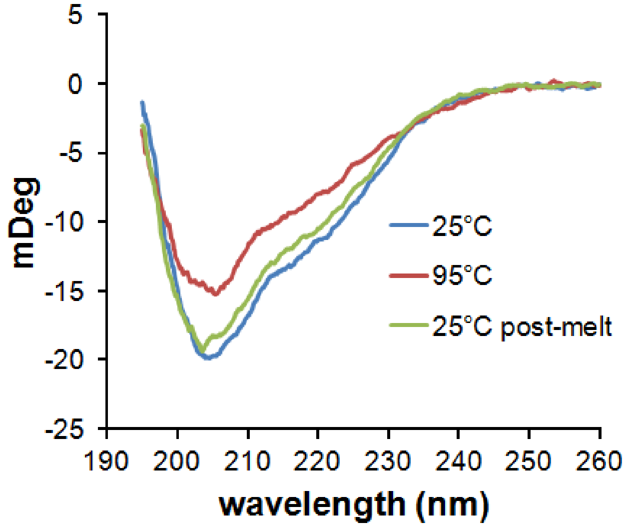


HEEE_2.1_05

amino acid sequence:
PCREEAKKRKEEAERKCTTLRVQCPSGCHFEIRCGNQIQEKC

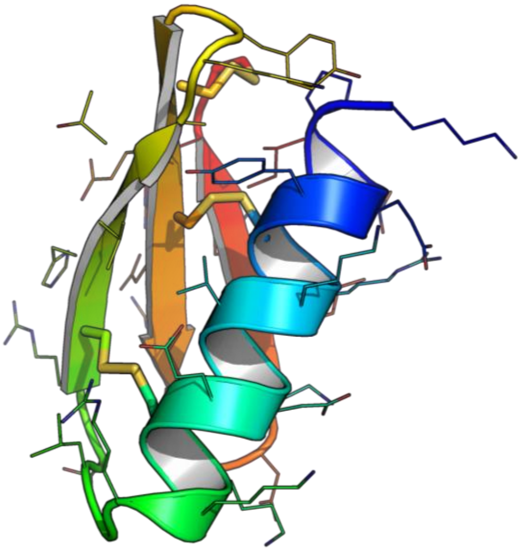


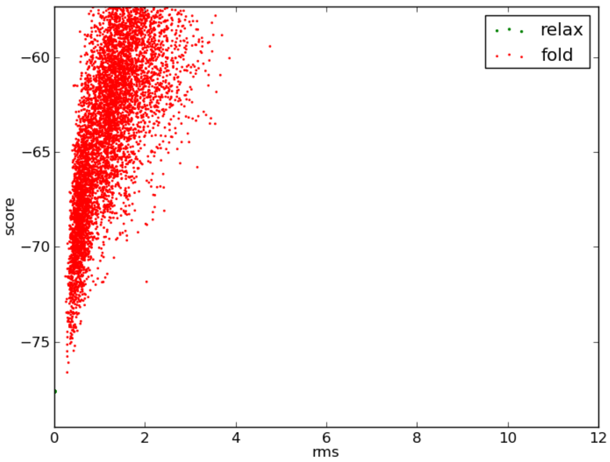




HEEE_2.1_06

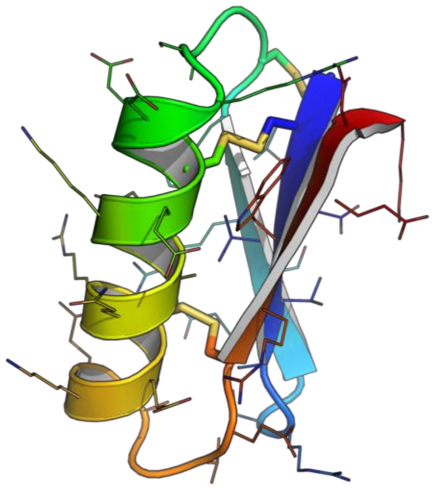
amino acid sequence:
KPEEYCRKVKDECKKRGLTRCHVTAKYGCECEVRGDTYQLRC

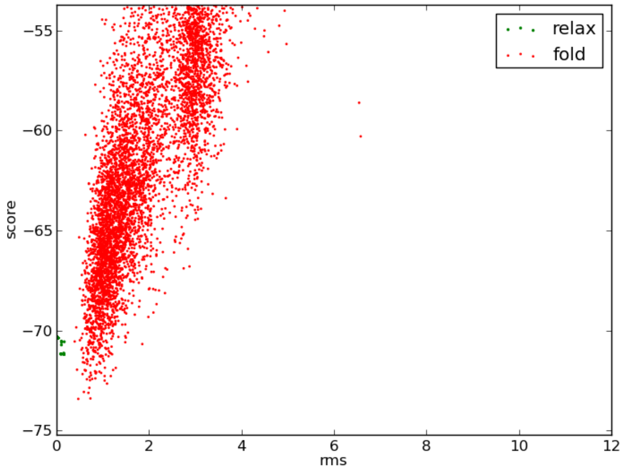


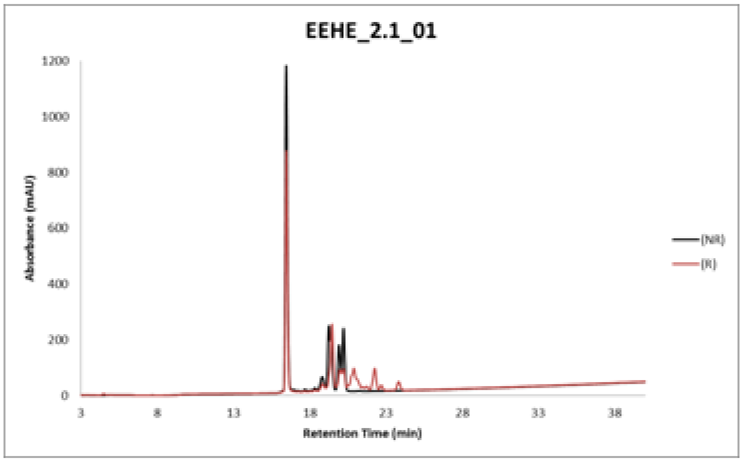


EEHE_2.1_01

amino acid sequence:
CCEVTSRSGESRTFCGASRDECEKEAQRCEKEAGVECRWEDK

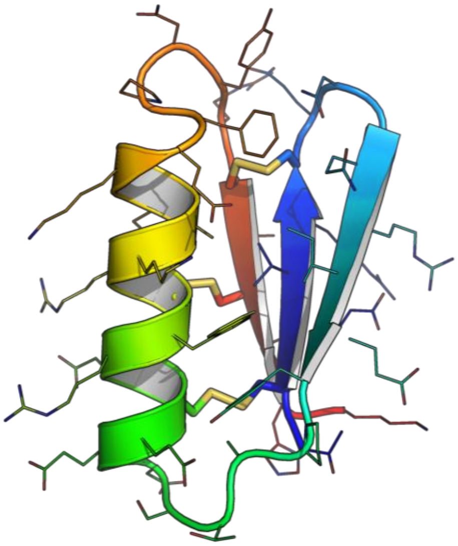


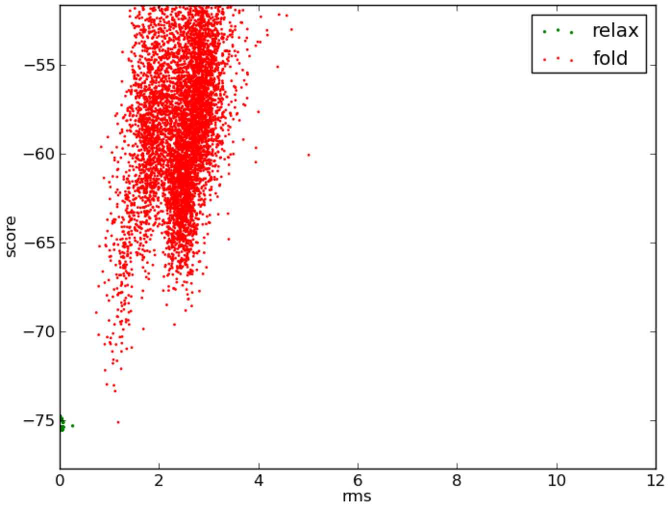


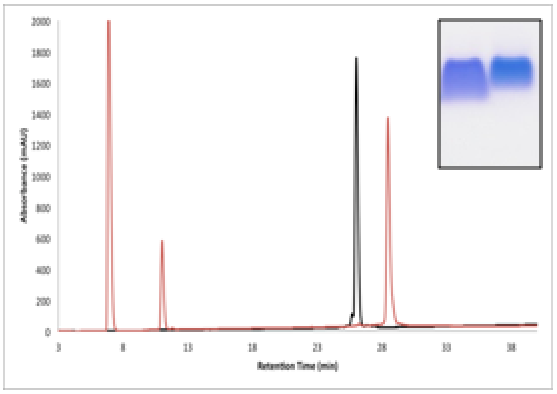


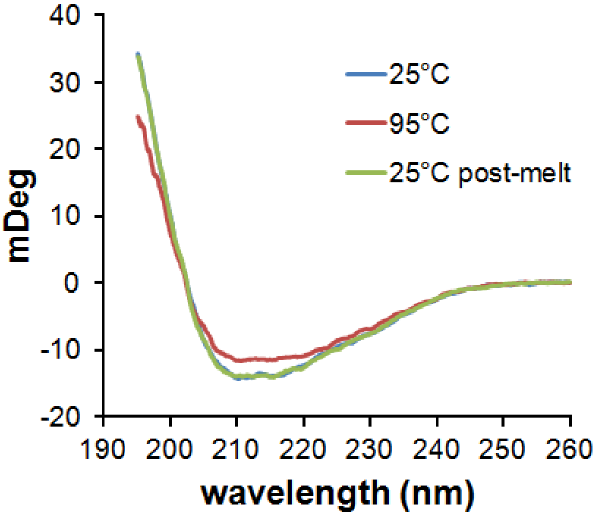
EEHE_2.1_02

amino acid sequence:
TCEVRCENGQRIEYPATSDEECERWCRKAKKEFPNYRCTCTHK



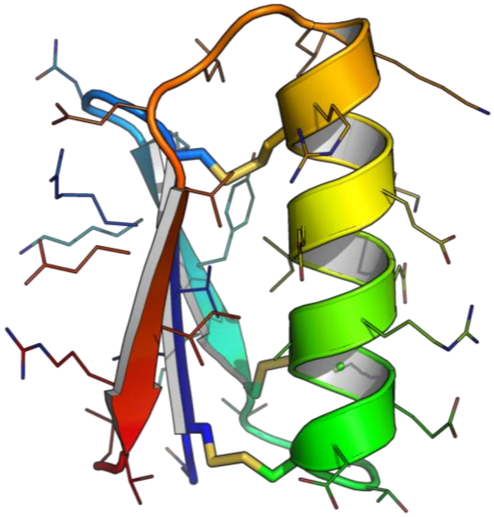


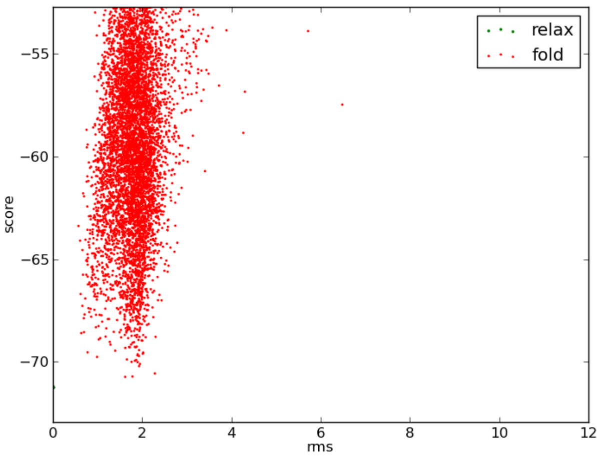




EEHE_2.1_03

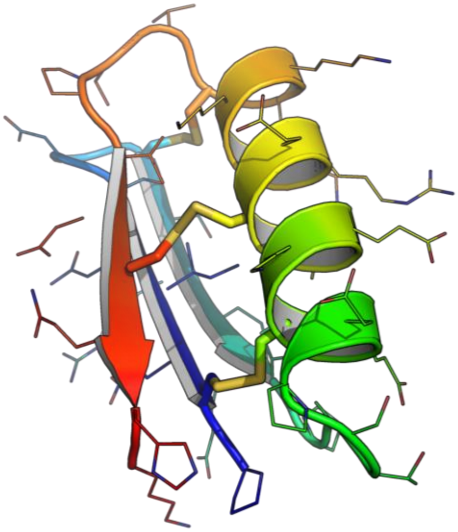
amino acid sequence:
ECTVRCGNQKYRCTTGTCDECAREIEEKCRKLGLEVEIRTL

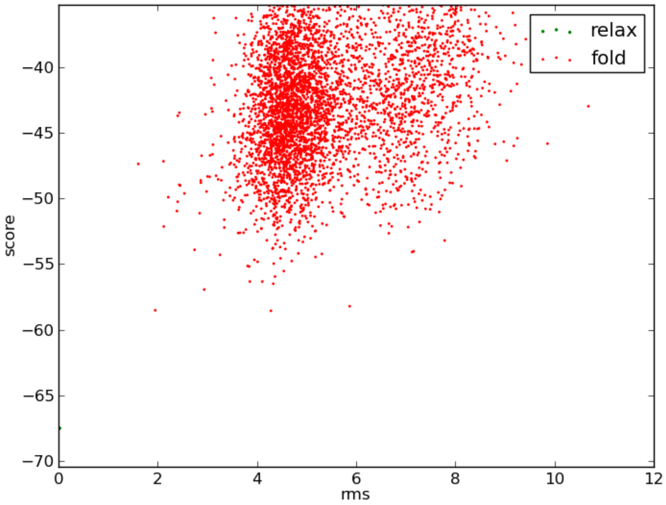




EEHE_2.1_04

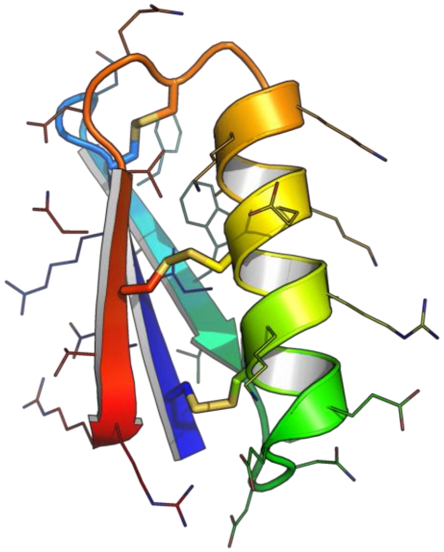
amino acid sequence:
PCEINSDGCTRQEIPATSPEECKEACERAKKKCTSPVDCQHK

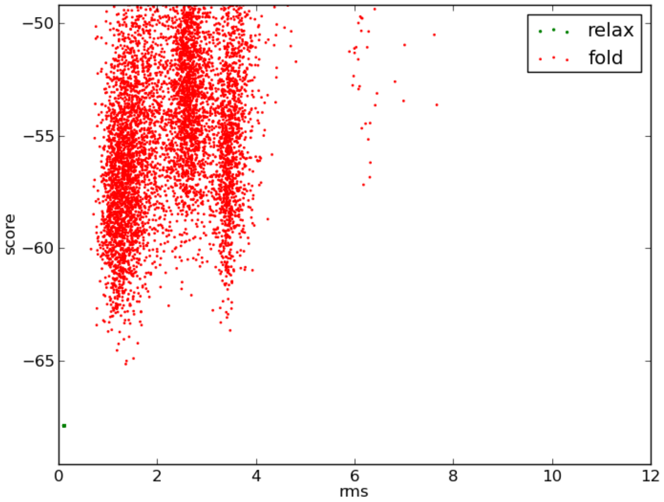


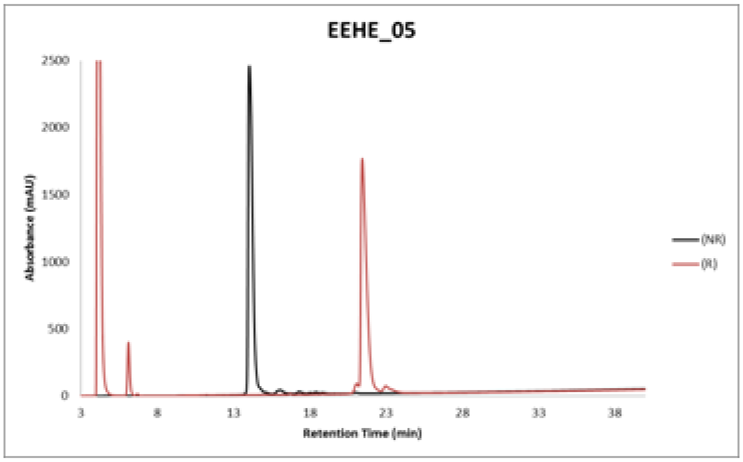


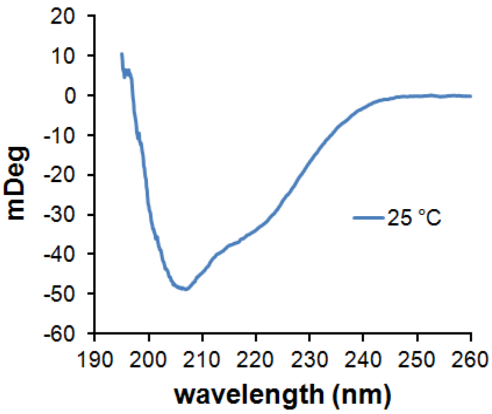
EEHE_2.1_05

amino acid sequence:
GCEIRCGNGYTWTVSDNEEKCKRECEKAKKSGCQDVNCTRR



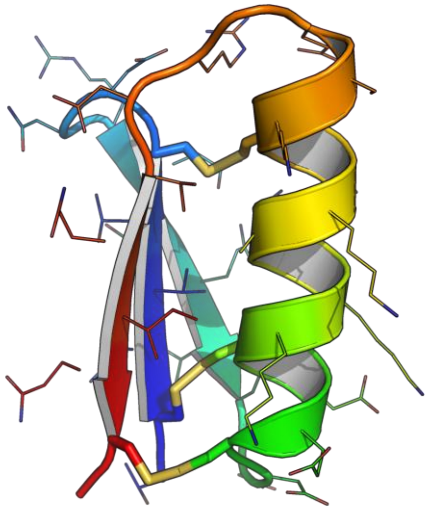


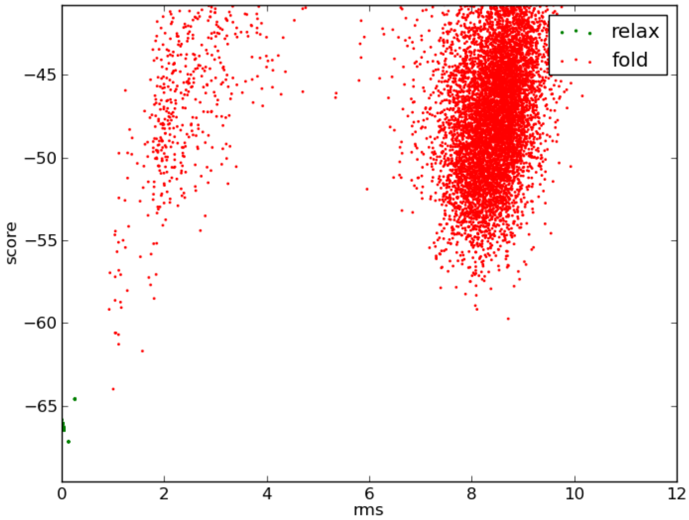




EEHE_2.1_06

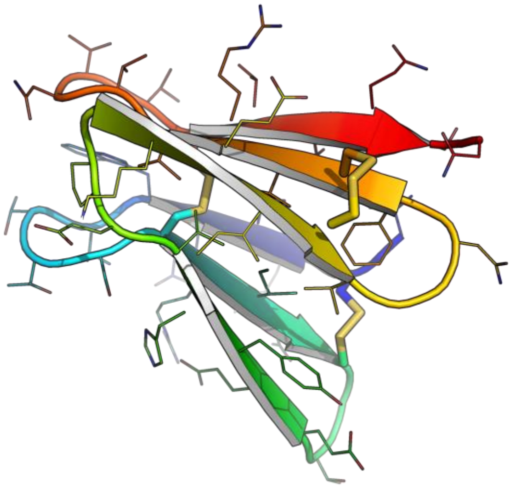
amino acid sequence:
TCNVTCDNRDTQTFDDCEECKKKAKECKSEGRDVQIQCG

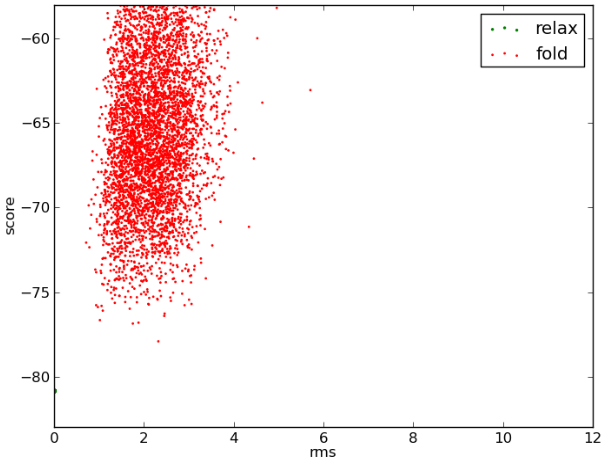




EEE_EEE_1.1_01

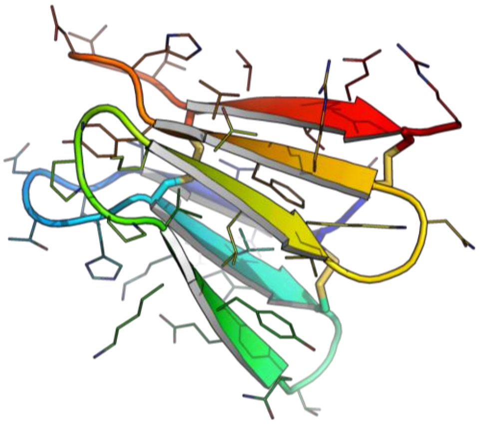
amino acid sequence:
VCRIEWTTTSCRIDCGTEEYHVEPGKEICVGNFCVRVTNTTCTVQSN

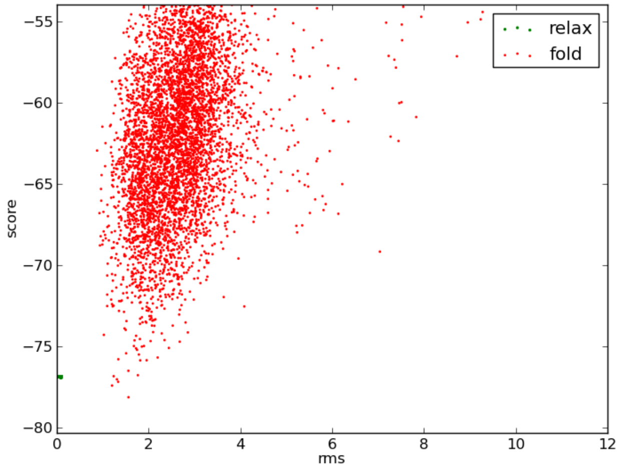


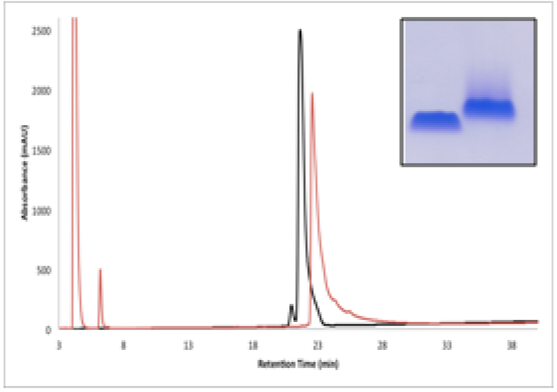


gEEEEEE_02

also known as: EEE_EEE_1.1_02
amino acid sequence:
TCEIRVTDTHCKVHCGTQEYKVPPGRTLKVGNCRFTYHDTTCTVECR

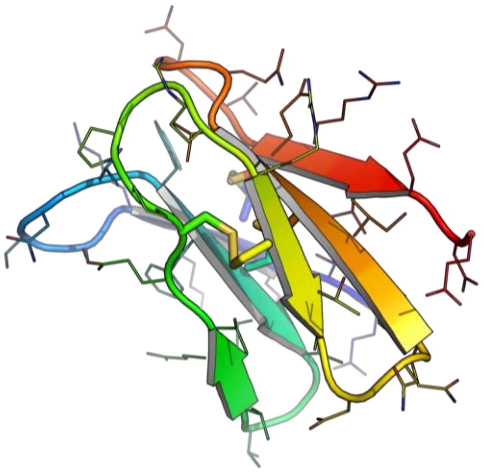


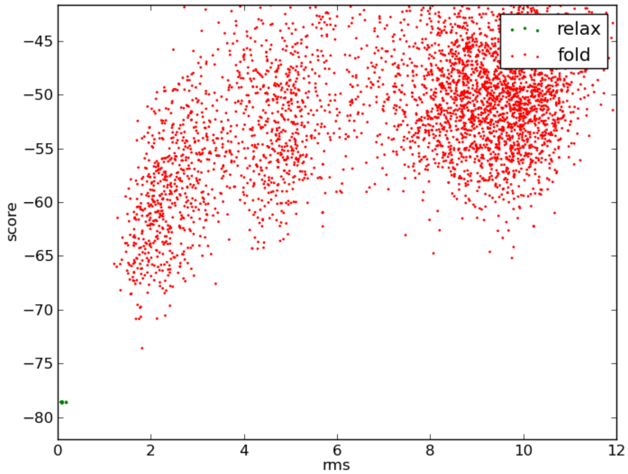


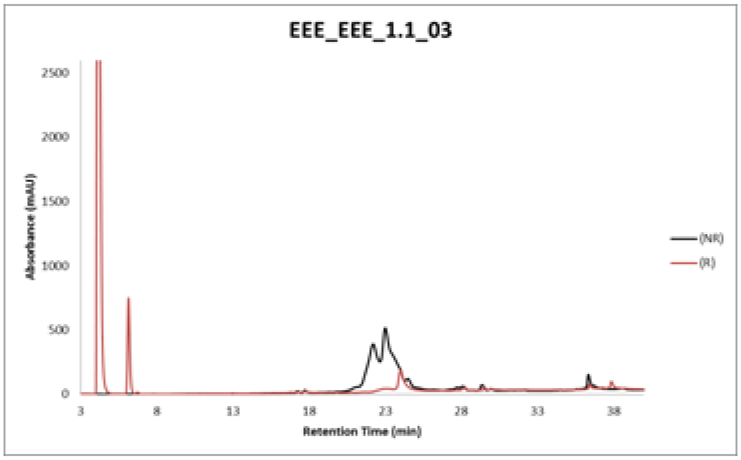


EEE_EEE_1.1_03

amino acid sequence:
ESRCKKSSNTWFCEVGTVQVECPPGRRCTINNQYICEVQGNTCRTENE

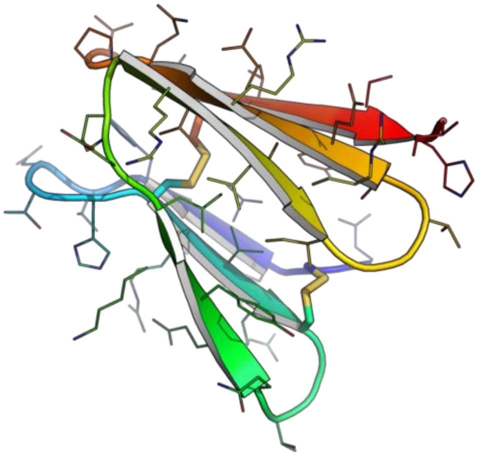


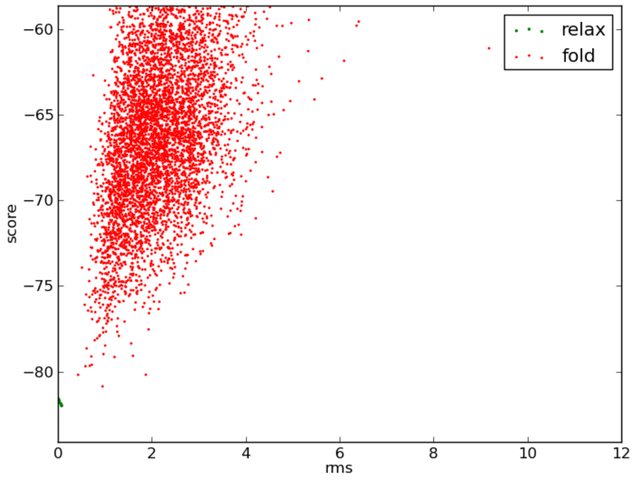




EEE_EEE_1.1_04

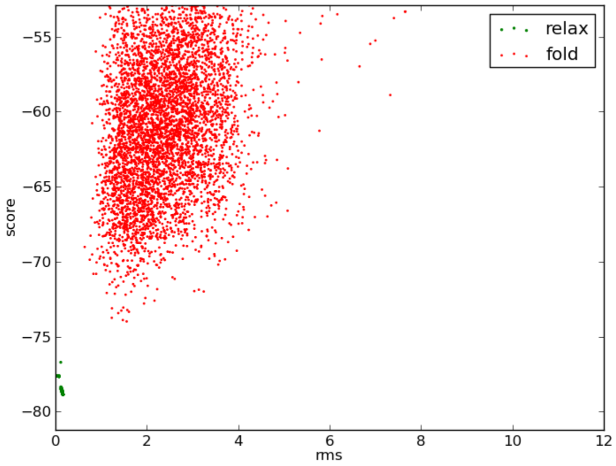
amino acid sequence:
QCEVRFTDTHCRVRCGTQEYKLEPGRRVRIGTSEFDVQPTTCTYSHI





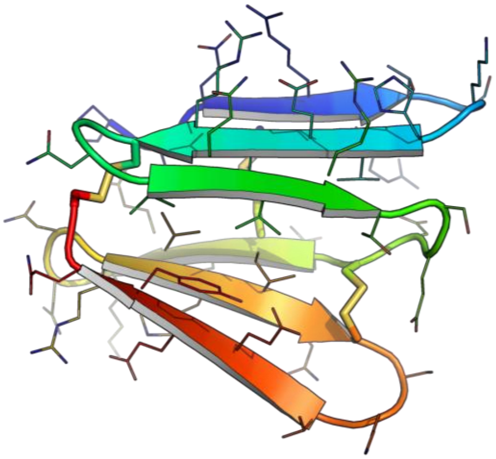
EEE_EEE_1.1_05

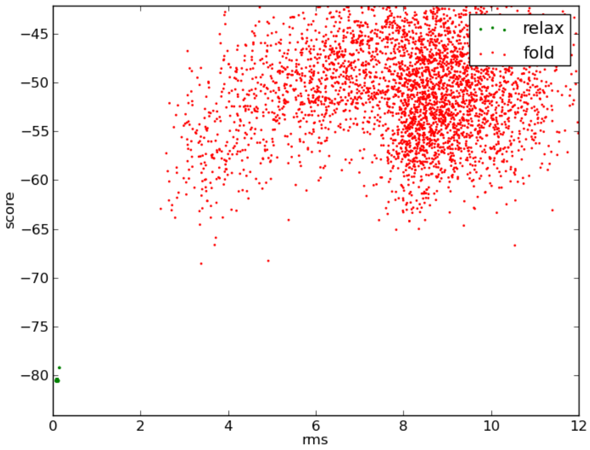
amino acid sequence:
TCEVRWTNTHCRIKCGTQEYECPPRRRCEIGNFHVDVHDTTCRLHSR

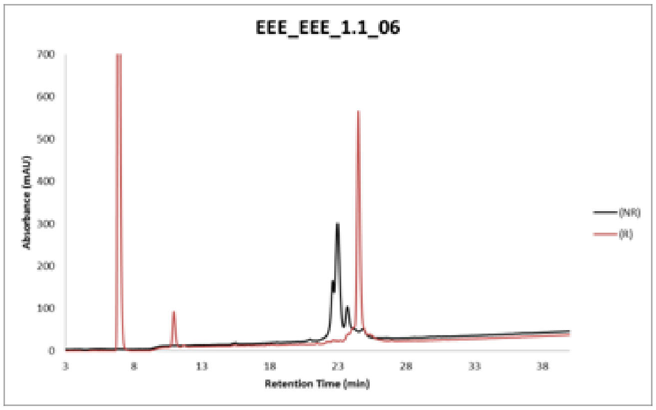


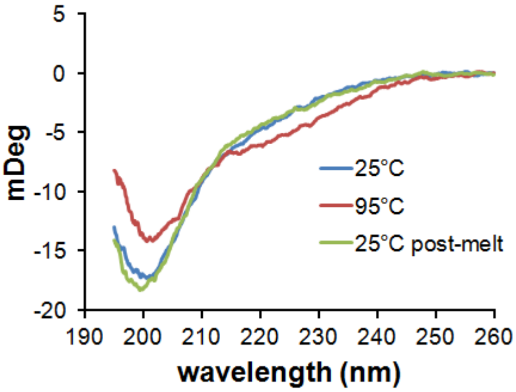
EEE_EEE_1.1_06

amino acid sequence:
PVECRRTSKHVEVRCGNVQVRTSEDCQCSEKNNRVHIQCSKTREEYQC



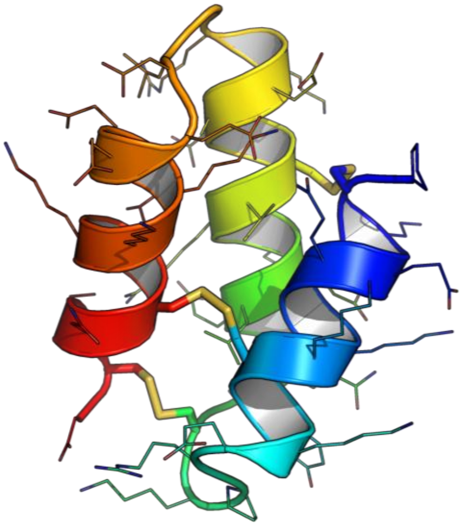


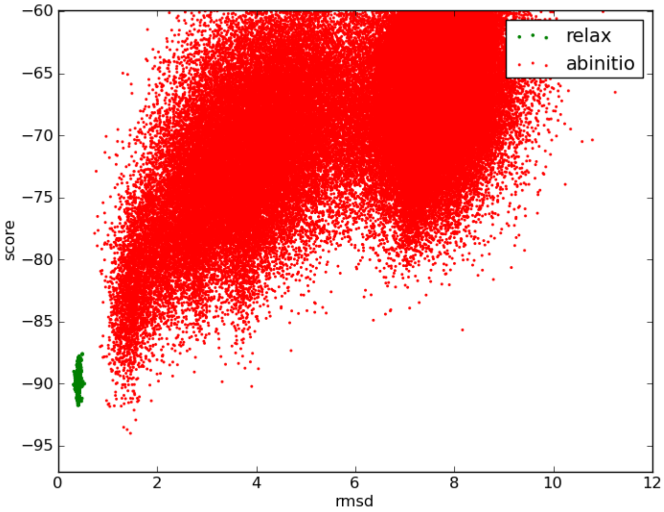




HHH_3.0_01

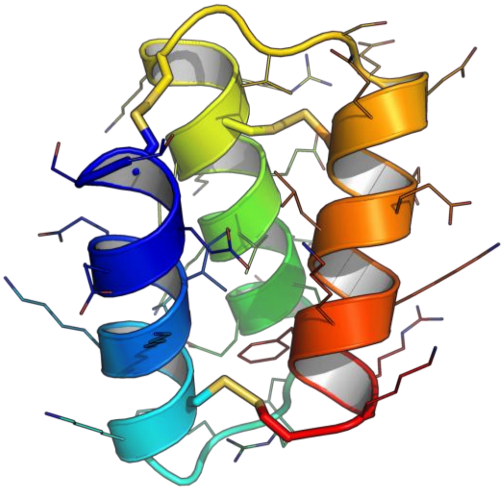
amino acid sequence:
PCKEQAKKCYKERPKCNQEELERRVCEAEKRGLDEEEKKKLCNSCD

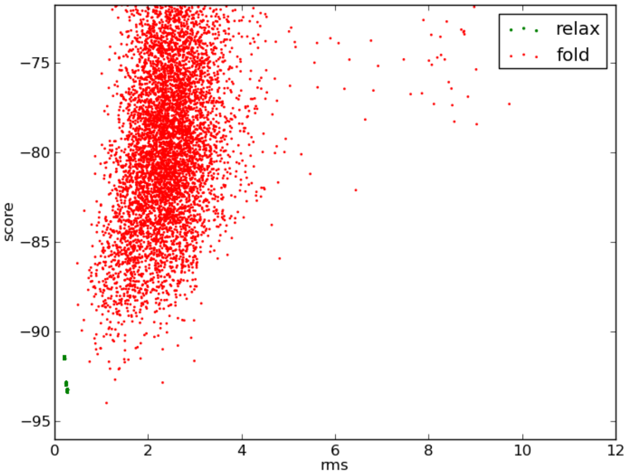




HHH_3.0_02

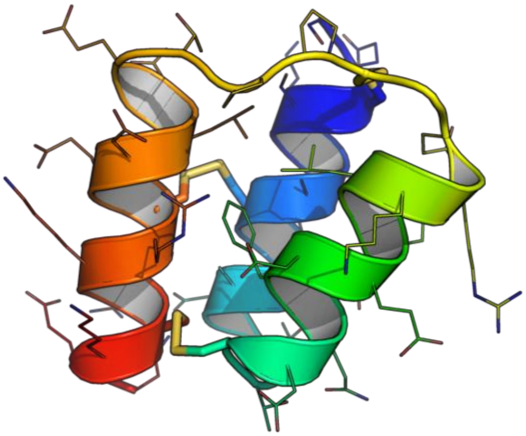
amino acid sequence:
TSCEEEIKKLCKSGKRDPEEEKKVEKICRKCGVSEDQCEELKKKFRKC

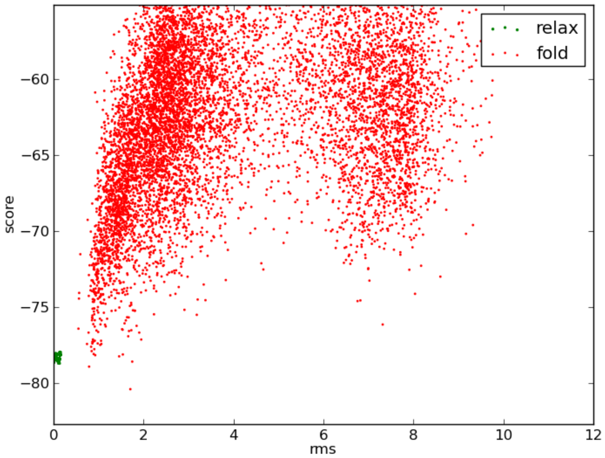




HHH_3.0_03

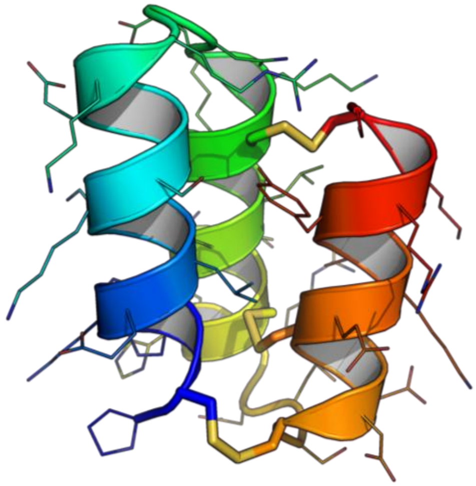
amino acid sequence:
SECSKEACKQAETGTCDQFDEWLKRQGCPPTEDLDECRKRCKEN

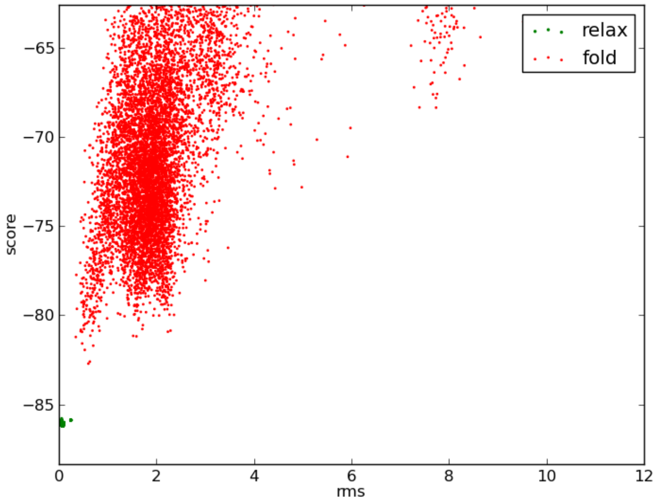




HHH_3.0_04

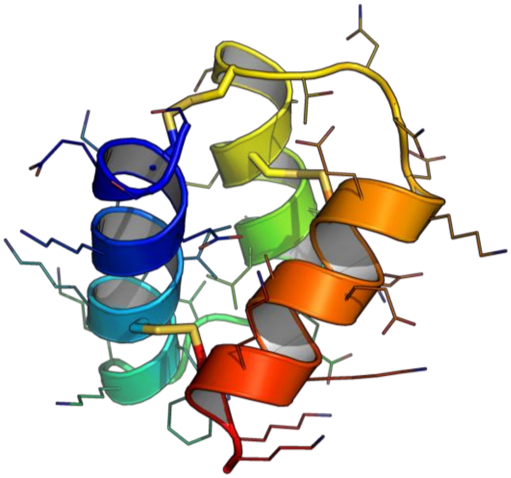
amino acid sequence:
PCWKELKKSAEKRGNEKCKKLAEECHRRNLSCDECEKLYRKCS

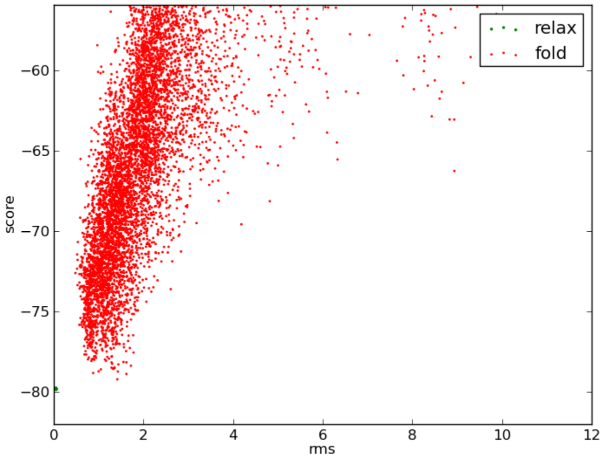




HHH_3.0_05

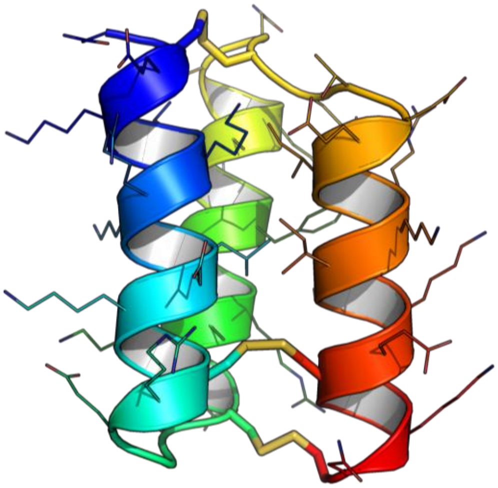
amino acid sequence:
SSECEKKICKEWKKGTSEDELRKLCSSCTNNDKECDEAIKKCKK

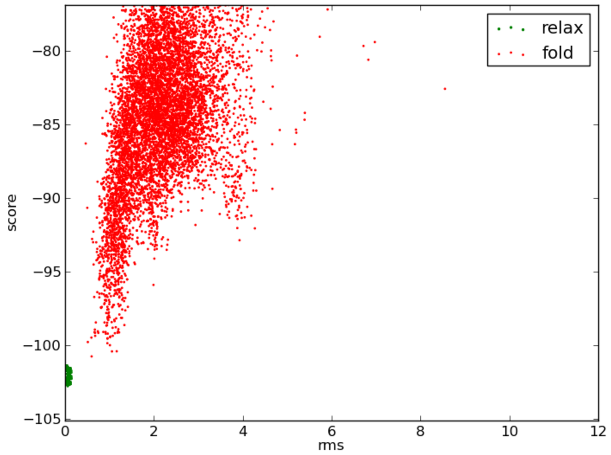


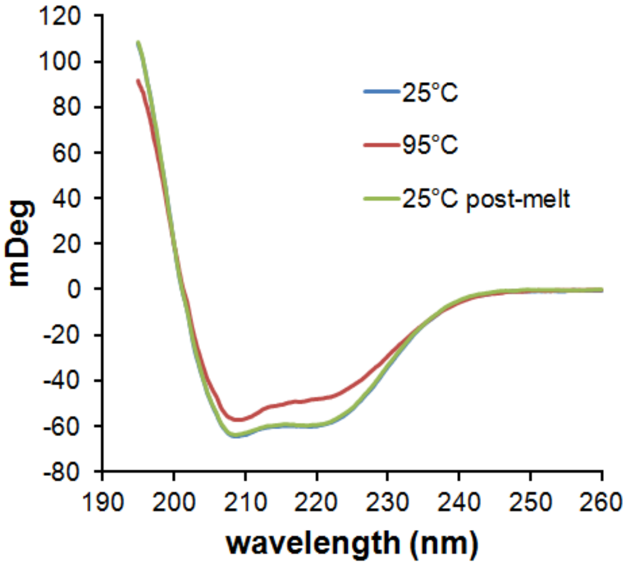


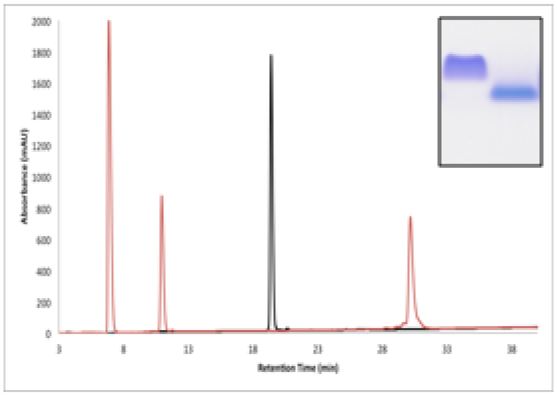
HHH_3.0_06

amino acid sequence:
NCEKLKRKLEKACREGNCDKARKAYEEAQRQNCETDEIRKIYKECEKNC









Round 5

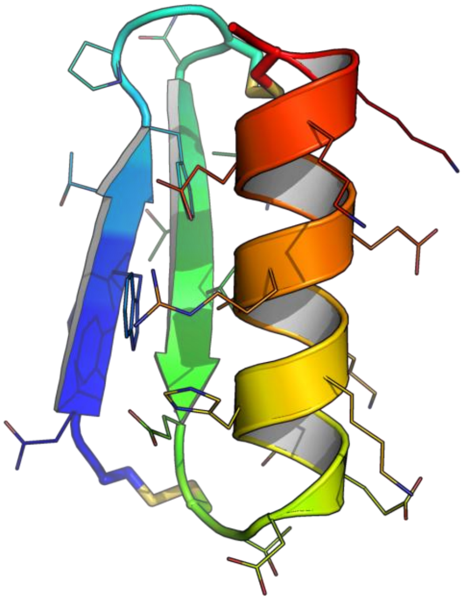
To test designs from this round, synthetic genes coding for designed proteins were cloned into the custom-made expression vector pCDB180, which was built by replacing the multiple cloning site of pET29b (Novagen) and carries a kanamycin resistance cassette, and expressed from BL21*(DE3) (Invitrogen) *E. coli*. Vector pCDB180 is functionally identical to pCDB26 (used in previous constructs) at the multiple cloning site, but the antibiotic resistance cassette has been exchanged to facilitate secretion of heterologously expressed proteins (the β-lactamase which confers ampicillin resistance is a secreted protein, and expression of this enzyme competes for access to the rate-limiting sec machinery with heterologous expression of designed proteins).

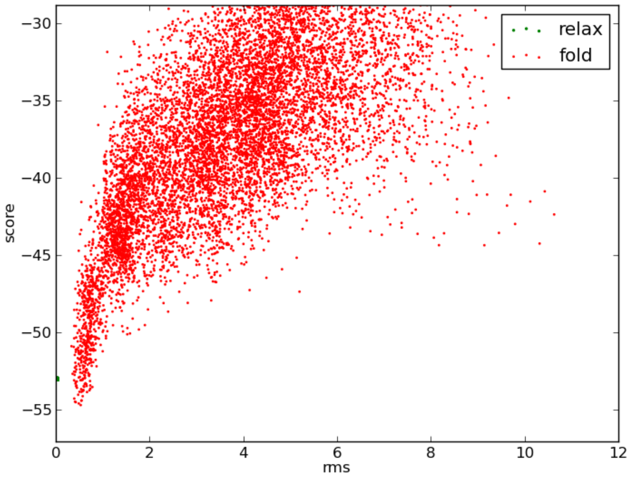
anatomy of the expression constructs

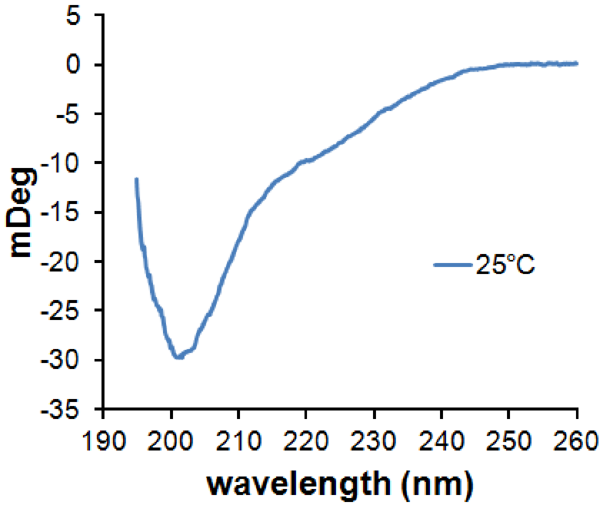
Designs were purified by IMAC from conditioned bacterial medium, and the fusion protein was removed using SUMO protease. Protease digestion is inhibited by a proline at the P1' position (first residue of the design). To facilitate removal of the fusion protein, an alanine residue was added to the N-terminus of any design beginning with a proline during generation of the synthetic gene.

EEH_1.0_01

amino acid sequence:
CQTWTFPGCNQTVTECTDEDHKKAREVEKKCG

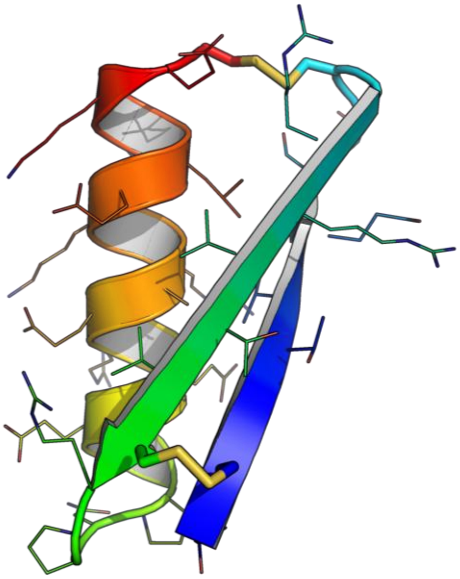


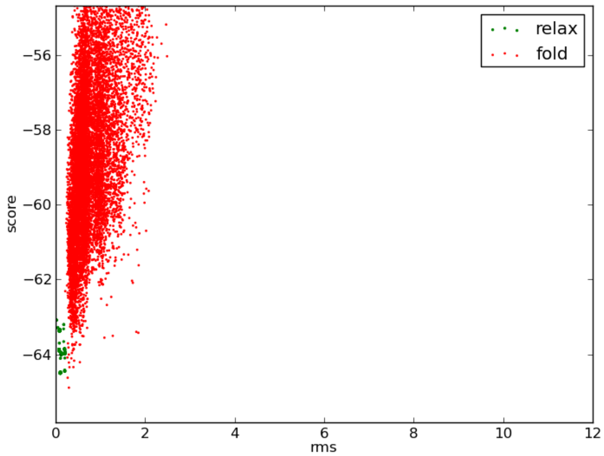




EEH_1.0_02

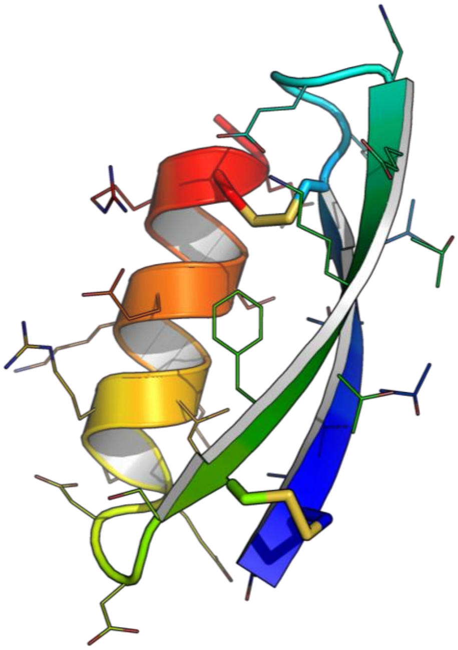
amino acid sequence:
TCVTVESSCGRRVTVCRPNPEEAEREARKELKKEC

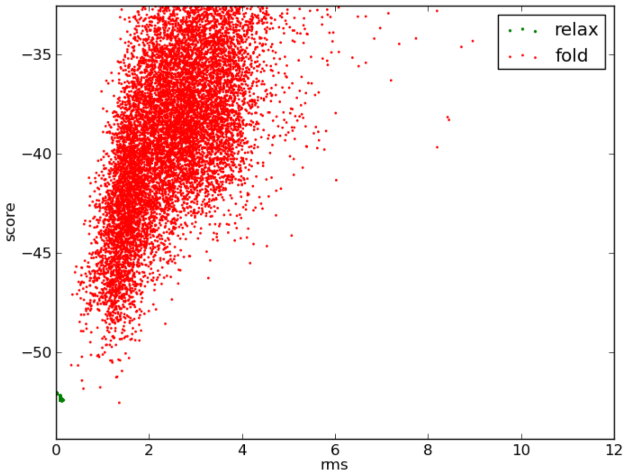




EEH_1.0_03

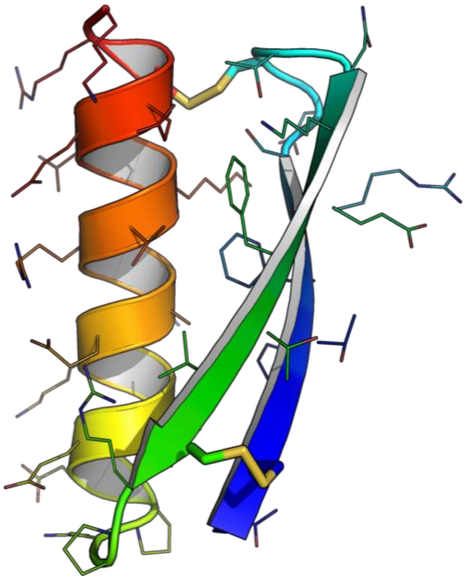
amino acid sequence:
TCITTTCKGENETKTFCSDDEERIKKESKRCEG

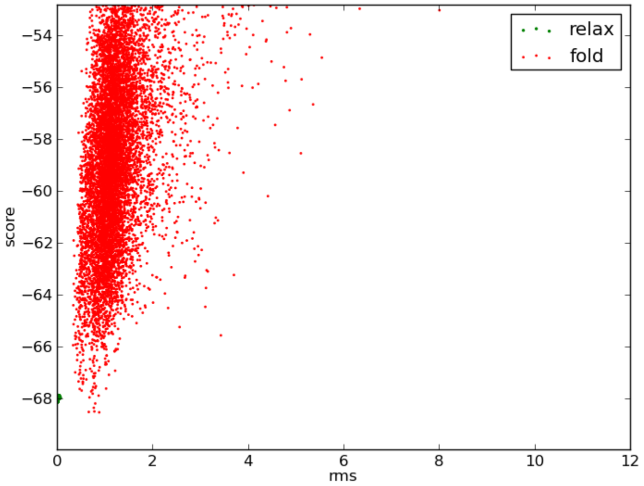


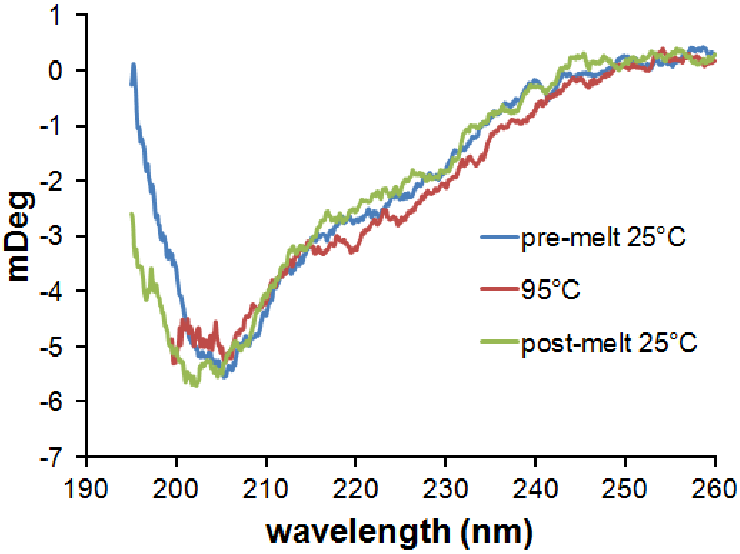


gEEH_04

also known as: EEH_1.0_04
amino acid sequence:
QCYTFRSECTNKEFTVCRPNPEEVEKEARRTKEEECRK

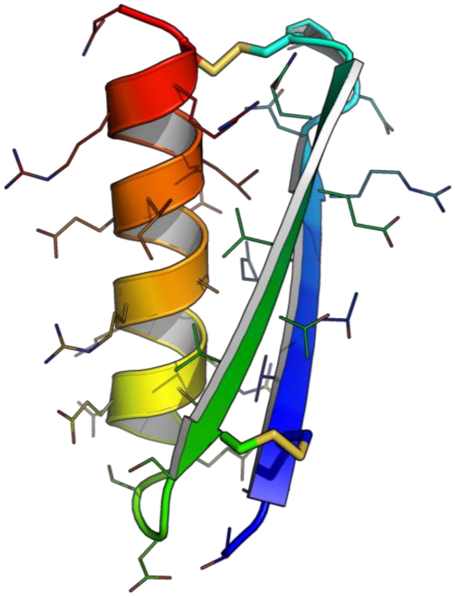


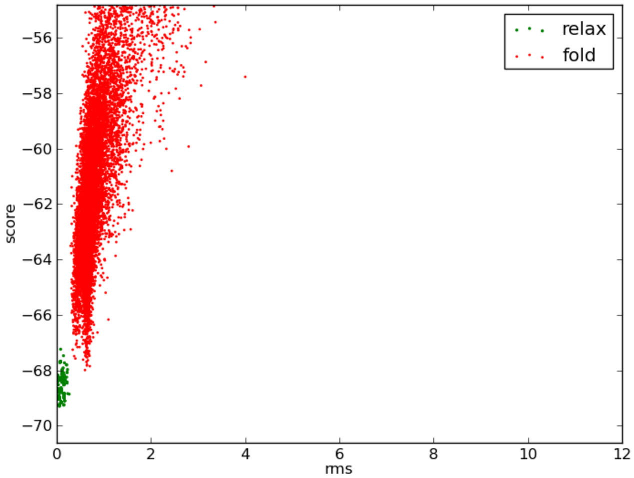




EEH_1.0_05

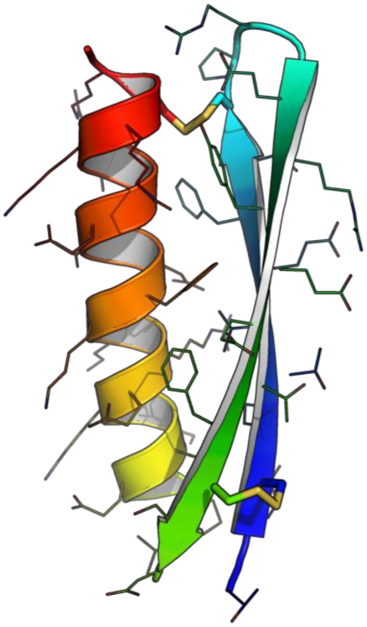
amino acid sequence:
TTCVTRRNDDCGQEVTVCSDSEEEARKRAEEILQRRCN

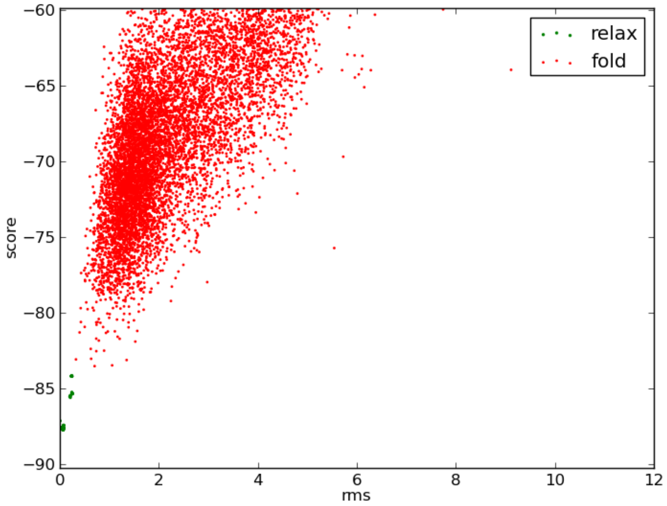


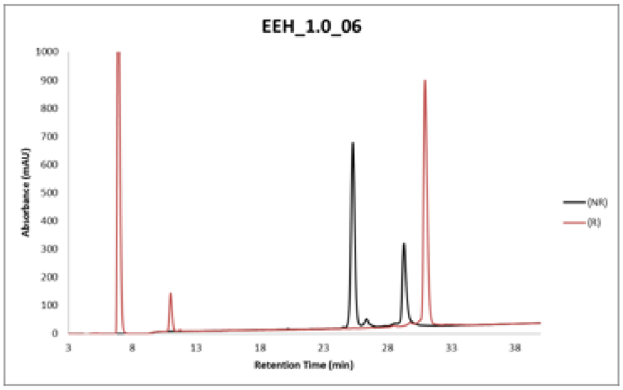


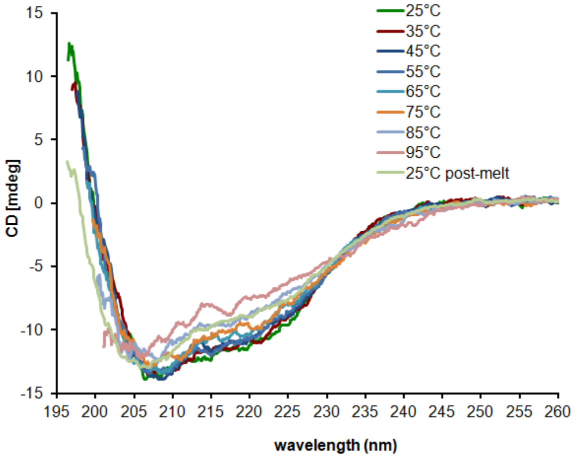
EEH_1.0_06

amino acid sequence:
TYCLTVEFTCPRGERYEETFCSDTPEEAKKERKKFETEAEKKCRG



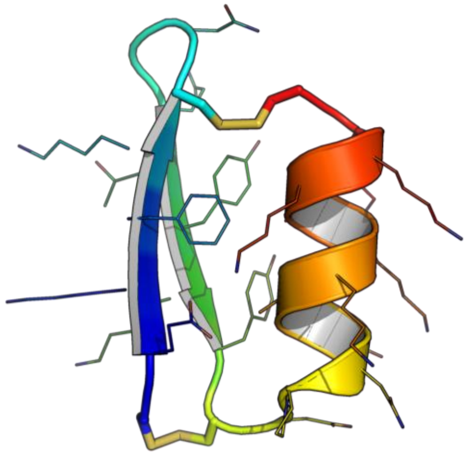


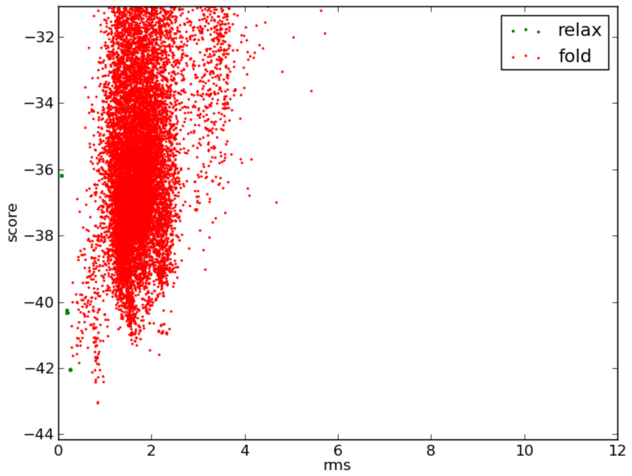


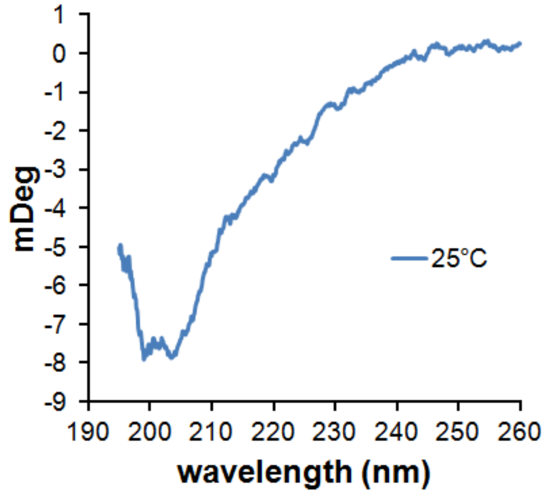


EEH_1.0_07

amino acid sequence:
CEKFKCNGQTYKYCDPNEAKKAKKKC

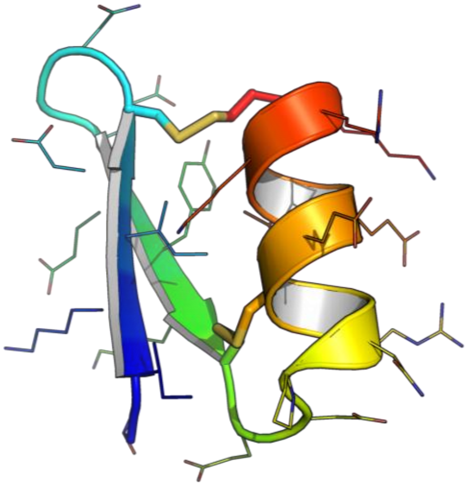


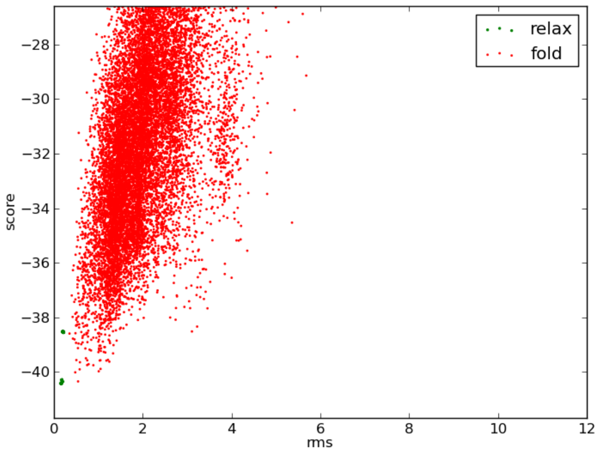




EEH_1.0_08

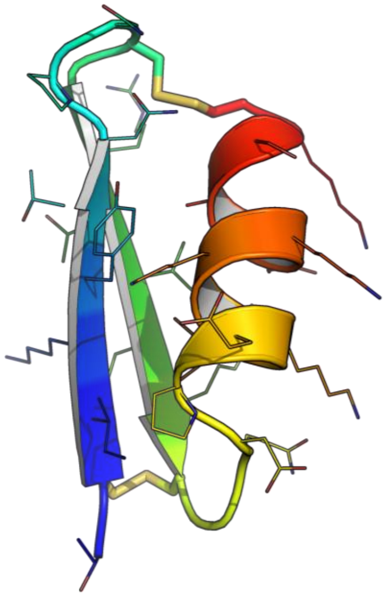
amino acid sequence:
TIKIDCNGEEYKCEDPNRCEEIKRKC

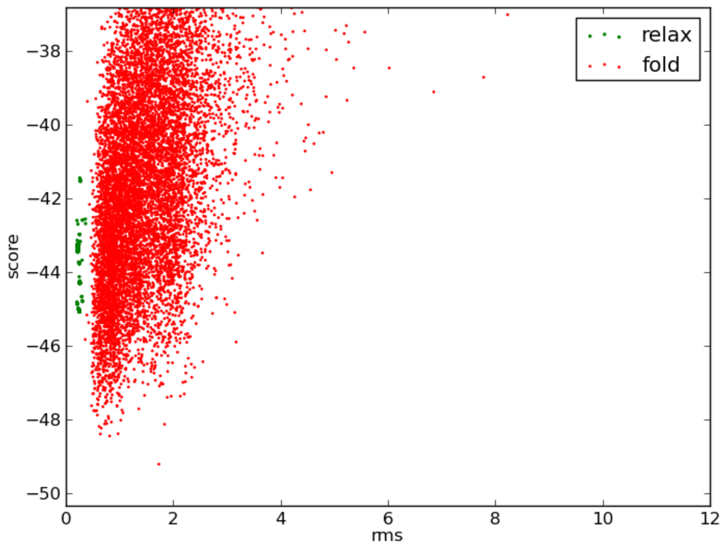


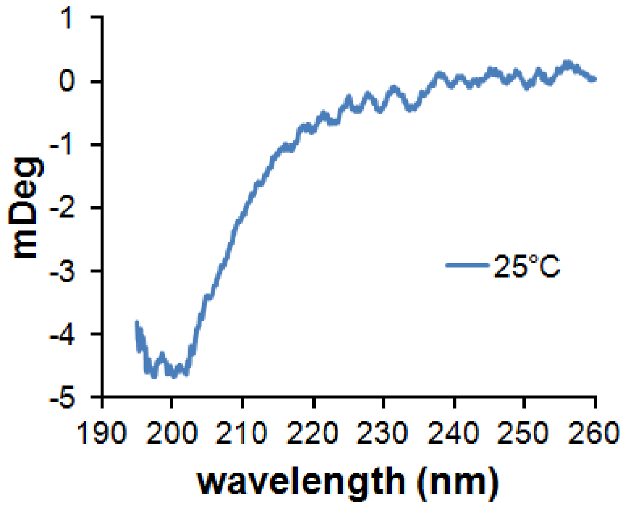


EEH_1.0_09

amino acid sequence:
TCIKYTNPNCGRTVERCGQDPEKIKKEASKC

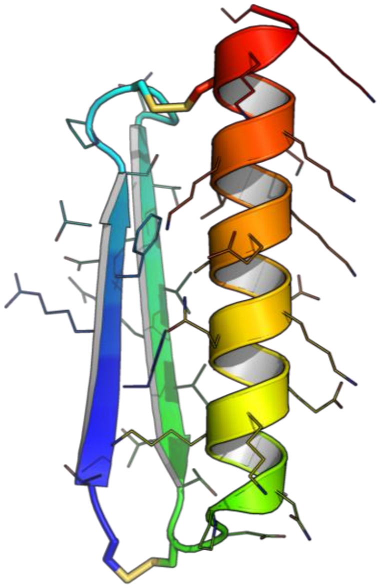


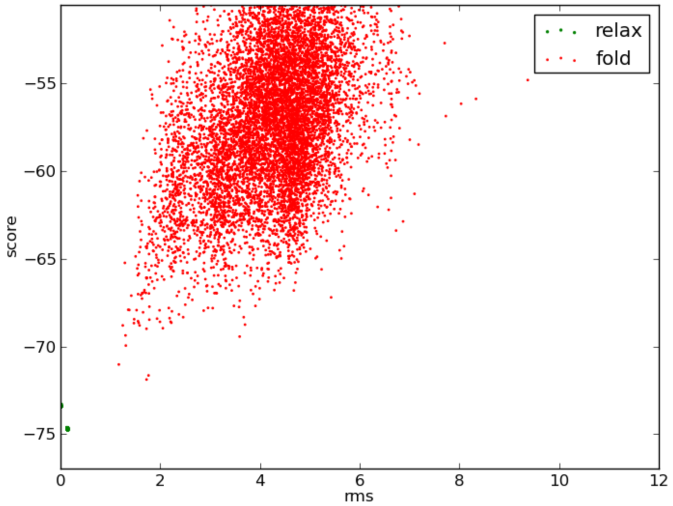


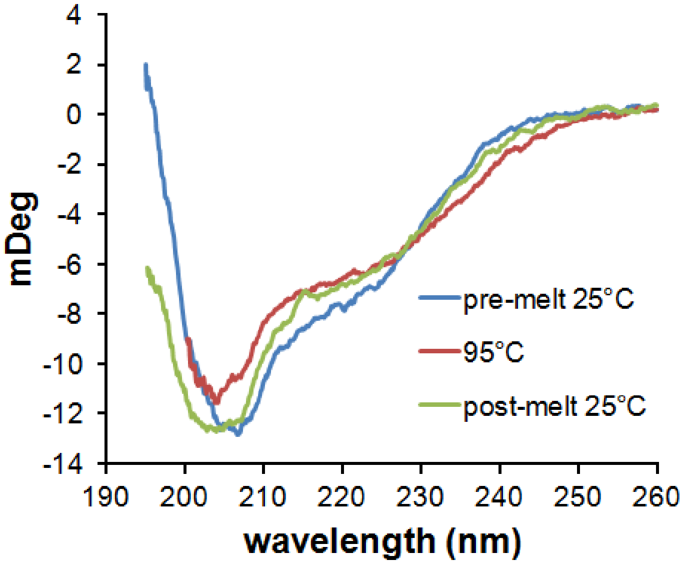


EEH_1.0_10

amino acid sequence:
CTTFRFTSPCGNTEVRVTTCDPNEKKEAQKEAEKLKKKCKKS

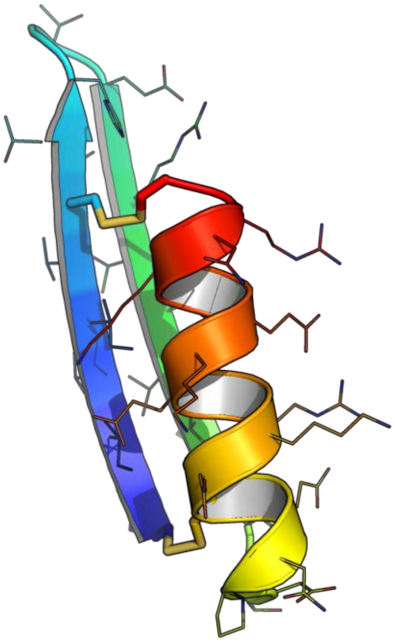


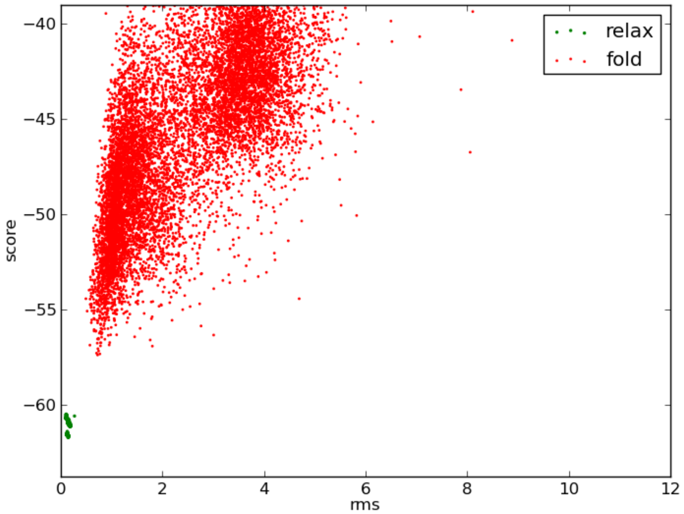


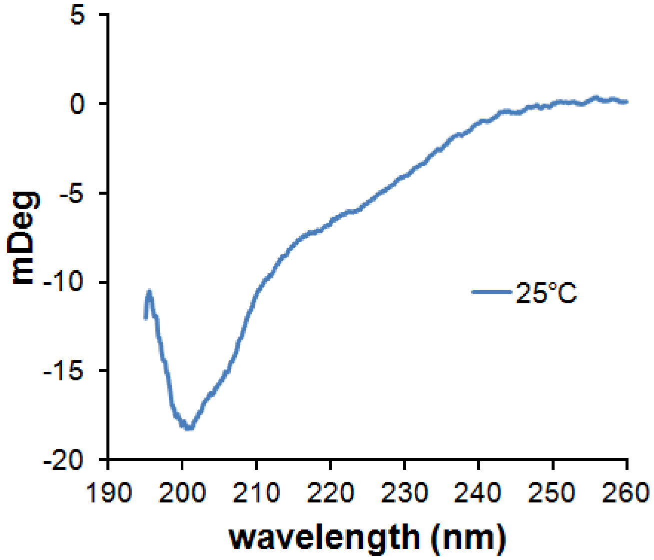


EEH_1.0_11

amino acid sequence:
CHITITCTHGTETRTETVKTTDPNECEKREKEIKNRC

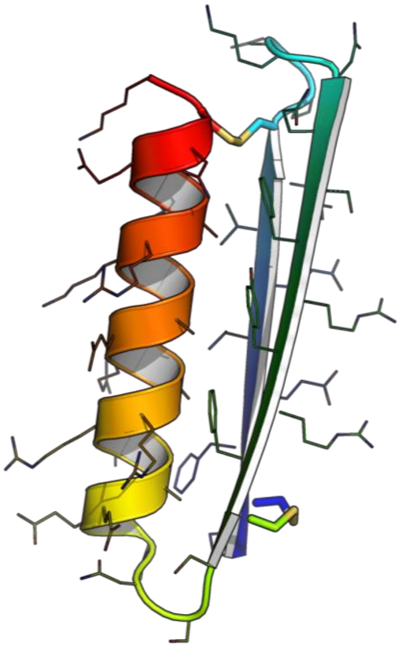


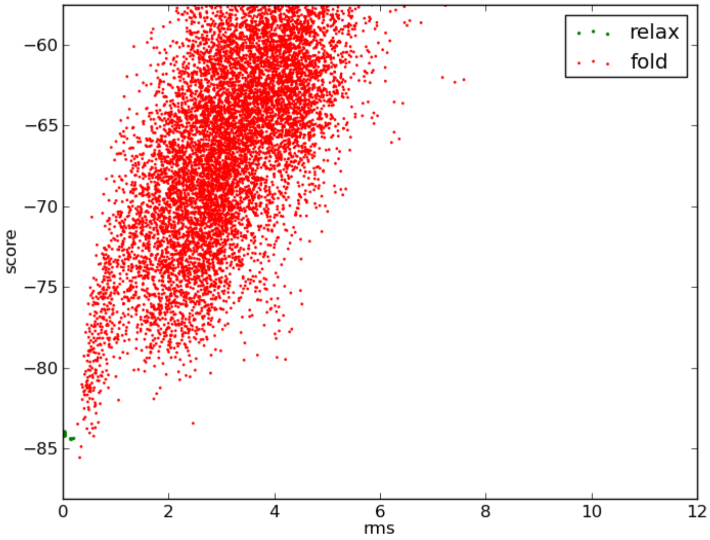




EEH_1.0_12

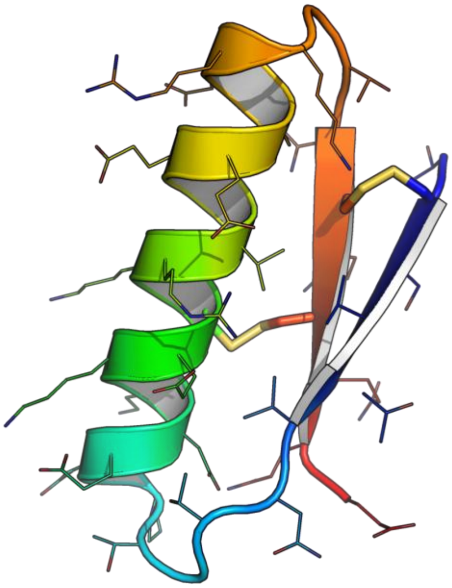
amino acid sequence:
QCFEVEVNCPDKNQSFRYRFCSSNPEEAERRAREAEKRARENCK

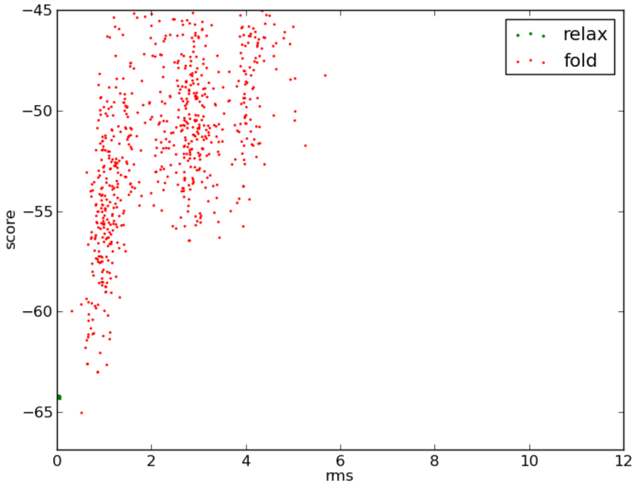




EHE_1.0_01

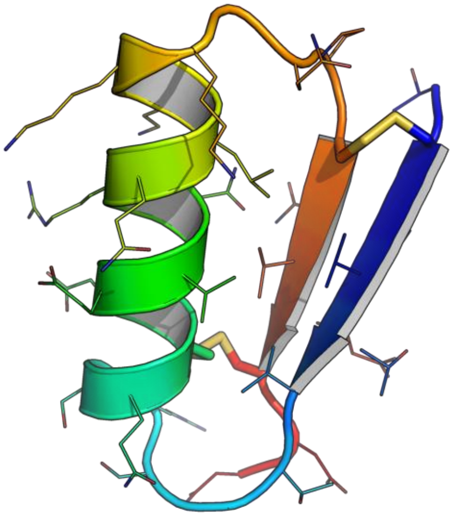
amino acid sequence:
TCSVTVNTGTPDEDKKECKRVQEEAERKGTQCQCQQE

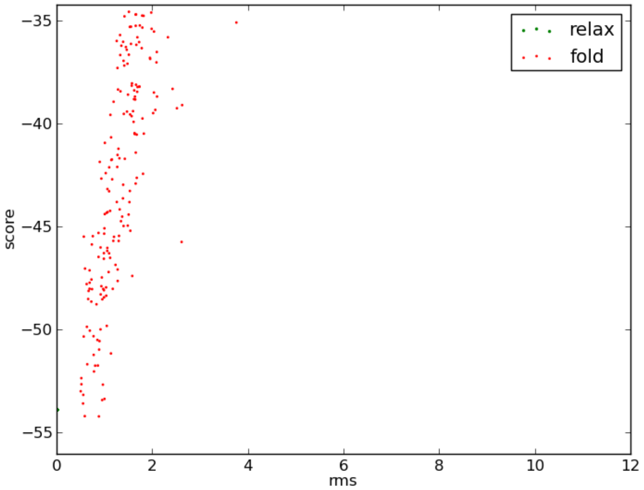




EHE_1.0_02

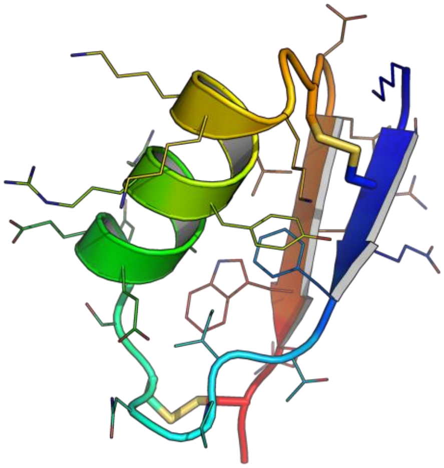
amino acid sequence:
TCSVTVTGSRSQCEEVQRQLKKKGQPCQVECDN

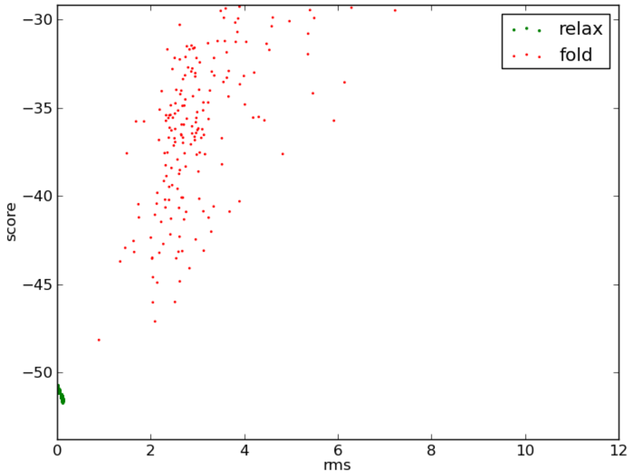


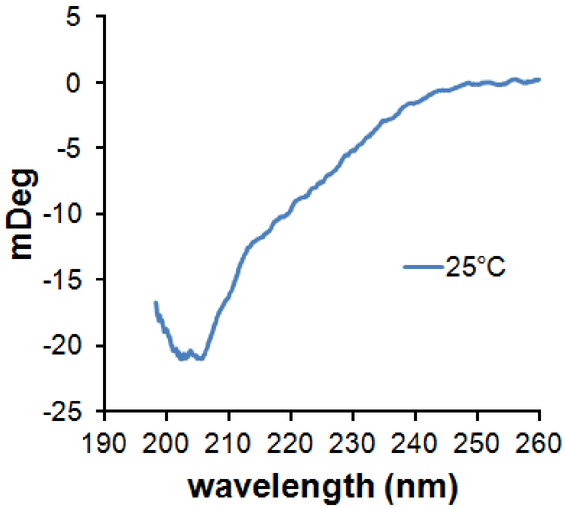


EHE_1.0_03

amino acid sequence:
KTCEFTIPNCSEEEARRYSKKKGCDETRWQCG

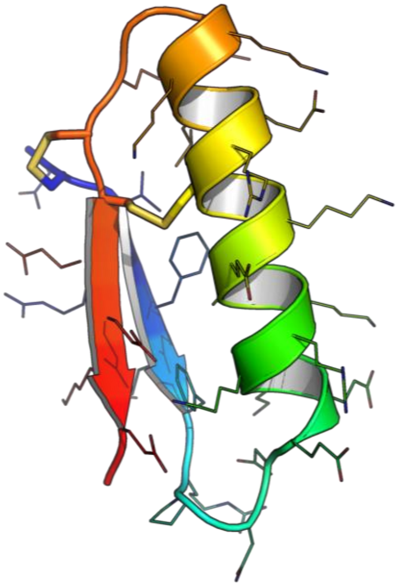


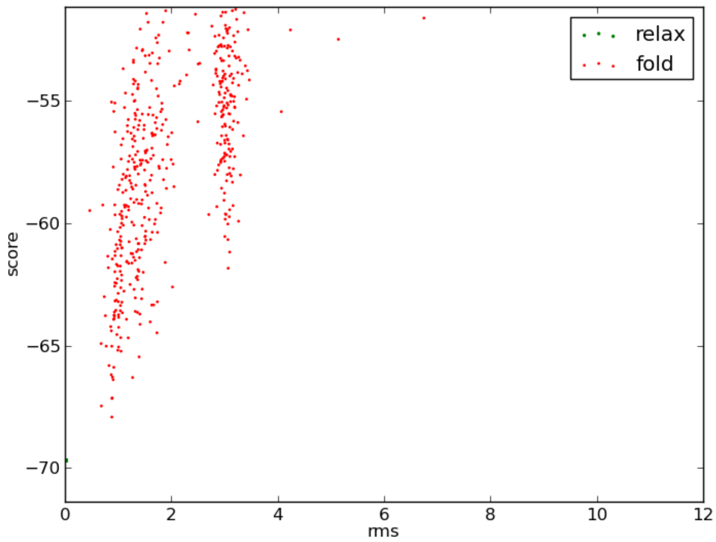


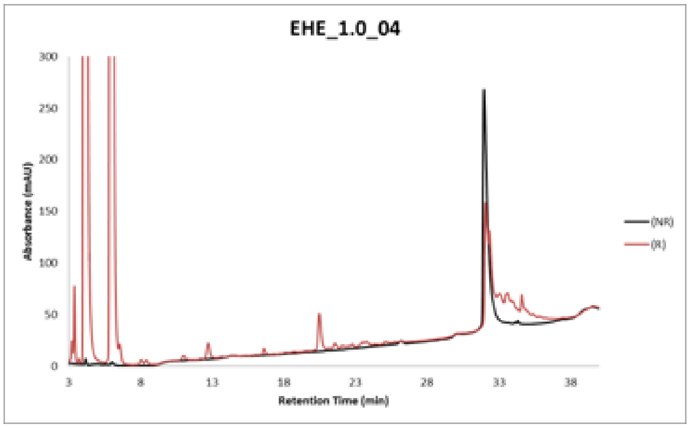


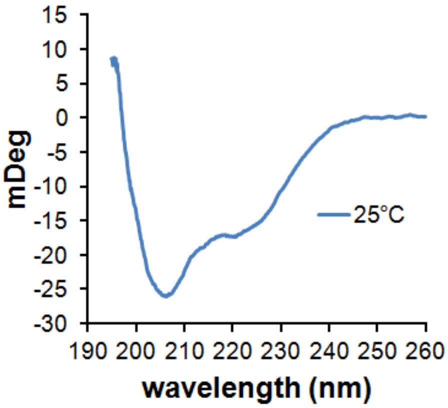
EHE_1.0_04

amino acid sequence:
QCVRFEFRPNDEEKKRKAEKACRELKKEGKCCEEKEG



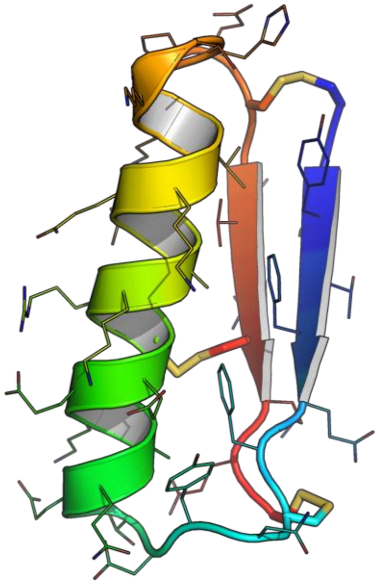


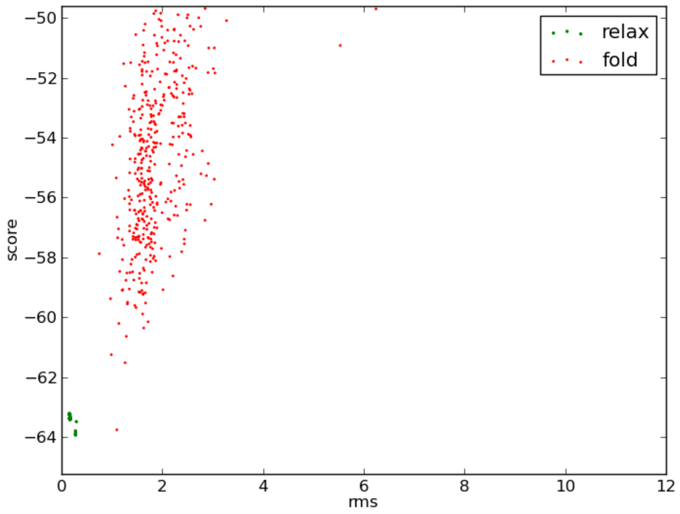


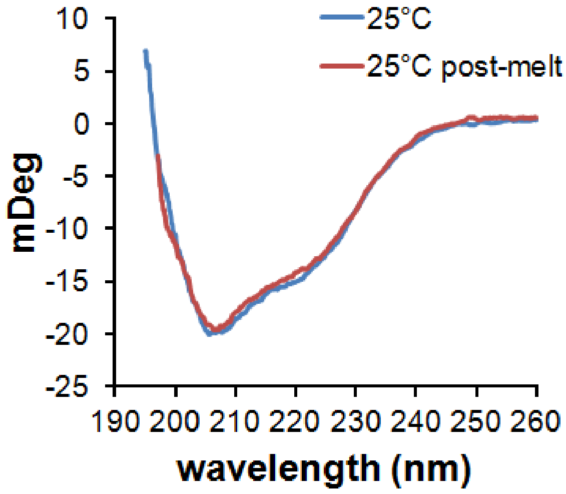


EHE_1.0_05

amino acid sequence:
CKYTFQFCNYDTEQAKEECRKAEEKVKKTHPECEVQCQEC

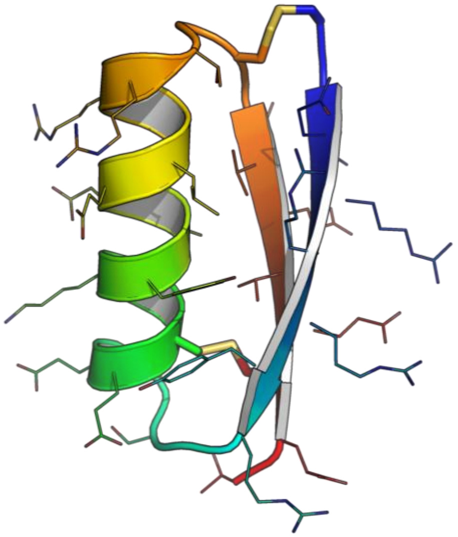


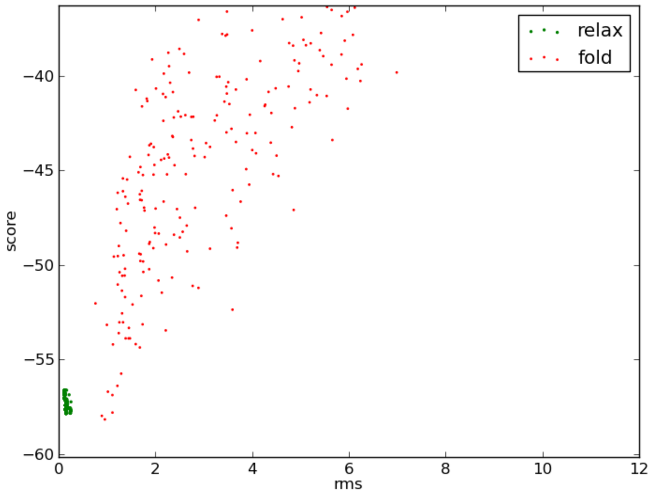


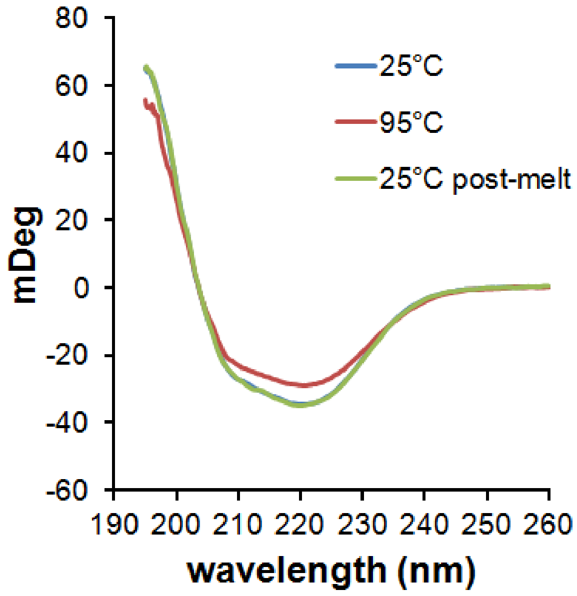


gEHE_06

also known as: EHE_1.0_06
amino acid sequence:
CKQRRRYRGSEEECRKYAEELSRRTGCEVEVECET

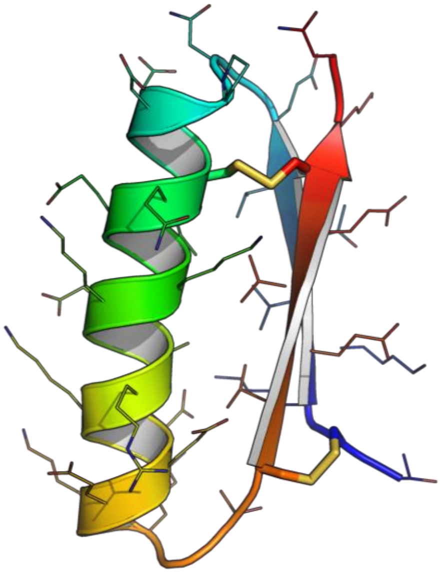


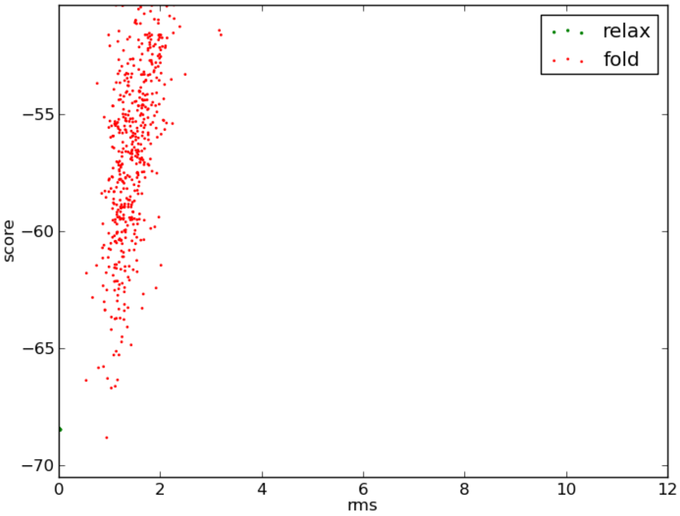


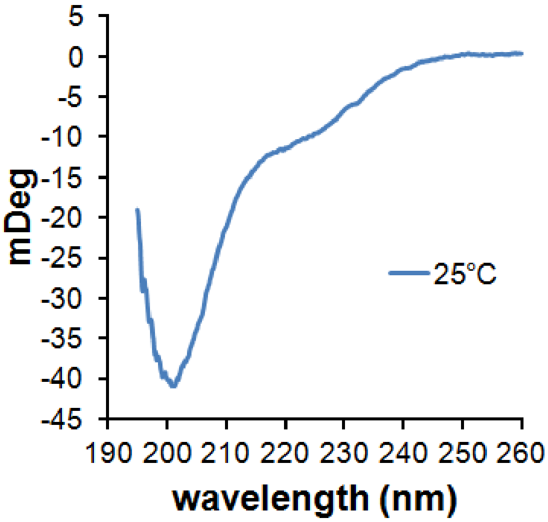


EHE_1.0_07

amino acid sequence:
TCKKVTVEGNPDECQEVKKEARKEEEKKGTCVEVECKN

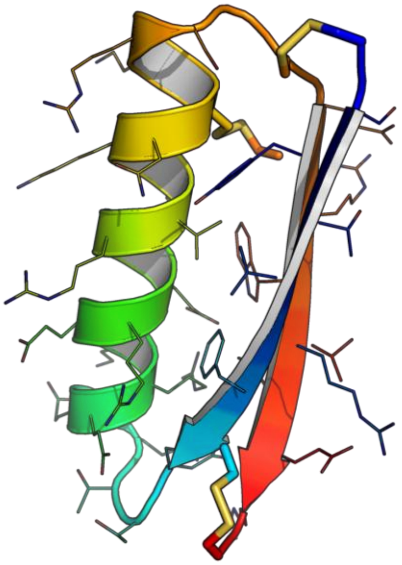


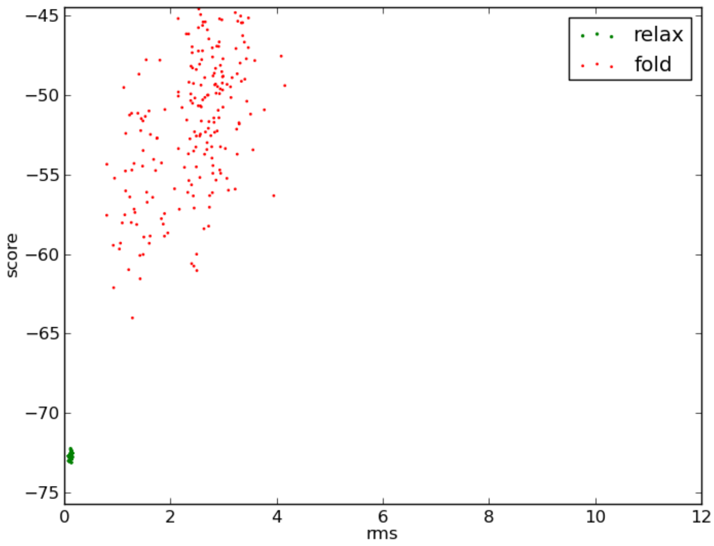


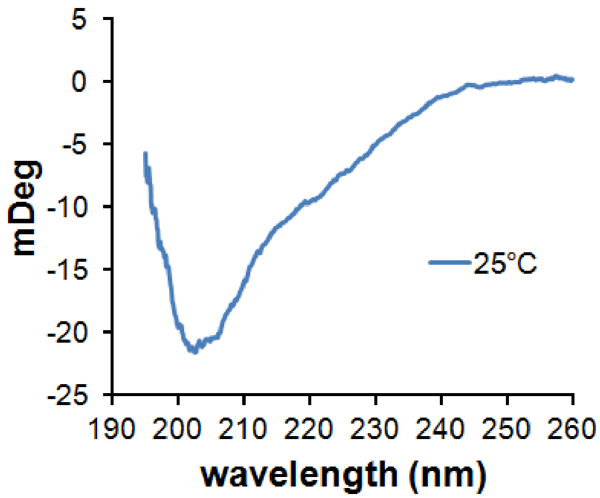


EHE_1.0_08

amino acid sequence:
CSYTVRFCYTTEEERKEREERVKKNCKRSGCECRWTNERC

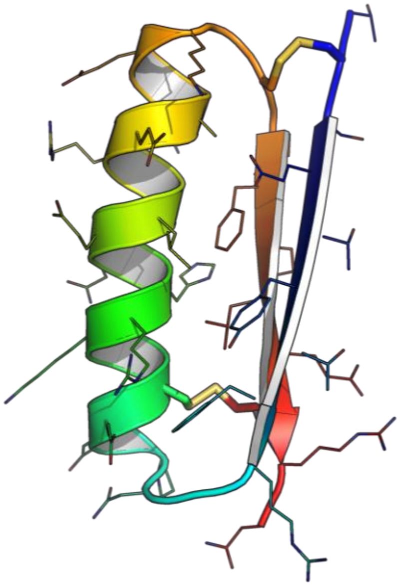


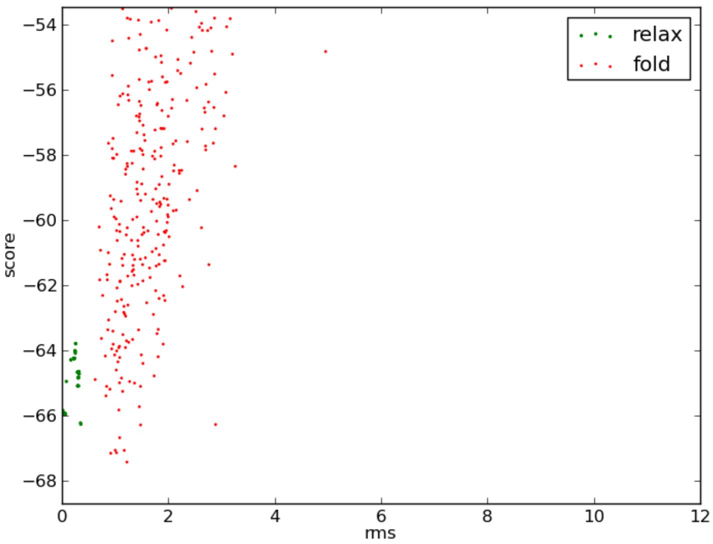


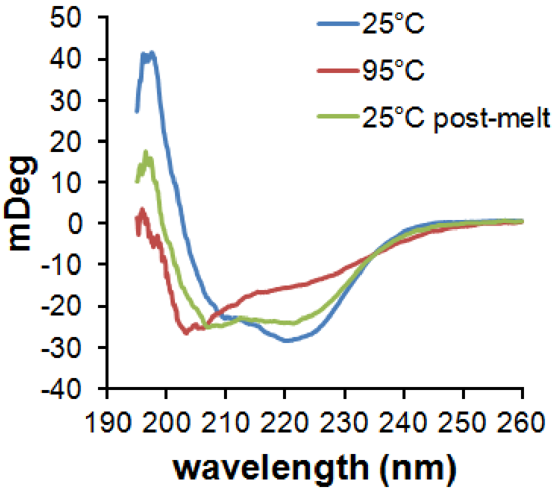


EHE_1.0_09

amino acid sequence:
TCSETYTFRGNPDECEKRHQELEREAREKGCQFQLECRN

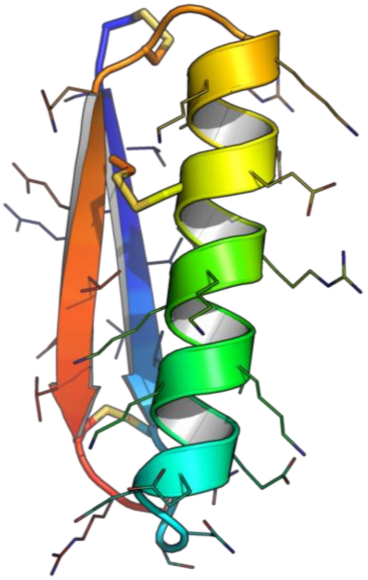


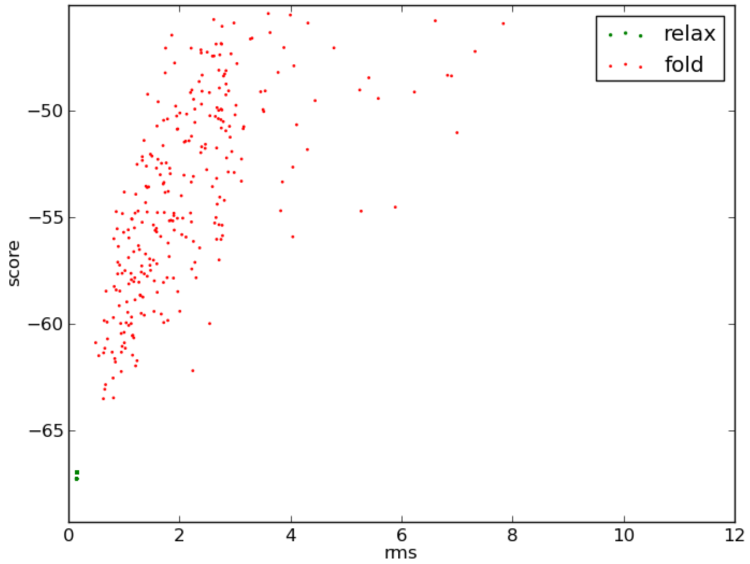


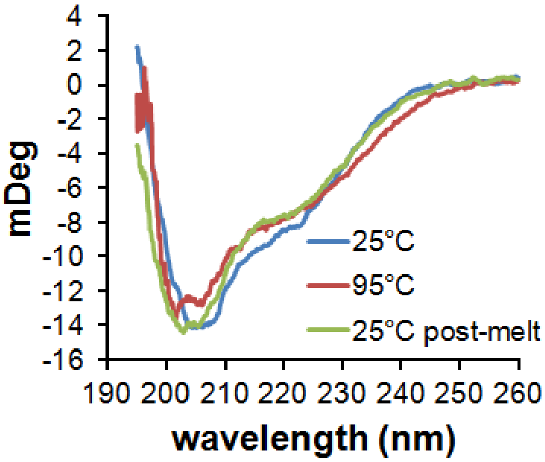


EHE_1.0_10

amino acid sequence:
CSTRVTVCNSNDEEAKKIKKRVCEEAKKRGCQCETETCRK

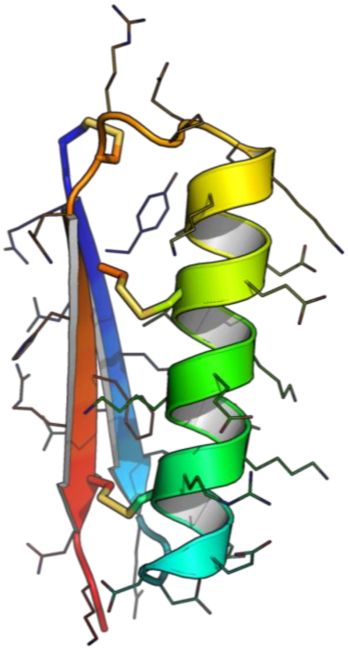


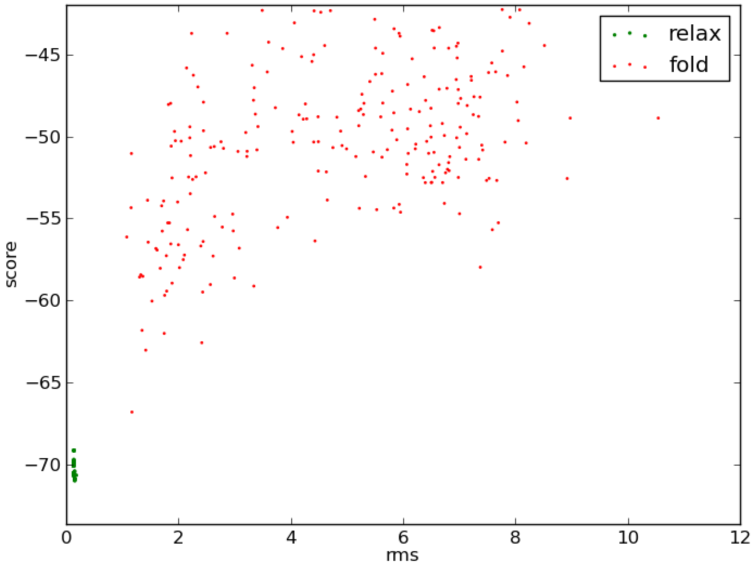


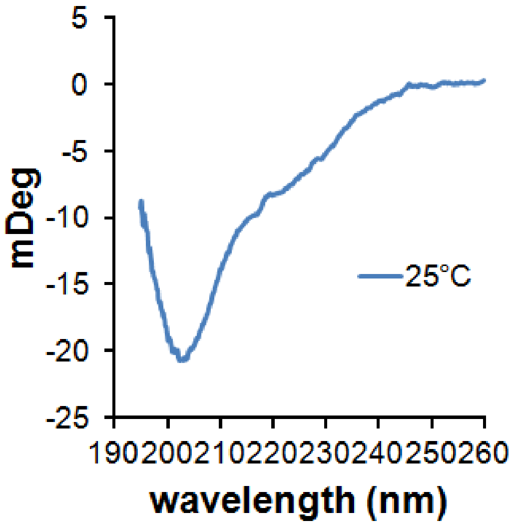


EHE_1.0_11

amino acid sequence:
CRYEKETRGDDEQCRKEKEKLCEEAKKEEPRCQCHFRCQKG

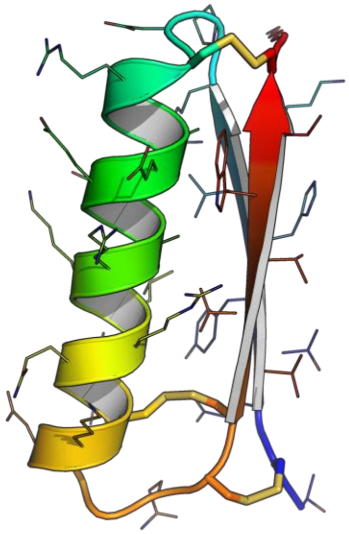


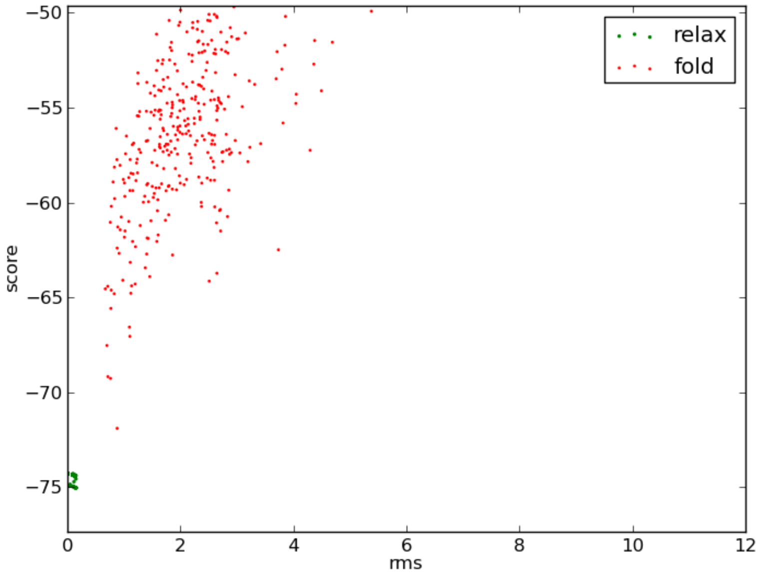


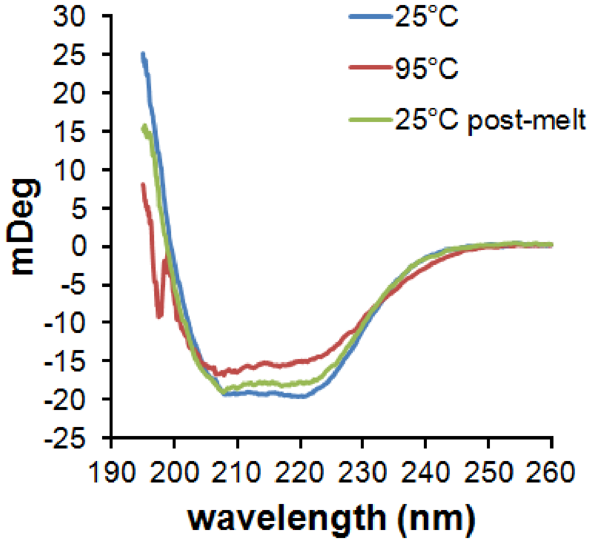


EHE_1.0_12

amino acid sequence:
TCETYHVKRPDCREAEEEARKLRQECKDRGQCCTVTWTCK





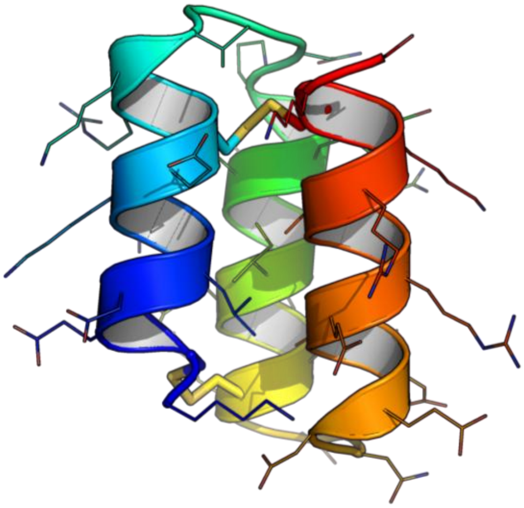


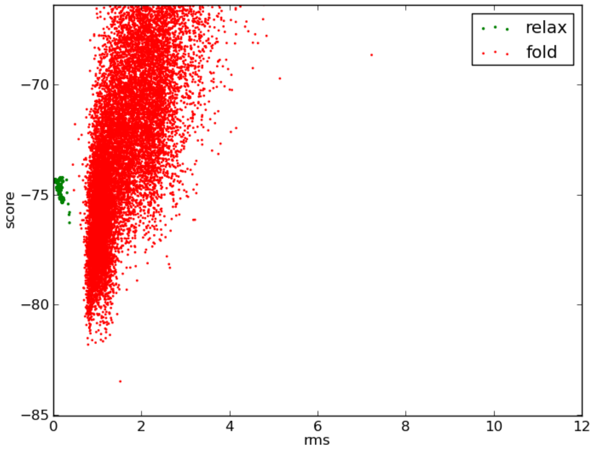
Round 6

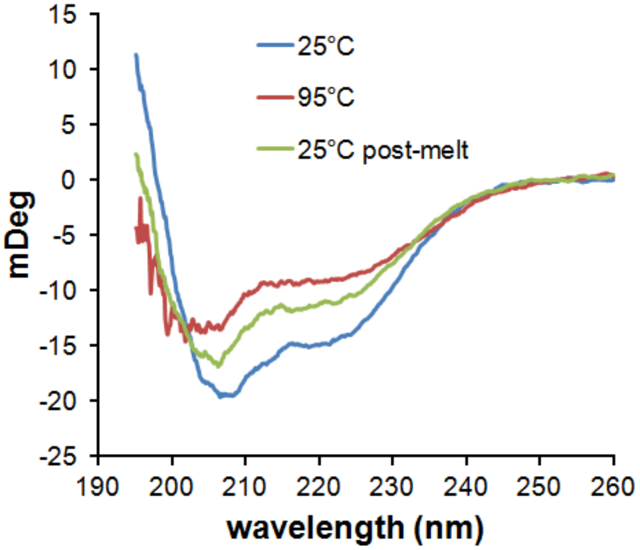
To test designs from this round, synthetic genes coding for designed proteins were cloned into pCDB180 and expressed from BL21*(DE3) (Invitrogen) *E. coli*. Designs were purified by IMAC from conditioned bacterial medium, and the fusion protein was removed using SUMO protease. Protease digestion is inhibited by a proline at the P1' position (first residue of the design). To facilitate removal of the fusion protein, an alanine residue was added to the N-terminus of any design beginning with a proline during generation of the synthetic gene.

HHH_4.0_01

amino acid sequence:
KCEDLRKECRKVGGNPEYEKRIEKMCRDGNDEEAERVARKCKS

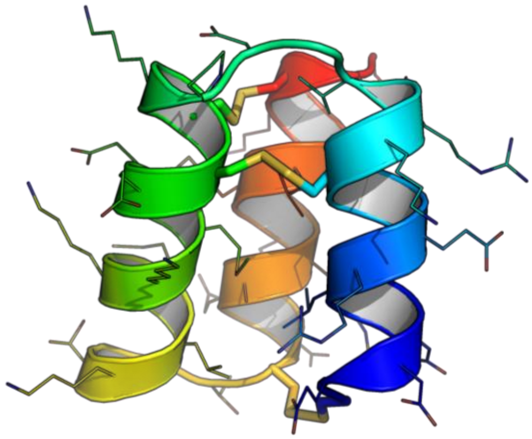


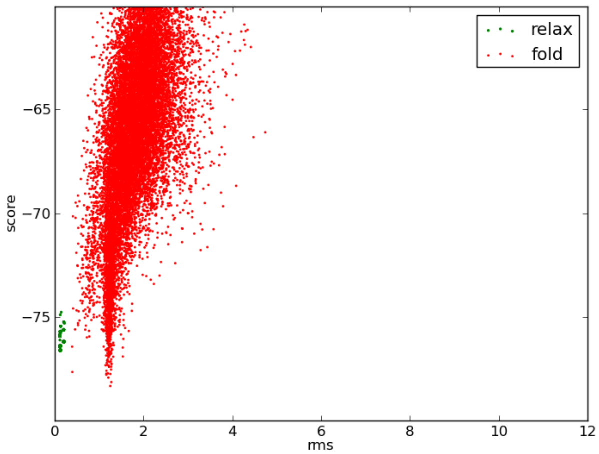


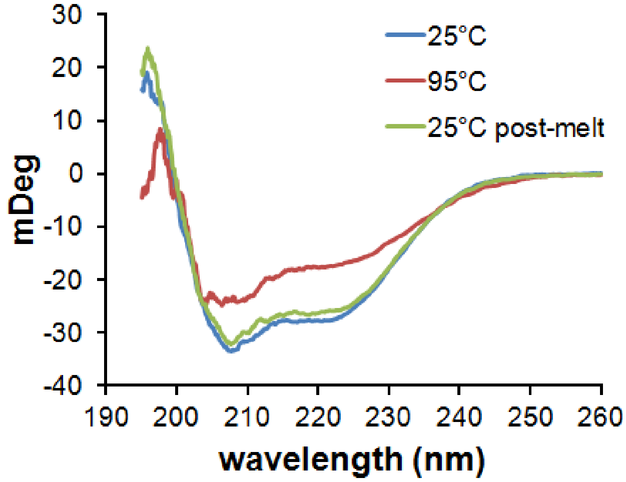


HHH_4.0_02

amino acid sequence:
CEDELRELCKRVGDPKCCEEMKKMLKTGTCDEARKMLEKCLK

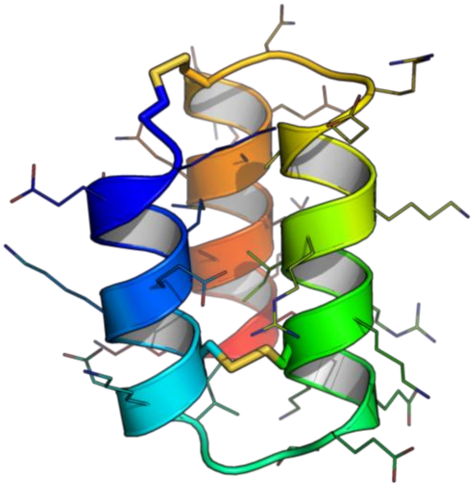


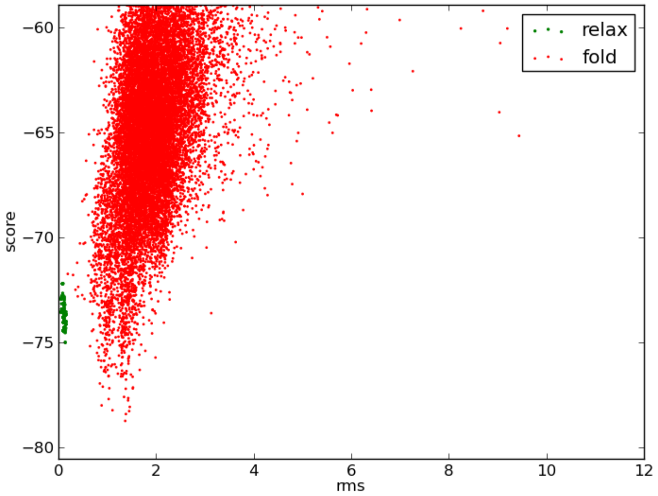


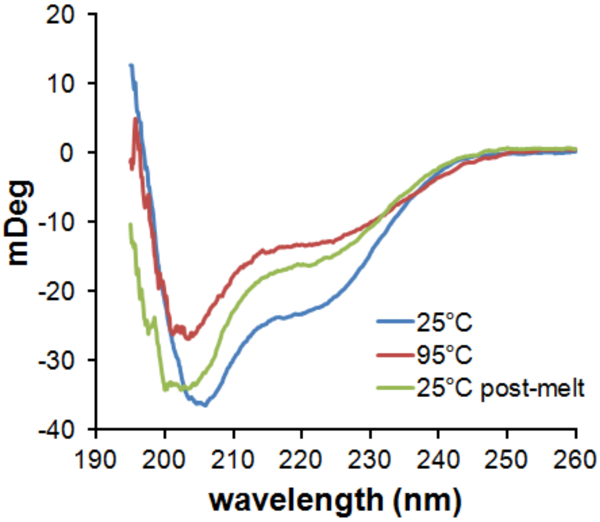


HHH_4.0_03

amino acid sequence:
CKEEMEKVCKEIGTEEKCKRIRKVAERGNCEEAQREAKRMKS

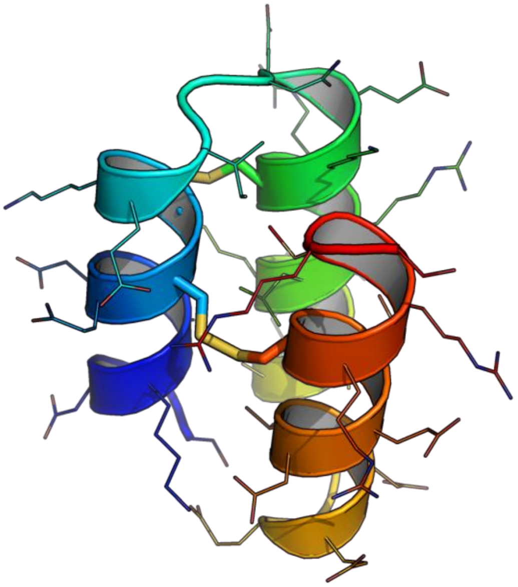


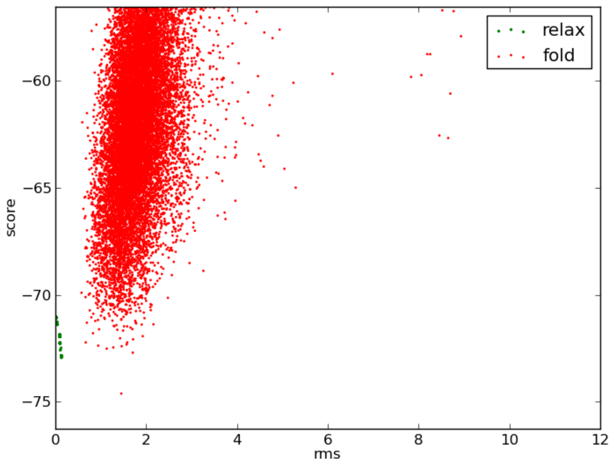


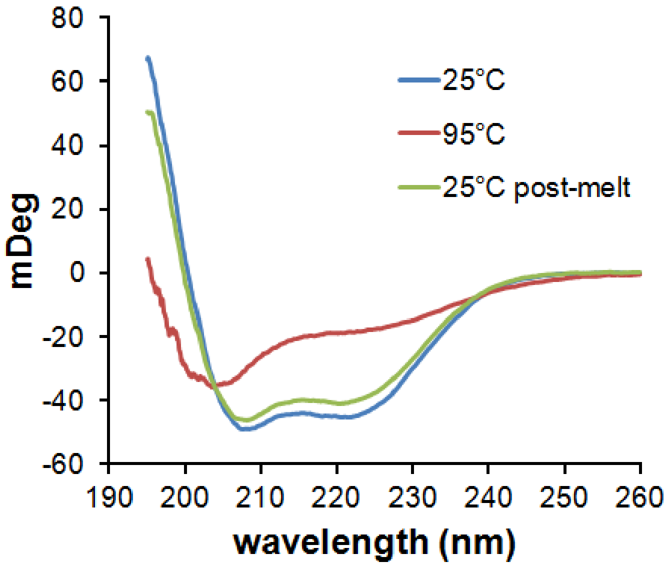


HHH_4.0_04

amino acid sequence:
SDDKAEQCCKEIGNEEKCRRLKEVAKDGSEEEVDEMCRRMRS

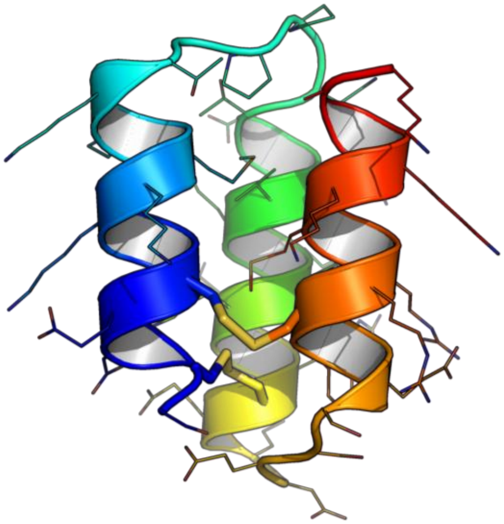


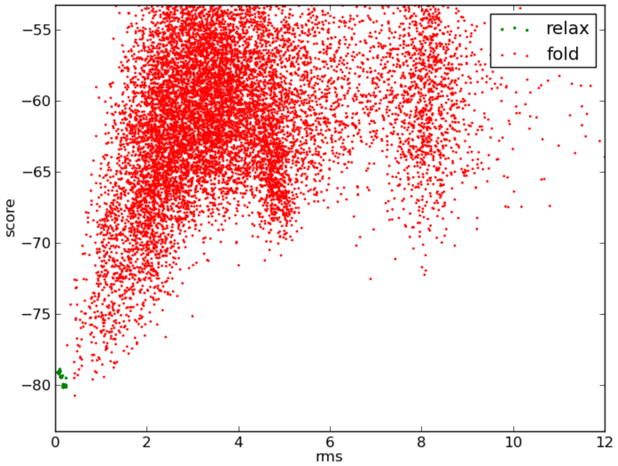


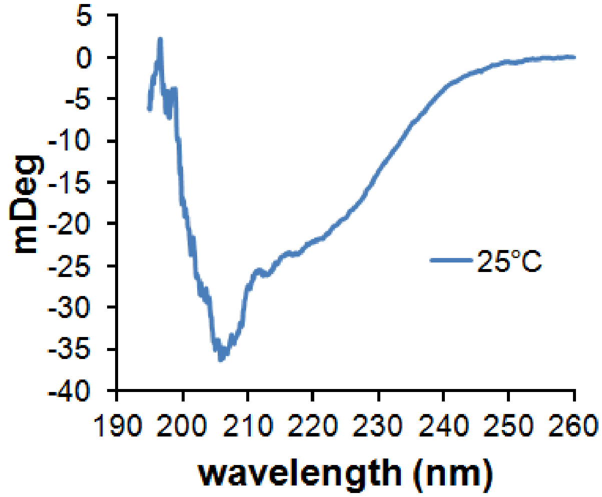


HHH_4.0_05

amino acid sequence:
SCDECYKKMQKTGPPNTEKVKELWKRCQKDESSEYCRRMKKMAK

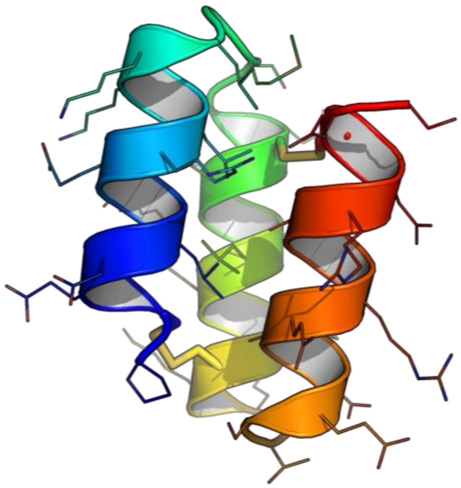


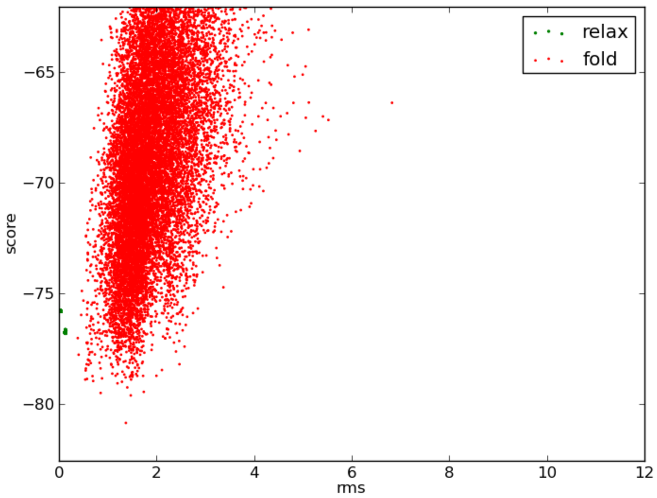


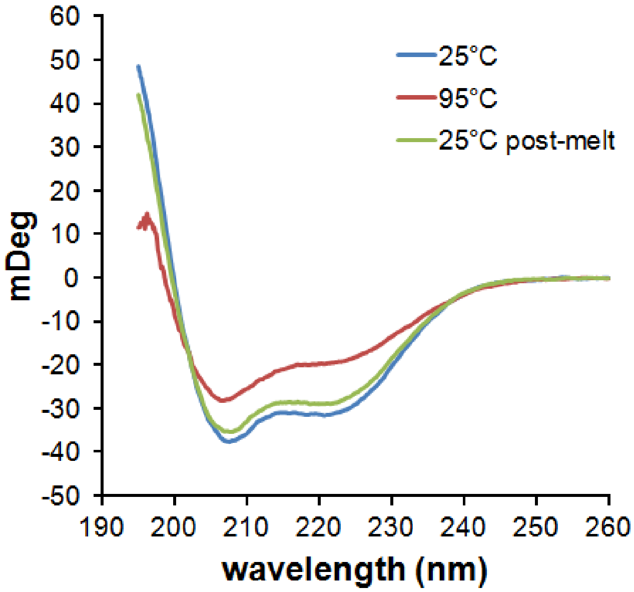


gHHH_06

also known as: HHH_4.0_06
amino acid sequence:
PCEDLKERLKKLGMSEECRQRLEKMCKEGTSEDAERMARNCES

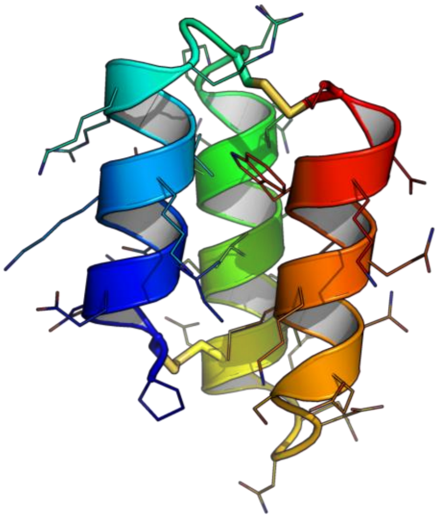


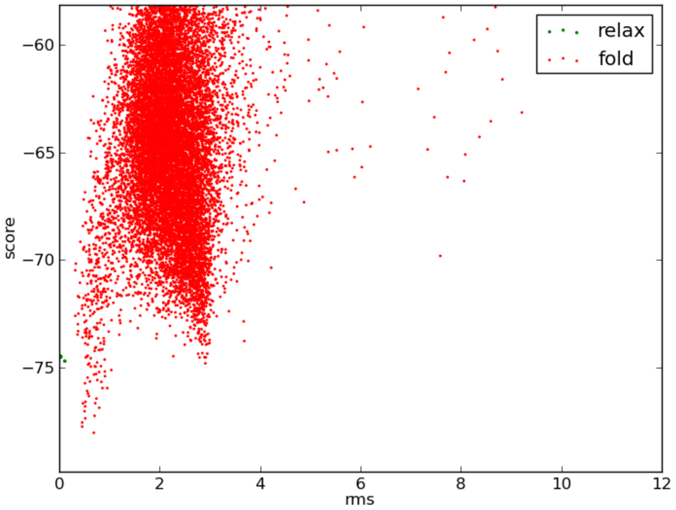


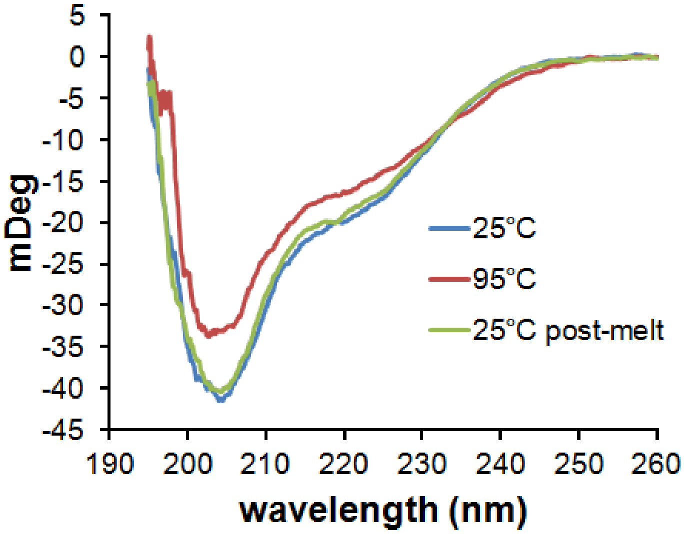


HHH_4.0_07

amino acid sequence:
PCDEIEKKVRKRGCDPQVEKEVRRVCEEQNDSEQMKQIWKDCS

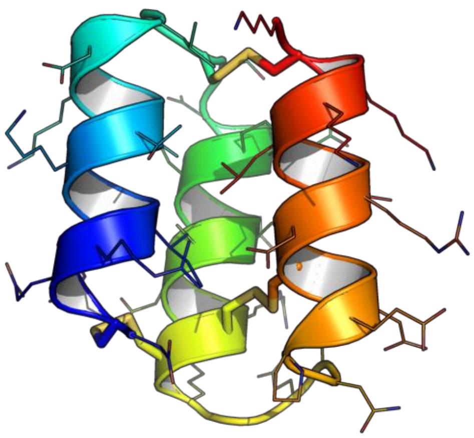


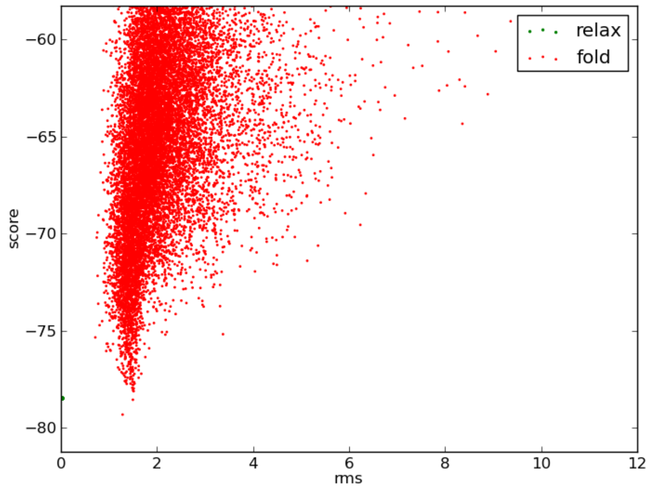


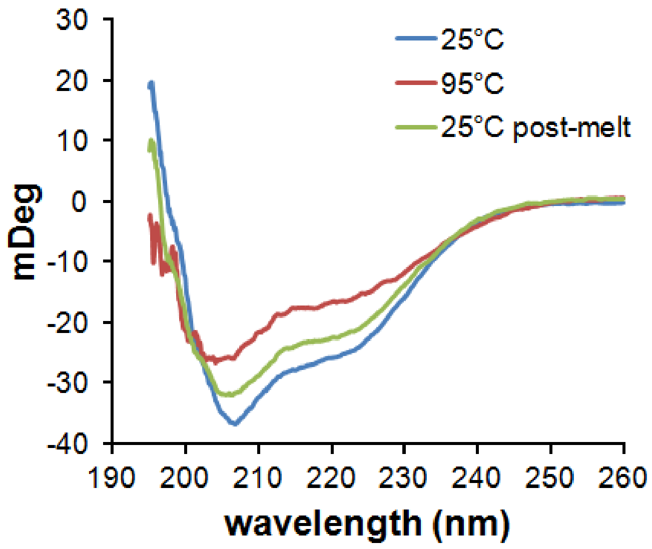


HHH_4.0_08

amino acid sequence:
DCERIRKTVKDLGCSDEMKEKAERCCRGEYNPEECDRELKKCK

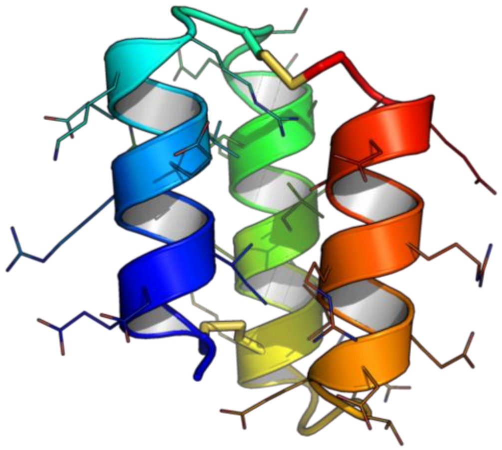


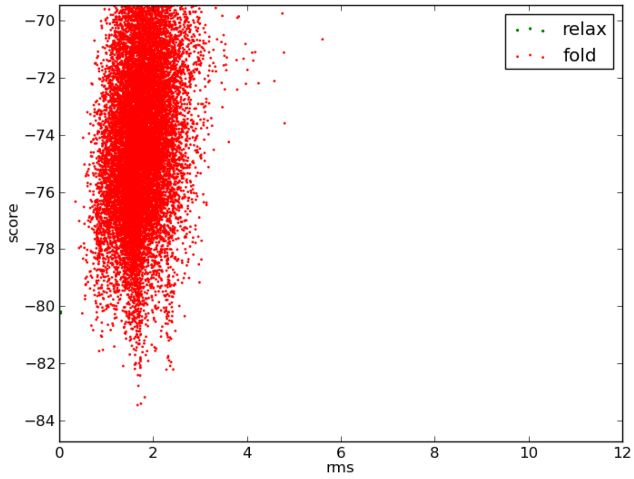


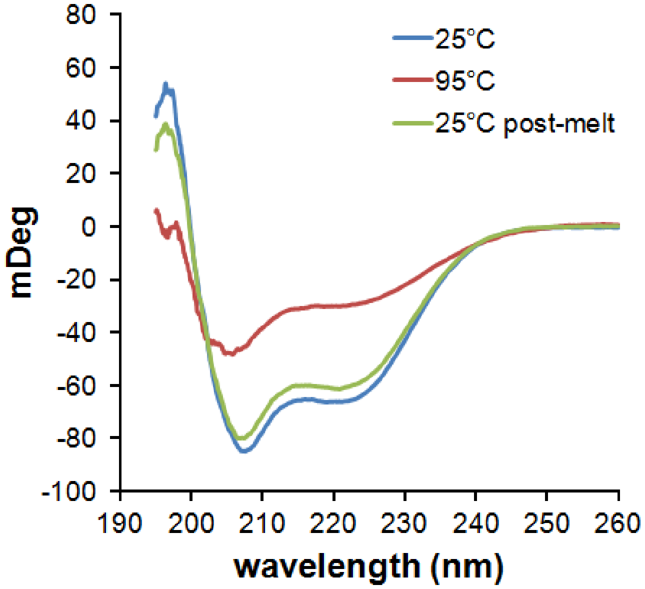


HHH_4.0_09

amino acid sequence:
GCEDIDREVEKRGCTEDARRELQKLCKNGQTEDEIRRAADELC

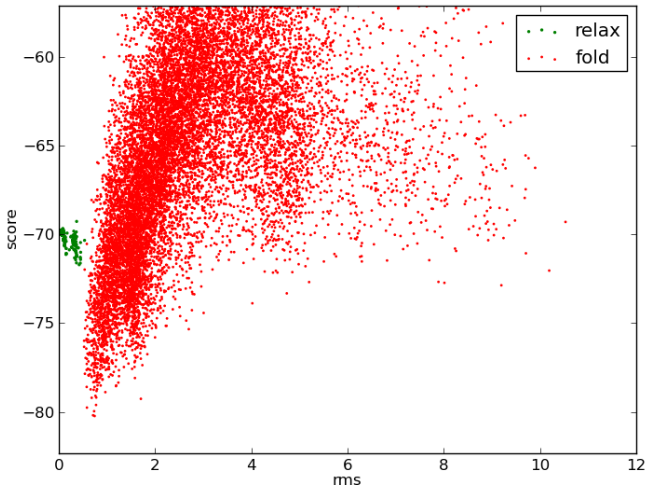


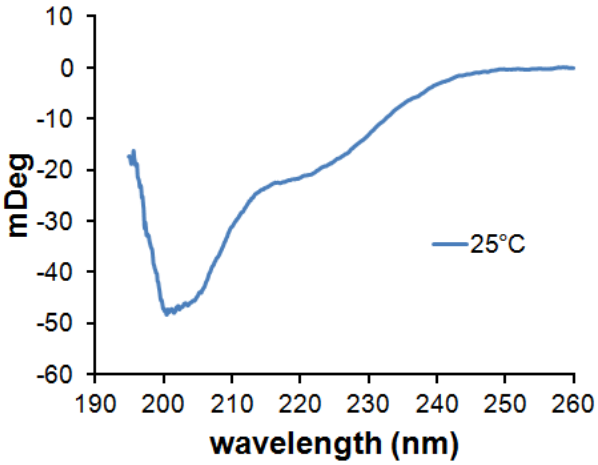




HHH_4.0_10

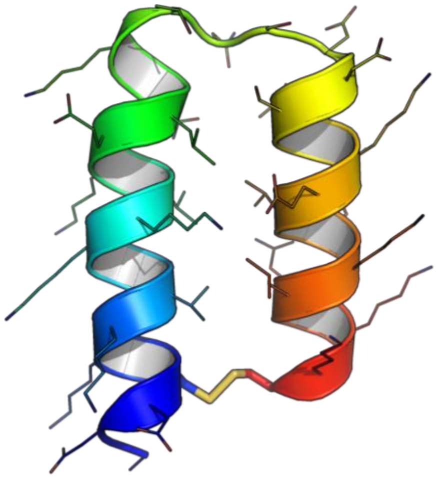
amino acid sequence:
PCDRCARELEEAYPNNPEVNEEARRVKKNCTDEMCKEVKKMKKR

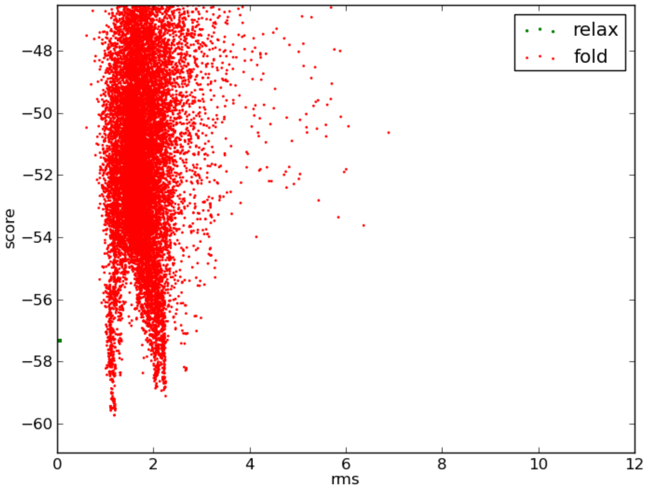


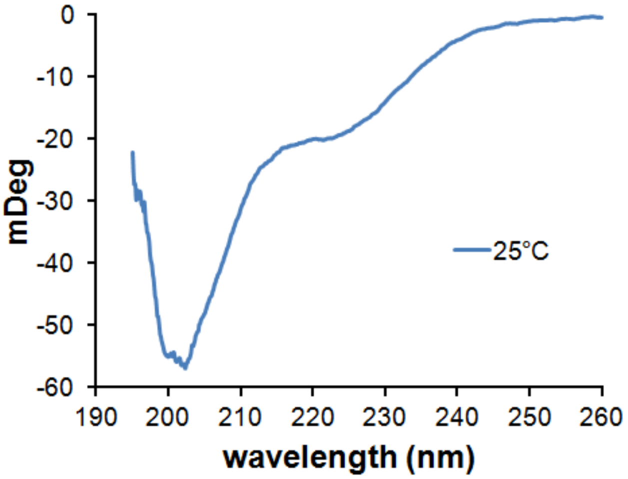


HH_2.0_01

amino acid sequence:
ADDCKKVQKKVKELNKTNSDDSLKEVKKLQKKCA

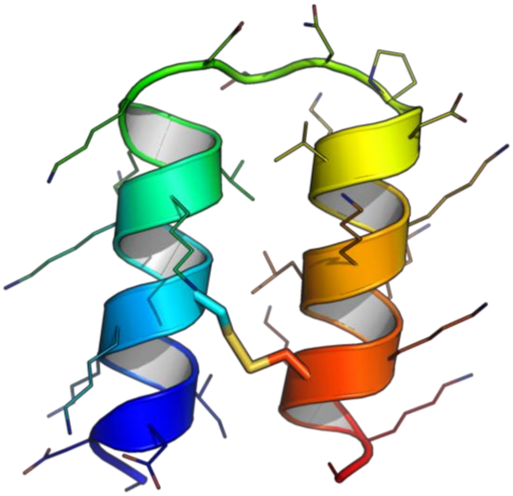


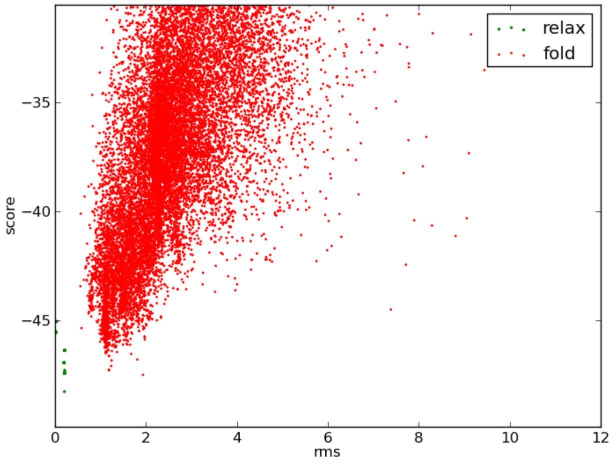


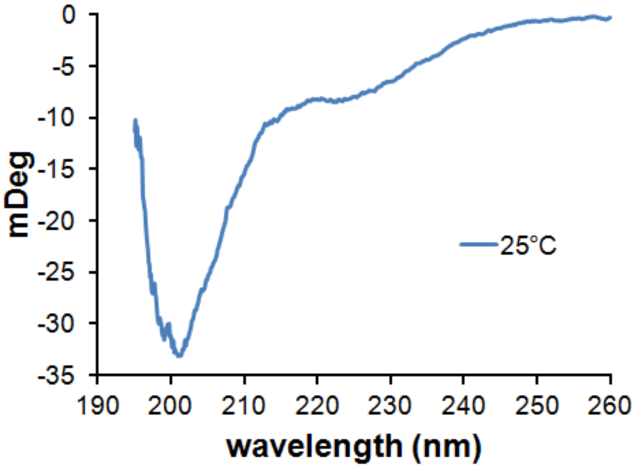


HH_2.0_26

amino acid sequence:
ADDIKKCEKKVRKDSNPDVKKKLKKCKKA

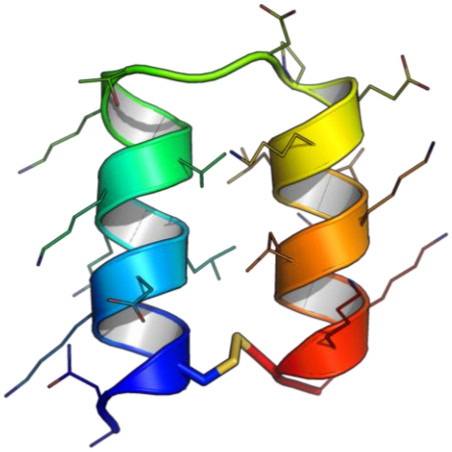


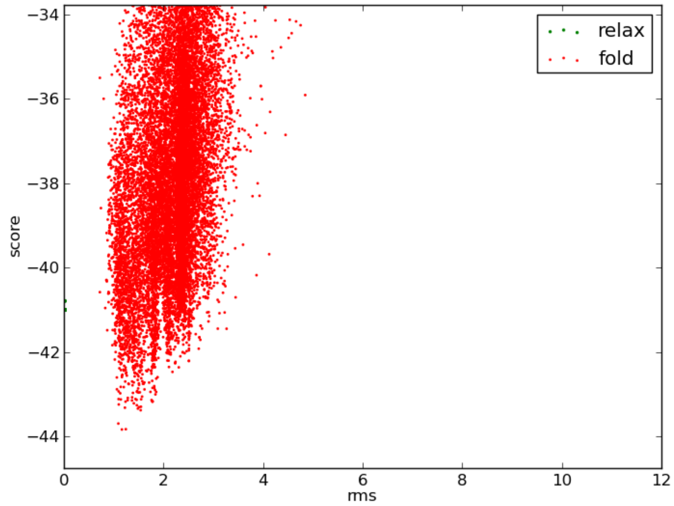


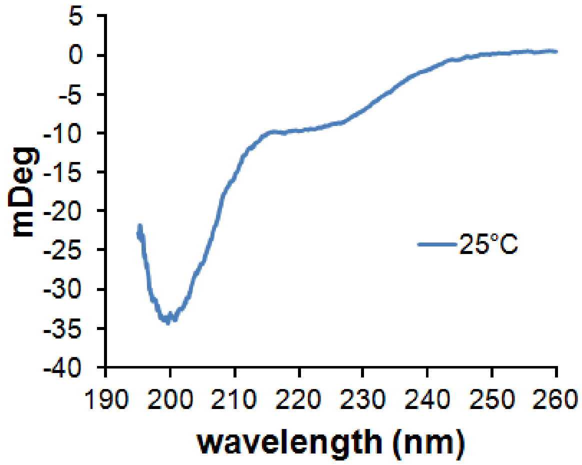


HH_2.0_29

amino acid sequence:
AQCEKDLKKVKKTGDPEKLDKIRKKCA

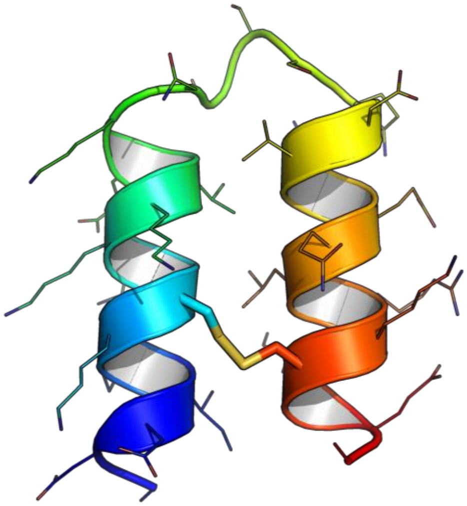


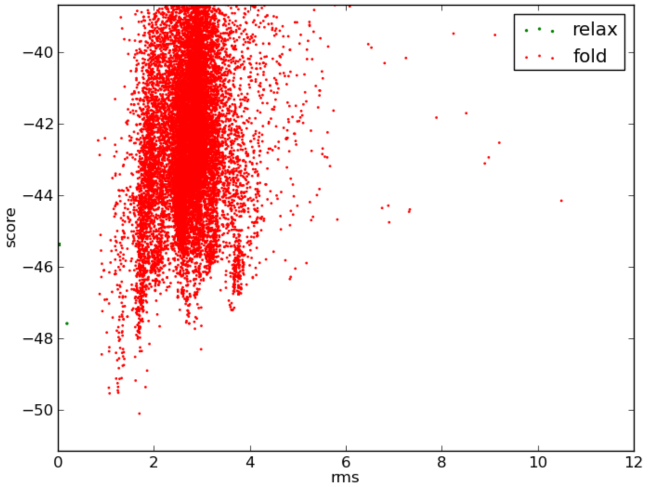


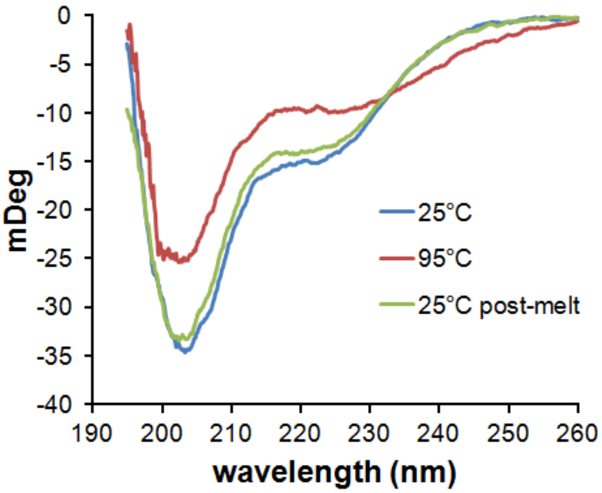


HH_2.0_34

amino acid sequence:
ADDIEKCRKKVEKNSSSQDVQEQLRKCKEA

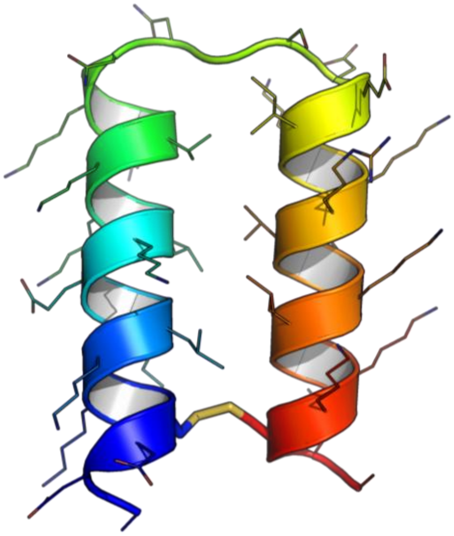


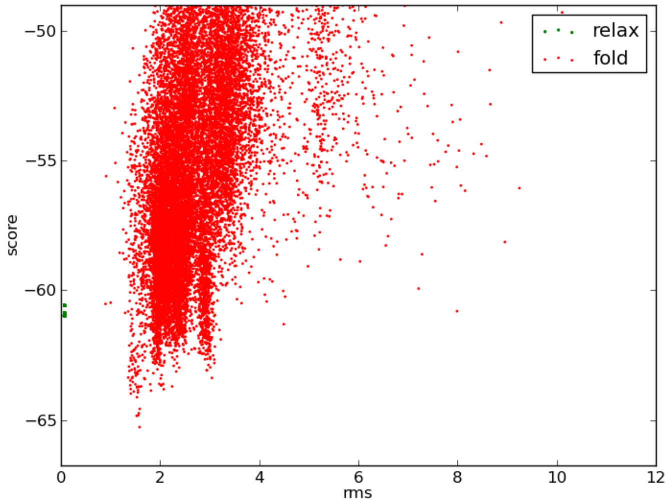


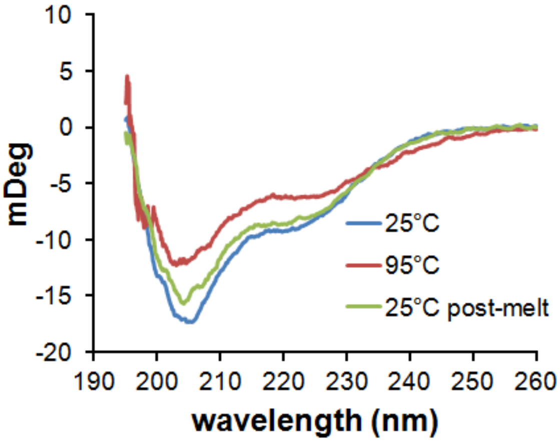


HH_2.0_35

amino acid sequence:
ADDCKKLKEKLKKVKKNNGSDEIKKRVEKLRKKCEA

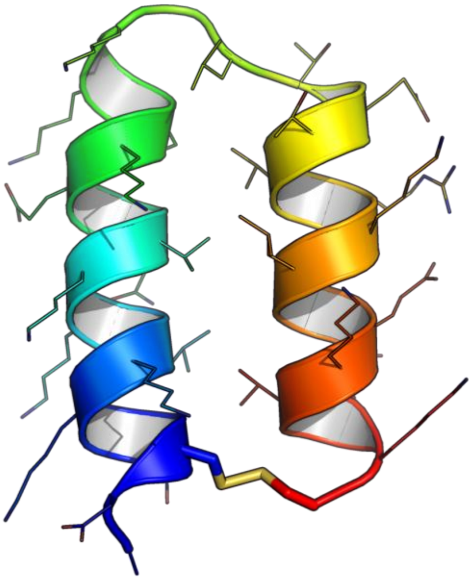


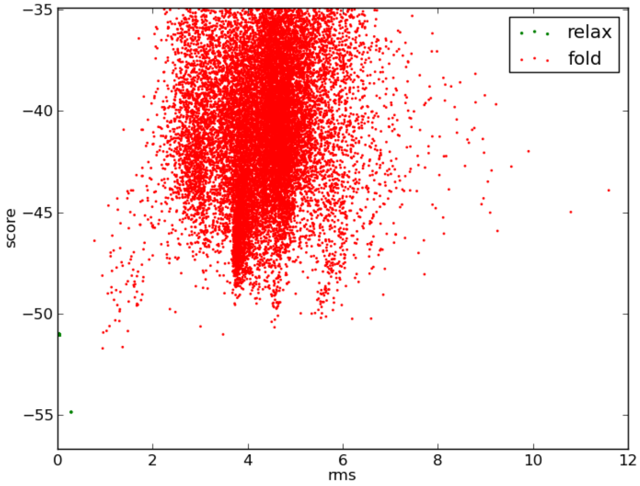


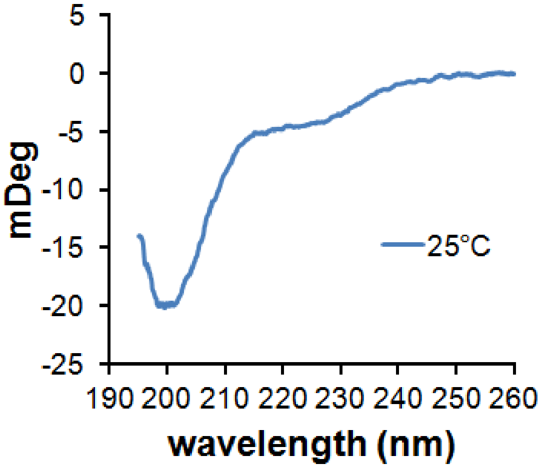


HH_2.0_36

amino acid sequence:
ADCDKKLKKVQEKSKKGLTETVRKLKEKVEKC

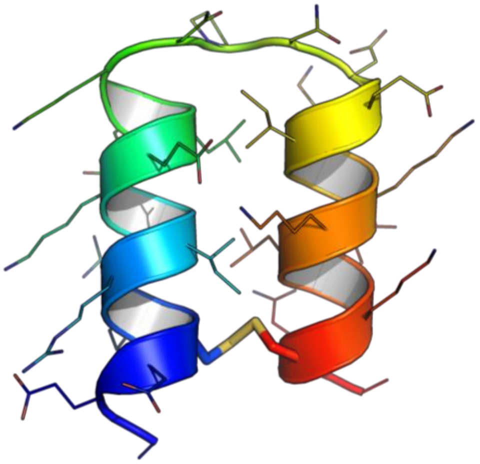


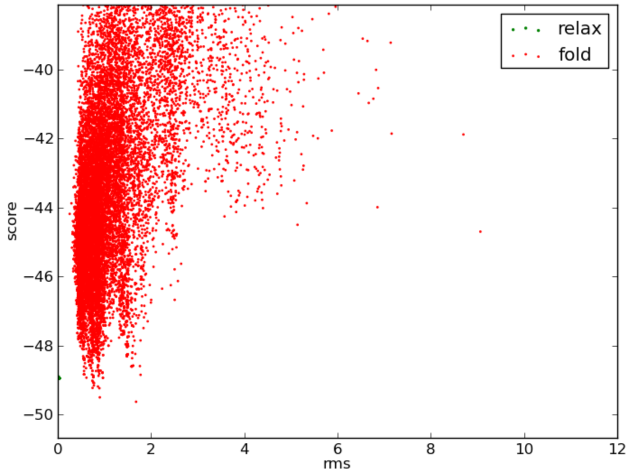


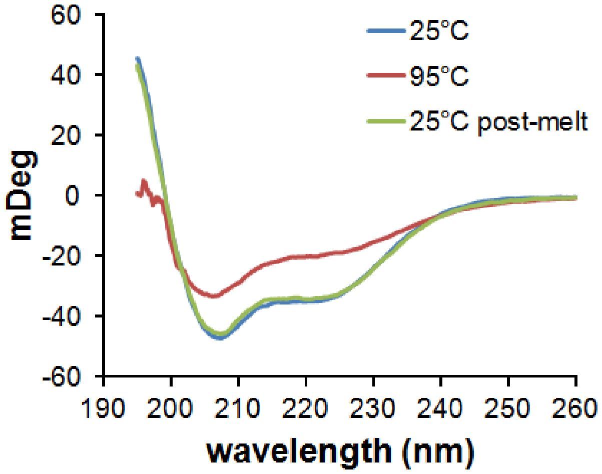


gHH_44

also known as: HH_2.0_44
amino acid sequence:
AEDCERIRKELEKNPNDEIKKKLEKCQA

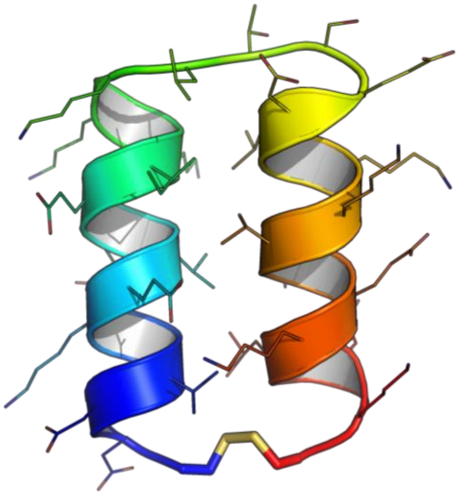


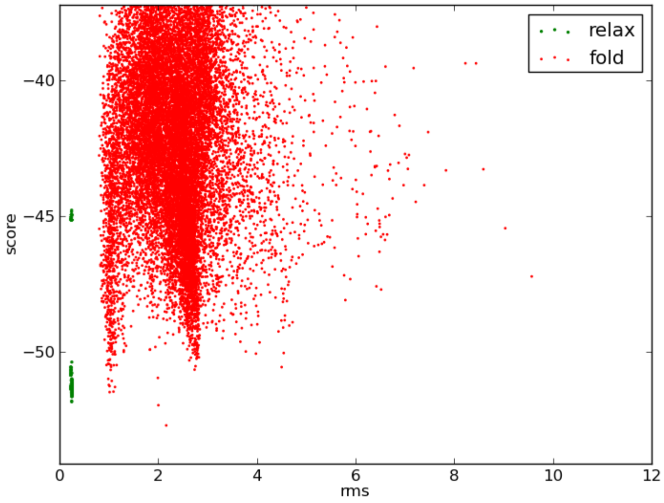


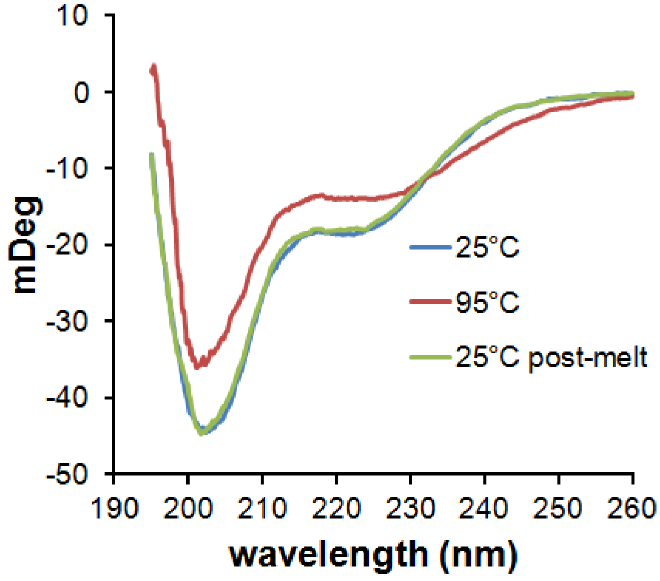


HH_2.0_45

amino acid sequence:
CDDVKKEVEEIKKKLTSEDLKKVQEKLDKC

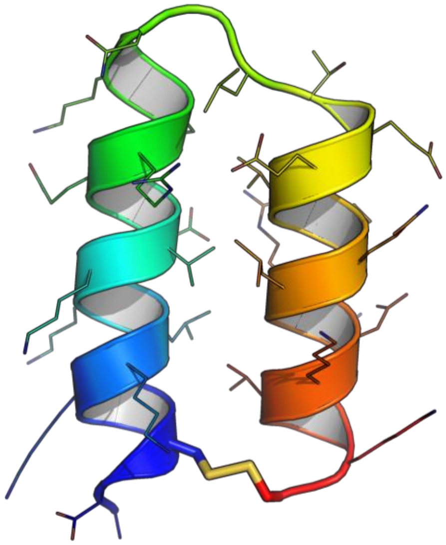


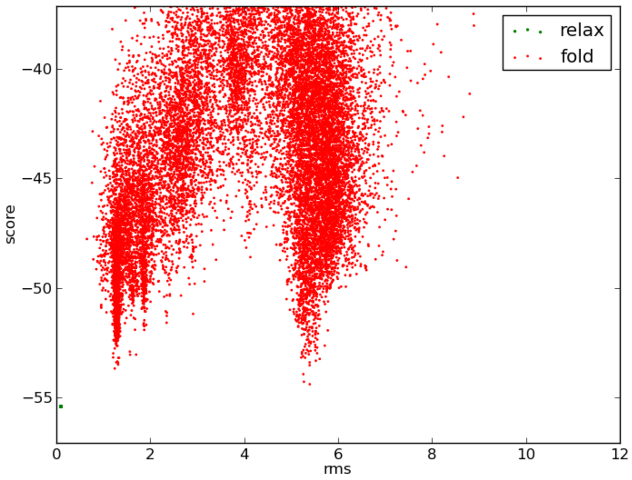


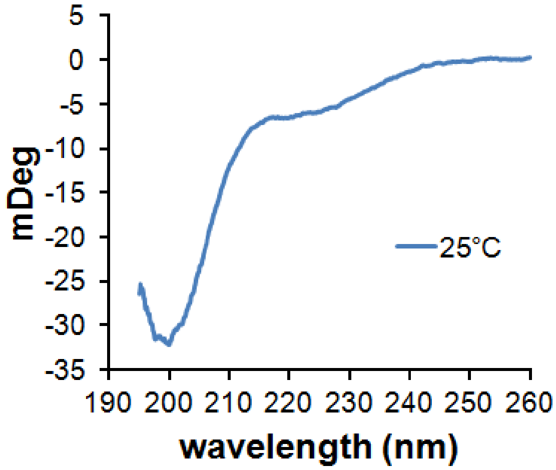


HH_2.0_47

amino acid sequence:
ADCDKKLKKVEERSKNGLTEEVQQLRDKVKKC

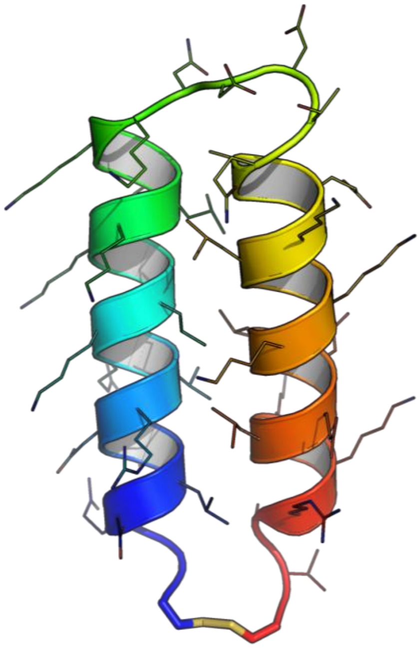


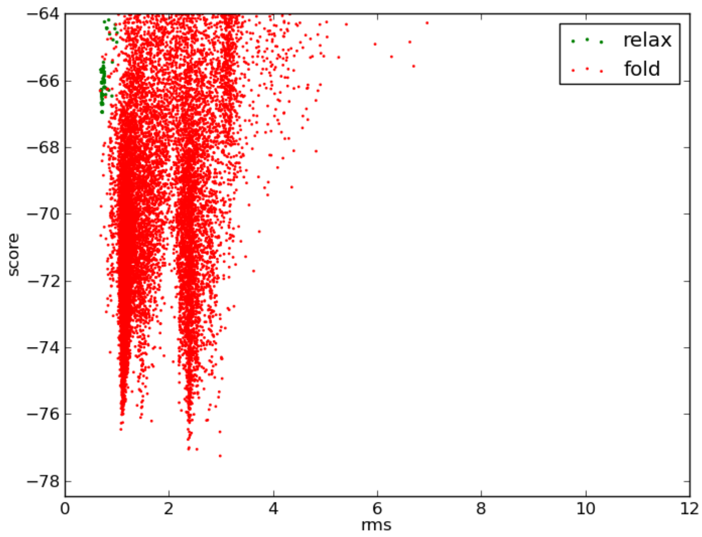


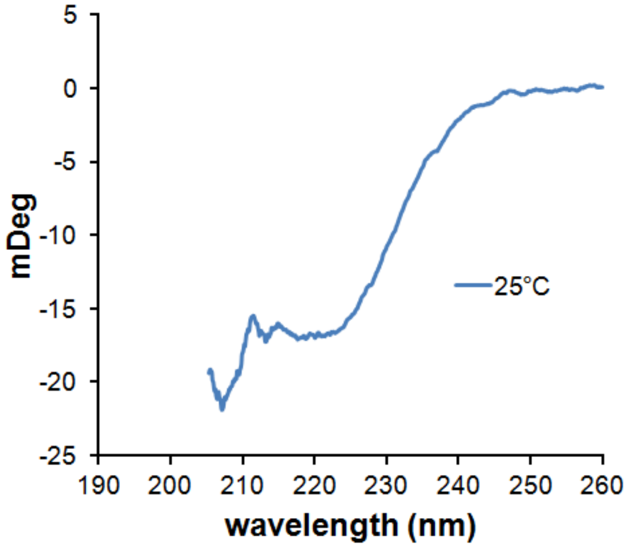


HH_2.0_48

amino acid sequence:
CAQELEDRVRKLEKKLRKKNDDTQVEKLQKKLDELKKRAVC

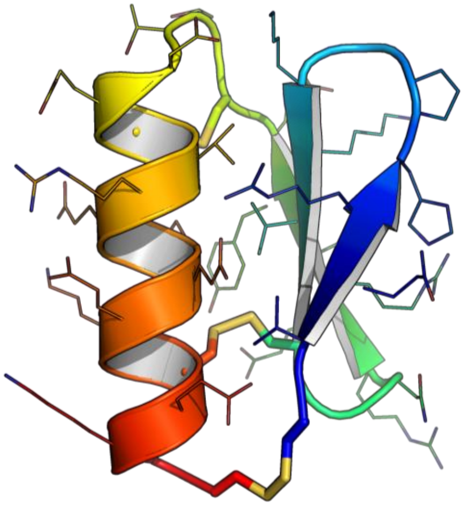


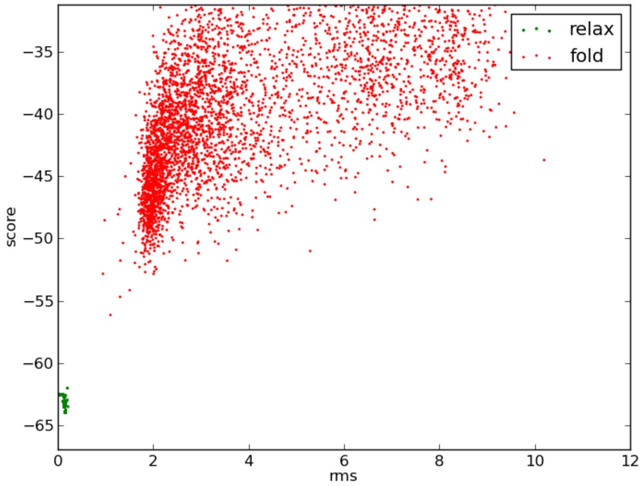


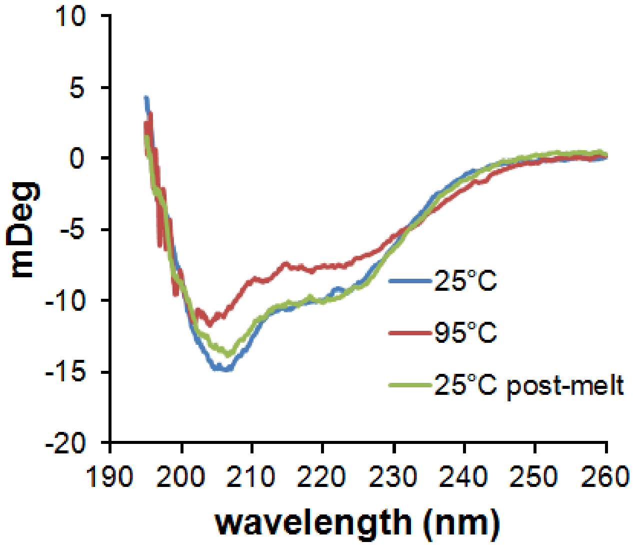


EEEH_3.2_01

amino acid sequence:
CVQRHPGKKVRCGNREEYQCTTDECVREMEEKCEKRC

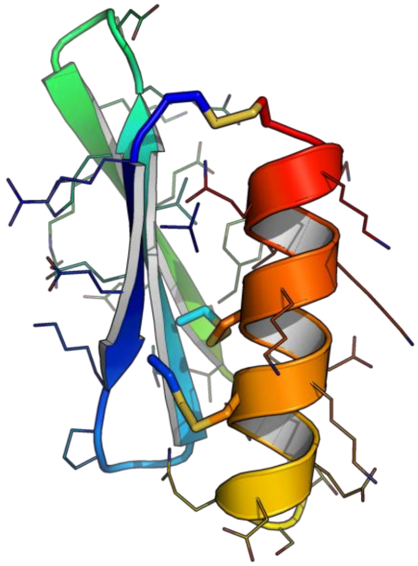


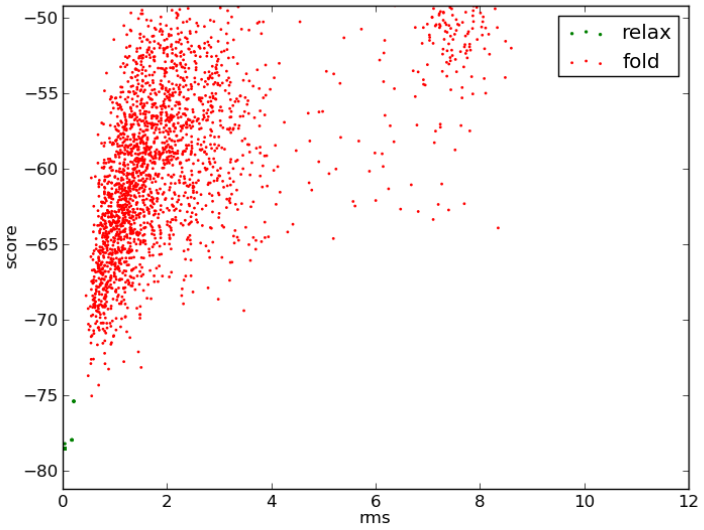


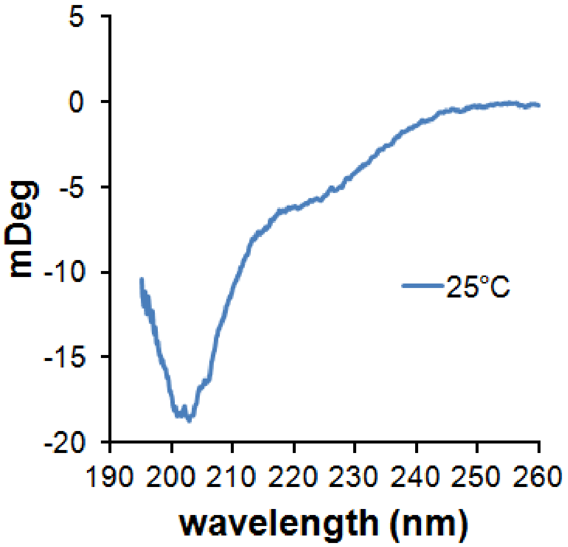


EEEH_3.2_02

amino acid sequence:
CRVECKPGGTCEVHRDSGKREEYTFPTSQDEVCKECKKLQKKC

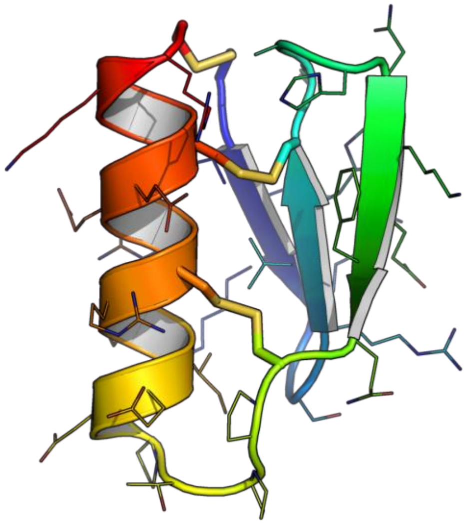


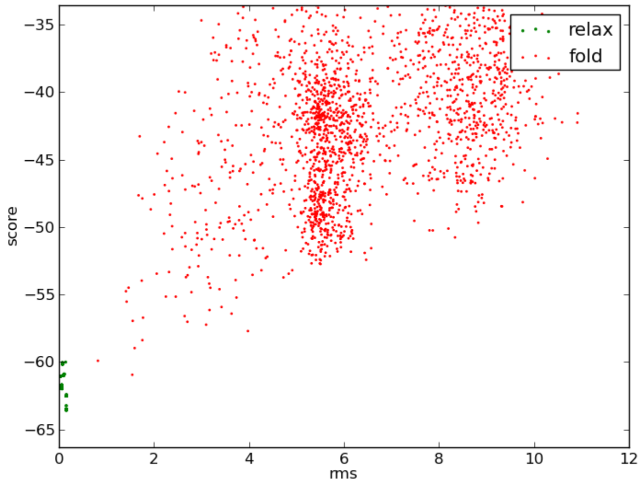


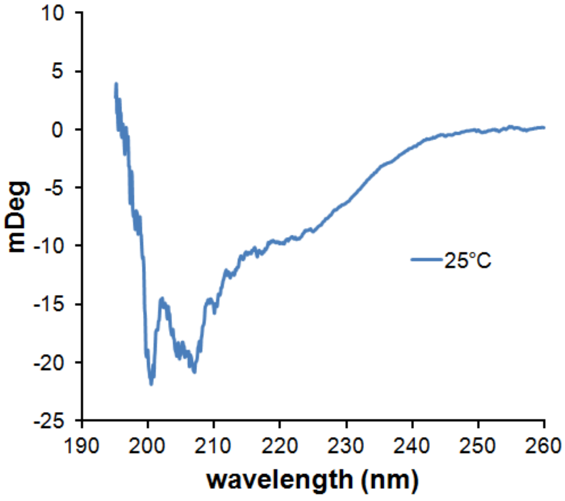


EEEH_3.2_03

amino acid sequence:
CVEKRGSRVHCKAHNKEFQCPPTPDEIERCREECEKRC

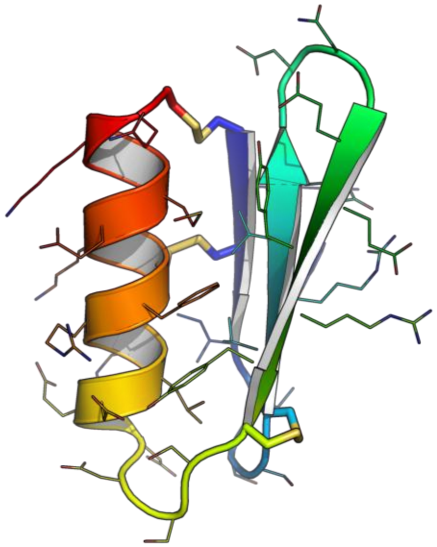


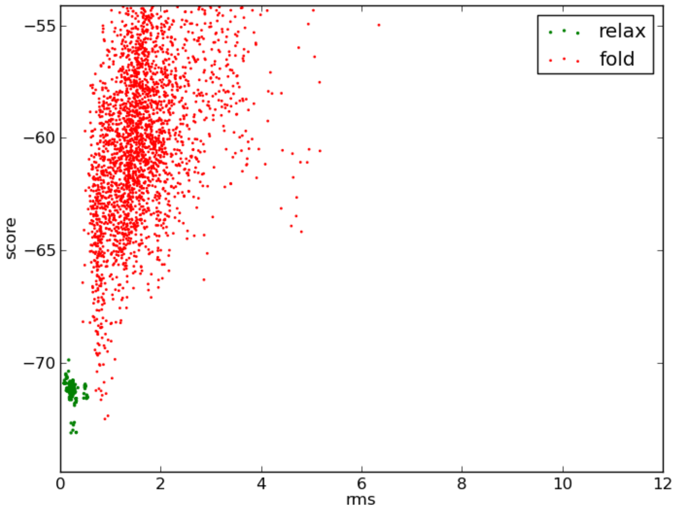


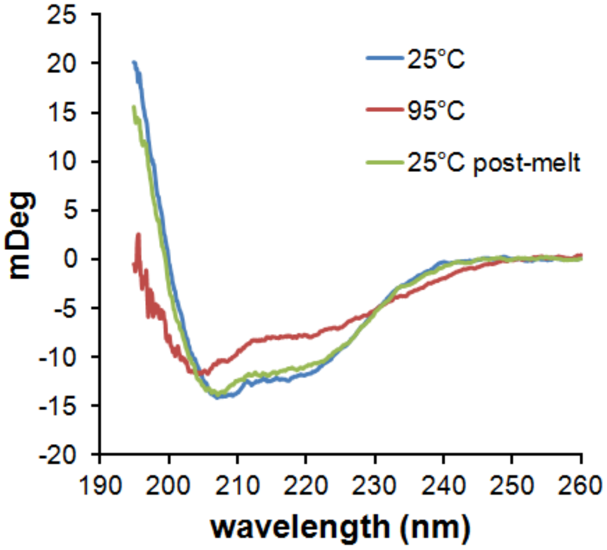


gEEEH_04

also known as: EEEH_3.2_04
amino acid sequence:
CRCHITSSCVRVEGDNGEEYRYCSSDEEDLRRFCKEMQKQC

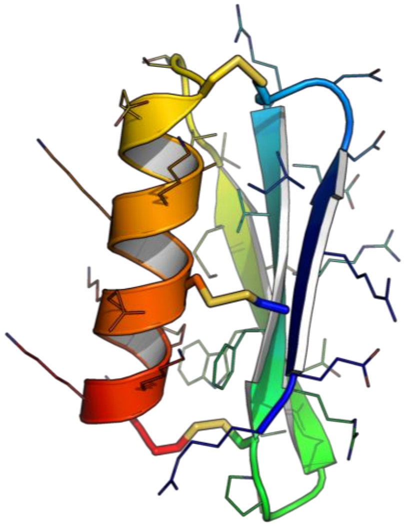


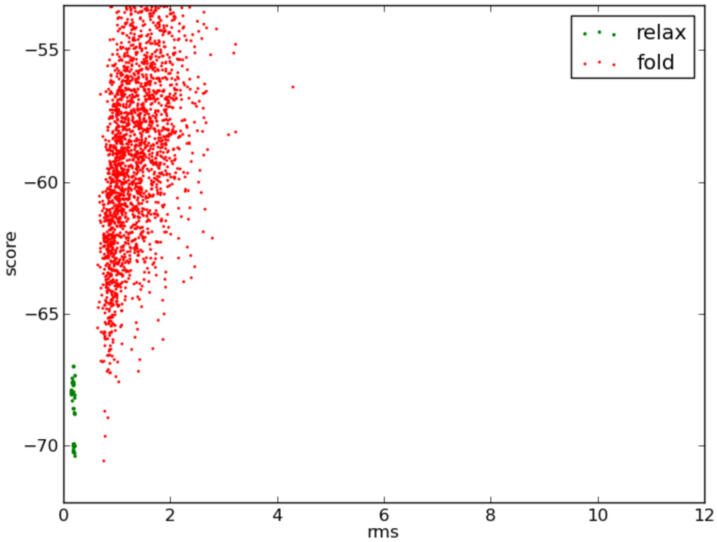


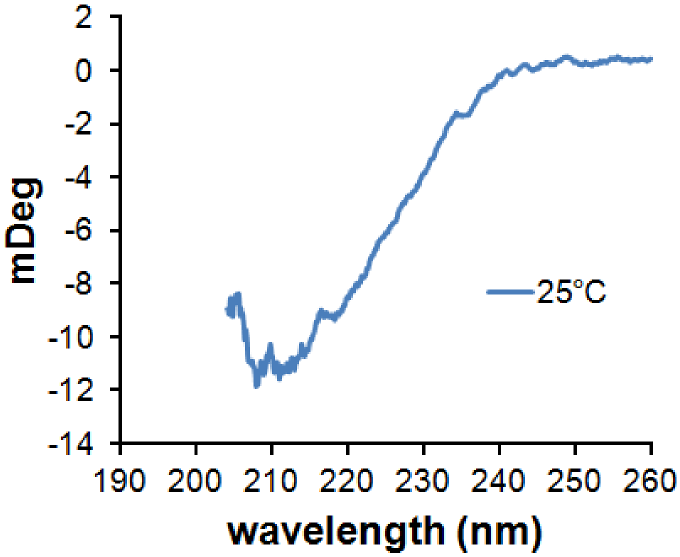


EEEH_3.2_05

amino acid sequence:
RECRINNCREVRFRCPSGQTWTMTVTSCEEAKKMCEKMKKQC

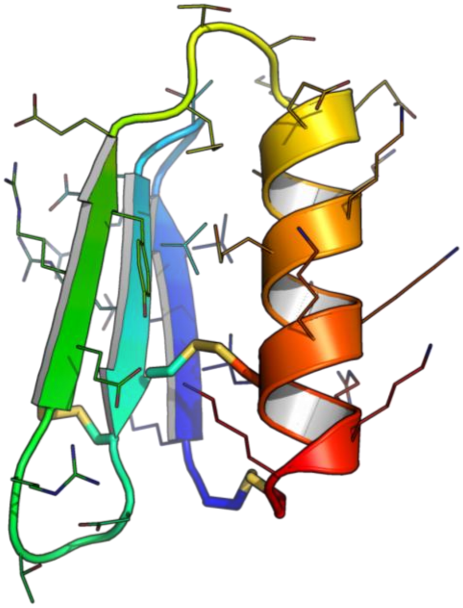


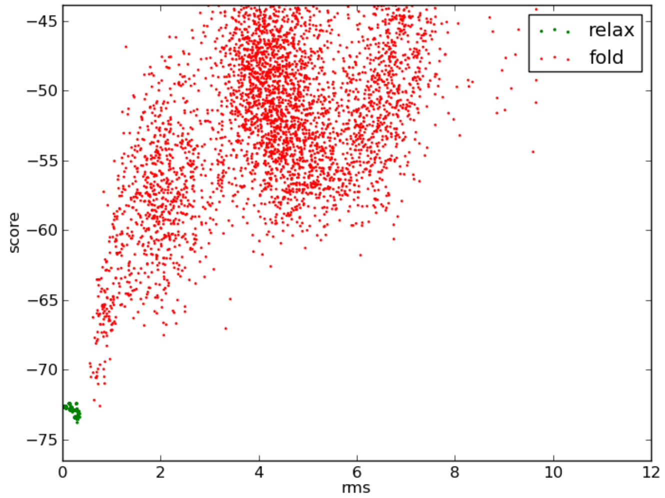


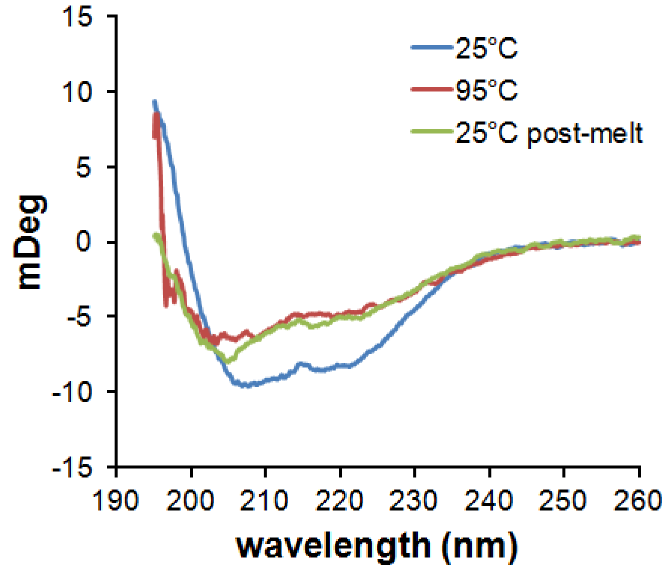


EEEH_3.2_06

amino acid sequence:
CRIEVRGTEVRCCDGTRCERYEMTSKEEAKKMEKKCRKKC

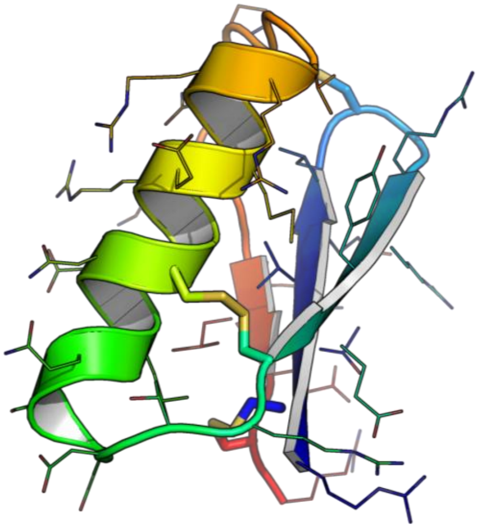


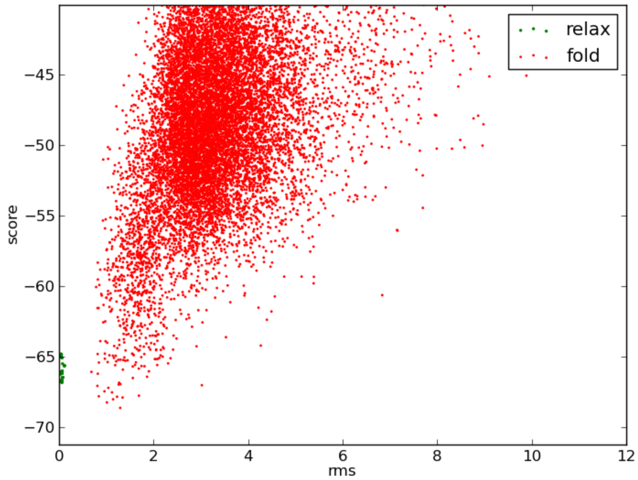


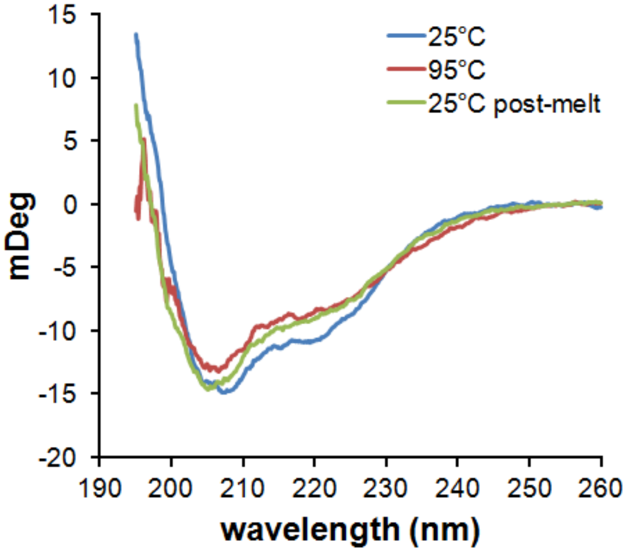


EEHE_2.2_01

amino acid sequence:
RCTVELCGRRYECRTDESQLENCAREMQRRVGCPQKPRLECR

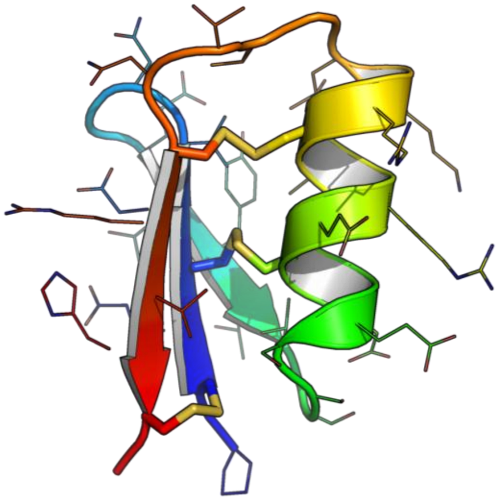


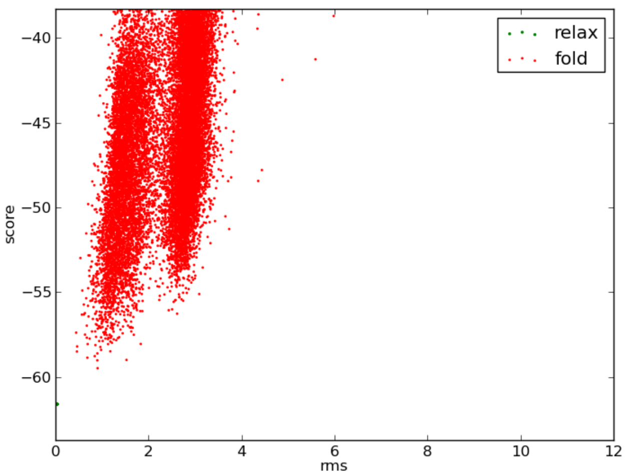


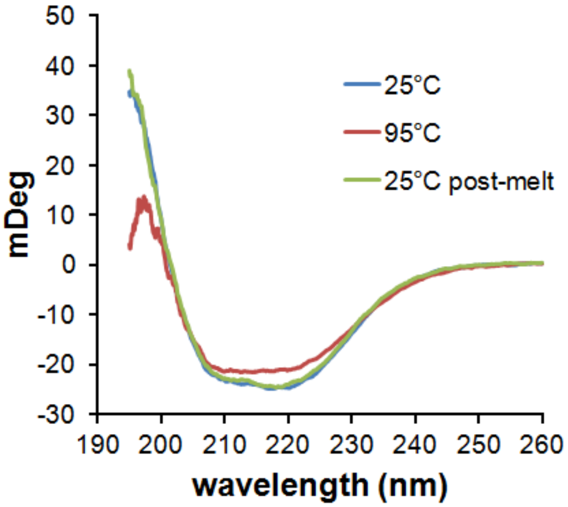


gEEHE_02

also known as: EEHE_2.2_02
amino acid sequence:
PCECDVNGETYTVSSSEECERLCRKLGVTNCRVHCG

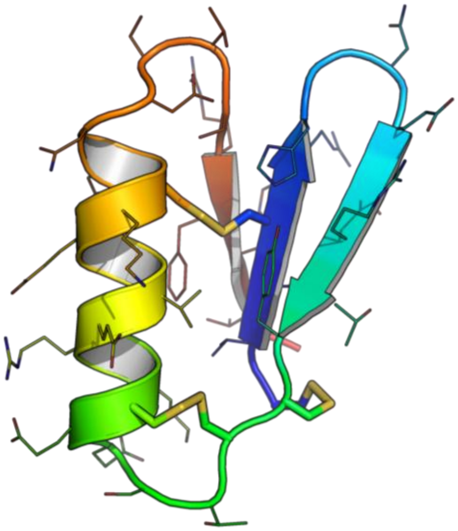


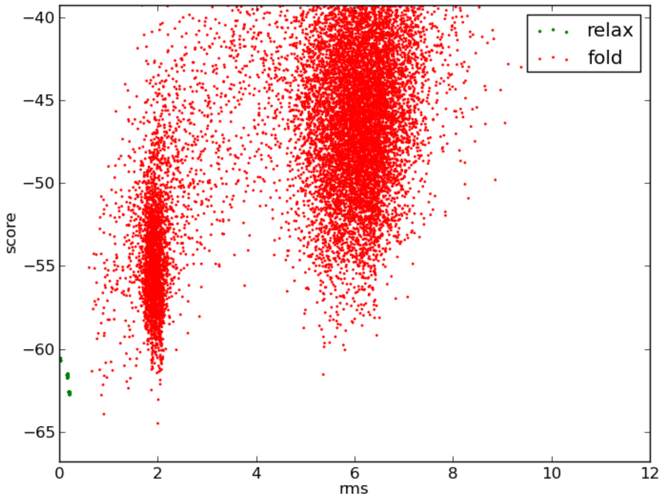


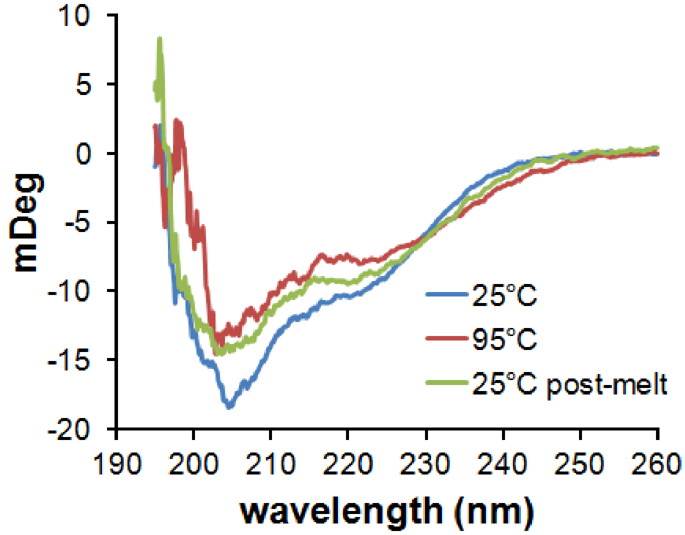


EEHE_2.2_03

amino acid sequence:
CVRCRHGNEERTYCCTSEECKREVKEKCDNDSTSRFHTG

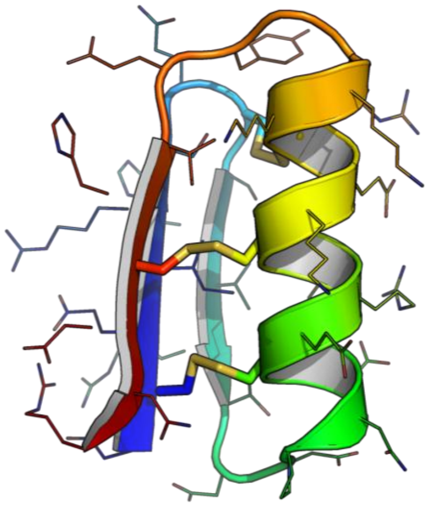


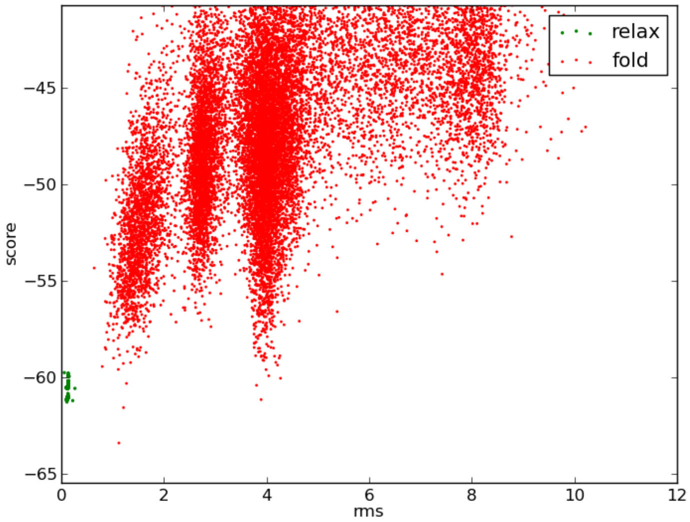


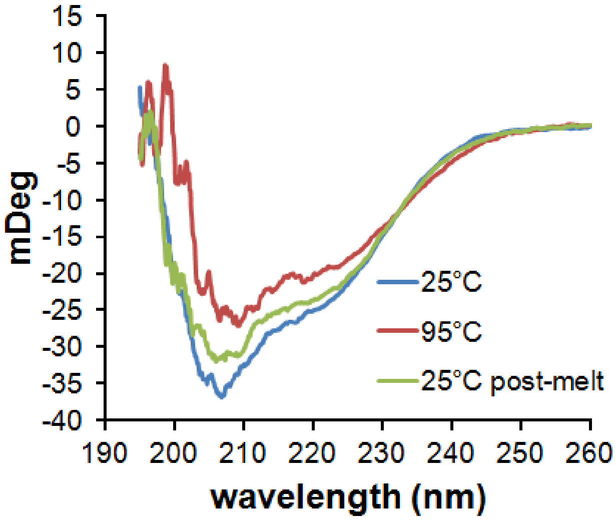


EEHE_2.2_04

amino acid sequence:
DCEIRSQCSHVRTDDPNECERICKECKKRGYEVHCDNR

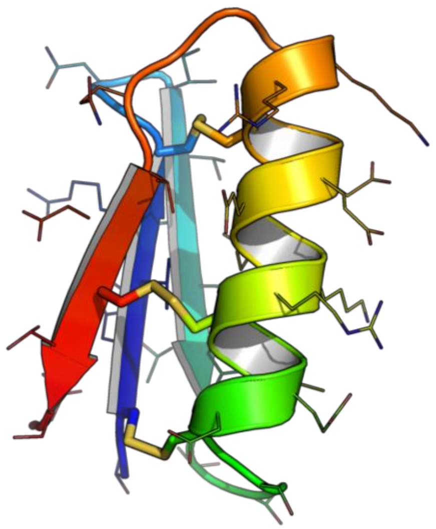


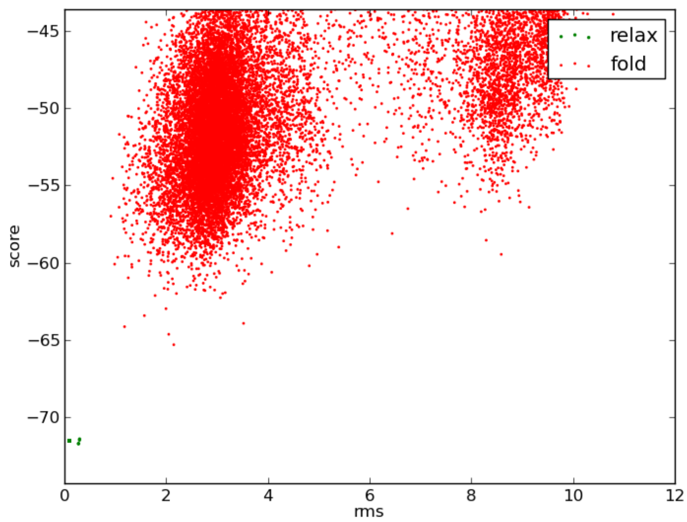


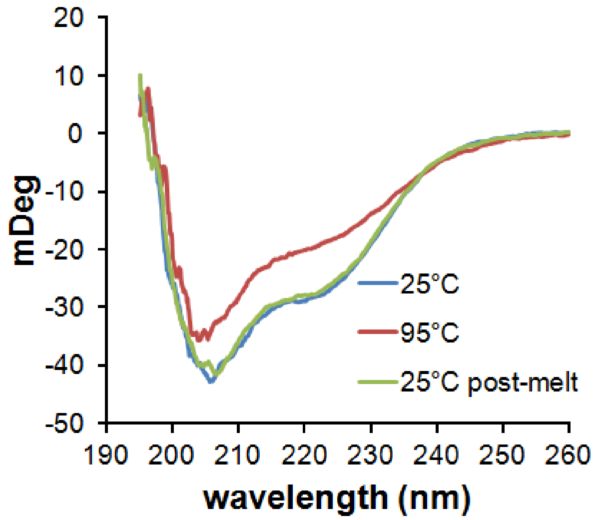


EEHE_2.2_05

amino acid sequence:
TCHVRCGNITEQTFTTGTCDEMCRKMEEECRKLGGQVDCTSL

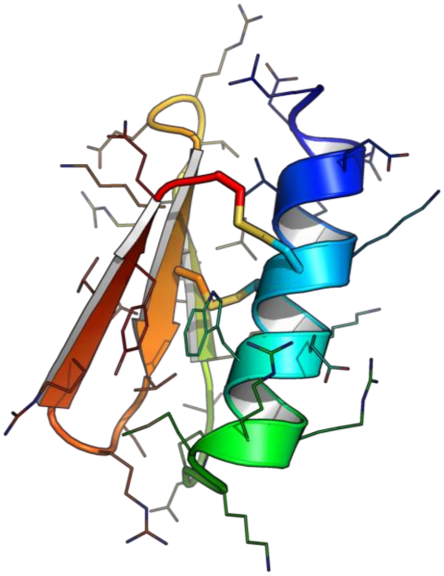


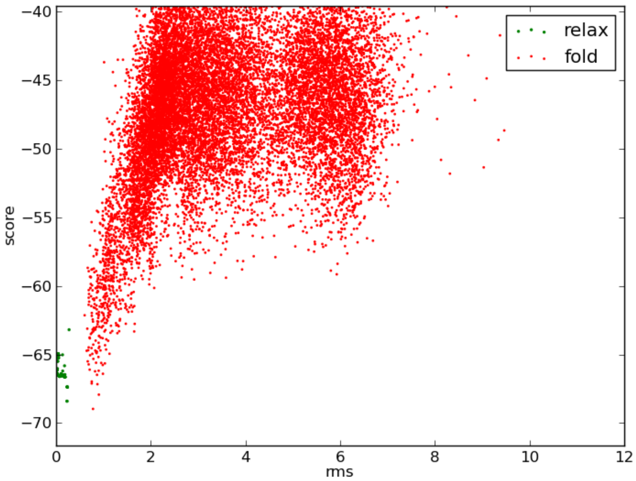


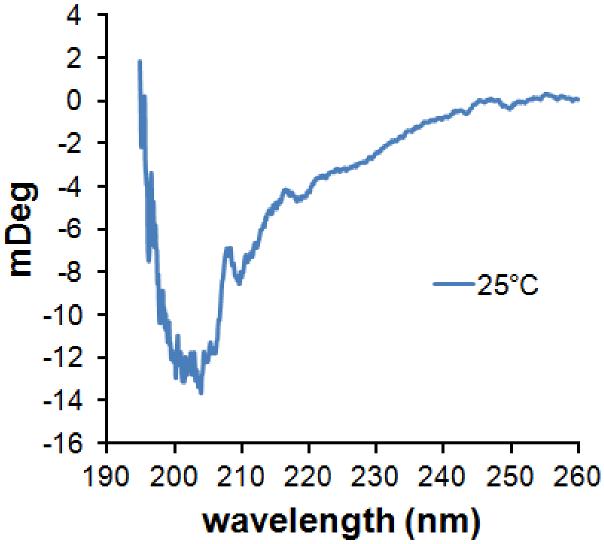


HEEE_2.2_01

amino acid sequence:
REEEVKKCCKEWHRRMKPDTFQVRTREGKCTVSRGRTYQC

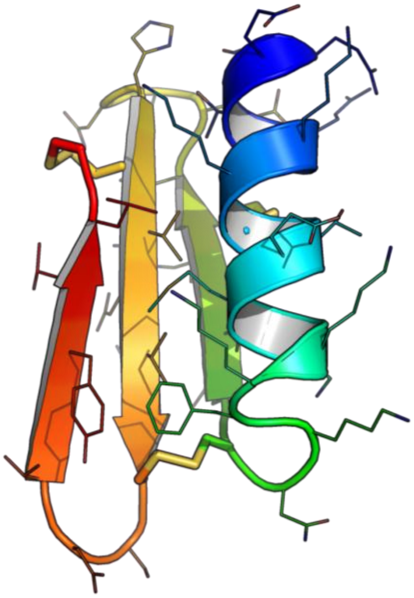


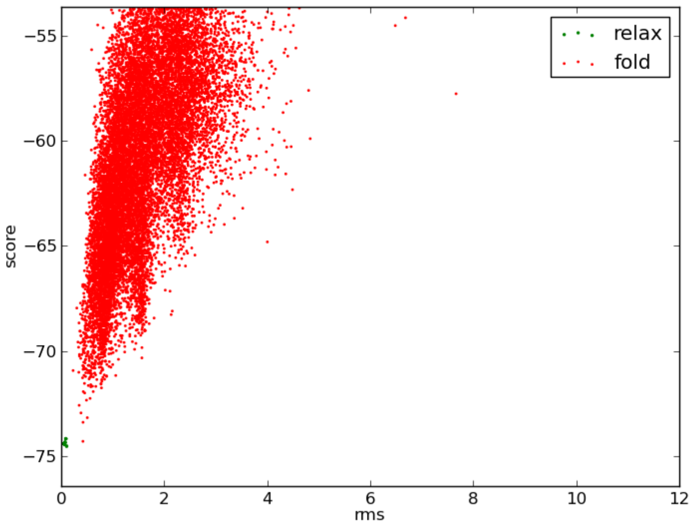


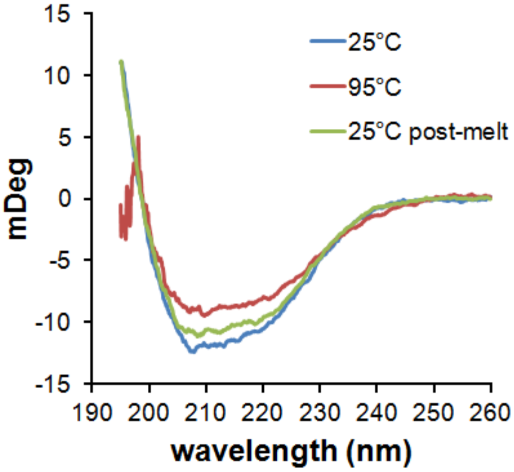


gHEEE_02

also known as: HEEE_2.2_02
amino acid sequence:
SQETRKKCTEMKKKFKNCEVRCDESNHCVEVRCSDTKYTLC

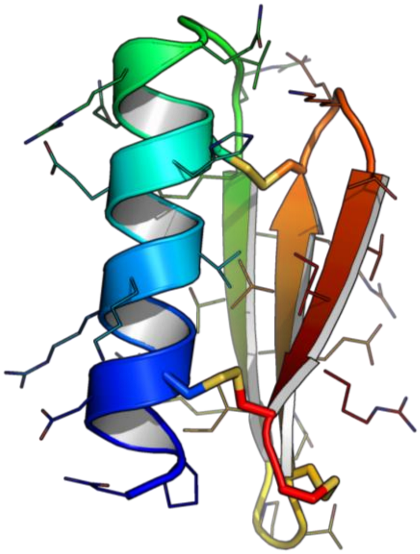


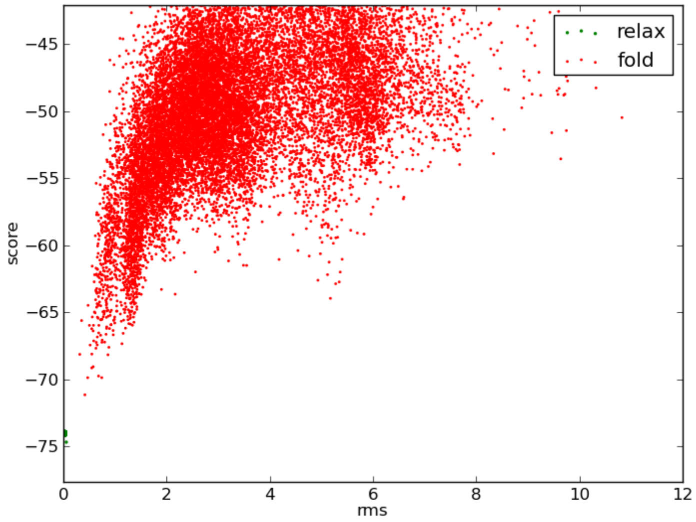




HEEE_2.2_03

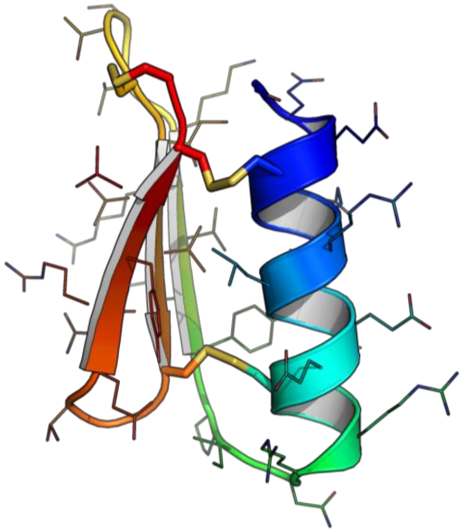
amino acid sequence:
NPEDCARKVEEHCQRQGVRYTTHRQPTCIEVRCEKTTIRCC

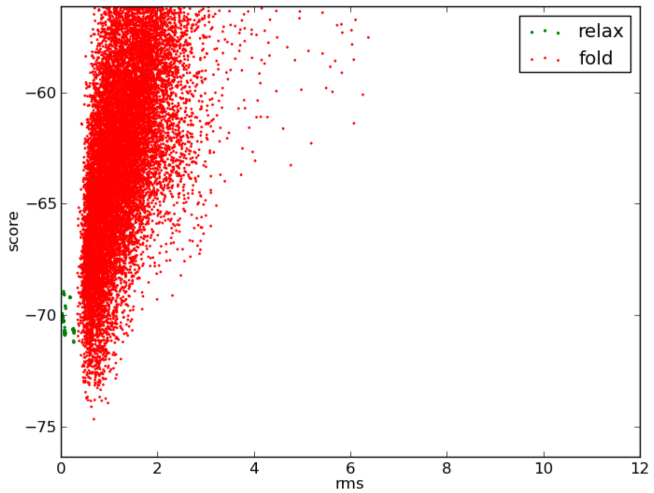


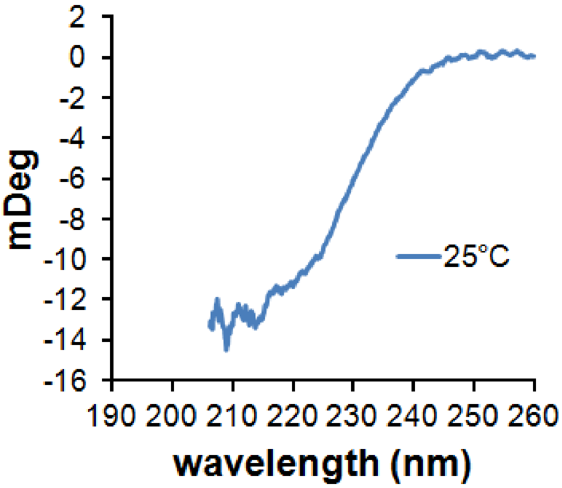


HEEE_2.2_04

amino acid sequence:
SEECAERLREECERRNIPYEVRKTSTCITVQCGTERYTCC

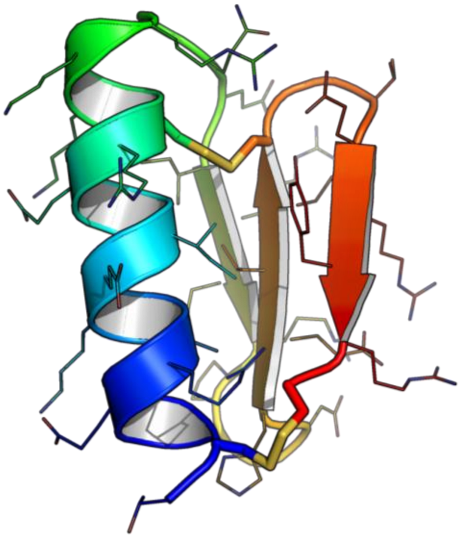


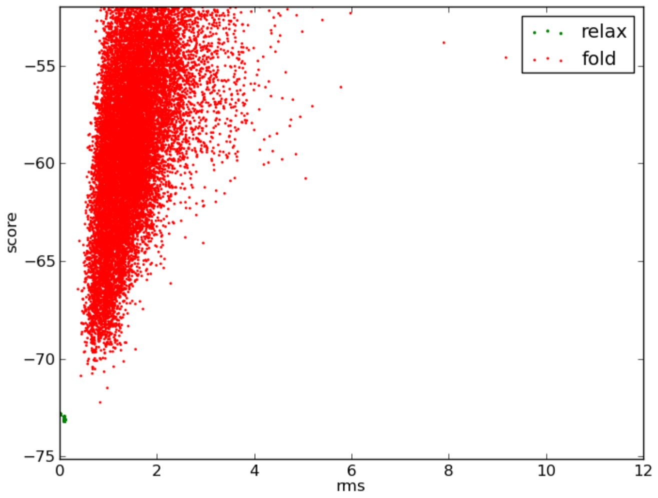


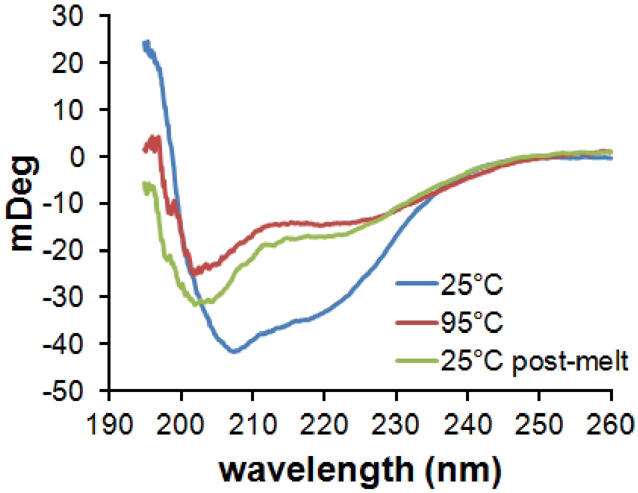


HEEE_2.2_05

amino acid sequence:
TCQERVKEIKERCKKRGQEIRERPGDHEVQCGTERYRC

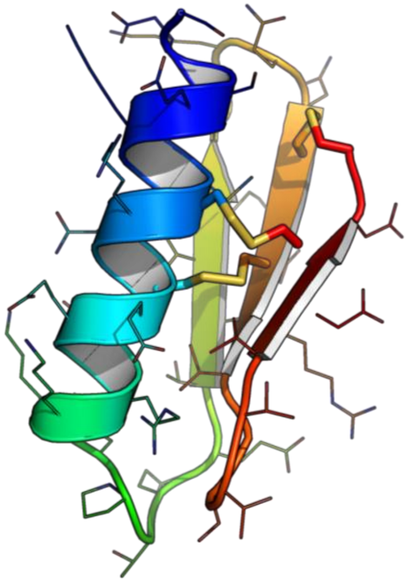


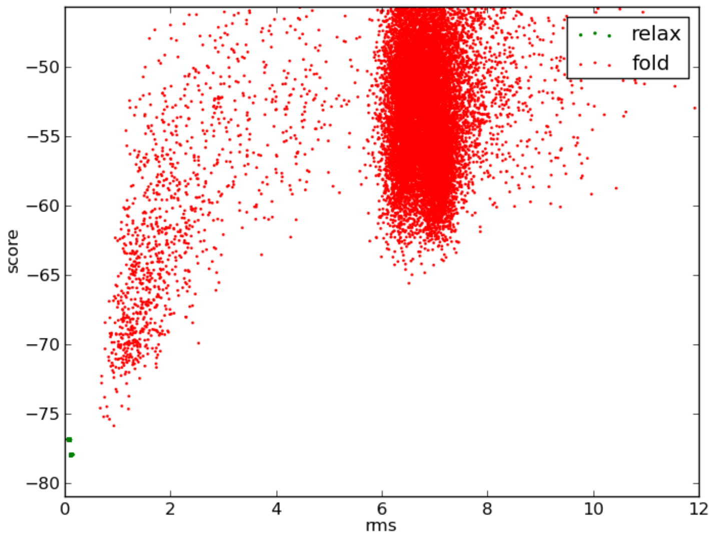


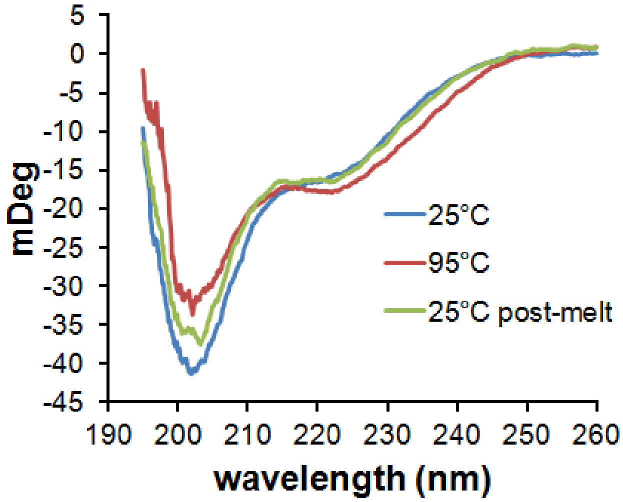


HEEE_2.2_06

amino acid sequence:
SESEKMCRQCEEERKKYPTQETSVRLPKQNCECRVGSTTVDCDC





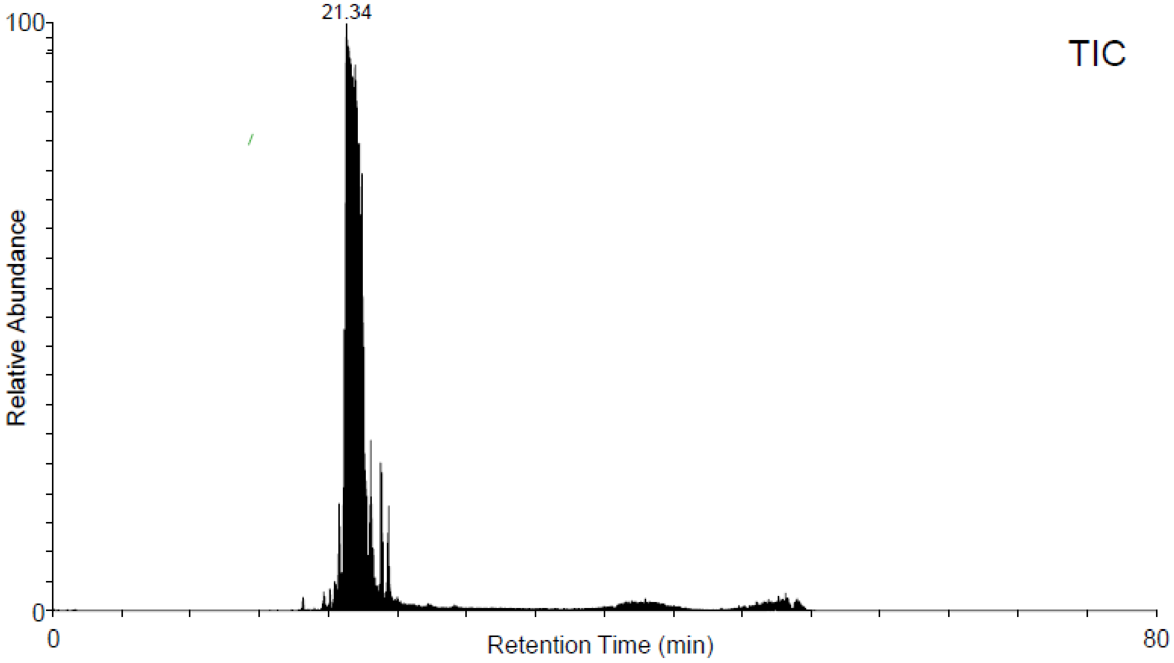


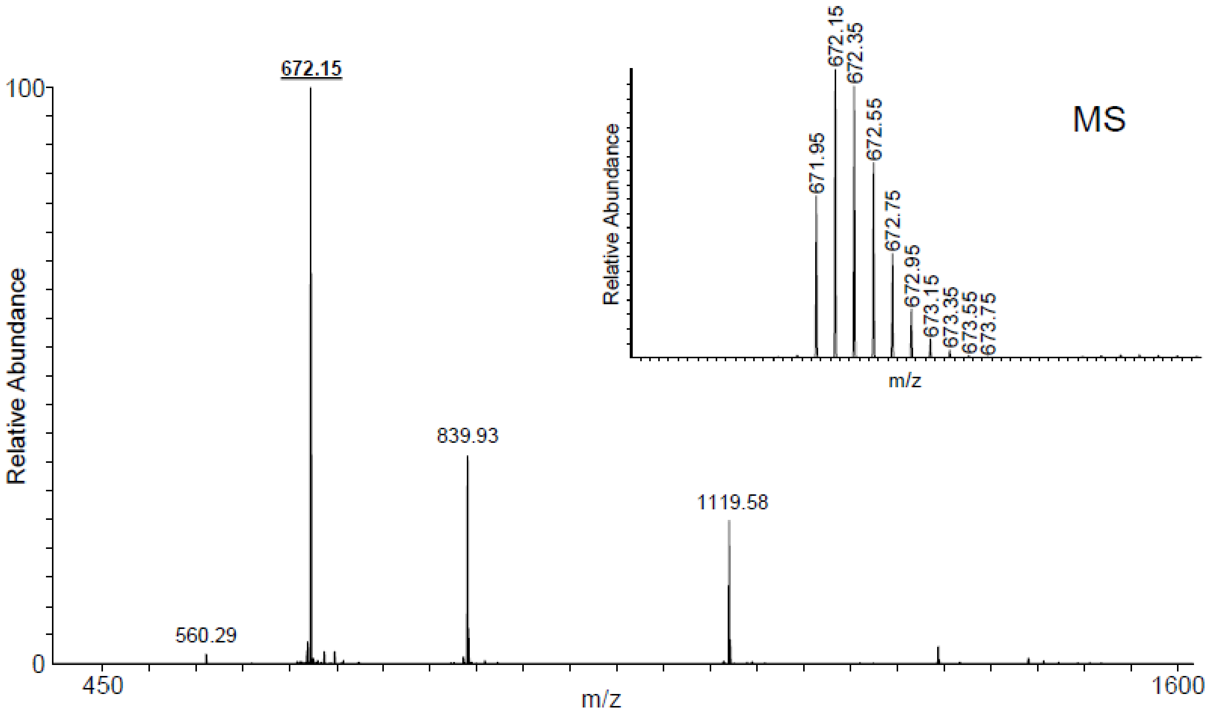
Supplementary information 4: HPLC and MS analysis of designed peptides

HPLC-MS traces for the genetically encodable peptides

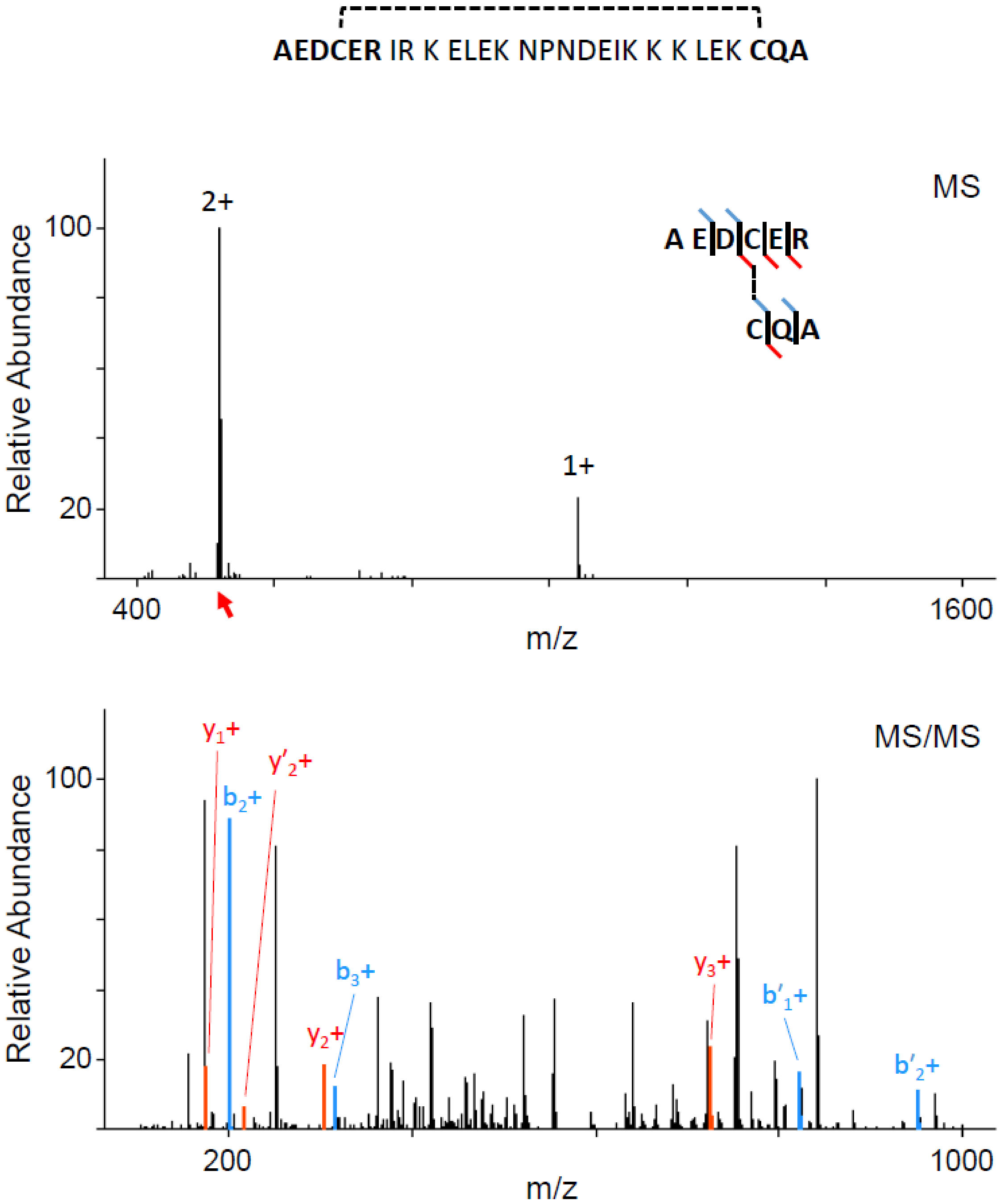
The total ion current (TIC) chromatogram for each design shows the retention time and intensity of ions detected over the course of the HPLC run. The retention time of the primary peak is indicated, and the averaged survey scans for the peak are shown below. The inset shows a narrow mass range around the observed ionic species peak, indicated in bold with underlining. The monoisotopic mass and isotopologues confirm the mass of the peptide with the expected number of disulfide bonds. We observed the spontaneous formation of pyroglutamic acid from the N-terminal glutamine of design gEEH_04, and our analysis accounts for the ~17 Da mass loss compared with the unmodified sequence. Following the intact analysis of each peptide, results of disulfide connectivity analysis are shown. For each, the primary sequence of the peptide is shown with spaces denoting tryptic cleavage sites and dashed lines indicating putative disulfide connections. The bolded peptides (linked by disulfides) shows one of the linked peptides searched for by targeted mass spectrometry of trypsin digested samples. The charge states detected in survey scans are indicated, and an annotated MS/MS scan corresponding to the peak marked with the red arrow is shown. For simplicity, only singly charged b and y ions are labeled in the MS/MS spectra. The detected ions are reflected in the inset of the MS survey spectra. To indicate the peptide origin of each ion, the second and third linked peptides (top to bottom) are denoted with ′ and ″, respectively.

gHH_44

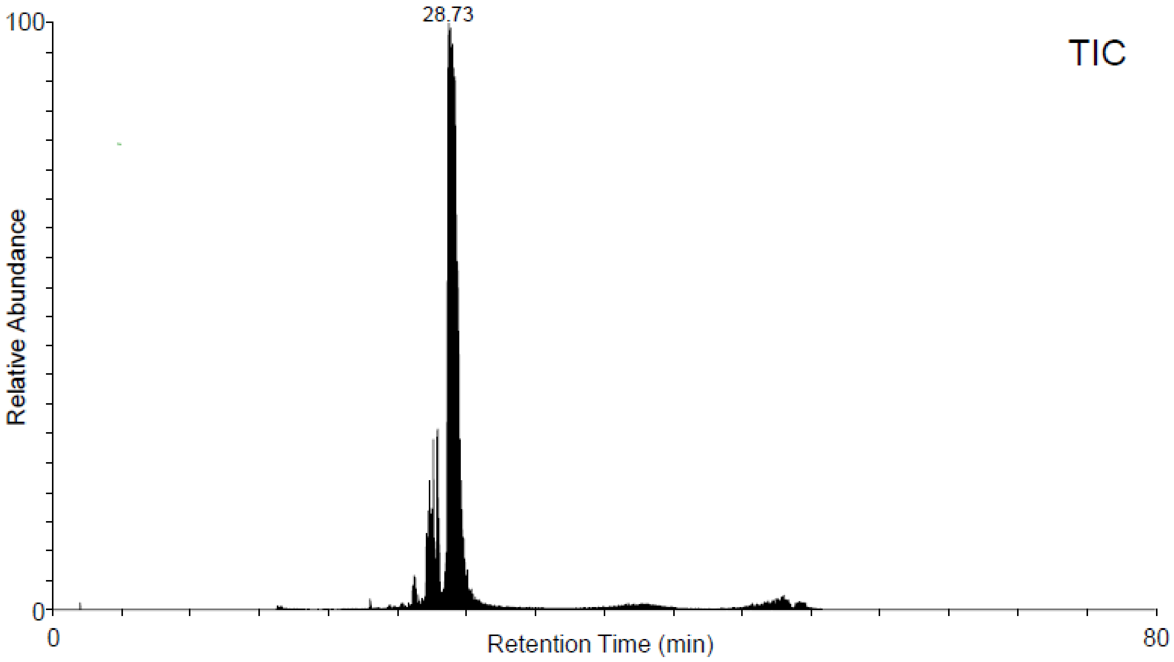

21.34
100
0
Relative Abundance
0
Retention Time (min)
80
TIC

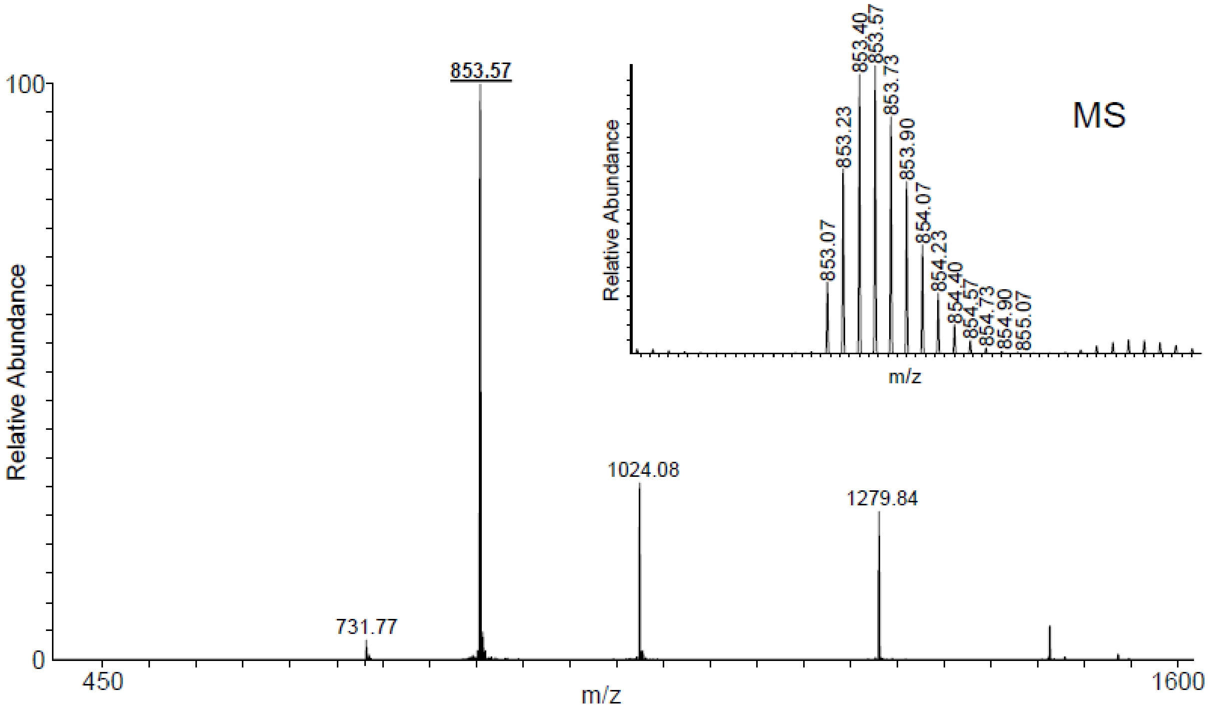

672.15
100
0
Relative Abundance
560.29
839.93
1119.58
450
m/z
1600
MS
Relative Abundance
671.95
672.15
672.35
672.55
672.75
672.95
673.15
673.35
673.55
673.75
m/z

gHH_44


AEDCER IR K ELEK NPNDEIK K K LEK CQA
MS
2+
1+
Relative Abundance
100
20
400
1600
m/z
A E D C E R
C Q A
MS/MS
y_1+
y'_2+
b_2+
b_3+
y_2+
y_3+
b'_1+
b'_2+
Relative Abundance
100
20
200
1000
m/z

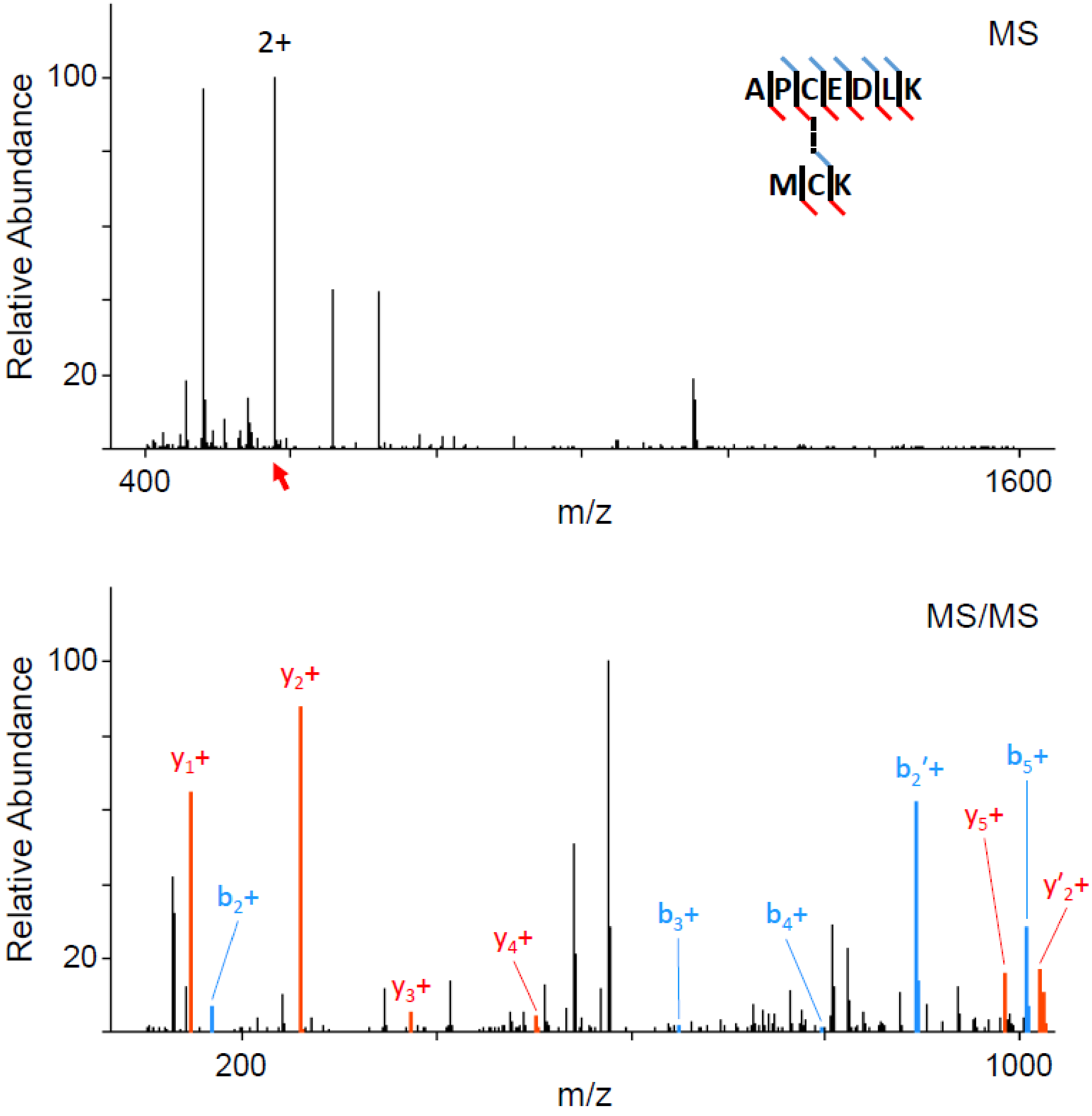
gHHH_06


28.73
TIC
100
0
Relative Abundance
0
Retention Time (min)
80

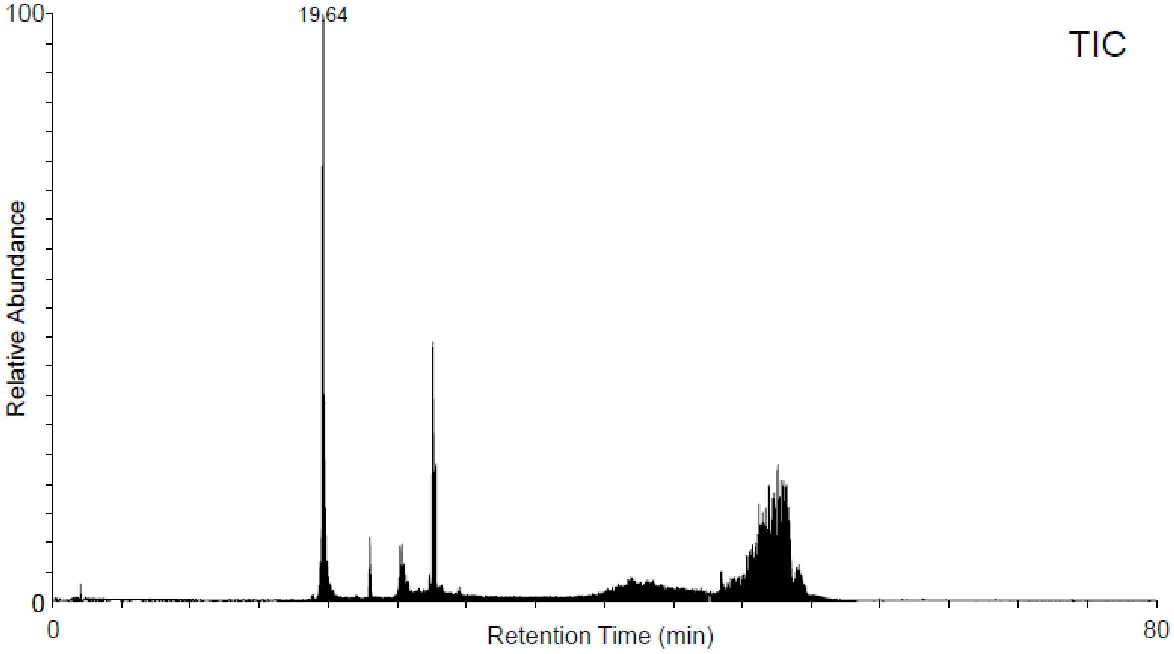

853.57
MS
100
0
Relative Abundance
450
m/z
1600
731.77
1024.08
1279.84
Relative Abundance
m/z
853.07
853.23
853.40
853.57
853.73
853.90
854.07
854.23
854.40
854.57
854.73
854.90
855.07

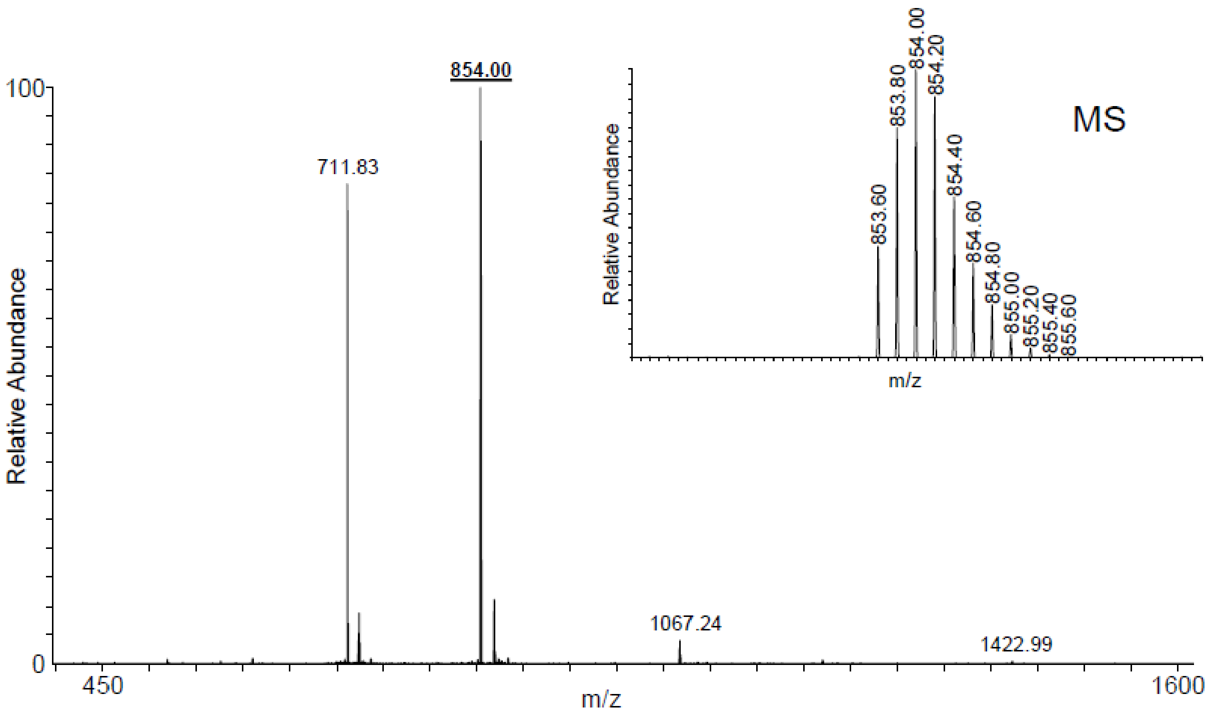
gHHH_06

APCEDLK ER LK K LGMSEECR QR LEK **MCK** EGTSEDAER MAR NCES

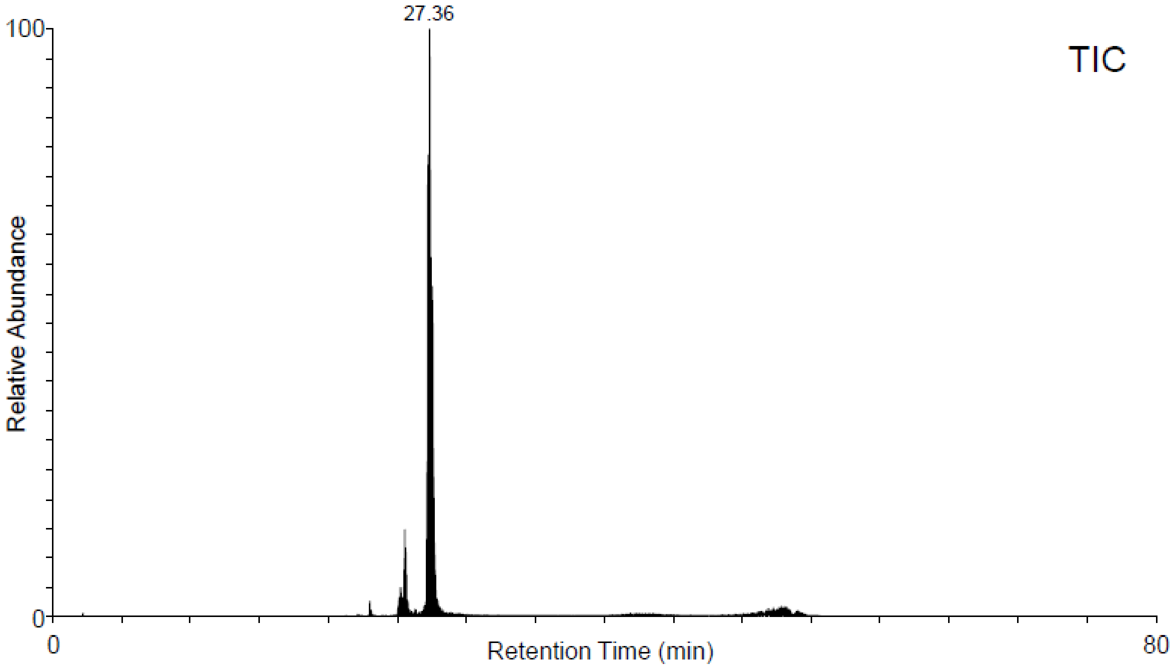


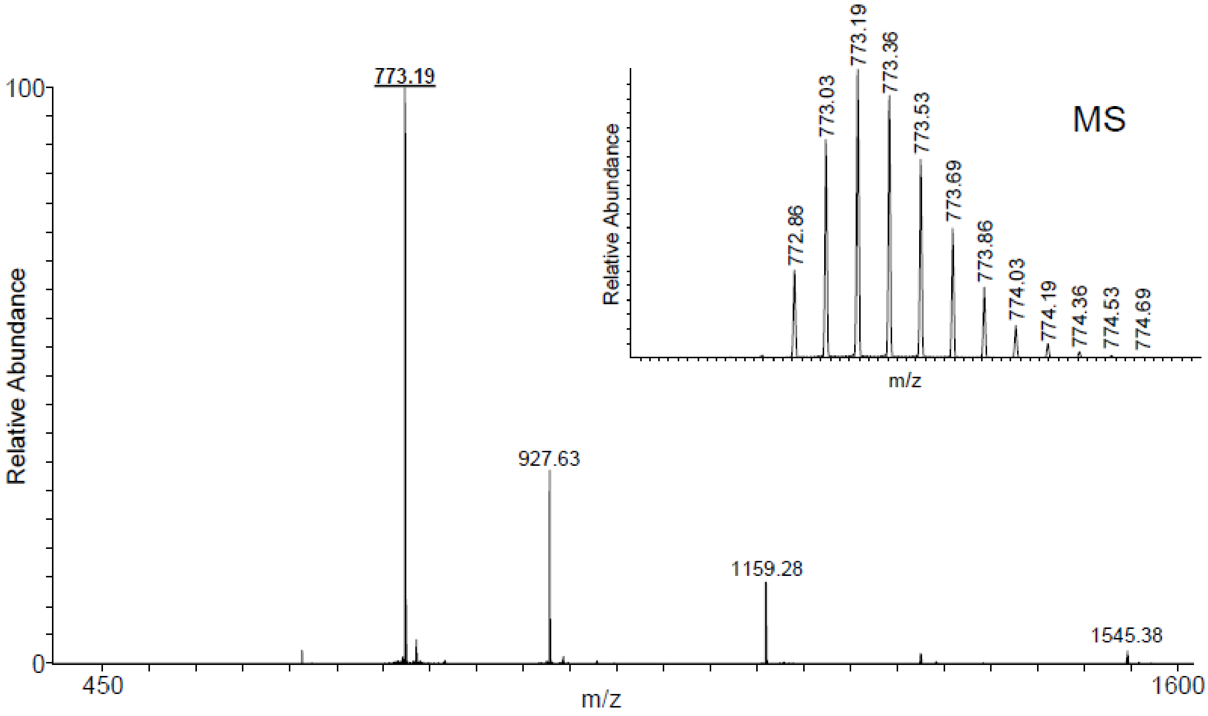
gEHE_06


TIC
100
0
Relative Abundance
19.64
0
80
Retention Time (min)


MS
100
0
Relative Abundance
854.00
711.83
1067.24
1422.99
450
1600
m/z
Relative Abundance
853.60
853.80
854.00
854.20
854.40
854.60
854.80
855.00
855.20
855.40
855.60
m/z

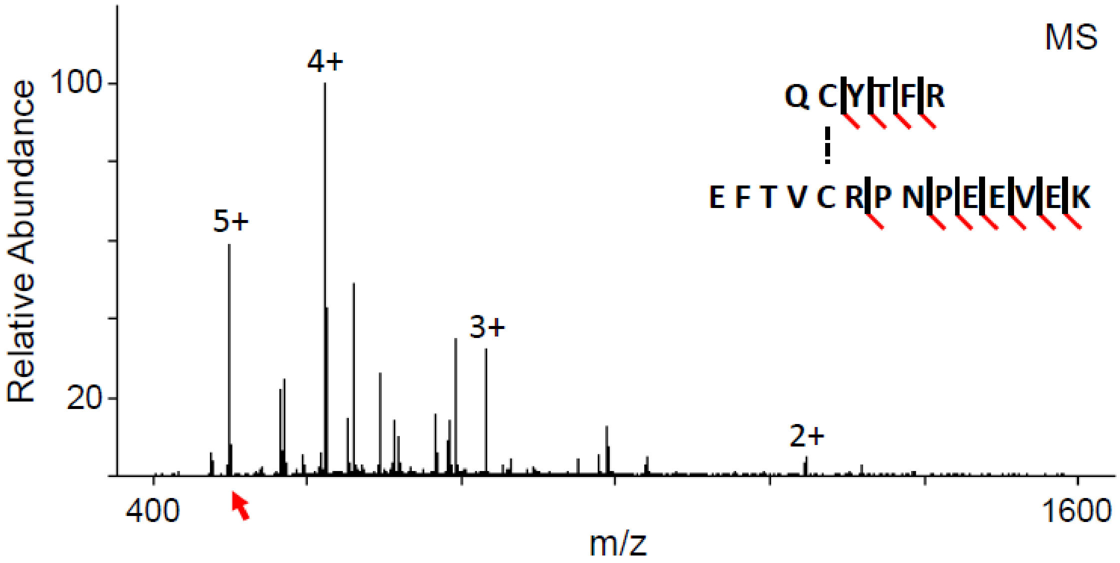
gEEH_04

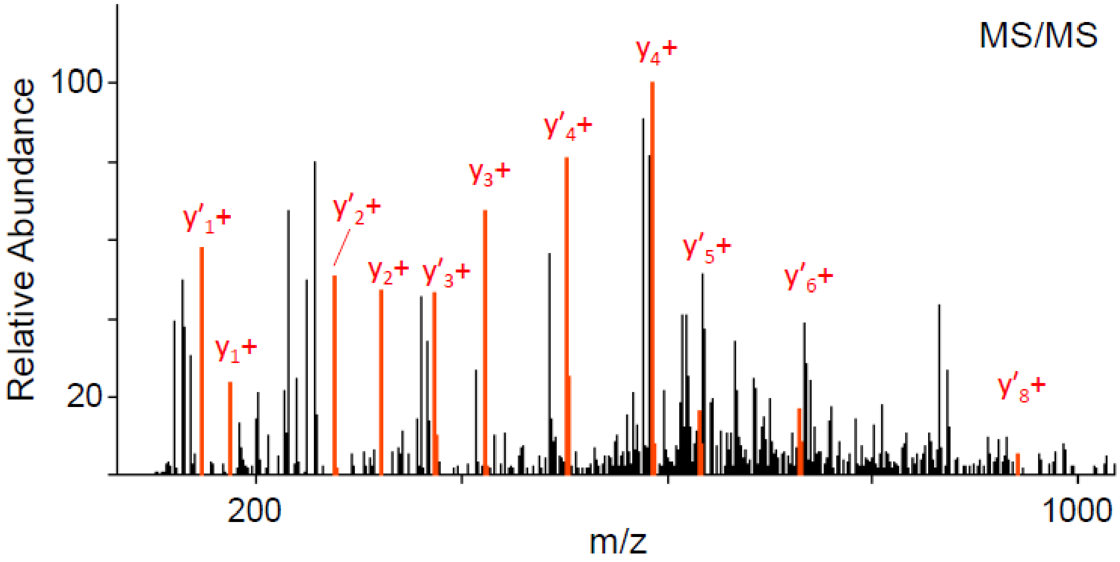

27.36
TIC
100
0
Relative Abundance
0
80
Retention Time (min)


773.19
100
0
Relative Abundance
927.63
1159.28
1545.38
450
1600
m/z
MS
Relative Abundance
772.86
773.03
773.19
773.36
773.53
773.69
773.86
774.03
774.19
774.36
774.53
774.69
m/z

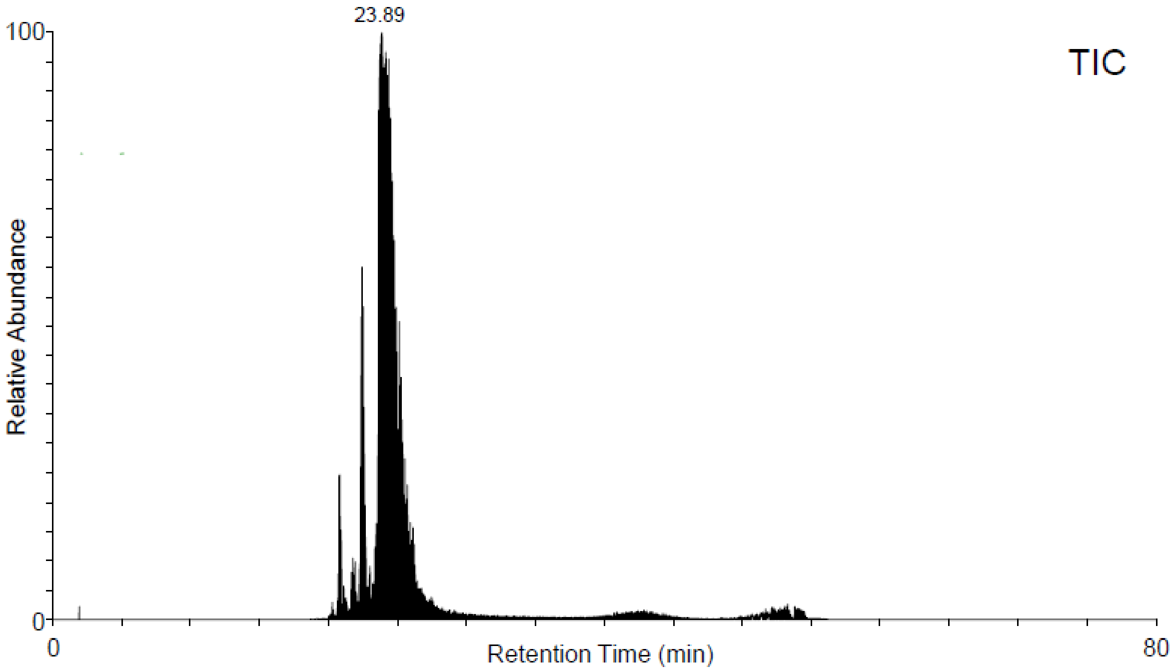
gEEH_04

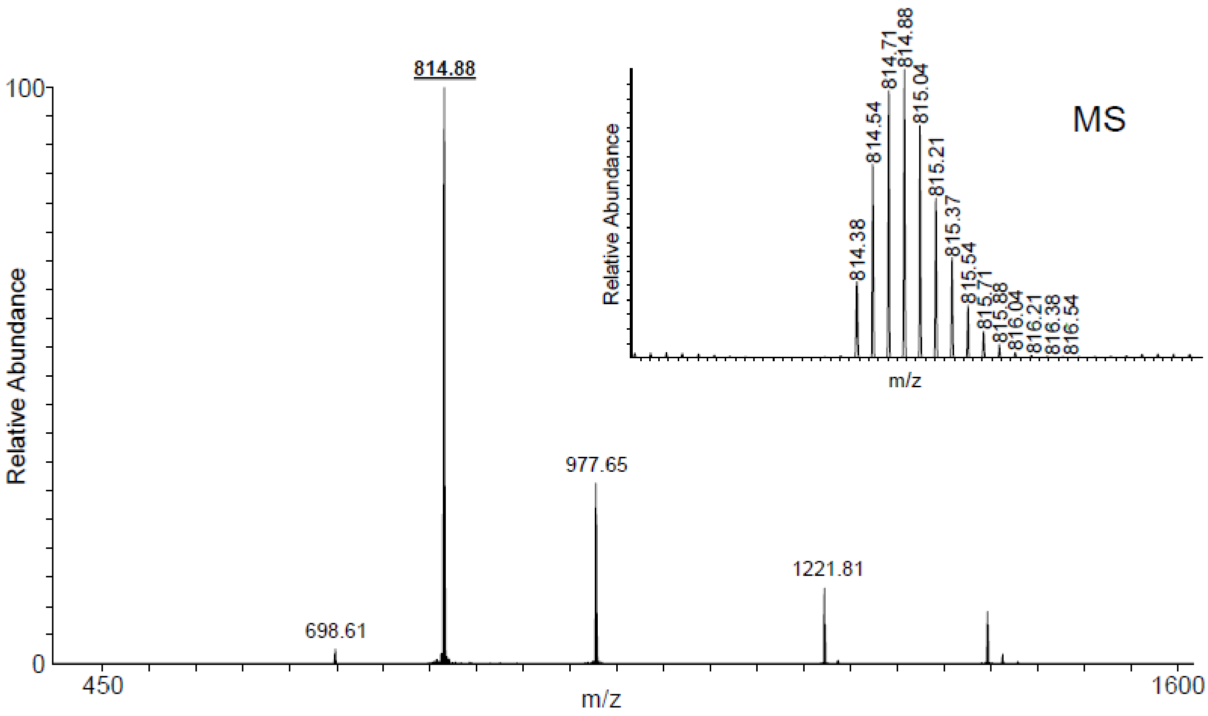

QCYTFR SECTNK EFTVCRPNPEEVEK EAR R TK EEECR K


MS
Relative Abundance
100
20
5+
4+
3+
2+
Q C|Y|T|F|R
E F T V C R|P N|P|E|E|V|E|K
400
1600
m/z

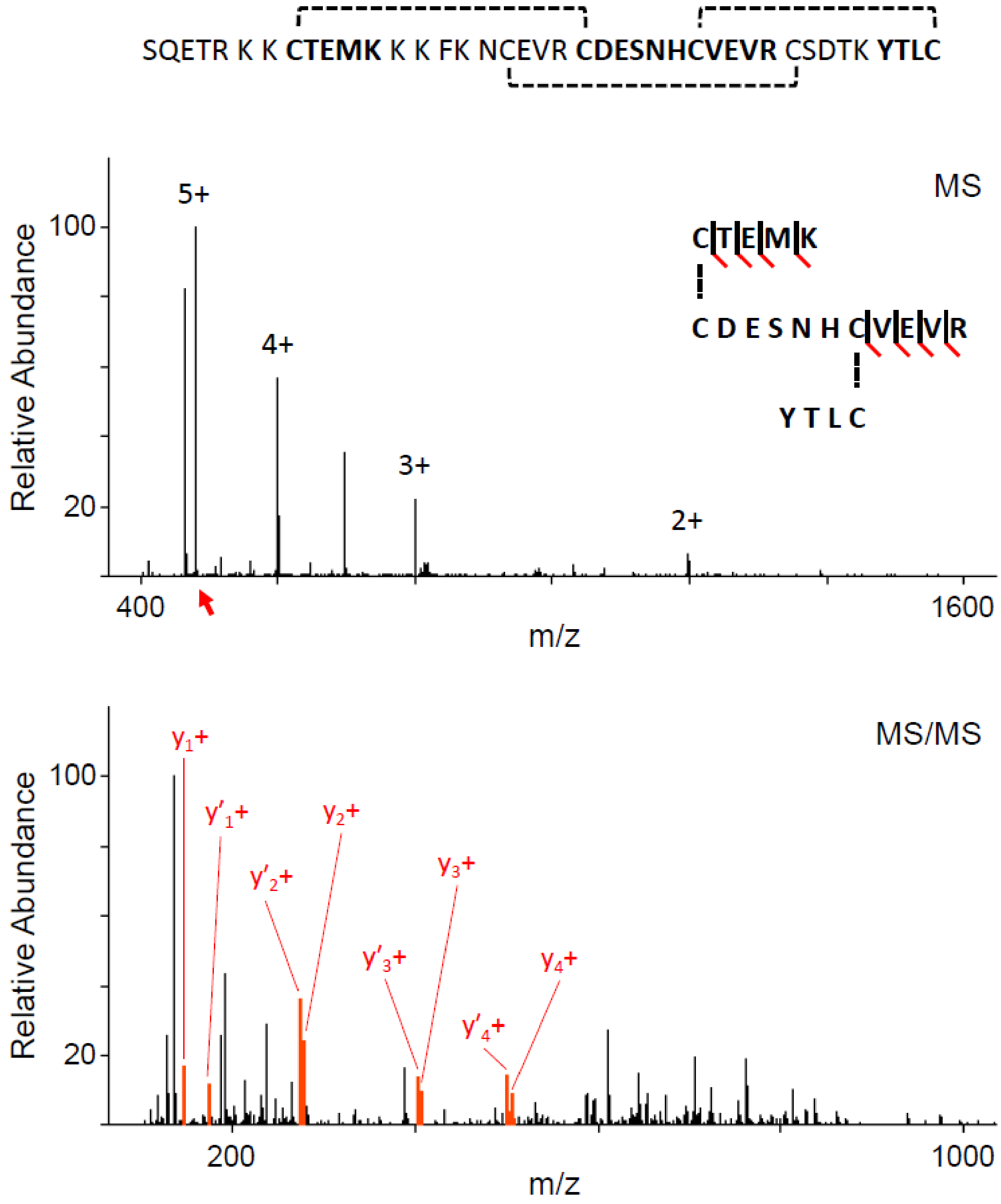

MS/MS
Relative Abundance
100
20
y'1+
y1+
y'2+
y2+
y'3+
y3+
y'4+
y4+
y'5+
y'6+
y'8+
200
1000
m/z

gHEEE_02

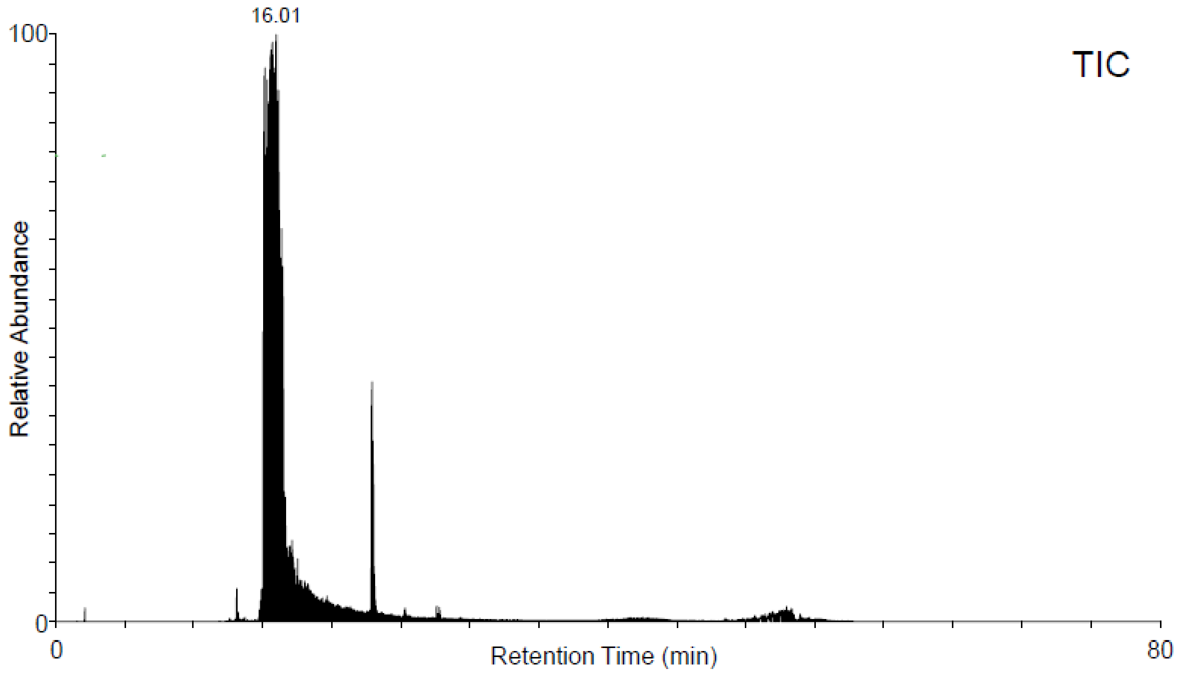

23.89
TIC
100
0
Relative Abundance
0
Retention Time (min)
80

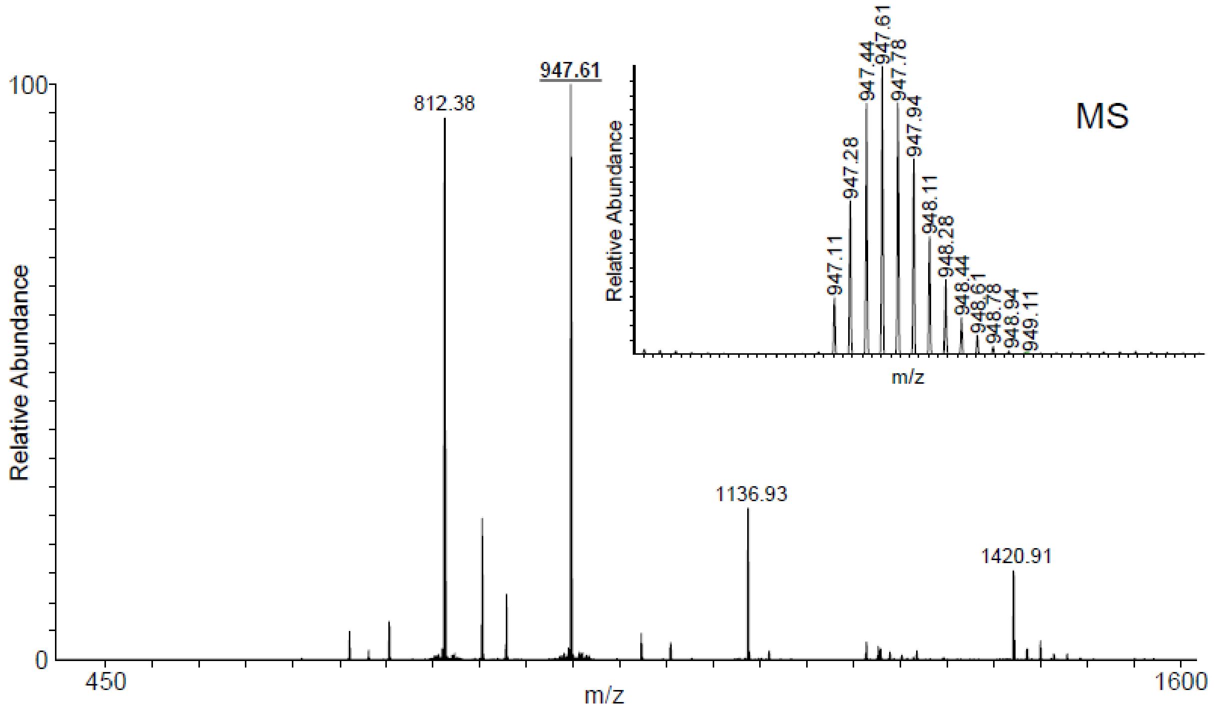

814.88
MS
100
0
Relative Abundance
698.61
977.65
1221.81
450
m/z
1600
Relative Abundance
814.38
814.54
814.71
814.88
815.04
815.21
815.37
815.54
815.71
815.88
816.04
816.21
816.38
816.54
m/z

gHEEE_02

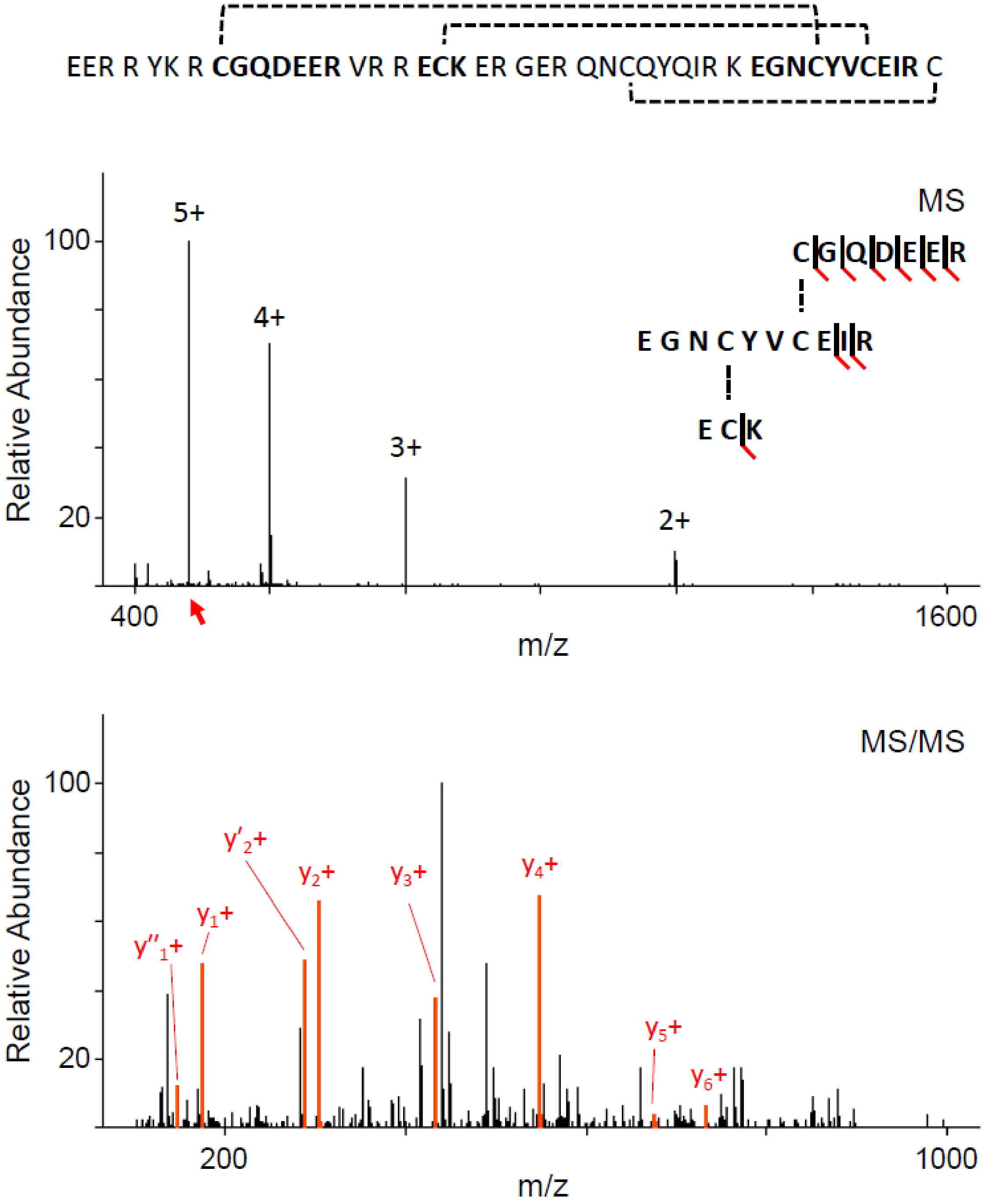

SQETR K K CTEMK K K FK NCEVR CDESNHCVEVR CSDTK YTLC
MS
5+
4+
3+
2+
100
20
Relative Abundance
CTEMK
CDESNHCVEVR
YTLC
400
1600
m/z
MS/MS
y_1+
y'_1+
y_2+
y'_2+
y_3+
y'_3+
y_4+
y'_4+
100
20
Relative Abundance
200
1000
m/z

gEHEE_06

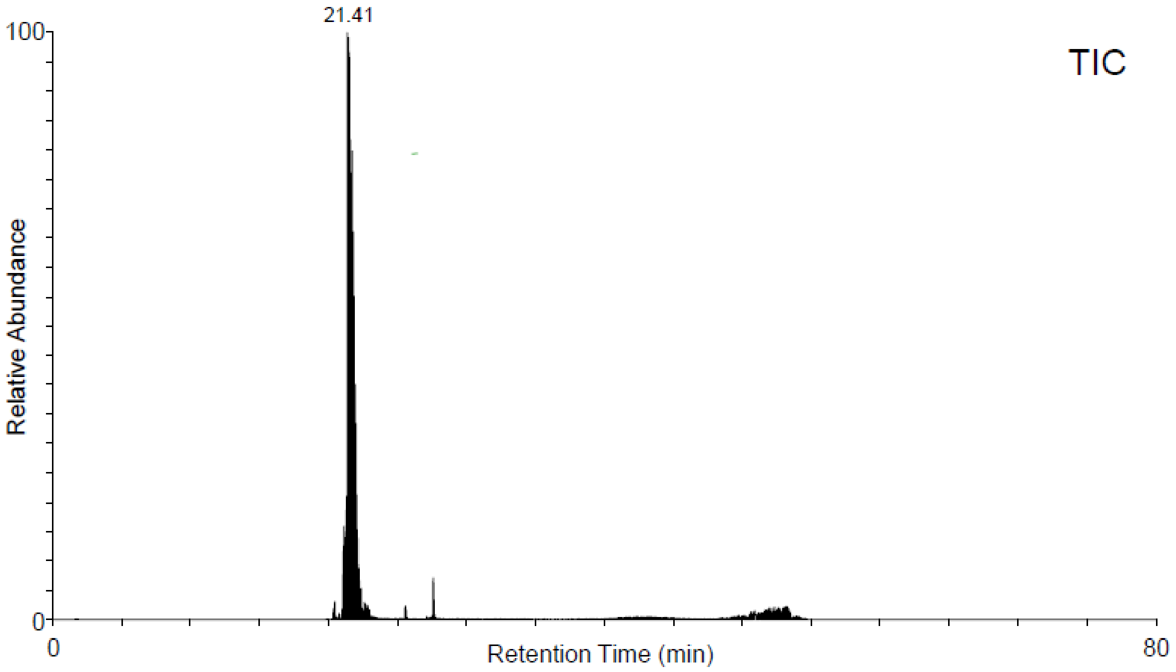

16.01
TIC
100
0
Relative Abundance
0
Retention Time (min)
80

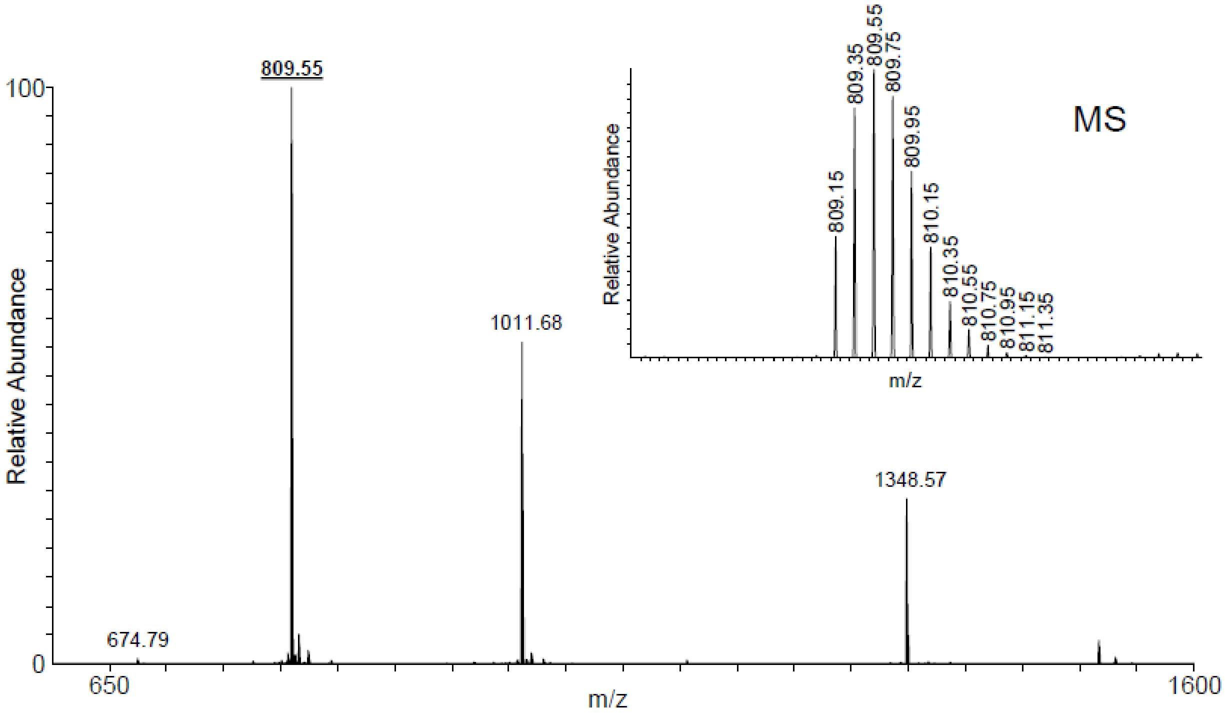

947.61
812.38
MS
100
0
Relative Abundance
1136.93
1420.91
450
m/z
1600
Relative Abundance
947.11
947.28
947.44
947.61
947.78
947.94
948.11
948.28
948.44
948.61
948.78
948.94
949.11
m/z

gEHEE_06

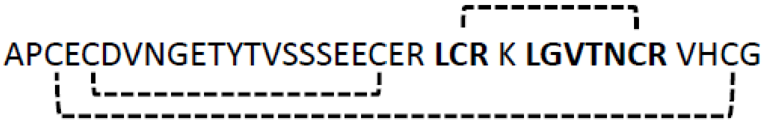

EER R YK R CGQDEER VR R ECK ER GER QNCQYQIR K EGNCYVCEIR C
MS
5+
4+
3+
2+
C G Q D E E R
E G N C Y V C E I R
E C K
Relative Abundance
100
20
400
1600
m/z
MS/MS
y′′1+
y1+
y′2+
y2+
y3+
y4+
y5+
y6+
Relative Abundance
100
20
200
1000
m/z

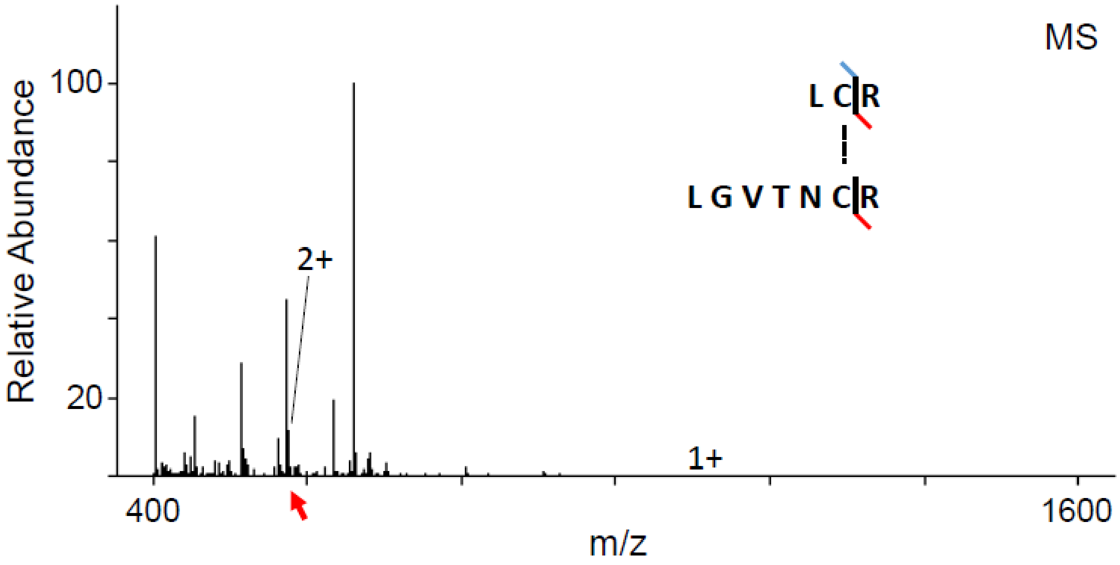
gEEHE_02

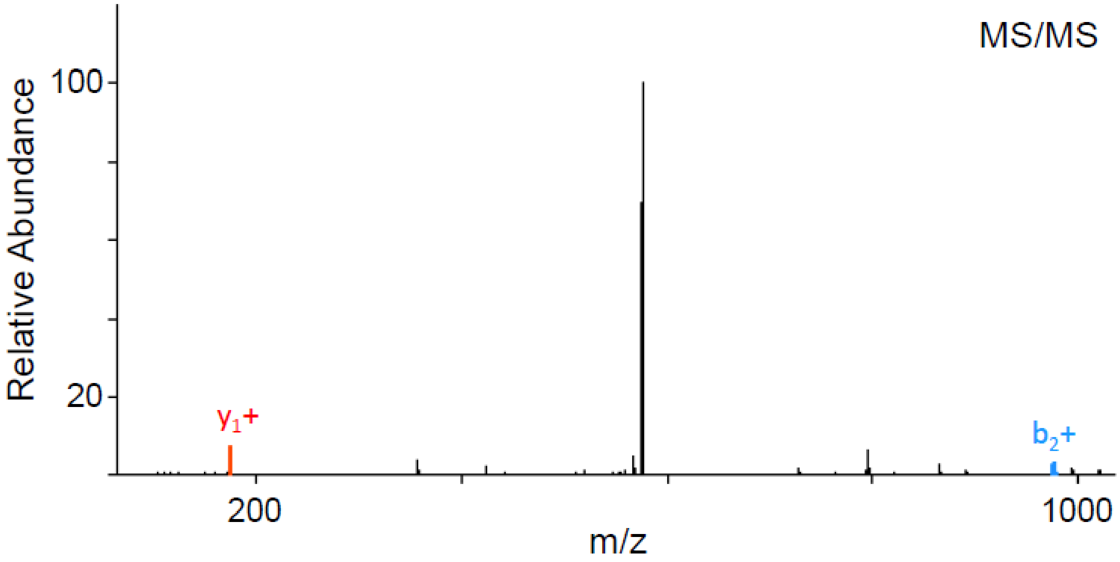

21.41
TIC
100
0
Relative Abundance
0
80
Retention Time (min)


809.55
MS
100
0
Relative Abundance
674.79
1011.68
1348.57
650
1600
m/z
809.15
809.35
809.55
809.75
809.95
810.15
810.35
810.55
810.75
810.95
811.15
811.35
Relative Abundance
m/z

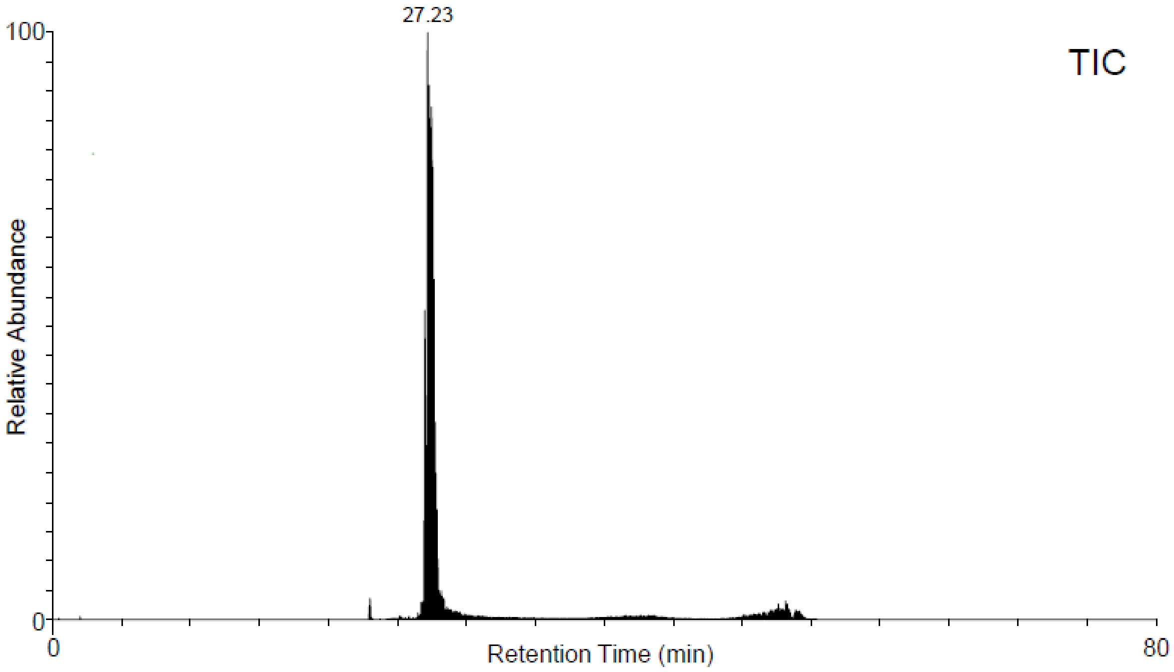
gEEHE_02

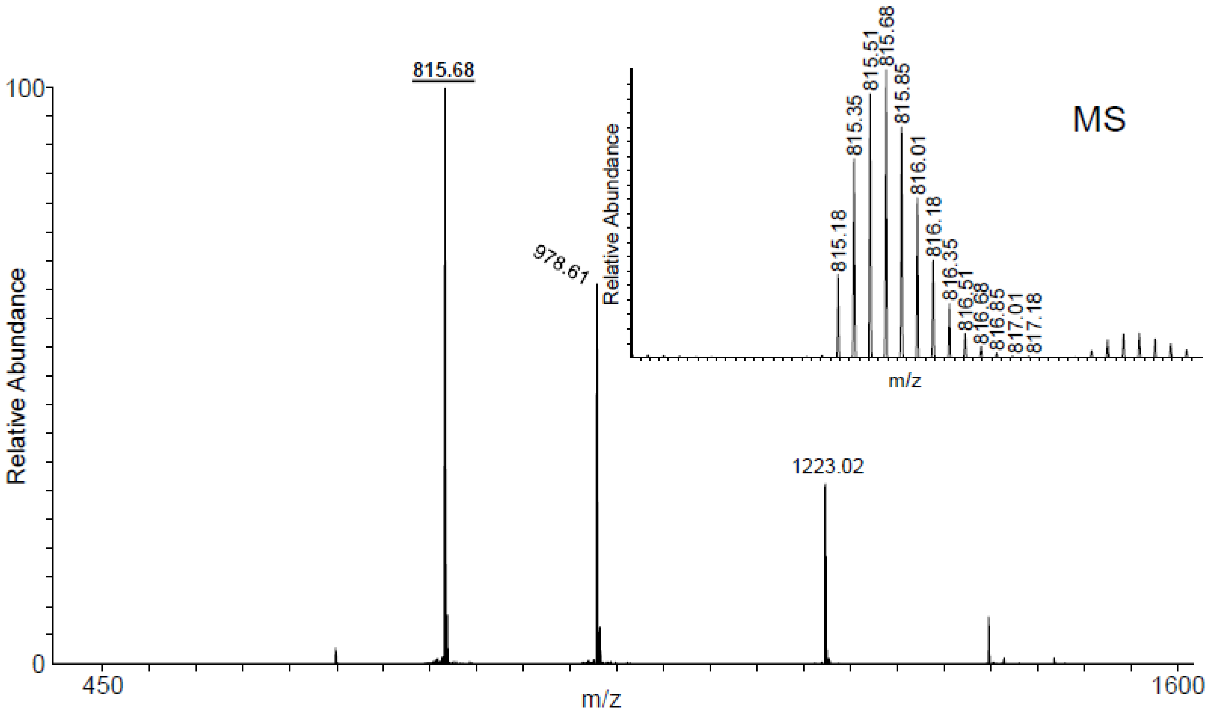

APCECDVNGETYTVSSSEECER **LCR** K **LGVTNCR** VHCG


MS
Relative Abundance
100
20
2+
1+
L C R
L G V T N C R
400
1600
m/z


MS/MS
Relative Abundance
100
20
y_1+
b_2+
200
1000
m/z

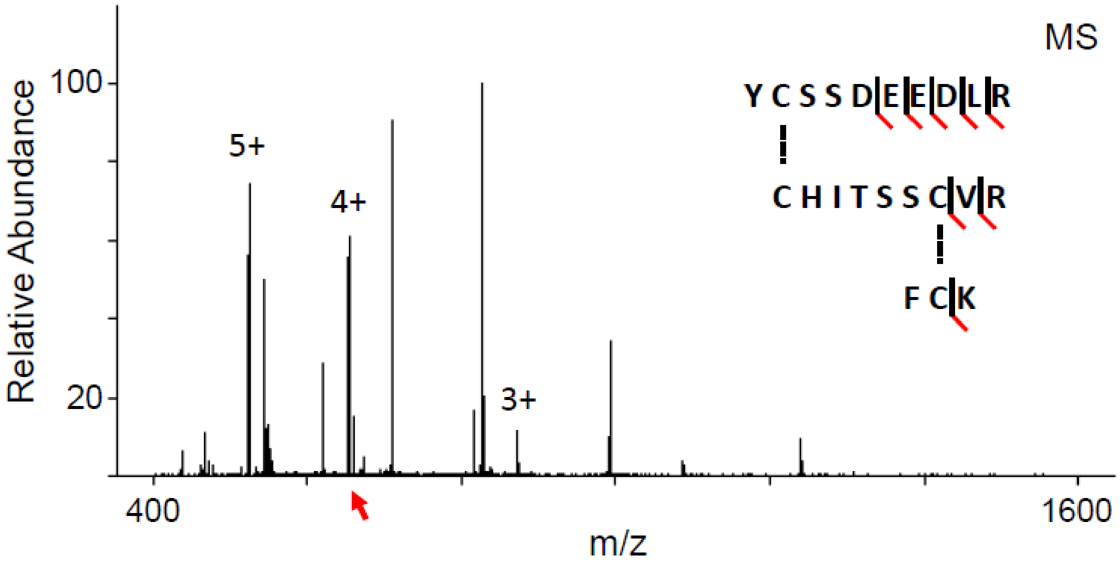
gEEEH_04

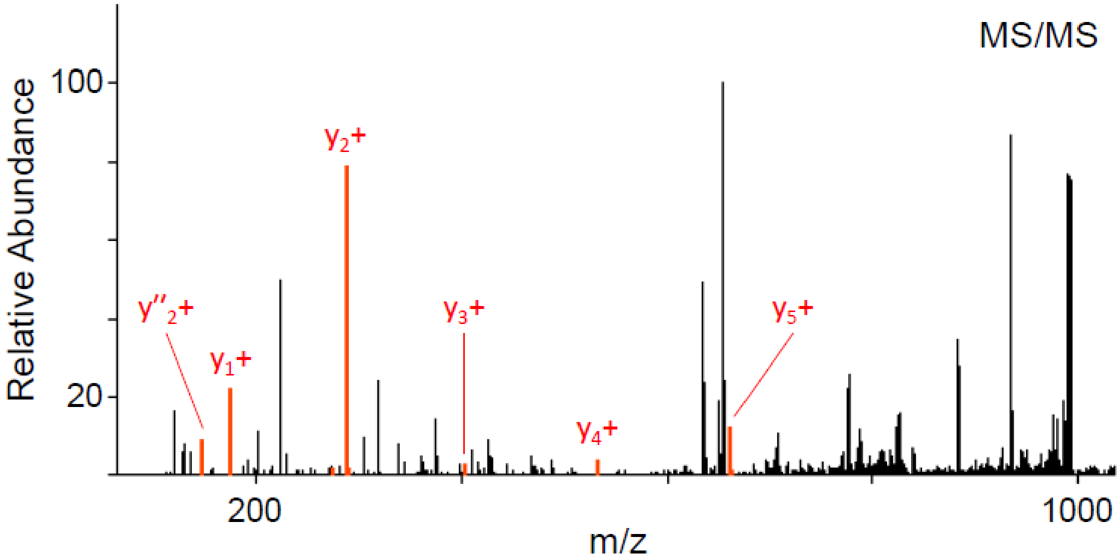

27.23
TIC
100
0
Relative Abundance
0
Retention Time (min)
80


815.68
978.61
1223.02
MS
100
0
Relative Abundance
450
m/z
1600
815.18
815.35
815.51
815.68
815.85
816.01
816.18
816.35
816.51
816.68
816.85
817.01
817.18
Relative Abundance
m/z

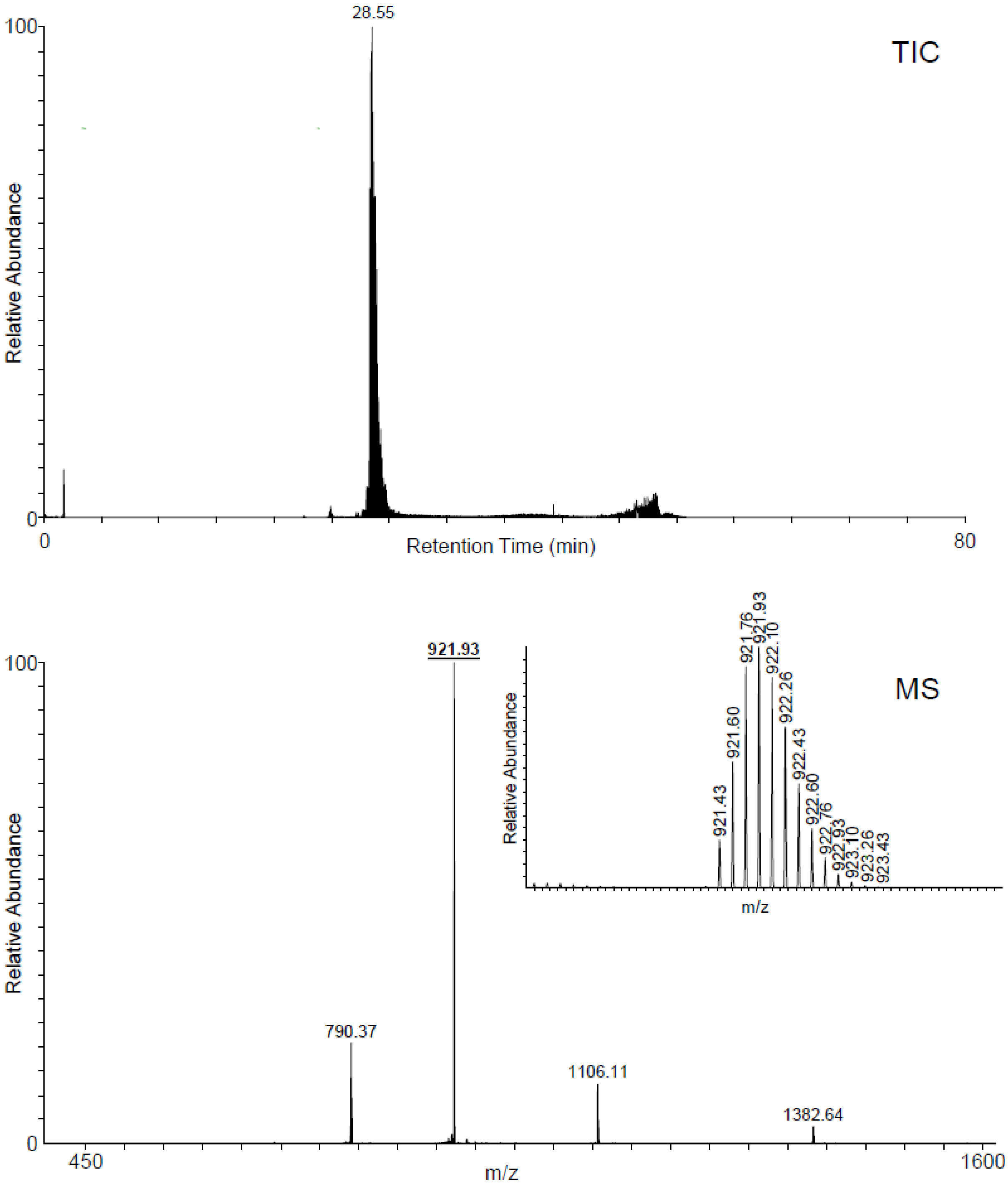
gEEEH_04


CR **CHITSSCVR** VEGDNGEEYR **YCSSDEEDLR** R **FCK** EMQK QC

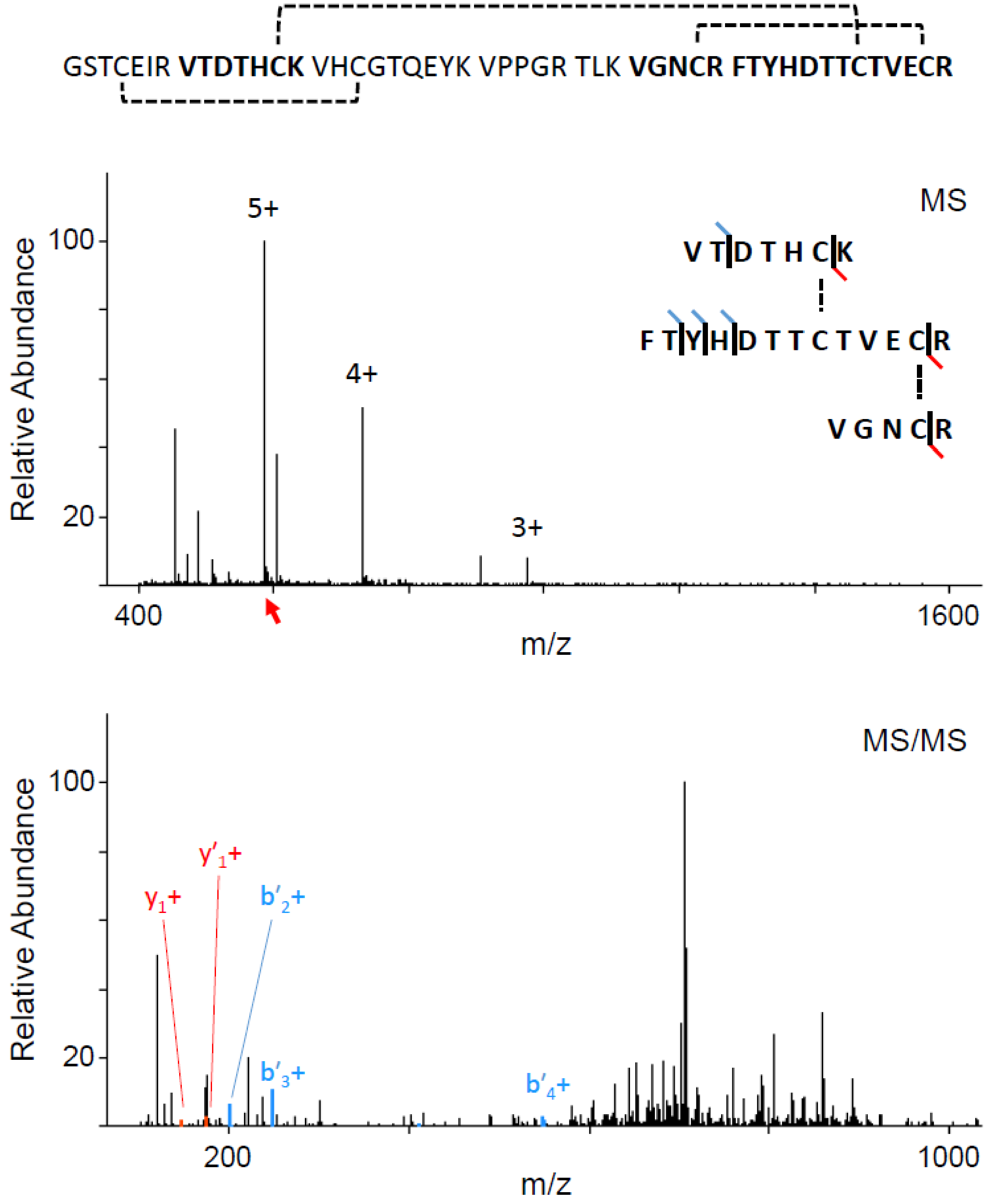

MS
Relative Abundance
100
20
5+
4+
3+
400
1600
m/z
Y C S S D E E D L R
C H I T S S C V R
F C K


MS/MS
Relative Abundance
100
20
y''₂+
y₁+
y₂+
y₃+
y₄+
y₅+
200
1000
m/z

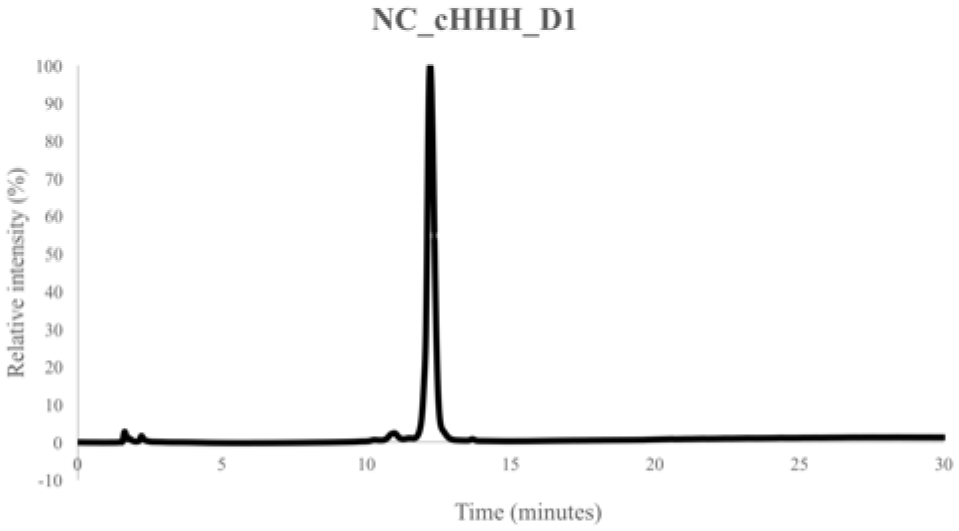
gEEEEEE_02

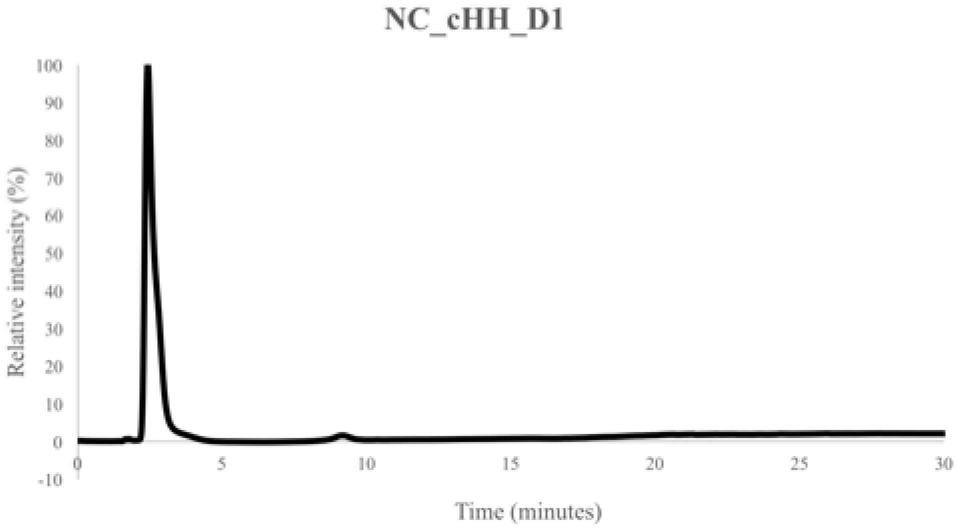

28.55
100
0
TIC
Relative Abundance
Retention Time (min)
0
80
921.93
100
0
MS
Relative Abundance
790.37
1106.11
1382.64
450
1600
m/z
Relative Abundance
921.43
921.60
921.76
921.93
922.10
922.26
922.43
922.60
922.76
922.93
923.10
923.26
923.43
m/z

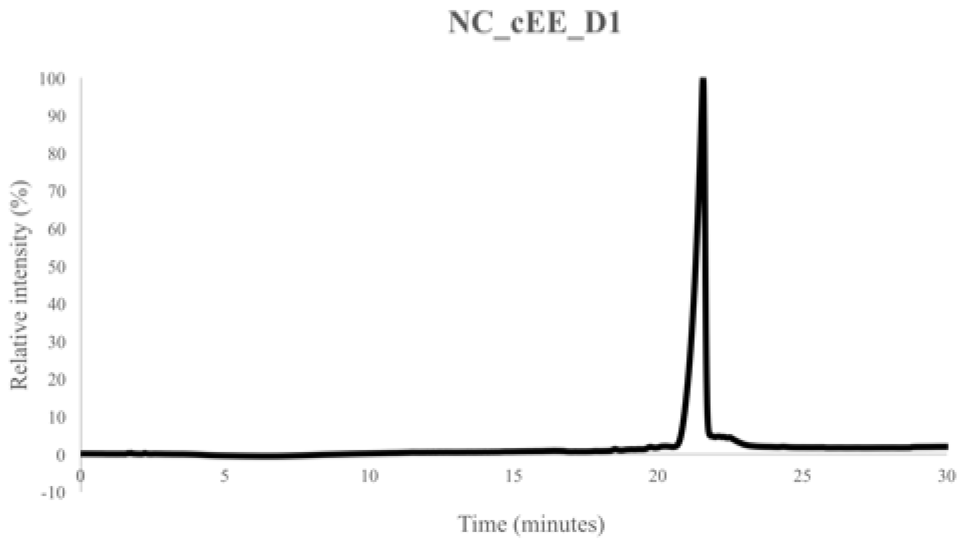
gEEEEEE_02

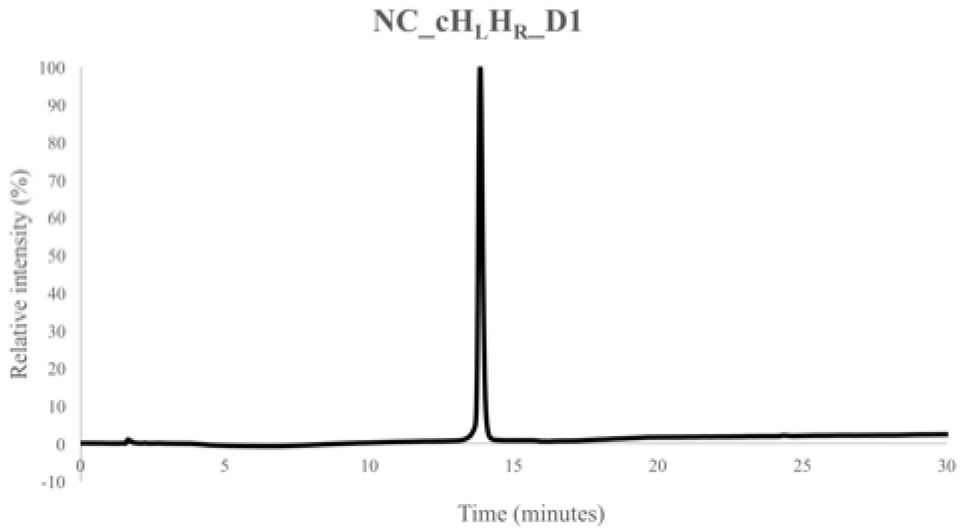

GSTCEIR **VTDTHCK** VHCGTQEYK VPPGR TLK **VGNCR FTYHDTTCTVECR**
MS
Relative Abundance
100
20
5+
4+
3+
400
1600
m/z
V T D T H C K
F T Y H D T T C T V E C R
V G N C R
MS/MS
Relative Abundance
100
20
y_1+
y'_1+
b'_2+
b'_3+
b'_4+
200
1000
m/z

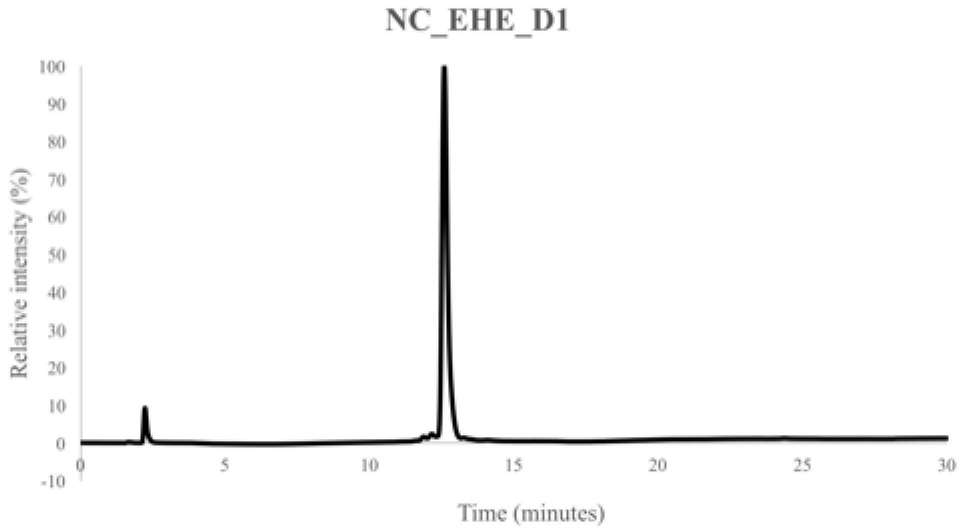
HPLC traces for the non-canonical peptides

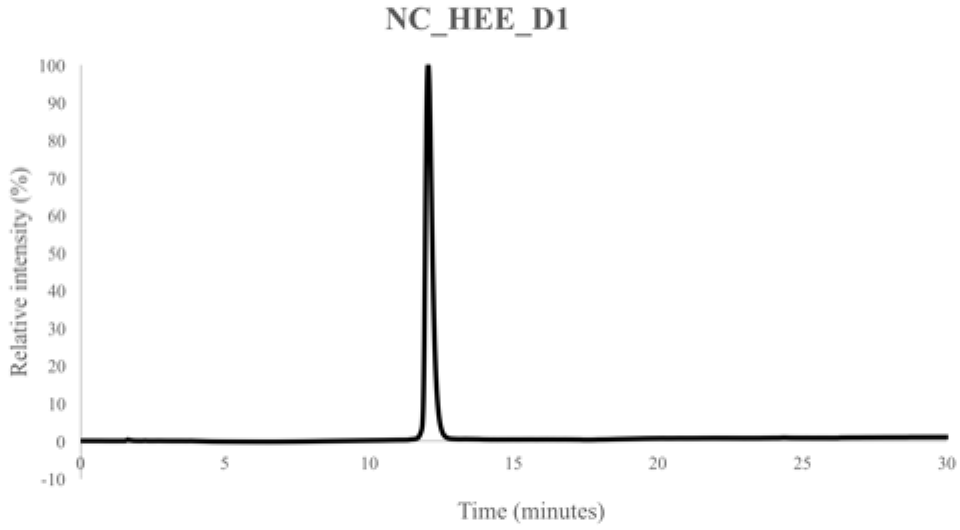


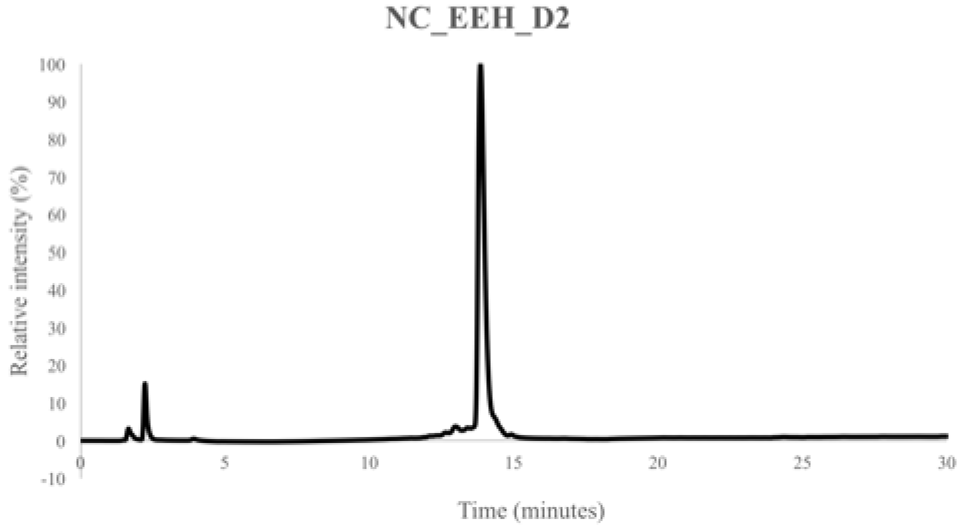



NC_cEE_D1
Relative intensity (%)
100
90
80
70
60
50
40
30
20
10
0
-10
0
5
10
15
20
25
30
Time (minutes)

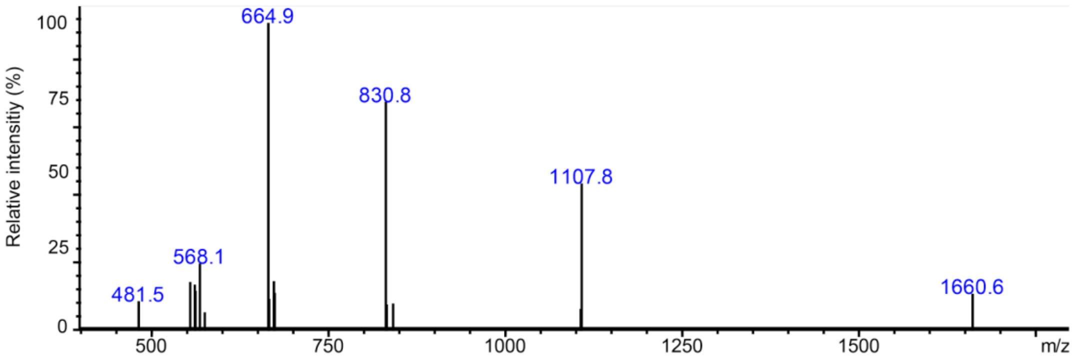

NC_cH$_{L}$H$_{R}$_D1
Relative intensity (%)
100
90
80
70
60
50
40
30
20
10
0
-10
0
5
10
15
20
25
30
Time (minutes)


NC_EHE_D1
Relative intensity (%)
100
90
80
70
60
50
40
30
20
10
0
-10
0
5
10
15
20
25
30
Time (minutes)

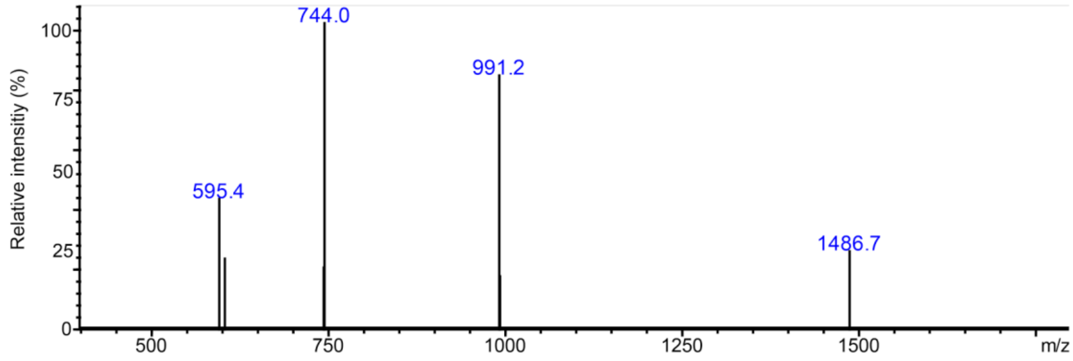




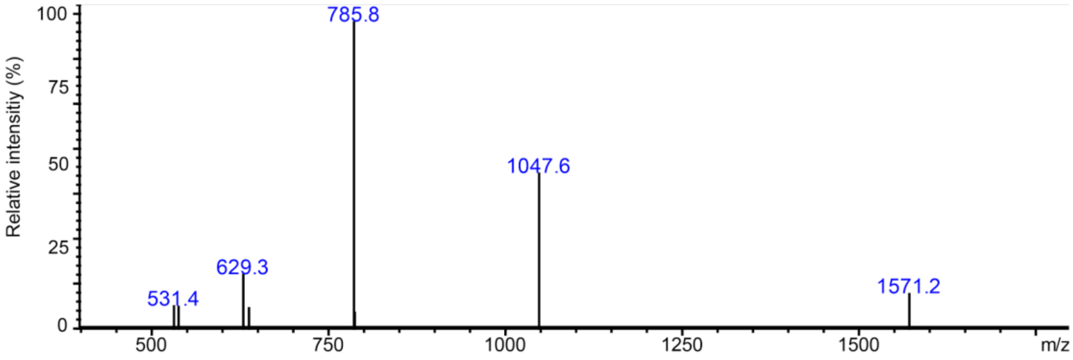
Mass Spectrometry (MS) traces for non-canonical peptides



NC_HEE_D1

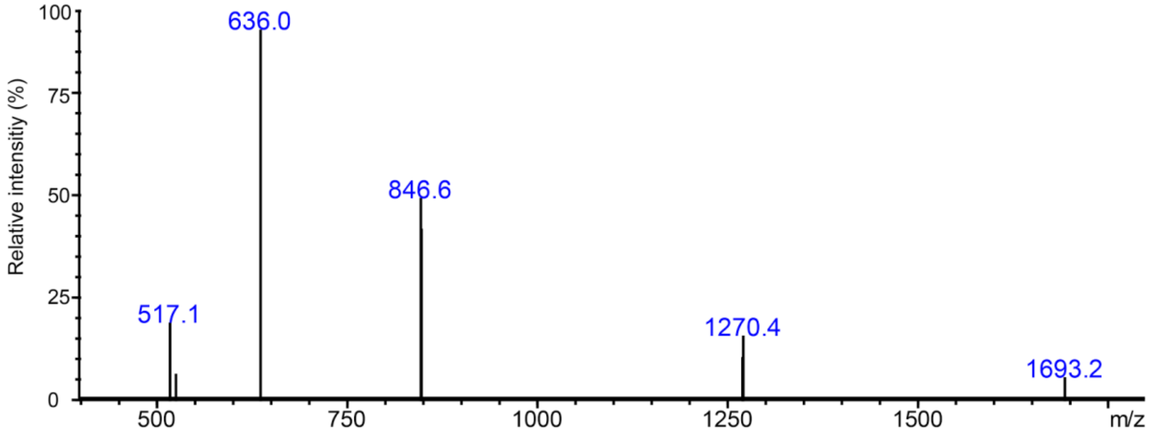


NC_EEH_D2

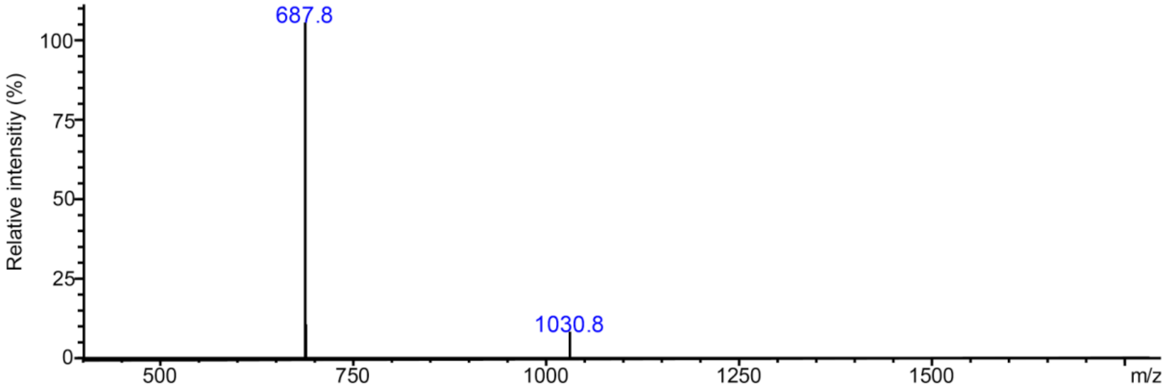


NC_cH_LH_R_D1

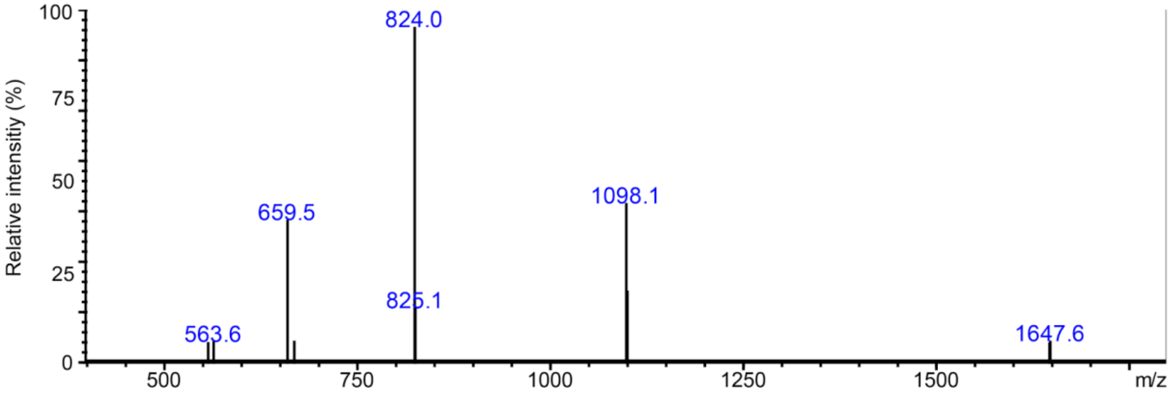
NC_cHHH_D1



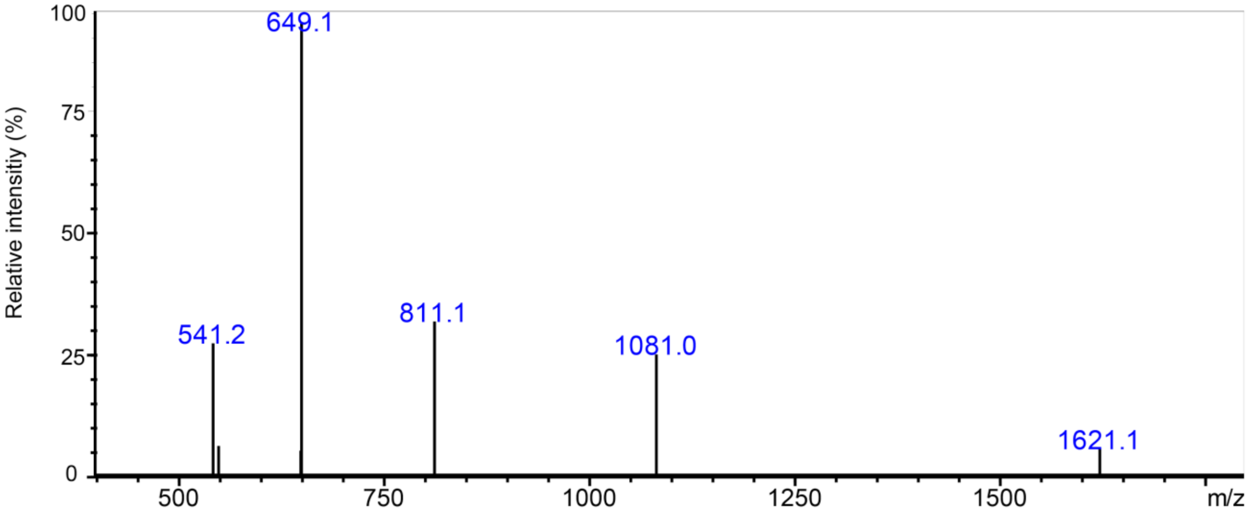
NC_cEE_D1



NC_EHE_D1



NC_cHH_D1


Relative intensity (%)
100
75
50
25
0
649.1
541.2
811.1
1081.0
1621.1
500
750
1000
1250
1500
m/z

Supplementary References

1. O'Meara, M. J. *et al.* Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. *J. Chem. Theory Comput.* **11,** 609–622 (2015).
2. Leaver-Fay, A. *et al.* Scientific benchmarks for guiding macromolecular energy function improvement. *Methods Enzymol.* **523,** 109–143 (2013).
3. Sievers, S. A. *et al.* Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. *Nature* **475,** 96–100 (2011).
4. Rohl, C. A., Strauss, C. E. M., Misura, K. M. S. & Baker, D. Protein structure prediction using Rosetta. *Methods Enzymol.* **383,** 66–93 (2004).
5. Mandell, D. J., Coutsias, E. A. & Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nat. Methods* **6,** 551–552 (2009).
6. Coutsias, E. A., Seok, C., Jacobson, M. P. & Dill, K. A. A kinematic view of loop closure. *J. Comput. Chem.* **25,** 510–528 (2004).
7. Khatib, F. *et al.* Algorithm discovery by protein folding game players. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 18949–18953 (2011).
8. Shapovalov, M. V. & Dunbrack, R. L., Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **19,** 844–858 (2011).
9. Yee, A. *et al.* An NMR approach to structural proteomics. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 1825–1830 (2002).