# Constraint Network Analysis - Thermostability

**Date:** 2022-08-30
**Tags:** *1_Drylab 3_CNA*
**Created by:** Fathoni Musyaffa



(_ Written by Christina Gohlke_)

(_Last update: 2022.08.30_)

<Section | Constraint Network Analysis>

{Rigidity analyses|operation} were performed using the {CNA|:software:} {5.0|version} (10.1002/elsc.200800043, 10.1021/ci400044m, 10.1093/nar/gkt292, 10.1002/wcms.1311). CNA functions as a front-end and back-end to the graph theory-based software {FIRST|:software:} {6.2|version} (Floppy Inclusions and Rigid Substructure Topography) (10.1002/prot.1081). Applying CNA to biomolecules aims at identifying their composition of rigid clusters and flexible regions, which can aid in the understanding of the biomolecular structure, stability, and function (10.1021/ci400044m, 10.1093/nar/gkt292, 10.1002/wcms.1311). As the mechanical heterogeneity of biomolecular structures is intimately linked to their diverse biological functions, biomolecules generally show a hierarchy of rigidity and flexibility (10.1002/elsc.200800043). In CNA, biomolecules are modeled as constraint networks in a body-and-bar representation, which has been described in detail by Hespenheide *et al.* (10.1088/0953-8984/16/44/003). A fast combinatorial algorithm, the pebble game, counts the bond rotational degrees of freedom and floppy modes (internal, independent degrees of freedom) in the constraint network (10.1103/PhysRevLett.75.4051). To monitor the hierarchy of rigidity and flexibility of biomolecules, CNA performs thermal unfolding simulations by consecutively removing noncovalent constraints (hydrogen bonds, including salt bridges) from a network in increasing order of their strength (10.1073/pnas.062492699, 10.1002/prot.20745, 10.1002/prot.22946).

To improve the robustness and investigate the statistical uncertainty, we carried out CNA on ensembles of network topologies generated from {MD|ensemble generation program} simulations ($ENT^{MD}$). {Average stability characteristics|calculated method} and {standard error of the means (SEM)|statistical validation method} were calculated over all ensembles.

The CNA software is available under academic license at https://cpclab.uni-duesseldorf.de/index.php/Software and the CNA web server is accessible at https://cpclab.uni-duesseldorf.de/cna .

<Section | Thermostability predictions>

# Constraint Network Analysis - Thermostability

**Date:** 2022-08-30
**Tags:** *1_Drylab 3_CNA*
**Created by:** Fathoni Musyaffa



A hydrogen bond energy $E_{HB}$ is computed by a modified version of the potential by Mayo *et al.* (10.1002/pro.5560060622). For a given network state  = f(*T*), hydrogen bonds with an energy {$E_{HB}$ lte $E_{cut}$()|retaining criteria} are removed from the network at temperature *T*. In the present study, thermal unfolding simulations were carried out by decreasing $E_{cut}$ from {-0.1 kcal mol$^{-1}$|$E_{start}$} to {-6.0 kcal mol$^{-1}$|$E_{end}$} with a step size of {0.1 kcal mol$^{-1}$|$E_{step}$}. As $E_{cut}$ can be converted to a temperature *T* using the linear equation introduced by Radestock *et al.* (Eq. 1)(10.1002/prot.22946, 10.1002/elsc.200800043), the range of $E_{cut}$ is equivalent to increasing the temperature from {302 K|$T_{start}$} to {420 K|$T_{end}$} with a step size of (2 K|$T_{step}$}. Because hydrophobic interactions remain constant or become even stronger as the temperature increases, the number of hydrophobic tethers was kept unchanged during the thermal unfolding simulation (10.1016/s0065-3233(08)60377-0).

$$T = \frac{-20K}{kcal*mol^{-1}} * E_{cut} + 300K \qquad \text{(eq. 1)}$$

<Section | Global and local indices for thermostability predictions>

From the thermal unfolding simulations, CNA computes a comprehensive set of indices to quantify biologically relevant characteristics of the protein's stability. Global indices are used for determining the rigidity and flexibility at a macroscopic level; local indices determine the rigidity and flexibility at a microscopic level of bonds (10.1002/jcc.23122). The cluster configuration entropy $H_{type2}$ is a global index that has been introduced by Radestock and Gohlke (10.1002/elsc.200800043). Here, we applied $H_{type2}$ {as a measure for global structural stability of proteins|reason for application}. The stability map $rc_{ij}$ is a local index that has been introduced by Radestock and Gohlke (10.1002/elsc.200800043). We applied $rc_{ij}$ {as a measure for local structural stability of proteins|reason for application}.

The cluster configuration entropy $H_{type2}$ was used to identify the phase transition temperature $T_p$ during the thermal unfolding simulation. At $T_p$, the protein switches from a rigid (structurally stable) to a floppy (unfolded) state. However, the percolation behavior of protein networks is usually more complex, and multiple phase transitions can be observed (10.1002/elsc.200800043, 10.1002/prot.22946, 10.1016/j.jbiotec.2012.01.027, 10.1371/journal.pone.0130289, 10.1038/srep17908, 10.1371/journal.pcbi.1004754, 10.1021/acs.jcim.9b00954). Initially, the protein network is dominated by a giant rigid cluster, and $H_{type2}$ is low because of the limited number of possible ways to configure a system with this cluster. When the giant rigid cluster starts to decay or stops to dominate the network, $H_{type2}$ jumps. There, the network is in a partially flexible state with many ways to configure a system consisting of many small clusters. In order to determine $T_p$, a double sigmoid fit was applied to an $H_{type2}$ versus $T(E_{cut})$ curve as done previously (10.1002/elsc.200800043, 10.1002/prot.22946, 10.1016/j.jbiotec.2012.01.027, 10.1371/journal.pone.0130289, 10.1038/srep17908,

# Constraint Network Analysis - Thermostability

**Date:** 2022-08-30
**Tags:** *1_Drylab 3_CNA*
**Created by:** Fathoni Musyaffa



10.1371/journal.pcbi.1004754, 10.1021/acs.jcim.9b00954, 10.1101/2020.06.02.129015), and $T_p$ taken as that $T$ value associated with the largest slope of the fit. The rigid cluster decomposition was visually inspected by {VisualCNA|:software:}(10.1093/bioinformatics/btv139), an easy-to-use {PyMOL|:software:} plug-in that allows setting up CNA runs and analyzing CNA results linking data plots with molecular graphics representations. VisualCNA is available under an academic license from https://cpclab.uni-duesseldorf.de/index.php/Software.

During a thermal unfolding simulation, the stability map $rc_{ij}$ indicates for all residue pairs the $E_{cut}$ value at which a rigid contact $rc$ between the two residues $i$ and $j$ (represented by their C atoms) is lost; $rc$ exists as long as $i$ and $j$ belong to the same rigid cluster $c$ of the set of rigid clusters (10.1002/jcc.23122). Thus, $rc_{ij}$ contains information about the rigid cluster decomposition cumulated over all network states during the thermal unfolding simulation. The sum over all entries in $rc_{ij}$ yields the chemical potential energy due to non-covalent bonding, obtained from the coarse-grained, residue-wise network representation of the underlying protein structure (10.1371/journal.pone.0130289). In the present study, we applied the neighbor stability map $rc_{ij,neighbor}$ to {investigate short-range rigid contacts|reason for application}. For this, as done previously (10.1371/journal.pone.0130289, 10.1021/acs.jcim.9b00954), $rc_{ij}$ was filtered such that only rigid contacts between two residues that are at most {5 Å | distance cutoff } apart from each other were considered.