PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Dynamics and thermodynamic properties of CXCL7 chemokine

Charles A. Herring,[1] Christopher M. Singer,[1] Elena A. Ermakova,[2] Bulat I. Khairutdinov,[2] Yuriy F. Zuev,[2] Donald J. Jacobs,[1,3] and Irina V. Nesmelova[1,3]*

[1] Department of Physics and Optical Science, University of North Carolina, Charlotte, North Carolina 28223

[2] Kazan Institute of Biochemistry and Biophysics, Kazan 40111, Russia

[3] Center for Biomedical Engineering, University of North Carolina, Charlotte, North Carolina 28223

ABSTRACT

Chemokines form a family of signaling proteins mainly responsible for directing the traffic of leukocytes, where their biological activity can be modulated by their oligomerization state. We characterize the dynamics and thermodynamic stability of monomer and homodimer structures of CXCL7, one of the most abundant platelet chemokines, using experimental methods that include circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy, and computational methods that include the anisotropic network model (ANM), molecular dynamics (MD) simulations and the distance constraint model (DCM). A consistent picture emerges for the effects of dimerization and Cys5-Cys31 and Cys7-Cys47 disulfide bonds formation. The presence of disulfide bonds is not critical for maintaining structural stability in the monomer or dimer, but the monomer is destabilized more than the dimer upon removal of disulfide bonds. Disulfide bonds play a key role in shaping the characteristics of native state dynamics. The combined analysis shows that upon dimerization flexibly correlated motions are induced between the 30s and 50s loop within each monomer and across the dimer interface. Interestingly, the greatest gain in flexibility upon dimerization occurs when both disulfide bonds are present, and the homodimer is least stable relative to its two monomers. These results suggest that the highly conserved disulfide bonds in chemokines facilitate a structural mechanism that is tuned to optimally distinguish functional characteristics between monomer and dimer.

INTRODUCTION

Chemokines are important regulatory proteins associated virtually with all physiologic or pathologic processes that involve immune system cell trafficking.[1] The interest to understand how chemokines function increased greatly after they were identified as key players in inflammation-related and infectious disease processes, including autoimmune disease, HIV/AIDS, and cancer.[2,3] While the assessment of chemokines in a variety of biological assays provided a wealth of information on their functional activity and new activities are continually being discovered, the structural biology approach allows for the mechanistic understanding of chemokine functioning.

There are 47 members in the family of human chemokines. Chemokine act on receptors that belong to the family of G-protein-coupled receptor (GPCR) superfamily.[4] Multiple chemokines bind a single receptor and a single chemokine binds multiple receptors, resulting in a broad range of both unique and shared signaling events.

Significant efforts are being made toward understanding the receptor binding affinity and specificity in the chemokine system to facilitate the development of targeted therapeutic agents.[5–7] It is expected that all chemokines have a similarly shaped receptor binding site, because chemokines are small molecules (8–10 kDa) with essentially the same three-dimensional fold.[8] Each chemokine monomer has a disordered N terminus of six to ten amino acid residues, followed by a long N-loop (about 10 residues), a 3_{10} α-helical turn, a three-stranded β-sheet, and a C-terminus α-helix. Furthermore, chemokines have two disulfide bonds joining four conserved cysteine residues per monomer (with only a few exceptions where chemokines have three or only one disulfide bond). The current model of receptor activation by a chemokine ligand, which is based on structure-function studies of several chemokines, proposes two consecutive interaction sites: the N-loop of chemokine ligand interacting with N-terminal residues of the receptor (Site-I) and the N-terminal residues of chemokine ligand interacting with extracellular loop/transmembrane residues of the receptor (Site-II),[5,9] although a recent crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine vMIP-II suggests that the binding interface may be even more extensive.[7] Accordingly, differences in amino acid composition of the N-terminal and N-loop regions contribute to the affinity and specificity of the receptor binding by a chemokine ligand. Additionally, experimental evidence showing that both the Site-I and Site-II of chemokine/receptor interaction comprise an extended and comparatively flexible region within the chemokine molecule, suggests a mechanistic role of protein dynamics in receptor binding.[10–20]

An interesting question is whether the differences in protein dynamics or flexibility in the apo structures through intermolecular interresidue couplings are important in the selection of chemokine-receptor interactions. Indeed, the importance of coupling between Site-I and Site-II in the chemokine ligand for receptor activation has been demonstrated for the CXCL8 chemokine,[21,22] suggesting similar effects are present in other chemokines because of the similarity of their structures and disulfide bond locations.[22]

Here, we employ a combination of experimental and computational methods to investigate the dynamics, the inter-residue couplings, and the role of disulfide bonds on chemokine stability, and flexibility in the monomer and dimer forms of CXCL7 chemokine. CXCL7 is a strong chemoattractant of neutrophils, and thus plays an important role in inflammation, blood clotting, and wound healing.[23] It activates neutrophils via the interactions with cell surface receptors CXCR1 and CXCR2, but has much higher affinity to CXCR2.[24–26] Similar to other chemokine ligands of CXCR2 receptor (for example, CXCL1-3 and CXCL5-8), CXCL7 has a characteristic three N-terminal amino acid residue motif, ELR, involved in receptor binding and cell activation[27] that

was shown to be highly dynamic.[16] By analogy to other CXCR2 chemokine ligands, the residues in the N-loop (Ile8-His15) and the 30s loop (Gly26-Val34) are expected to be implicated in receptor binding as they define relatively dynamic regions of the protein.[14,16] Besides the chemotactic activity regulated through the receptor, it was found that the CXCL7 variant thrombocidin-1 (TC-1), missing just the two C-terminus amino acids, possesses strong antimicrobial activity.[28] While a positive patch on protein surface formed by several lysine and arginine residues (Lys17, Lys41, Arg54, Lys56, Lys57, Lys 60, and Lys61) was found to be essential for the antimicrobial activity of folded TC-1,[29,30] functional differences (CXCL7 being inactive vs. TC-1 being active) of the two proteins with essentially the same monomer structure were explained by a higher and less restricted mobility of C-terminal residues in TC-1, leading to the increased possibility of interactions with the negatively charged bacterial membranes.[14] Thus, protein dynamics plays an important role in both the receptor activation and antimicrobial activity of CXCL7.

This report presents our initial effort in establishing a workflow that can be extended to the entire chemokine family. We demonstrate the utility of the phenomenological minimal distance constraint model (mDCM) that has emerged as a fast and robust all-atom ensemble-based computational method for elucidating thermodynamic and mechanical properties of proteins[31–33] by investigating the dynamics and stability of CXCL7 chemokine. For the first time, we extend the mDCM approach to account for solvation effects semiempirically by extracting free energy differences upon dimerization from experimental circular dichroism (CD) experiments. This latter connection provides a route for quantitatively comparing backbone flexibility and residue pair couplings between the monomer and dimer forms to elucidate the effect of the protein-protein interface. We compare mDCM predictions to low-frequency protein motions determined by the anisotropic network model[34–36] (ANM) and ANM, molecular dynamics (MD), and mDCM results are compared with experimental nuclear magnetic resonance (NMR) relaxation data. Taken together, we arrive at a consistent picture of CXCL7 dynamics and stability in regards to the effect of dimerization and role played by the highly conserved disulfide bonds.

MATERIALS AND METHODS

Experimental procedures

CXCL7

CXCL7 was expressed as 6-His-Trx-DDDDK fusion protein in Origami B (DE3) PLysS *E. coli* cells. Metal affinity chromatography was used to purify 6His-tagged fusion protein. The cleavage of 6-His-Trx-tag was performed using enterokinase light chain (Applied

Biological Materials Inc., Canada). CXCL7 chemokine was finally purified using heparin affinity chromatography. The molecular weight of CXCL7 has been confirmed by mass spectrometry. The purity of protein was more than 98% according to SDS-PAGE analysis.

CD spectroscopy

CXCL7 samples were prepared in H_2O at pH 3.8. Two concentrations of CXCL7, 43 and 12 μM, were used. CD measurements were performed on a Jasco-715 spectropolarimeter, equipped with a Peltier temperature control system, using quartz glass cell with a path length, l, of 1 mm. Far-UV CD spectra were recorded in the range of 190 to 260 nm The corresponding buffer baseline was subtracted from each spectrum. Spectra were recorded using a 50 nm/min scan rate with a 2 s response and 1 nm bandwidth. To follow the thermal unfolding of CXCL7, CD spectra were collected at 3–5°C intervals in the temperature range from 25°C to 90°C. At each temperature, the sample was equilibrated for 10 min. The unfolding of CXCL7 was only partially reversible. The increase of temperature above 85°C caused partial precipitation of CXCL7. Therefore, the temperature-induced unfolding of CXCL7 was monitored within 25–85°C temperature range. Reported spectra are averages of four to six scans and are expressed as mean-residue molar ellipticity, $[\theta]$, calculated by using the relation

$$[\theta] = \frac{M_0 \theta_\lambda}{100 \cdot C \cdot l} \qquad (1)$$

in which M_0 is the mean residue molar mass, θ_λ is the measured ellipticity in degrees, and C is the total concentration of protein. The value of $M_0 = 109.3$ g/mol was obtained by dividing the molecular weight of CXCL7 with the number of amino acid residues in it.

CD data analysis

CD data were analyzed assuming a two-state dissociation and a two-state unfolding model describing the equilibrium between CXCL7 dimers, D, monomers, M, and unfolded monomers, U, as shown in Eq. (2),

$$D \overset{K_D}{\leftrightarrow} 2M \overset{K_U}{\leftrightarrow} 2U \qquad (2)$$

Here, K_D is the equilibrium dissociation constant for dimers and K_U is the equilibrium unfolding constant given by Eq. (3),

$$K_D = \frac{[M] \cdot [M]}{[D]}, \quad K_U = \frac{[U]}{[M]} \qquad (3)$$

At each temperature, the observed ellipticity is a weighted average of the ellipticities of unfolded, $[\theta]_U$, and folded, $[\theta]_F$, states:

$$[\theta] = f_U [\theta]_U + f_F [\theta]_F, \qquad (4)$$

where f_U and f_F are the fractions of unfolded and folded CXCL7, respectively. Using Eq. (3), the observed ellipticity can be expressed as a function of CXCL7 monomer concentration and the equilibrium unfolding constant:

$$[\theta] = [\theta]_U + ([\theta]_F - [\theta]_U) \cdot \left(1 - \frac{K_U[M]}{C}\right) \qquad (5)$$

The total molar concentration of CXCL7 in solution at each temperature is given by Eq. (6),

$$C = [M] + 2[D] + [U] = \frac{2}{K_D}[M]^2 + (1 + K_U)[M] \qquad (6)$$

The monomer concentration can be readily obtained by solving quadratic Eq. (7),

$$[M] = -\frac{K_D}{4} - \frac{K_D \cdot K_U}{4} + \sqrt{\frac{K_D^2(1+K_U)^2}{16} + \frac{C \cdot K_D}{2}} \qquad (7)$$

The dependence of $[\theta]_{obs}$ on temperature, T, is determined by the temperature dependence of equilibrium dissociation and unfolding constants, which can be expressed using Gibbs free energies of dimerization, ΔG_D, and unfolding, ΔG_U, as follows:

$$K_D = \exp\left(-\frac{\Delta G_D}{RT}\right), \quad K_U = \exp\left(-\frac{\Delta G_U}{RT}\right) \qquad (8)$$

Here, R is the gas constant, 1.986 cal·K^{-1} mol^{-1}. For a two-state process, the Gibbs free energy is given by a Gibbs-Helmholtz equation,

$$\Delta G(T) = \Delta H\left(1 - \frac{T}{T_R}\right) + \Delta C_P(T - T_R) + \Delta C_P T \ln\frac{T_R}{T}, \qquad (9)$$

where T_R is the temperature of the transition, which in our model is the dissociation of CXCL7 dimers or the unfolding of CXCL7 monomers. ΔH and ΔC_P are the changes of enthalpy and heat capacity, respectively, associated with the transition.

The melting temperature, T_m, i.e. the transition temperature of unfolding, was determined by fitting the observed ellipticity $[\theta]$ at different temperatures to Eq. (5) taking into account Eqs. (7–9). $[\theta]_F$ was approximated as linear function of temperature. The ellipticity of unfolded CXCL7 $[\theta]_U$ was determined from the CD spectrum acquired in 6M GuHCl $[\theta]_U$.

NMR spectroscopy

Uniformly ^{15}N-enriched CXCL7 was prepared using a 95% H_2O/5% D_2O mixture at pH 3.8. 8% 2-

chloroethanol was added to dissociate the CXCL7 into monomers.[14,16,37] NMR experiments were carried out at 37°C on a Bruker Avance 950 and 700 MHz spectrometers at David H. Murdock Institute (Kannapolis, NC). The assignments were obtained by the comparison to published data[14,16] and using an ^{15}N-HSQC-TOCSY data. A total of 63 residues out 66 were definitively assigned and used for NMR relaxation analysis. For the longitudinal R1 spin-relaxation experiments, the relaxation delays were 10, 50, 100, 200, 300, 400, 600, 800, and 1000 ms at 950 MHz and 10, 50, 100, 300, 600, and 800 ms at 700 MHz. For the transverse R2 spin-relaxation experiments, the relaxation delays were 17, 33, 51, 68, 85, 102 ms at 950 MHz. ^{1}H-^{15}N heteronuclear NOE data were acquired in an interleaved manner.

NMR spectra were processed using NMRPipe[38] and analyzed with the Sparky 3.113 program.[39] R1 and R2 relaxation rates were obtained either by fitting the peak intensities or peak volumes to a single exponential decay. The heteronuclear NOE were calculated as the ratio of intensities estimated from spectra acquired with or without the NOE effect.

The ^{15}N R1, R2, and NOE relaxation data were analyzed to describe the internal dynamics of CXCL7 using the model-free Lipari-Szabo approach[40] by the relax-GUI program[41] and built-in automated protocol of model selection and optimization.[42] Relaxation data for each residue were fitted to the different models, where the appropriate model was chosen for each residue by evaluating the quality of the fit.[43] The optimal overall tumbling time, τ_c, was found on a per residue basis using the isotropic spectral density function to be 4.2 ± 0.6 ns.

Theory and computation

CXCL7 structure

The X-ray crystallographic structure of CXCL7 tetramer[44] was used (pdb code 1NAP). Monomer structures in the X-ray structure have small differences leading to an overall asymmetric structure. Chains A and B were selected from the tetramer to represent the CXCL7 monomers or dimer. For the mDCM analysis, the monomer and dimer structures were processed using the program MOE from Chemical Computing Group, Inc. Protonation states were then determined using the H++ web server[45] held fixed for all subsequent analysis. To study the effects of the disulfide bonds, each monomer (A and B) was prepared in four cases: both Cys5-Cys31 and Cys7-Cys47 or none of the disulfide bonds form, or only one of the two disulfide bonds form.

ANM analysis

The ANM[34–36] was used to quantify low-frequency vibrational modes that correspond to dominant large-scale motions. In ANM, the three-dimensional structure of a protein is represented as an elastic network where nearby Cα atoms are connected by harmonic springs with equilibrium lengths defined by the input structure. In this work, a cutoff distance of 15 Å was adopted to define nearest neighbors, and the force constant was identical for all springs. Based on the packing of Cα atoms, ANM provides information about the direction of fluctuations in residue positions. As a general characteristic, core regions of a protein characterized by higher density of Cα − Cα nearest neighbors will exhibit less mobility, while surface regions and especially protruding loops will tend to have greater mobility. As such, the large-scale protein motions captured by ANM are because of the overall shape of the protein structure, and not sensitive to the disulfide bond configurations. Consequently, the nature of the global motions of the backbone exhibited in a monomer is expected to depend on whether the monomer chain is isolated or joins the dimer. Moreover, by the coarse graining nature of the ANM, the global motions are insensitive to perturbations in molecular interactions that do not change atomic packing.

The ANM vibrational modes of the protein are determined by a Hessian matrix H.[34,36]

$$H = \begin{pmatrix} h_{11} & h_{12} & \ldots & h_{1N} \\ h_{21} & h_{22} & \ldots & h_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ h_{N1} & h_{N1} & \ldots & h_{NN} \end{pmatrix} \qquad (10)$$

The elements of H are obtained from the 2nd order derivative of the residue interaction harmonic potential V with respect to coordinates x, y, and z:

$$h_{ij} = \begin{pmatrix} \dfrac{\partial^2 V}{\partial x_i \partial x_j} & \dfrac{\partial^2 V}{\partial x_i \partial y_j} & \dfrac{\partial^2 V}{\partial x_i \partial z_j} \\[2mm] \dfrac{\partial^2 V}{\partial y_i \partial x_j} & \dfrac{\partial^2 V}{\partial y_i \partial y_j} & \dfrac{\partial^2 V}{\partial y_i \partial z_j} \\[2mm] \dfrac{\partial^2 V}{\partial z_i \partial x_j} & \dfrac{\partial^2 V}{\partial z_i \partial y_j} & \dfrac{\partial^2 V}{\partial z_i \partial z_j} \end{pmatrix} \qquad (11)$$

The diagonalization of Hessian matrix derived from the ANM potential provides $3N-6$ non-zero modes characterized by the eigenvalues and the corresponding eigenvectors. The mean square fluctuations of the displacement of the i-th residue is given by the value of $<\Delta R_i^2>$ obtained by taking a pseudo inverse of the Hessian matrix as a singular value decomposition.[36] Note that the x, y, and z components of a given residue are grouped together. The Debye-Waller or B-factor for residue i, which is correlated to the mean-

square fluctuation, is calculated using the following relation:

$$B_i = 8\pi^2/3 < \Delta R_i^2 > \qquad (12)$$

The cross correlation coefficient between residue fluctuations is calculated using the first 20 modes according to the Eq. (13):

$$C_{ij} = \frac{<\Delta R_i \cdot \Delta R_j>}{(<\Delta R_i \cdot \Delta R_i><\Delta R_j \cdot \Delta R_j>)^{1/2}} \qquad (13)$$

All calculations were carried out using the automated ANM server[36] for the monomer and dimer structures separately. The results for monomers were averaged over chains A and B.

mDCM analysis

The mDCM[31–33] is used to characterize thermodynamic stability and flexibility of CXCL7. The mDCM maps the three-dimensional all-atom structure of the native state of a protein onto a graph, where vertices represent atoms and edges represent different interactions, such as covalent bonding, hydrogen bonds (H-bonds), salt bridges and atomic packing through torsion interactions. A salt bridge is treated as a special type of H-bond. A free energy decomposition scheme is employed, where an energy and entropy contribution is assigned to each interaction, and moreover, each interaction is modeled as one or more distance constraint. Note that covalent bonds are always present in the graph because they do not fluctuate, whereas torsion distance constraints and H-bond distance constraints fluctuate, which generates a large ensemble of accessible constraint networks. The total energy and conformational entropy must be calculated for each network. Although correlations between degrees of freedom (DOF) invalidate an additive reconstitution of free energy components defined by a free energy decomposition,[46,47] a nonadditive reconstitution is possible[48] by applying the pebble game algorithm[49] to a constraint network to identify independent distance constraints. A lowest upper bound estimate for the total conformational entropy is obtained by preferentially adding the most restrictive independent interactions present in the network, while energy contributions are added over all interactions present. This approach proves to be sufficient to provide high accuracy within the context of a phenomenological model comprising three free parameters.

The H-bonds identified from the native structure are allowed to break and reform, but no non-native H-bonds enter into the calculation. Torsion constraints fluctuate between native-like and disordered states to model good or poor atomic packing. Good atomic packing lowers energy and conformational entropy, while poor atomic packing increases energy and conformational entropy. To generate an ensemble of constraint networks that characterize conformational fluctuations about the native structure, and to calculate the free energy of the protein, the number of H-bonds, N_{hb}, and the number of good packing constraints, N_{gp}, are order parameters used to specify a macrostate of the protein. Given the maximum number of H-bonds, N_{hb}, and the maximum number of good packing torsion constraints, N_{hb}^{max}, as determined from an input structure, the number of possible microstates, Ω_m, for a given macrostate (N_{hb}, N_{gp}) is given as $\left(N_{hb}^{max}! N_{gp}^{max}!\right) / \left(N_{hb}!\left(N_{hb}^{max} - N_{hb}\right)! N_{gp}!\left(N_{gp}^{max} - N_{gp}\right)\right)$. The mixing entropy S_m is then given by $S_m = R\ln(\Omega_m)$.

By employing Monte Carlo sampling, the conformational entropy for a macrostate is estimated as:

$$S_c\left(N_{hb}, N_{gp} | \delta_{gp}\right)$$
$$= R\left[\delta_{gp} Q_{gp} + \delta_{pp} Q_{pp} + \delta_{hb}^{max} \sum_{k=1}^{N_{hb}^{max}} \left(1 + \frac{1}{8} E_k\right)\langle q_k n_k\rangle\right], \qquad (14)$$

where $R\delta_{gp}$ is the entropy of an independent good packing constraint, $R\delta_{pp}$ is the entropy of an independent poor packing constraint and δ_{hb}^{max} is the entropy of the weakest possible independent H-bond constraint. The dimensionless parameters δ_{pp} and δ_{hb}^{max} have been determined previously to be equal to 2.53 and 1.89, respectively,[31] and they are transferable across globular proteins. Q_{gp} and Q_{pp} represent the average number of independent good packing or poor packing constraints found in the sub-ensemble of constraint networks specified by (N_{hb}, N_{gp}). The variable n_k equals 0 when the k-th native H-bond is broken or 1 when it is present. The quantity q_k counts the number of independent distance constraints associated with the k-th H-bond when present. The energy, E_k, for the k-th H-bond (or salt bridge) is limited to the range from 0 to −8 kcal/mol,[50] with corresponding entropy being a linear function of its energy. The Q_{gp}, Q_{pp}, and q_i have unique values because of the preferential ordering imposed on Eq. (14) where the lowest entropy constraints are placed in the network before other constraints with higher entropy. It is worth mentioning that the good/poor packing constraints adapted in this work were previously referred to as native/disordered torsion constraints. However, due to using multiple input structures from MD simulations, some of the non-native fluctuations that occur in the native state basin are captured.

The free energy landscape is calculated by combining total energy and entropy contributions for each macrostate using a free energy functional. Accordingly, the free energy of a macrostate of the protein is constructed as:
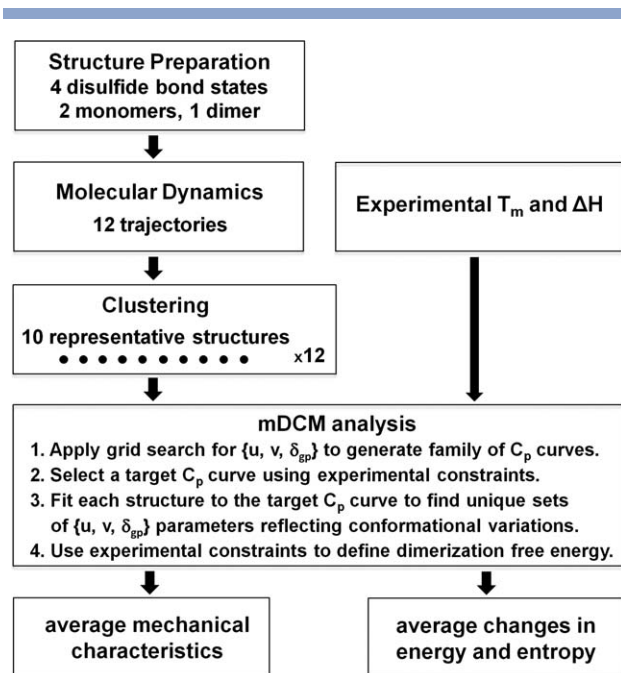
Figure 1
Flowchart of steps of the combined CD/MD/mDCM analysis.

$$G(N_{hb}, N_{gp})$$
$$= U(N_{hb}) - uN_{hb} + vN_{gp} - T\lfloor S_m(N_{hb}, N_{gp}) \qquad (15)$$
$$+ S_c(N_{hb}, N_{gp}|\delta_{gp})\rfloor,$$

where U is the average intramolecular H-bond energy given as $U = \sum_{i=1}^{N_{hb}} E_i n_i$. In Eq. (15) the three parameters, (u, v, δ_{gp}) are adjusted for thermodynamic predictions to best agree with experimental measurements. The u parameter represents a favorable (i.e. a negative value) effective protein-solvent H-bond energy, which competes against the formation of intramolecular H-bonds. The v parameter is a favorable energy that occurs when good atomic packing forms, and $R\delta_{gp}$ was described above as the entropy of an independent good packing constraint. For an exhaustive calculation on the CXCL7 dimer, about 2^{720} constraint networks are required, ignoring non-native interactions for a given input structure. Only a few hundred random samples of constraint networks per macrostate are sufficient to obtain accurate estimates of the free energy function given in Eq. (15). Previously, the phenomenological parameters u, v, and δ_{gp}, were usually determined by fitting the predicted heat capacity curve based on one input structure to the target heat capacity curve from differential scanning calorimetry (DSC). In this work, the melting temperature, T_m, is known from CD, and multiple input structures from MD simulation are considered. The flowchart shown in

Figure 1 summarizes the steps of the combined CD/MD/mDCM analysis as explained below.

Quantitative stability/flexibility relationships

The backbone Flexibility Index (FI) and Cooperativity Correlation (CC) are two types of Quantitative Stability/Flexibility Relationship (QSFR) metrics used to characterize mechanical properties of a protein.[32,33,51] The DOF that are of interest are φ and ψ backbone torsion angles, and χ side-chain torsion angles corresponding to rotatable bonds. A rotatable bond is mechanically equivalent to a hinge, but it may become rigid due to constraints that arise from atomic packing and intramolecular H-bonds. In a flexible region, we denote the number of rotatable bonds that rotate like a hinge as H, and the number of independent disordered torsions as A. Then $h_i = A/H$ is the density of independent DOF within that region with all of its H rotatable bonds assigned the same A/H value. Within a rigid region, we denote the number of locked rotatable bonds as L, and the number of redundant constraints as B. Then $l_i = B/L$ is the density of redundant constraints within this over-constrained region with all of its L bonds assigned the same B/L value. For a given microstate that defines a particular constraint network, FI is calculated as $f_i = (h_i - l_i)$ for the i-th rotatable bond. FI indicates when an a priori rotatable bond remains flexible within a region having an excess DOF, or it becomes locked within an over-constrained region where there are an excess number of constraints. An isostatic rigid region has $f_i = 0$ indicating no excess DOF or constraints. Since a rotatable bond can either rotate or be locked, any localized region of a protein cannot be simultaneously flexible and rigid. However, the reported FI is an ensemble average of f_i over different constraint networks. Therefore, an FI$\approx$0 could reflect a region that is rarely isostatic within any particular constraint network, but over an ensemble of possible constraint networks, the region could fluctuate between under- and over-constrained. In this case, the rotatable bond on average reflects a region close to being isostatic.

The CC matrix is calculated similarly to FI and quantifies the rotatable bonds that flex together in a continuous motion, or lock together to form a rigid cluster. If the i-th and j-th rotatable bonds fall within in the same flexible region, the matrix elements CC_{ij} and CC_{ji} are assigned the value of h_i (also equal to h_j). Likewise, when the i-th and j-th rotatable bonds fall within the same over-constrained region, the matrix elements CC_{ij} and CC_{ij} are assigned $-l_i$ (also equal to $-l_j$). If a pair of rotatable bonds falls within an isostatic region, or they are not members of the same flexible or rigid region, the matrix elements CC_{ij} and CC_{ji} are assigned 0.

To average FI and CC, each constraint network is assigned a thermodynamic statistical weight based on

solving the mDCM for the free energy landscape as described above. In proteins with clear two-state behavior, it is often convenient to calculate the average mechanical properties within the native state ensemble (a restricted set of macrostates) at the melting temperature (i.e. a single temperature). In this report, we average over the full ensemble of macrostates because this avoids technical complexities in merging results from different representative structures. The resulting temperature dependence of the average mechanical properties (e.g. FI and CC) allows questions about CXCL7 stability to be addressed.

Filtering is applied to CC-difference plots using a signal beyond noise ratio (SBNR) to quantify statistically significant results, rather than show raw differences. SBNR is defined as follows. The standard deviation, σ, is first calculated for all raw differences between corresponding pixels, x. If $x > \sigma$, then $SBNR(x) = (x - \sigma)/\sigma$. If $x < -\sigma$ then $SBNR(x) = (x + \sigma)/\sigma$ otherwise x is below noise, and $SBNR(x) = 0$. According to this definition, when the magnitude of SBNR is more than 1 the results are statistically significant, while variations within one standard deviation are completely filtered out.

### mDCM parameterization on multiple input structures

Residue pair couplings are sensitive to the H-bond network within a protein structure. The asymmetric X-ray crystal structure of a CXCL7 tetramer (pdb code 1NAP) showed differences in H-bonding details among the four monomers. Therefore, MD simulation was used to generate multiple input structures for the mDCM.[52] The MD simulation was performed using Gromacs 4.5.5[36,37] in the NVT ensemble with the AMBER99SB-ILDN force field.[38] The protein systems were solvated by adding 10.0 Å of TIP3P water[39] in a periodic cubic box, with counter ions added to neutralize the net charge. Before production, the starting structure was obtained by minimizing the potential energy of the system, followed by 1 ns of NPT and 1 ns of NVT equilibration. Pressure (1 atm) was regulated using the extended ensemble Parrinello-Rahman approach[40] and temperature (300 K) was controlled by a Nose-Hoover temperature coupling.[41,42] A cutoff distance of 10.0 Å was used for van der Waals interactions, and the Particle-Mesh-Ewald[43] method was employed to account for the long-range electrostatic interactions. All bonds to hydrogen atoms in proteins were constrained using LINCS,[44] and bonds and angles of water molecules were constrained by SET-TLE[45] allowing for a time step of 0.002 ps. A 100 ns trajectory was collected for each structure. A total of 2,000 evenly spaced frames from each trajectory were clustered using the KCLUST module[49] from the MMTSB tool set[50] based on the root-mean-square deviation (RMSD) of all heavy atoms. The cluster radius was adjusted so that the ten largest clusters represented 85% or more of the 2000 conformations. This method allows some of the non-native fluctuations that occur in the native state basin to be captured. We found that non-native H-bonds appear <8% of the time over an MD trajectory, suggesting errors in predicted in quantities should be within typical experimental errors of binding energy estimates.

The number of input structures to mDCM consisted of 40 dimers and 80 monomers in total by considering each of the disulfide bond states for the dimer and monomer (A and B chains) separately. A simple grid search over the three dimensional parameter-space was performed to obtain different sets of parameters u, v, and δ_{gp} that would yield heat capacity curves with a peak at the experimentally determined T_m, while consistent with experimental ΔH of unfolding of a monomer. Initially, the grid searches were performed for 20 monomer structures and 10 dimer structures to identify a consensus target heat capacity curve for the monomer and dimer forms. The standard deviation in heat capacity peak height relative to the mean value was within 25%. Each input structure (with all disulfide bonds present) produced a unique set of parameters u, v, and δ_{gp}, needed to fit the target heat capacity in order to account for conformational differences. The parameters u, v, and δ_{gp}, were determined for structures with one or both disulfide bonds removed by computationally *reforming* the missing disulfide bonds so that the same target heat capacity for when all disulfide bonds are present can be used again. Following prior works,[53,54] these new parameters are kept fixed when the disulfide bonds that were just added are subsequently removed because the global features of the protein conformation are unaltered.

Estimating free energy of solvation upon dimerization

The mDCM phenomenological parameters, which in part reflect solvation contributions to the free energy, are robust across all structures from a MD trajectory because quantities such as heat capacity, average number of H-bonds and mechanical properties, all being a function of temperature, are not affected by arbitrary constant shifts in the energy or entropy parameters. However, because differences in free energy, energy and entropy are sought between the monomer/dimer states and disulfide bonding states, relative parameter shifts between representative structures from the MD simulation must be accounted for. Importantly, changes in solvent exposed surface area due to displacement of solvent at the dimer interface must be considered. To circumvent modifying the mean-field treatment of solvation effects in the mDCM, a model independent approach is employed that uses the empirical values of enthalpy, entropy, and temperature, $\{\Delta H_a, \Delta S_a, T_a\}$, of dimer association based on fitting to the CD measurements assuming dimer association is a two state process with $\Delta H_a = T_a \Delta S_a$.
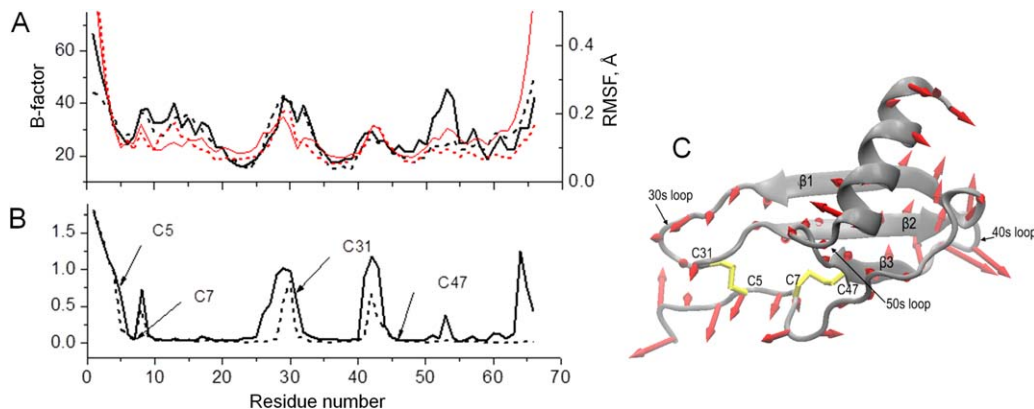
Figure 2

ANM description of CXCL7 dynamics. **A**: Crystallographic (black solid line) and ANM-calculated using all modes (black dashed line) isotropic B-factors for the dimer. The red line shows the RMSF for monomer (solid line) and dimer (dashed line) from MD simulation. **B**: Low-frequency vibrations in CXCL7 monomer (solid line) and dimer (dashed line) calculated as a sum over the first twenty modes. **C**: Directions of residue fluctuations are shown as red arrows that annotate the CXCL7 monomer structure based on the 20 slowest modes. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

The free energy of a given representative structure relative to an arbitrary reference state is given by $G_k(T) = H_k(T) - TS_k(T) + (h_k - Ts_k)$, where $\{h_k, s_k\}$ are constant shifts in (energy, entropy) that reflect small variations in solvation free energy between MD frames. Since we have 80 different free energy curves for monomers with all disulfide bonds present and 40 such curves for dimers, we average over the respective numbers (80 and 40) and obtain two average free energy curves that serve as reference targets for the monomer and dimer structures. Note that deviations in free energy for a specific structure relative to the reference free energy curve at any given temperature is partly because the $\{h_k, s_k\}$ shifts are poorly modeled by mDCM in addition to structural differences. Therefore, uncertainty in mDCM predictions are obtained by adjusting the $\{h_k, s_k\}$ parameters so that each free energy curve derived from a MD frame matches the reference curve the best it can through a least squares error fitting. Any deviation remaining reflects the intrinsic uncertainty in the mDCM prediction combined with the conformations explored by the MD simulation. This congruency process yields two free energy curves (for monomer and dimer) resulting in mDCM predictions with appropriate error bars. Although the change in free energy upon formation of the dimer interface is not accounted for (yet), we know $\Delta H_a = H^{(D)} - 2H^{(M)}$ and $\Delta S_a = S^{(D)} - 2S^{(M)}$, where (D) and (M) respectively represent the dimer and monomer forms. At T_a let $\Delta H_a^{(0)} = H_{mDCM}^{(D)}(T_a) - 2H_{mDCM}^{(D)}(T_a)$ and $\Delta S_a^{(0)} = S_{mDCM}^{(D)}(T_a) - 2S_{mDCM}^{(D)}(T_a)$, and using the freedom to add a constant energy and entropy shift between the monomer and dimer, it follows $\Delta H_{pred}^{(0)} = H_{mDCM}^{(D)} - \Delta H^{(0)} + \Delta H_a$ and $S_{pred}^{(0)} = S_{mDCM}^{(D)} - \Delta S^{(0)} + \Delta S_a$. Then the pre-

dicted free energy curve for the dimer is given by $G_{mDCM}^{(D)} = H_{pred}^{(D)} - TS_{pred}^{(D)}$. Although this procedure fixes the difference in free energy curves for the dimer minus two monomers, it is still possible to shift the overall energy and entropy of the monomer because this arbitrariness does not modify any response property, or any measurable differences in energy and entropy of interest.

RESULTS

Global and local dynamics of CXCL7 predicted by anisotropic network model

It has been shown that the ANM describes low frequency global motions in proteins that are often functionally relevant,[55,56] including chemokine proteins.[57] Here, we applied the ANM to elucidate and compare the principal motions of the protein backbone of CXCL7 in monomer and dimer forms. Figure 2(A) compares the crystallographic isotropic B-factors to that calculated by Eq. (12). A correlation coefficient of 0.78 between the experimental and theoretical B-factors indicates that the ANM model captures the large-scale dynamics of CXCL7 well. For residues Asp51-Arg54 located at the end of the loop connecting the C-terminal α-helix to beta-sheet structure and at the beginning of the C-terminal α-helix, the calculated B-factors are larger than crystallographic. We also show root-mean-square fluctuation (RMSF) values obtained from MD trajectory for comparison. The RMSF trace from MD closely resembles the trend shown by B-factors, and there is little difference between the monomer and dimer. Because the first twenty modes account for almost half of the structural variance of the backbone (48% in a monomer and 44% in a dimer),

they are used for subsequent ANM analysis related to low-frequency vibrations.

The ANM identifies the N- and C-termini and the loops connecting secondary structure elements as the most mobile regions [Fig. 2(B)] within the monomer. Of these regions, the loops connecting the first to the second or the second to the third beta-strands, called the 30s and the 40s loops respectively, are almost as mobile as the N- or C-terminal residues. Figure 2(C) shows the direction of residue motions within the monomer. Comparing the low-frequency vibrations in the CXCL7 monomer with that in the dimer in Figure 2(B) shows that the slow motions in the 30s and 50s loops diminish upon dimerization. It is interesting that the behavior of two cysteine residues located on the N-terminus of CXCL7, Cys5 and Cys7, is different. While Cys5 fluctuates with large amplitude, Cys7 demonstrates negligibly small fluctuations in the first 20 modes. The behavior of disulfide bonds formed by these cysteines is also different. The fluctuations of the Cys7-Cys47 disulfide bond are minimal in accordance with a general view that the disulfide bond stabilizes the structure and limits the motion in a protein, while the fluctuations of the Cys5-Cys31 disulfide bond are large. Since Cys5 and Cys7 are located on the unstructured N-terminus of CXCL7, the difference in their behavior likely results from the different properties of their disulfide bond partners. Cys31 is located on a relatively large loop, connecting beta-strands 1 and 2, i.e. the 30s loop, whereas Cys47 is located on the beta-strand 3, which is part of a beta-sheet structure. Therefore, Cys47 has a limited ability to fluctuate and limits the mobility of Cys7 via the disulfide linkage. In contrast, being positioned on a flexible loop, Cys31 has the ability to undergo large fluctuations without limiting the mobility of Cys5. Residues within the 30-s loop, the residue Cys5, and the residues proximal to Cys5, all move in the same direction as indicated by arrows representing the direction of motion [Fig. 2(C)].

The dimerization of CXCL7 essentially eliminates the fluctuations of C-terminal residues and of the loop connecting the C-terminal α-helix and beta-strand 3, the 50s loop. This result is expected because the C-terminal helices have extensive contacts in the dimer that limit each other's ability to undergo large fluctuations. Interestingly, the mobility of the 40s loop and some residues in the 30s loop in a dimer is also reduced. The likely reason for reduced fluctuations of Gly26 and Lys27 in the 30s loop upon dimer formation is the electrostatically favorable interaction of Lys27 with Asp66 of the second monomer. Reduced fluctuation of the 40s loop likely results from long-range correlations, because it is not located at the intermonomer interface.

To capture these correlations, the cross correlation for the slow motions (first 20 modes) were calculated using Eq. (13). The results for CXCL7 monomer and dimer are given in Supporting Information [Fig. S1(A,B)]. The cross-correlation value ranges from −1 to 1, and are positive when the residues move in the same direction and negative when they move in the opposite direction. The correlation coefficient $C_{ij} = 0$ corresponds to uncorrelated fluctuations. The cross correlation map of monomer [Supporting Information Fig. S1(A)] shows highly positively intracorrelated regions including the three beta-strands and the α-helix. Additionally, a strong positive correlation exists between the residues in the second and third beta-strands and the N-terminal residues Leu6-Lys7 or N-loop residues Lys9-Thr11, and between the N-loop residues Ser12-Gly13 and the residues Asp49-Pro53 in the 50s loop. Strong negative correlations are observed between the residues His15-Lys17 forming a 3_{10} alpha-helical turn and the N-terminal residues Leu6-Lys7, N-loop residues Lys9-Thr11, or the residues located in beta-strands. The residues of the C-terminal α-helix are also negatively correlated with the rest of the molecule, suggesting that the C-terminus α-helix is moving back and forth like a local hinge motion.

Overall features of the cross correlation map of the monomer in the CXCL7 dimer remain similar to that of the monomer alone, with one significant change [Supporting Information Fig. S1(B)]. A new positive correlation appears between residues Ile14-Lys17, forming the 3_{10} alpha-helical turn, and residues Ile37-Ile46 that comprise the 40s loop and adjacent residues on the second and third beta-strands, which explains the reduced fluctuation on the 40s loop upon dimerization. In addition, strong positive correlations of the first and second beta-strands in one monomer and the C-terminal α-helix of the second monomer appear.

CXCL7 dynamics determined from NMR relaxation analysis

The internal dynamics of CXCL7 was assessed experimentally by using NMR relaxation data R_1, R_2, and NOE (Nuclear Overhauser Effect) at ^1H frequency of 950 MHz and R_1 at 700 MHz. Experimental relaxation data are shown in Figure 3. The data were analyzed using a Lipari-Szabo model-free approach.[40] The S^2 order parameters, describing the amplitude of the fast time scale (ps to ns) internal motion, and the R_{ex} term, indicative of slow time scale (μs to ms) conformational rearrangements, derived from this analysis are shown in Figure 3 below the experimental NMR relaxation data. The average generalized order parameter is 0.73 ± 0.18 for all residues and 0.82 ± 0.13 for the residues of secondary structure elements. The termini residues and the residues in the N-loop, 30s loop, and some residues in the 50s loop, representing flexible parts of the molecule, show small S^2 values. However, several residues in the N-terminus and N-loop show relatively high S^2 values and
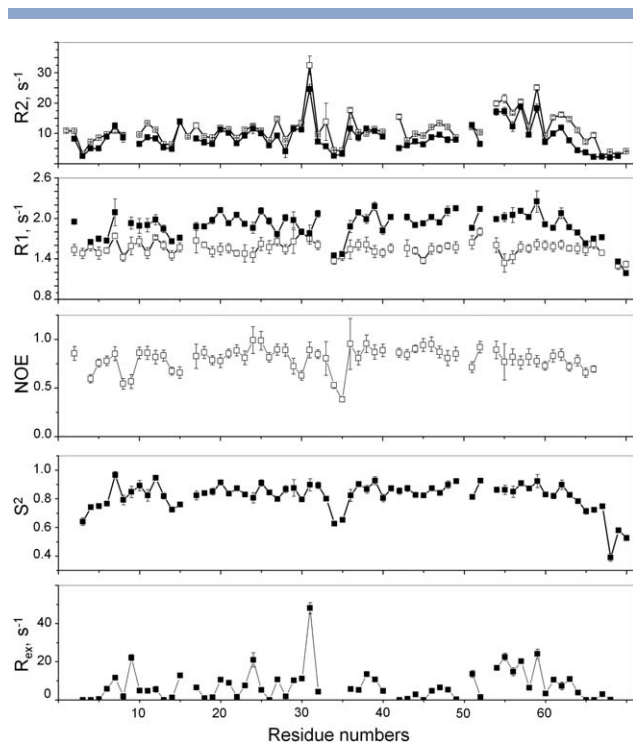
Figure 3

NMR relaxation analysis of CXCL7 dynamics. ^{15}N R_1, R_2, and NOE values and motional order parameters, S^2, and conformational exchange terms, R_{ex} for monomeric CXCL7 are plotted against the amino acid sequence of CXCL7.

all residues in the 40s loop show high S^2 values, comparable to those of neighboring secondary structure elements.

CXCL7 residues in the 30s loop and C-terminal α-helix showed particularly high R_{ex} values. It is worth noting that in CXCL12 and CCL27 chemokines, large R_{ex} contribution were caused by monomer-dimer equilibrium.[17,18] We exclude the possibility of residual dimerization as a source of R_{ex} in CXCL7 because the model fitting for the residues in the first beta-strand that participate in the intermonomer contact did not require a significant R_{ex} contribution. A large R_{ex} contribution observed for the residues in the C-terminal α-helix, having the S^2 parameter above average at the same time, may be related to the fluctuations of the helix as a whole about its average position. In the dimer, the motion of helices is restricted because the two helices form contacts with each other. Apparently, in the monomer, when this restriction is absent, the C-terminal α-helix experiences a slow (μs-ms) motion relative to the beta-sheet structure. The values of B-factors and RMSF found for the residues in the helix [Fig. 2(A)] are relatively higher than in the beta-sheet region, particularly in beta strands 1 and 2, indicating that the helix residues undergo larger fluctuations.

The residue Cys31 is particularly interesting, because it participates in the disulfide bond and shows the largest R_{ex} term at the same time. Similarly, large R_{ex} contribution has been earlier observed for Cys31 in CXCL7[14,16] and for the corresponding cysteine and surrounding residues in other chemokines (CXCL8,[13] CCL11,[12,58] CCL23,[15] CCL24,[10] and viral chemokine VMIP-II[59]), but not in CCL26[11] or CCL4.[60] In CCL11, disulfide bond isomerization was proposed as a possible source of the significant dynamics on the μs-ms time scale.[12] In the case of CXCL7, it is likely that Cys31 participates in a slow motion because of the fact that it is located in the 30s loop region, rather than due to the isomerization of the disulfide bond, because Cys5 forming the disulfide bond with Cys31, does not show significant exchange contribution. This result fits well with the ANM observation of the movement of the 30s loop and N-terminal part of the protein in the same direction.

CD spectroscopic analysis of CXCL7 unfolding

The temperature-induced unfolding of CXCL7 was monitored by measuring far-UV CD spectra (190–260 nm) in the range of temperatures from 25 to 85°C. The ellipticity at 222 nm was used to access structural changes in CXCL7 with temperature and to derive the value of T_m.

Figure 4(A) shows CD spectra of 43 μM CXCL7 recorded at different temperatures. The trace of CD spectrum collected at 25°C (folded protein) demonstrates a mixture of features typical of alpha-helical (negative bands at 208 nm and 222 nm and a positive band at approximately 192 nm) and beta-sheet conformations (a negative band at 218 nm and a positive band at 195 nm), as well as of the unstructured protein (a minimum at 200 nm). This is consistent with CXCL7 having a mixed alpha-helical and beta-sheet structure typical to all chemokines.[8] Increasing temperature from 25 to 55°C has almost no effect on the CXCL7 far-UV CD spectra, indicating the stability of its structure within this temperature range. However, the further increase of temperature above 55°C causes significant changes in CD spectra [Fig. 4(A)]. These changes include a gradual decrease of the peak intensity at 218–222 nm and the increasing ellipticity around 200 nm. Since there is no significant structural difference between CXCL7 monomers and dimers, the observed changes reflect the structural transition in CXCL7 related to temperature-induced unfolding of the protein. Nonetheless, CXCL7 is remarkably stable. A significant amount of secondary structure remains present even at the highest accessible temperature. CD spectra collected using a 12 μM CXCL7 sample showed similar changes (data not shown).
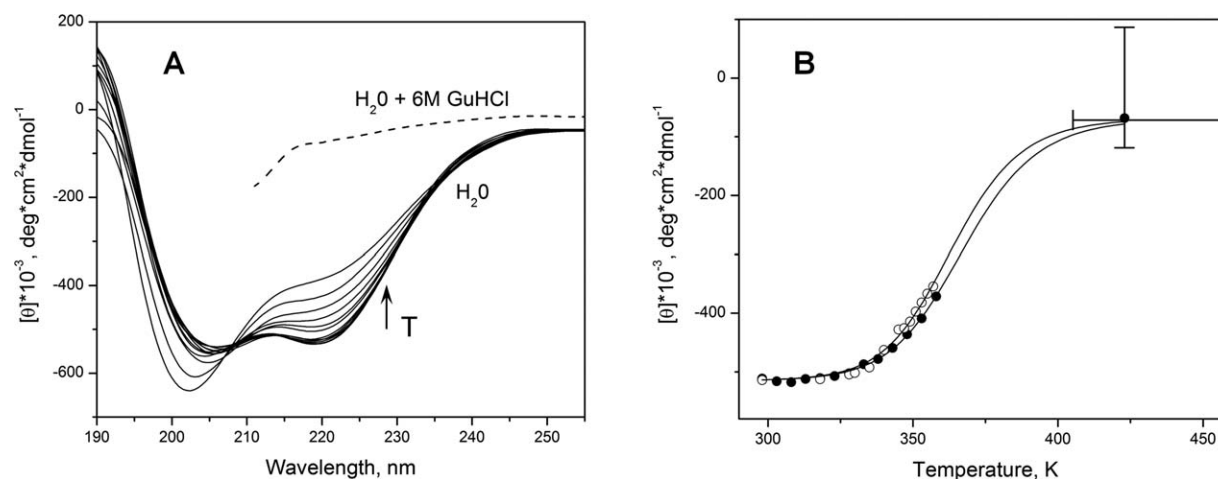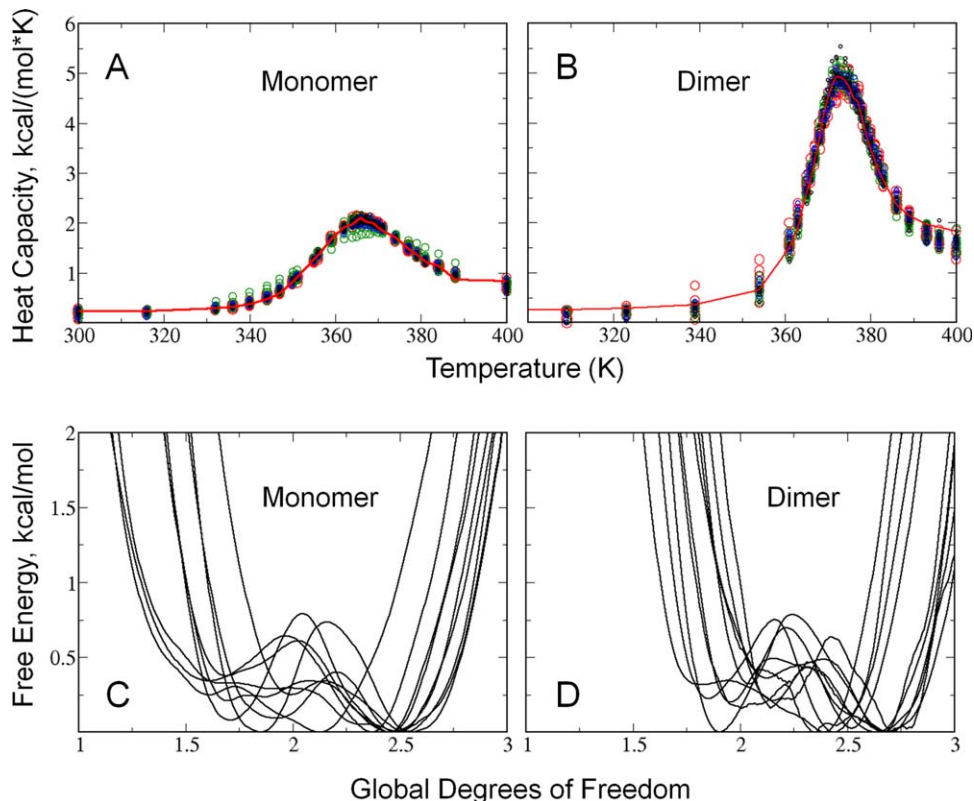
Figure 4

Temperature-induced dissociation and unfolding of CXCL7. **A**: Far-UV CD spectra of 43 μM CXCL7 dissolved in water (solid lines) at pH 5.0 acquired at different temperatures ranging from 25°C to 85°C. Dashed line represents the CD spectrum recorded on the same sample upon addition of 6 M GuHCl. **B**: Ellipticity at 222 nm as a function of temperature, determined from the CD spectra recorded at the CXCL7 concentration of 43 μM (solid circles) and 12 μM (open circles). Solid lines represent global fits of experimental data according to the Eqs. (5, 7, 8, and 9). Error bars represent the ranges of the ellipticity of unfolded protein and of the temperature, at which CXCL7 becomes fully unfolded that were used to determine the error in the determination of T_m.

The value of $[\theta]$ at 222 nm was used to follow the unfolding of CXCL7. This choice of wavelength allowed accessing a cumulative change of alpha-helical and beta-sheet secondary structure. Figure 4(B) shows the temperature dependence of the ellipticity at 222 nm measured at two concentrations of CXCL7, 43 μM (solid circles) and 12 μM (open circles). Although the thermal unfolding of CXCL7 is incomplete, both dependences clearly show the onset of the unfolding transition.

Additionally, a weak dependence on protein concentration is observed, such that the unfolding curve recorded at the smaller concentration of CXCL7 is shifted toward lower temperatures. This effect is likely related to the association-dissociation of CXCL7 dimers. The association constant of $5.3 \pm 2.0 \times 10^{-5}$ M has been reported for CXCL7.[61] Consequently, CXCL7 exists in monomer–dimer equilibrium at concentrations used here, with approximately 50% protein in a monomeric form at 43 μM and 74% monomeric protein at 12 μM. The decrease of transition temperature as the protein concentration decreases is consistent with the idea that the unfolding transition is coupled to the dissociation of CXCL7. Note that the reported association constant provides a lower estimate for the concentration of monomers, because the experimental conditions at which it was determined (50 mM phosphate buffer at pH = 7.0 and in the presence of 100 mM NaCl) favor the association of CXCL7 as compared to the experimental conditions used here (H_2O at pH = 5.0). The weak dependence of unfolding curves on the concentration of CXCL7 may be because of the abundance of CXCL7 monomers.

The temperature dependence of CD spectra shows an isodichroic point at 208 nm that is indicative of a two-state process. CD data were analyzed using the three-state model described by Eq. (2), in which we assume that the unfolding of CXCL7 is a two-state process and the dissociation of dimers into monomers occurs with no structural changes. The assumption of a two-state unfolding process is reasonable, because no clear intermediate states are observed, although we cannot fully eliminate the possibility of a continuous unfolding based on our experimental data. However, the purpose of CD measurements was to obtain an estimate of melting temperature of the protein. We define the melting temperature as the temperature corresponding to a midpoint of transition, and in this regard either two-state or continuous unfolding yield the same result. The melting temperature of CXCL7 was determined by the global fitting experimental data to Eqs. (5) and (7–9). The value of $[\theta]_U$, included in Eq. (7), was determined from CD spectrum of CXCL7 unfolded in the presence of 6 M GuHCl [Fig. 4(A), dashed line]. Additionally, we varied $[\theta]_U$ and the temperature, at which CXCL7 becomes fully unfolded, in the range indicated in Figure 4(B) by vertical and horizontal bars, respectively, to estimate the error in determining T_m. Solid lines in Figure 4(B) show the best fit results. The T_m determined by the described procedure is found to be equal to 367 ± 4 K at 43 μM and 364 ± 4 K at 12 μM CXCL7. We also extracted the ΔH of unfolding of a monomer to be 30 kcal/mol, and −30 kcal/mol is the value we estimated for the enthalpy of association at the temperature of $T_a = 314$ K.

Figure 5

Heat capacity and free energy of CXCL7. Heat capacity curves for the monomer (**A**) and dimer (**B**). The solid red line is the target heat capacity curve that was initially generated for the case of all disulfide bonds present. The data points are the fits to this target curve for 80 representative structures for monomers and 40 representative structures for the dimer. The corresponding free energy landscapes that result from good fits to the heat capacity are shown for the top 10 representative structures for monomer (**C**) and dimer (**D**). The lowest point in each of the free energy landscapes have been set to be zero so that all the curves can be easily compared. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

## CXCL7 stability/flexibility predicted by the minimal distant constraint model

Heat capacity curves and free energy landscapes of CXCL7 by mDCM

The target heat capacity curves and spread of the various heat capacity fits to the targets for the monomer and dimer are shown in Figure 5(A,B). The mDCM parameters are $u = -1.72 \pm 0.26$ kcal/mol, $v = -0.47 \pm 0.19$ kcal/mol and $\delta_{gp} = 1.20 \pm 0.04$ for the monomer, and $u = -1.44 \pm 0.27$ kcal/mol, $v = -0.11 \pm 0.20$ kcal/mol and $\delta_{gp} = 1.43 \pm 0.23$ for the dimer. These parameters indicate that the native state of the dimer has greater conformational entropy per residue than that of the monomer (as seen before packing is not as strong,[62]) while the destabilizing effect of H-bonding from protein to solvent plays a lesser role. The maximum heat capacity, Cp_{max}, of CXCL7 dimer is more than twice that of the monomers [Fig. 5(B)], suggesting that the dimer is more cooperative. Typical free energy landscapes are shown in Figure 5(C,D) for the monomer and dimer. The landscapes with barriers are indicative of a first order transition; however a small fraction of the landscapes do not show a barrier that separates the native and unfolded basins. Moreover, when present, the barriers are low. While the model parameters can be adjusted to produce a clear two-state behavior in all representative structures, it would require a larger change in enthalpy of folding than estimated from CD experiments; hence the unfolding transition is not very cooperative.

Backbone flexibility of CXCL7 monomer and dimer

The backbone flexibility, quantified by the FI at temperatures 300K and 350K, is shown in Figure 6(A) for the monomer and in Figure 6(B) for the dimer. The temperature 350K is chosen to identify mechanically less stable regions in CXCL7 at elevated temperatures because at this temperature the protein remains folded with a weakened H-bond network. In general, the backbone flexibility gradually increases in a nearly uniform fashion throughout the protein as temperature increases. In the monomer, the N-terminus and the 50s loop are flexible, while the three beta-strands, C-terminal α-helix, 3_{10}
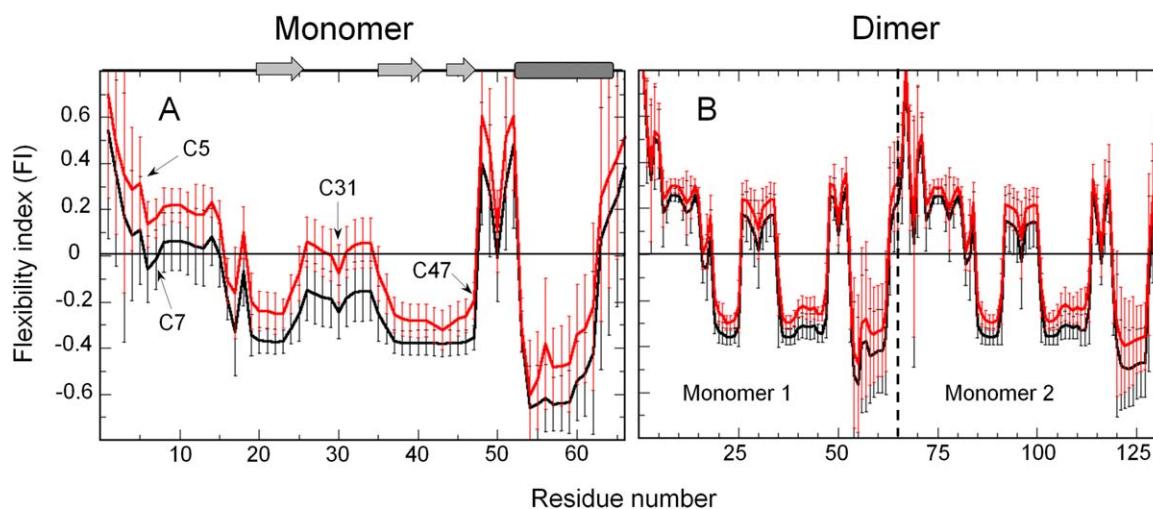
Figure 6

Flexibility index of CXCL7. The backbone flexibility is plotted using the flexibility index at $T = 300K$ and 350K for the monomer (**A**) and dimer (**B**). The secondary structure elements[16] are shown above the plots. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

alpha-helical turn and the 40s loop have a negative FI indicating that they are over-constrained. To a lesser degree, the 30s loop is over-constrained at biological temperatures and it becomes flexible just before the protein unfolds. This suggests that the two disulfide bonds mechanically stabilize the 30s loop region. The specific role of disulfide bonds in stabilizing the protein will be addressed below.

Dimerization enhances mechanical stability in an interesting way. Most H-bonds formed at the interface connect pre-existing rigid substructures within the monomers, which lowers the FI in regions already over-constrained in the monomer. A histogram showing where H-bonds form across the interface in relation to the flexibility index is given in Figure 7 for the case of both disulfide bonds are present. On average, about 13 H-bonds are formed at the interface. Despite these additional interfacial H-bonds, backbone flexibility increases in the 30s loop and for several residues at the N-terminus, including Cys5. Backbone flexibility along the C-terminal α-helix is not significantly affected by dimerization, even though the two α-helices align in the CXCL7 dimer and new side-chain to side-chain contacts form between the two α-helices across the dimer interface.

Cooperativity correlation in the CXCL7

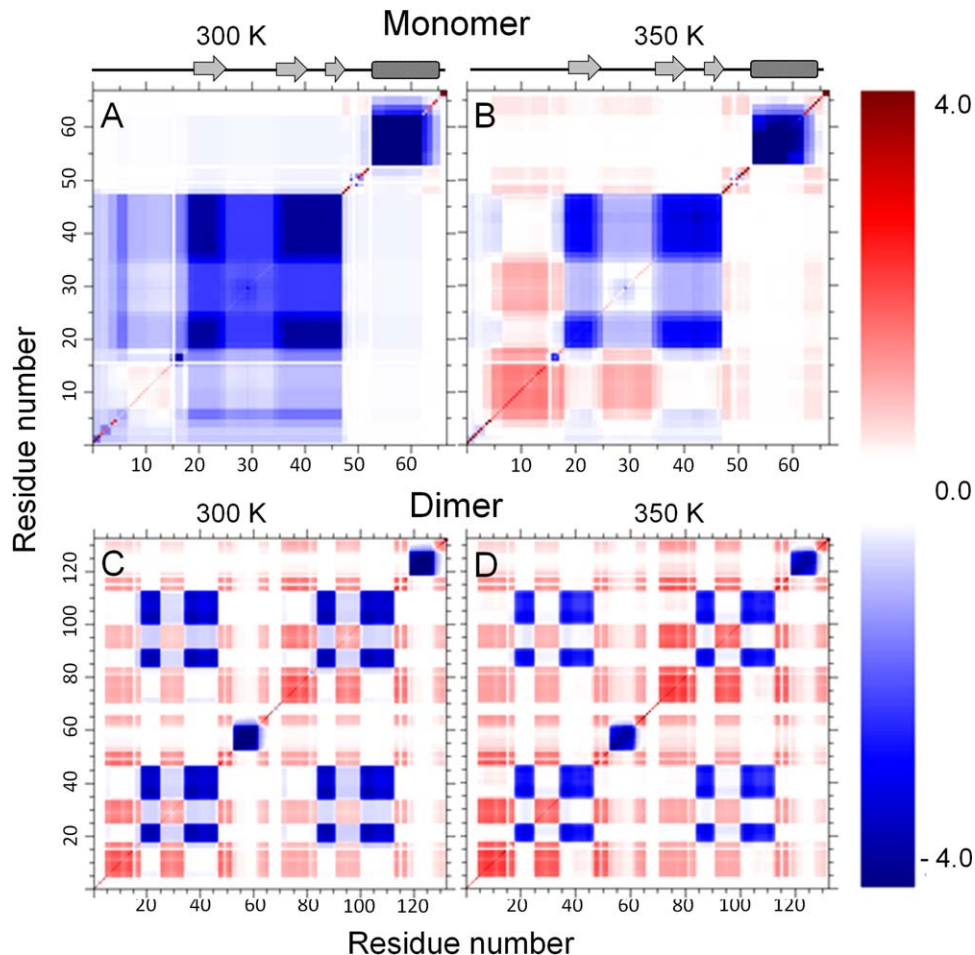Residue-to-residue mechanical couplings (the correlations in DOF of atomic motions) within CXCL7 are shown at $T = 300K$ and 350K in Figure 8(A,B) for the monomer, and in Figure 8(C,D) for the dimer. At $T = 300K$ [Fig. 8(A)], the most rigidly correlated residues are located in the beta-sheet and α-helix secondary struc-tures, but little rigidity propagates between these two secondary structures. Although some correlated flexibility is observed in the N-terminal region between residues Cys5 and His15, overall the monomer forms a fairly rigid molecule at 300K. At 350K the H-bond network weakens, and there is an increase in flexibly correlated motion in the 50s loop hinge region where it couples to the Cys5 to His15 N-terminus section, the 30s loop, and a small section at the end of the C-terminus [Fig. 8(B)]. Surprisingly, the C-terminal α-helix is not strongly coupled to other parts of the molecule, including the beta-sheet on top of which it folds. As such, only at low temperature (300K) does the helix have a weak propensity to
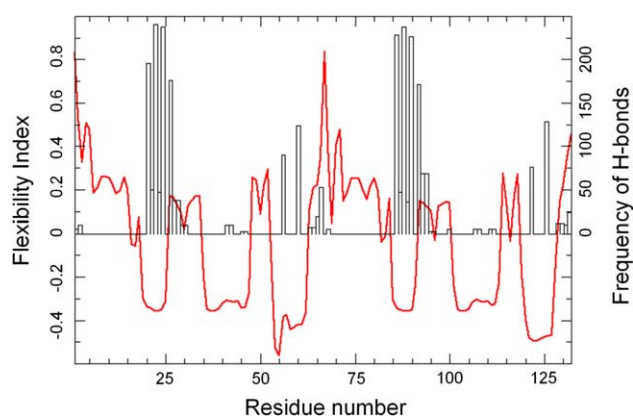
Figure 7

Location of interfacial H-bonds in CXCL7. A histogram showing where H-bonds form across the interface in relation to the flexibility index is shown for the case where both disulfide bonds are present. [Color figure can be viewed in the online issue, which is available at wileyonline-library.com.]

Figure 8

Cooperativity correlation plots for CXCL7 monomer and dimer. The residue-residue mechanical couplings are plotted for the monomer at $T = 300$ K (**A**) and at $T = 350$ K (**B**). This quantity is also plotted for the dimer at $T = 300$K (**C**) and at $T = 350$K (**D**). The coloring scheme is defined by the color bar and is the same for all panels, where white represents no correlation, blue indicates a degree of rigidity correlation, and red indicates a degree of flexibility correlation. The data is shown for the case when all disulfide bonds are present.

intermittently stick to the beta-sheet, but otherwise, it moves independently from the rest of the molecule. We observed a similar behavior in a closely related chemokine CXCL4 before.[63]

Figure 8(C,D) show that the dimerization of CXCL7 greatly increases flexibility correlations. The 30s loop is flexibly correlated with the 50s loop, a large section in the N-terminus, and a small end section of the C-terminus. The rigidity correlation is nearly the same within the beta-sheets and alpha helices as in the monomer, but extends across the monomers, forming a larger rigid core. The dimerization has also enabled new flexibly correlated motions, that is, the 30s and 50s loops in each monomer, as well as the flexible N-terminus and C-terminus regions are flexibly correlated in different monomers in the dimer. At 350K, the correlations pattern is similar with weaker levels of rigidity correlation morphing into flexible correlations. This results from

loop regions having lower density of H-bonds weakening more readily than in the monomer as temperature increases.

The effect of disulfide bonds

The characteristic feature of all chemokines is the presence of two disulfide bonds between highly conserved cysteine residues. To investigate the structural effects of each bond on the dynamics and mechanical couplings in CXCL7 monomer and dimer, we carried out the MD/ mDCM analysis for protein states where one or both bonds were removed. The melting temperature only slightly drops with the removal of one or two disulfide bonds, and destabilization is less in the dimer compared with the monomer (Supporting Information Fig. S2), suggesting that disulfide bonds do not play a critical role in maintaining structural integrity in either the

Table I
Free Energy, Enthalpy, Entropy, and H-Bond Energy of CXCL7 Dimerization Calculated Using the mDCM Approach

Disulfide bonds present	$\Delta G_{D\text{-}2*M}$ ($\Delta H - T\Delta S$)	$\Delta H_{D\text{-}2M}$	$-T\Delta S_{D\text{-}2M}$	$\Delta HB_{D\text{-}2M}$	Mean number of interfacial H-bonds
Cys5-Cys31 Cys7-Cys47	-1.1 ± 0.0 (+ +)	-27.9 ± 0.0 (+ +)	26.8 ± 0.0 (− −)	-1.8 ± 0.2 (+)	13.4
Cys7-Cys47	-2.9 ± 0.4 (−)	-35.9 ± 2.3 (−)	33.0 ± 2.3 (+)	-1.4 ± 0.4 (+)	14.0
Cys5-Cys31	-3.4 ± 0.3 (− −)	-33.6 ± 1.6 (+)	30.2 ± 1.6 (−)	-3.0 ± 0.3 (−)	14.6
None	-2.1 ± 0.4 (+)	-36.8 ± 2.9 (− −)	34.7 ± 2.9 (+ +)	-4.1 ± 0.4 (− −)	16.0

All energies are expressed in kcal/mol. The last column gives the mean numbers of interfacial hydrogen bonds calculated over the MD trajectory. The data is shown for the cases when two disulfide bonds present, either Cys5-Cys31 or Cys7-Cys47 bond removed, or both disulfide bonds removed. (+ +) and (− −) indicates the most and the least favorable case, respectively.

monomer or dimer. This result agrees with reported experimental data on a closely related chemokine, CXCL4.[64]

To quantify changes in thermodynamic stability for each disulfide bond configuration, the change in Gibbs free energy, enthalpy, entropy, total energy of the H-bond network, and the number of interfacial H-bonds formed upon dimerization, is compared at 300K. Table I shows that dimerization of CXCL7 is energetically favorable, and enthalpically driven for all disulfide bond configurations. Because the overall shape of the protein does not change significantly when the disulfide bond is removed, the quantity $-T\Delta S$ listed in Table I is dominated by differences in conformational entropy because of different disulfide bond configurations. Although the conformational entropy per residue is higher in the dimer than in the monomer, the total entropy is reduced in the dimer for all disulfide bond configurations due to solvation effects. Interestingly, differences in $-T\Delta S$ for different disulfide bond configurations span about 7 kcal/mol, which can easily be accounted for by subtle differences in the H-bond network, especially at the inter-monomer interface. However, it is also clear from Table I that the free energy differences do not track differences in the total energy of the H-bond network. The correlation coefficient of 0.11 between the change in free energy and change in total H-bond energy indicates that the changes in conformational entropy (which depends on the location of H-bonds and their microenvironments, i.e. the entire H-bond network), play a critical role in stabilizing the dimer

The end result is just what one would expect: the dimerization is favorable at low temperature because a small number of interfacial H-bonds lower the energy enough to overcome the conformational entropy reduction. At elevated temperature, the entropic component to the free energy becomes more important, and overcomes the lower energy, and hence the dimer dissociates. Because different disulfide bond configurations affect molecular packing and H-bond arrangements, the effect of removing a disulfide bond is non-additive and context dependent. For example, the dissociation temperature is predicted to be the lowest when two disulfide bonds are

present, next lowest when both are removed, and the highest when only Cys7-Cys47 is removed.

Removing one or both disulfide bonds leads to interesting and often non-obvious effects, which are revealed by the change in FI. In the monomer, removing either one or both disulfide bonds will generally increase backbone flexibility throughout the protein. In all cases, the main increase in flexibility appears in the C-terminal α-helix. Overall, no substantial FI differences are noticed when either the first or second disulfide bond is removed, although they do affect different regions differently. In all cases, the 50s loop becomes less flexible. The overall characteristics in the differences are insensitive to temperature (data not shown). In the case of the dimer, removing the disulfide bonds has an opposite effect (decrease in flexibility) in many regions throughout the protein with increases in a few localized regions and in α-helices. Amongst the three possible combinations, removing the Cys7-Cys47 disulfide bond produces the greatest increases in backbone flexibility throughout the protein. FI for both disulfide bonds present and three cases of FI differences for monomer and dimer is shown in Supporting Information Figures S3 and S4, respectively.

These effects are further analyzed in Figures 9 and 10 that show differences in residue–residue couplings for cases where one or both disulfide bonds are removed relative to when both disulfide bonds are present. Figures 9 and 10 show the data points outside one standard deviation (for example, SBNR), hence present only statistically significant changes because of the removal of disulfide bonds. For the monomer at $T = 300$K, Figure 9 shows a dramatic increase in flexibly correlation in the C-terminus α-helix, and regardless of which disulfide bond is removed residues Cys5 through Cys7 are more flexibly correlated with residues from Asn18 through Cys47. However, when only the Cys7-Cys47 disulfide bond is removed, flexibility correlations that couple to residues Cys5 through Cys7 further extend to residues in the N-terminal loop and in the 30s loop. At elevated temperature, little statistically significant correlations remain, and no new patterns of flexibility correlations emerge. The overall increase of flexibility correlations seen in Figure 9 is consistent with an overall increase in conformational entropy within a
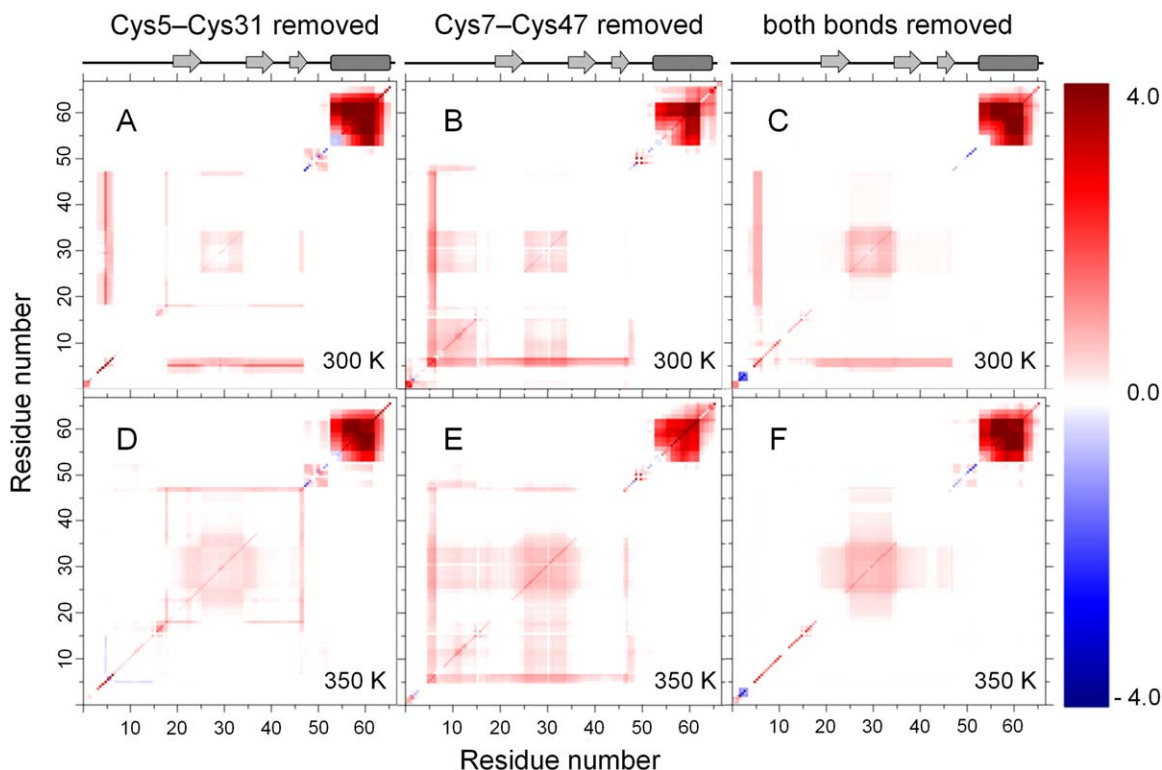
Figure 9

Cooperativity correlation difference plots for CXCL7 monomer. The difference in residue-residue mechanical couplings is shown for the monomer at $T = 300$ K in panels (**A–C**) and at $T = 350$ K in panels (**D–F**). To construct these plots, the residue-residue mechanical couplings for each case of disulfide bond configurations were subtracted from the residue-residue mechanical couplings when both disulfide bonds are present: the Cys5-Cys31 disulfide bond removed (A,D); the Cys7-Cys47 disulfide bond removed (B,E); and both disulfide bonds are removed (C,F). Shown data is filtered based on the signal beyond noise ratio (SBNR) as explained in the text. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

monomer. In the dimer (Fig. 10) removal of one or both disulfide bonds generally increases rigidity correlations. However, as mentioned above, removing the Cys7-Cys47 disulfide bond increases flexibility and increased flexibility correlations, as shown in Figure 10B for $T = 300$K, which further increases [Fig. 10(E)] at $T = 350$K. Flexibility correlation between the 30s and 50s loops is much less in the monomer (c.f. Fig. 8). As seen in Figure 10 there is a net structural stabilization by removing disulfide bonds within the dimer, which is consistent with the thermodynamic analysis (Table I) showing conformational entropy decreases upon removal of disulfide bonds. Interestingly, the 30s and 50s loops are less flexibly correlated to one another when both disulfide bonds are missing than compared to when both disulfide bonds are present at $T = 300$K, and this difference gradually decreases as temperature increases.

DISCUSSION

The results from a variety of computational (ANM, MD and mDCM) and experimental (CD and NMR spec-troscopy) methods taken together lead to a consistent picture for CXCL7 dynamics and stability. However, we first highlight results that have greatest differences between the methods. The ANM analysis shows that the 30s loop and the 40s loop undergo significantly larger fluctuations than the residues in the 50s loop. The NMR relaxation analysis shows that residues Val34-Val36, at the end of the 30s loop and at the beginning of the second beta-strand undergo substantial high-frequency motions, but not other residues in the 30s loop or in the 40s and 50s loops. At the same time, for monomers, the mDCM predicts that the residues in the 50s loop, but not in the 30s and 40s loop, are highly flexible. Some of these differences can be further understood not to be contradictory. For example, the 50s loop is less flexible in the dimer compared with the monomer (using MD/mDCM) and the contributions to mobility from low frequency modes (using ANM) in the monomer is little and reduces to vanishing levels in the dimer. However, there is mobility in this loop region as evidenced by the ANM using all modes including higher frequency motions, and the RMSF from MD, which also partly
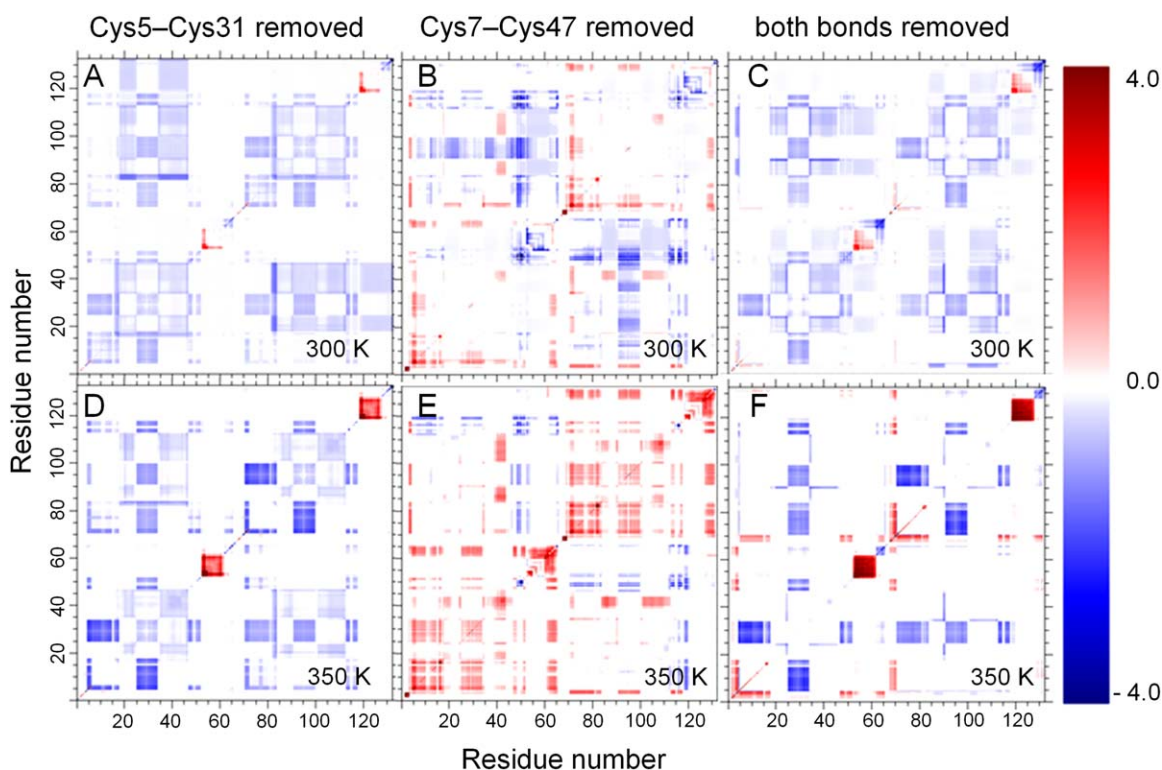
Figure 10

Cooperativity correlation difference plots for CXCL7 dimer. The difference in residue-residue mechanical couplings is shown for the monomer at $T = 300$ K in panels (**A–C**) and at $T = 350$K in panels (**D–F**). These plots were constructed similarly to the plots shown for CXCL7 monomer and demonstrate the SBNR of the difference in residue-residue mechanical couplings for the case of Cys5-Cys31 disulfide bond removed (A,D), the Cys7-Cys47 disulfide bond removed (B,E), and both disulfide bonds removed (C,F). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

covers the first few residues of the C-terminus α-helix. While the 50s loop is flexible (Figs. 6 and 7), consistent with the high frequency motions, residue–residue coupling [Fig. 8(A)] indicates a small degree of rigidity propagates between the C-terminus alpha-helix and the beta-sheet core. This rigidity correlation morphs into a weak flexibility correlation as temperature increases [Fig. 8(B)]. Because these correlations are averages over representative structures that individually have typical correlations that are either flexible or rigid due to thermal fluctuations in the H-bond network, this hinge is viewed as being sticky, having appreciable mobility, but mainly from higher frequency motions.

In the absence of experimental structure of chemokine-receptor complex, mutagenesis studies have shown that the N-terminal residues preceding the first cysteine residue, the N-loop preceding the first beta-strand, and the 30s loop connecting the first and second beta-strands are implicated in receptor binding and activation for most chemokines studied to date[5,9,65] (and references therein). The two-step model of chemokine-receptor binding proposes two sites of interactions between the chemokine ligand and cognate receptor. At Site I, the N-loop of the chemokine interacts with N-terminal residues of the receptor. At Site II, the N-terminal residues preceding the first cysteine and the 30s loop of the chemokine interact with extracellular loop/transmembrane residues of the receptor.[5,65] In this regard, CXCL7 is no exception as the triad of N-terminal residues preceding Cys5 has been shown to be critical for neutrophil activation.[66] Furthermore, these residues form an ELR motif that is critical for the function of chemokine ligands binding the same receptor as CXCL7.[5,9,27] The ANM analysis has shown that usually the residues participating in substrate recognition have high mobility.[67] Accordingly, in the case of CXCL7, we find that the ELR N-terminal residues are highly mobile in slow modes [Fig. 2(B)]. Similarly, our mDCM and NMR data show that these residues are flexible (Figs. 3, 6, and 7) as expected for the N-terminus residues.

It is more interesting, however, to analyze the dynamic properties of the 30s loop and the N-loop residues. Although the role of the 30s loop in CXCL7 has not been investigated previously, its importance has been shown experimentally for other CXCR2 chemokine ligands.[22] The ANM analysis shows that in slow modes,

the 30s loop of CXCL7 undergoes large fluctuations. The NMR analysis shows a significantly large R_{ex} contribution for Cys31 and a relatively small order parameter S^2 for residues 34–35 located in the 30s loop. The mDCM analysis shows that at 300K the 30s loop is prone to being flexible in the monomer, and very flexible in the dimer. Interestingly, the ANM analysis does not detect large amplitude motions within the N-loop, and the NMR analysis finds relatively large conformational flexibility (relatively small S^2 values) only for some of the N-loop residues (residues 13–15, at the end of N-loop), indicating that the loop adopts a defined conformation. While the mDCM indicates the N-loop residues are flexible and highly correlated, this does not imply high mobility. Furthermore, the ANM correlation analysis shows that residues in the 30s loop move collectively within the loop, and Val34 at the end of this loop has statistically significant correlations with other parts of the protein. Similarly, residues 9–12 and 15–17 in the N-loop show correlations with the residues distant in protein sequence. The mDCM shows the whole N-terminus of CXCL7 is flexibly correlated to the 30s loop, suggesting that a perturbation at site I (for example, receptor binding) propagates to the site II through conformational rearrangement. It also suggests that residue substitutions introduced in these parts of the molecule may perturb the dynamics by altering existing mechanical couplings, ultimately changing the receptor-binding properties. Note that the recent study of another CXCR2/CXCR1 ligand, CXCL8, revealed that perturbing the GP motif in the 30s loop causes changes in dynamics and conformational rearrangements altering receptor binding by perturbing the equilibrium between binding competent and incompetent conformations of CXCL8 within the dynamic ensemble.[22]

The coupling between N-terminus residues and the 30s loop can be intuitively understood because of the disulfide bond between Cys5 and Cys31 linking these two regions. In support, ours and literature data[14,16] show large R_{ex} contribution for Cys31, indicative of significant slow motion that involves this residue. In other chemokines, similar slow motions in the 30s loop have also been detected for the corresponding cysteine or for the residues in its close proximity.[10,12,68] In application to the receptor binding, this would mean that the N-terminus of CXCL7 samples a significantly larger ensemble of conformations and takes longer to find the correct binding conformation. This result is in agreement with observations made in other chemokines, where perturbations of the disulfide bond linking the N-terminus and 30s loop had deleterious effect on receptor binding and function.[69,70]

Importantly, the reduction of one or both disulfide bonds was found not to affect the antimicrobial activity of Thrombocidin-1 (TC-1), which differs from CXCL7 only by a two-amino acid C-terminal deletion.[29] This result is consistent with the MD/mDCM predictions that disulfide bonds are not critical for maintaining protein stability in CXCL7. The MD/mDCM results also show that the C-terminal α-helix is weakly coupled to the 50s loop, and is otherwise independent of the protein regardless of the disulfide bond states (Figs. 8–10). Moreover, the set of residues {Lys17, Lys41, Arg54, Lys56, Lys57, Lys61, Lys62} that were identified experimentally as being critical to function[14,29,30] fall in the correlated rigid regions of the monomer and dimer. In contrast, the helix has relatively large contributions from slow motions (Fig. 3) in agreement with previous studies.[14,16] This result indicates the C-terminal helix is able to move relative to the rest of the protein, which was observed in CXCL4 where the helix exhibited large-amplitude motions relative to the beta-sheet in a MD simulation.[63] These results suggest it may be possible to truncate the C-terminal helix and retain receptor-binding capability. Conversely, residue substitution/deletion into highly correlated parts of a protein often dramatically alters stability,[54] and would likely affect receptor binding by shifting the equilibrium conformational ensemble.[21,22,71–73]

The MD/mDCM results clearly show that dimerization of CXCL7 induces correlations in flexibility between the 30s and 50s loop. At the same time, there is an increase in thermal and structural stability due to interfacial H-bonding. Recently, the Le Châtelier's principle was invoked[52] to explain in general terms why flexibility tends to redistribute throughout the protein upon a local perturbation to oppose a net shift in rigidity or flexibility. Specifically, to regain a new equilibrium conformational ensemble, a protein attempts to restore a balance in DOF, such that due to the action of rigidifying one region, this drives another region to become more flexible. From this viewpoint, dimerization perturbs two monomers. Each monomer becomes rigidified within beta strand 1 of the beta-sheet, and other areas of the beta-sheet due to interfacial H-bonds. Both monomer structures snap together and extend rigidity across the beta-sheets of each monomer. As this increase in rigidity takes place, a large increase in flexibility in loop regions is induced that become flexibly correlated.

While the CXCL7 fold prevents the removal of disulfide bonds from being detrimental to structural stability, the Le Châtelier tendency to restore equilibrium drives the H-bond network to relax in a way that shifts the CXCL7 monomer-dimer equilibrium toward dimerization. Moreover, a redistribution of flexibility occurs to maintain the number of DOF about the same. In CXCL7, the structure changes in such a way that more H-bonds form at the dimer interface when disulfide bonds are removed (Supporting Information Fig. S5). For example, about 3 more interfacial H-bonds form when no disulfide bonds are present compared to when both disulfide bonds are present (Table I). Based on these features, a "snap on" mechanism seems to occur

during the process of dimerization. That is, monomers are predisposed to propagate rigidity and flexibility through distinct channels. Upon dimerization, the rigidity channels within the monomers snap together as new interfacial H-bonds form, leading to an extended rigid region (extended beta-sheet, Supporting Information Fig. S4). At the same time, the DOF are released in the flexibility channels connecting the 30s and 50s loops in both monomers. Interestingly, the presence of both disulfide bonds creates maximal correlated flexibility between these loops, and at the same time the dimer is least thermally stable. Because the disulfide bonds are highly conserved, this suggests that chemokines function best when there is marginal stability between monomer and dimer forms. In addition to securing marginal stability, the presence of both disulfide bonds provides maximum contrast in flexible correlations between the 30s and 50s loops.

Among the three computational methods, the combined MD/mDCM approach provides detailed information about the native state ensembles, the role of the disulfide bonds in the monomer and dimer and the role of the interface. This work establishes the MD/mDCM approach as a viable means to study the thermodynamic and mechanical properties across the family of chemokine proteins within humans.

ACKNOWLEDGMENTS

REFERENCES

1. Baggiolini M. Chemokines and leukocyte traffic. Nature 1998;392:565–568.
2. Raman D, Sobolik-Delmaire T, Richmond A. Chemokines in health and disease. Exp Cell Res 2011;317:575–589.
3. Gerard C, Rollins BJ. Chemokines disease. Nat Immunol 2001; 2:108–115.
4. Bonecchi R, Galliera E, Borroni EM, Corsi MM, Locati M, Mantovani A. Chemokines and chemokine receptors: an overview. Front Biosci (Landmark Ed) 2009; 14:540–551.
5. Rajagopalan L, Rajarathnam K. Structural basis of chemokine receptor function—A model for binding affinity and ligand selectivity. Biosci Rep 2006; 26:325–339.
6. Kufareva I, Salanga CL, Handel TM. Chemokine and chemokine receptor structure and interactions: implications for therapeutic strategies. Immunol Cell Biol 2015; 93:372–383.
7. Qin L, Kufareva I, Holden LG, Wang C, Zheng Y, Zhao C, Fenalti G, Wu H, Han GW, Cherezov V, Abagyan R, Stevens RC, Handel TM. Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. Science 2015; 347:1117–1122.
8. Clark-Lewis I, Kim KS, Rajarathnam K, Gong JH, Dewald B, Moser B, Baggiolini M, Sykes BD. Structure-activity relationships of chemokines. J Leukoc Biol 1995; 57:703–711.
9. Allen SJ, Crown SE, Handel TM. Chemokine: receptor structure, interactions, and antagonism. Annu Rev Immunol 2007; 25:787–820.
10. Mayer KL, Stone MJ. Backbone dynamics of the CC-chemokine eotaxin-2 and comparison among the eotaxin group chemokines. Proteins 2003; 50:184–191.
11. Ye J, Mayer KL, Mayer MR, Stone MJ. NMR solution structure and backbone dynamics of the CC chemokine eotaxin-3. Biochemistry 2001; 40:7820–7831.
12. Ye J, Mayer KL, Stone MJ. Backbone dynamics of the human CC-chemokine eotaxin. J Biomol NMR 1999; 15:115–124.
13. Grasberger BL, Gronenborn AM, Clore GM. Analysis of the backbone dynamics of interleukin-8 by 15N relaxation measurements. J Mol Biol 1993; 230:364–372.
14. Nguyen LT, Kwakman PH, Chan DI, Liu Z, de Boer L, Zaat SA, Vogel HJ. Exploring platelet chemokine antimicrobial activity: nuclear magnetic resonance backbone dynamics of NAP-2 and TC-1. Antimicrob Agents Chemother 2011; 55:2074–2083.
15. Rajarathnam K, Li Y, Rohrer T, Gentz R. Solution structure and dynamics of myeloid progenitor inhibitory factor-1 (MPIF-1), a novel monomeric CC chemokine. J Biol Chem 2001; 276:4909–4916.
16. Young H, Roongta V, Daly TJ, Mayo KH. NMR structure and dynamics of monomeric neutrophil-activating peptide 2. Biochem J 1999; 338: 591–598.
17. Baryshnikova OK, Sykes BD. Backbone dynamics of SDF-1alpha determined by NMR: interpretation in the presence of monomer-dimer equilibrium. Protein Sci 2006; 15:2568–2578.
18. Jansma AL, Kirkpatrick JP, Hsu AR, Handel TM, Nietlispach D. NMR analysis of the structure, dynamics, and unique oligomerization properties of the chemokine CCL27. J Biol Chem 2010; 285: 14424–14437.
19. Kim KS, Rajarathnam K, Clark-Lewis I, Sykes BD. Structural characterization of a monomeric chemokine: monocyte chemoattractant protein-3. FEBS Lett 1996; 395:277–282.
20. Liou JW, Chang FT, Chung Y, Chen WY, Fischer WB, Hsu HJ. In silico analysis reveals sequential interactions and protein conformational changes during the binding of chemokine CXCL-8 to its receptor CXCR1. PLoS One 2014; 9:e94178.
21. Joseph PR, Sarmiento JM, Mishra AK, Das ST, Garofalo RP, Navarro J, Rajarathnam K. Probing the role of CXC motif in chemokine CXCL8 for high affinity binding and activation of CXCR1 and CXCR2 receptors. J Biol Chem 2010; 285:29262–29269.
22. Joseph PR, Sawant KV, Isley A, Pedroza M, Garofalo RP, Richardson RM, Rajarathnam K. Dynamic conformational switching in the chemokine ligand is essential for G Protein coupled-receptor activation. Biochem J 2013;456:341–251.
23. von Hundelshausen P, Petersen F, Brandt E. Platelet-derived chemokines in vascular biology. Thromb Haemost 2007; 97:704–713.
24. Moser B, Schumacher C, von Tscharner V, Clark-Lewis I, Baggiolini M. Neutrophil-activating peptide 2 and gro/melanoma growth-stimulatory activity interact with neutrophil-activating peptide 1/interleukin 8 receptors on human neutrophils. J Biol Chem 1991; 266:10666–10671.
25. Gear AR, Camerini D. Platelet chemokines and chemokine receptors: linking hemostasis, inflammation, and host defense. Microcirculation 2003; 10:335–350.
26. Ludwig A, Petersen F, Zahn S, Gotze O, Schroder JM, Flad HD, Brandt E. The CXC-chemokine neutrophil-activating peptide-2 induces two distinct optima of neutrophil chemotaxis by differential interaction with interleukin-8 receptors CXCR-1 and CXCR-2. Blood 1997; 90:4588–4597.

27. Veenstra M, Ransohoff RM. Chemokine receptor CXCR2: physiology regulator and neuroinflammation controller? J Neuroimmunol 2012; 246:1–9.

28. Krijgsveld J, Zaat SA, Meeldijk J, van Veelen PA, Fang G, Poolman B, Brandt E, Ehlert JE, Kuijpers AJ, Engbers GH, Feijen J, Dankert J. Thrombocidins, microbicidal proteins from human blood platelets, are C-terminal deletion products of CXC chemokines. J Biol Chem 2000; 275:20374–20381.

29. Kwakman PH, Krijgsveld J, de Boer L, Nguyen LT, Boszhard L, Vreede J, Dekker HL, Speijer D, Drijfhout JW, te Velde AA, Crielaard W, Vogel HJ, Vandenbroucke-Grauls CM, Zaat SA. Native thrombocidin-1 and unfolded thrombocidin-1 exert antimicrobial activity via distinct structural elements. J Biol Chem 2011; 286: 43506–43514.

30. Nguyen LT, Chan DI, Boszhard L, Zaat SA, Vogel HJ. Structure-function studies of chemokine-derived carboxy-terminal antimicrobial peptides. Biochim Biophys Acta 2010; 1798:1062–1072.

31. Jacobs DJ, Dallakyan S. Elucidating protein thermodynamics from the three-dimensional structure of the native state using network rigidity. Biophys J 2005; 88:903–915.

32. Verma D, Jacobs DJ, Livesay DR. Changes in lysozyme flexibility upon mutation are frequent, large and long-ranged. PLoS Comput Biol 2012; 8:e1002409.

33. Mottonen JM, Xu M, Jacobs DJ, Livesay DR. Unifying mechanical and thermodynamic descriptions across the thioredoxin protein family. Proteins 2009; 75:610–627.

34. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001; 80:505–515.

35. Doruker P, Atilgan AR, Bahar I. Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to alpha-amylase inhibitor. Proteins 2000; 40:512–524.

36. Eyal E, Yang LW, Bahar I. Anisotropic network model: systematic evaluation and a new web interface. Bioinformatics 2006; 22:2619–2627.

37. Mayo KH, Yang Y, Daly TJ, Barry JK, La Rosa GJ. Secondary structure of neutrophil-activating peptide-2 determined by 1H-nuclear magnetic resonance spectroscopy. Biochem J 1994; 304: 371–376.

38. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995; 6:277–293.

39. Goddard TD, Kneller DG. SPARKY 3. San Francisco: University of California, 2008.

40. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules.II. Analysis of experimental results. J Am Chem Soc 1982; 104:4559–4570.

41. Bieri M, d'Auvergne EJ, Gooley PR. relaxGUI: a new software for fast and simple NMR relaxation data analysis and calculation of ps-ns and mus motion of proteins. J Biomol NMR 2011; 50:147–155.

42. d'Auvergne EJ, Gooley PR. Optimisation of NMR dynamic models I. Minimisation algorithms and their performance within the model-free and Brownian rotational diffusion spaces. J Biomol NMR 2008; 40:107–119.

43. Bieri M, Gooley PR. Automated NMR relaxation dispersion data analysis using NESSY. BMC Bioinformatics 2011; 12:421.

44. Malkowski MG, Wu JY, Lazar JB, Johnson PH, Edwards BF. The crystal structure of recombinant human neutrophil-activating peptide-2 (M6L) at 1.9-A resolution. J Biol Chem 1995; 270:7077–7087.

45. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res 2005; 33:W368–W371.

46. Mark AE, van Gunsteren WF. Decomposition of the free energy of a system in terms of specific interactions. Implications for theoretical and experimental studies. J Mol Biol 1994; 240:167–176.

47. Dill KA. Additivity principles in biochemistry. J Biol Chem 1997; 272:701–704.

48. Jacobs DJ, Dallakyan S, Wood GG, Heckathorne A. Network rigidity at finite temperature: relationships between thermodynamic stability, the nonadditivity of entropy, and cooperativity in molecular systems. Phys Rev E 2003; 68:061109.

49. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. Proteins 2001; 44:150–165.

50. Dahiyat BI, Gordon DB, Mayo SL. Automated design of the surface positions of protein helices. Prot Sci 1997; 6:1333–1337.

51. Livesay DR, Jacobs DJ. Conserved quantitative stability/flexibility relationships (QSFR) in an orthologous RNase H pair. Proteins 2006; 62:130–143.

52. Li T, Tracka MB, Uddin S, Casas-Finet J, Jacobs DJ, Livesay DR. Redistribution of flexibility in stabilizing antibody fragment mutants follows Le Chatelier's principle. PLoS One 2014; 9:e92870.

53. Mottonen JM, Jacobs DJ, Livesay DR. Allosteric response is both conserved and variable across three CheY orthologs. Biophys J 2010; 99:2245–2254.

54. Jacobs DJ, Livesay DR, Hules J, Tasayco ML. Elucidating quantitative stability/flexibility relationships within thioredoxin and its fragments using a distance constraint model. J Mol Biol 2006; 358:882–904.

55. Bahar I, Lezon TR, Yang LW, Eyal E. Global dynamics of proteins: bridging between structure and function. Annu Rev Biophys 2010; 39:23–42.

56. Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc Natl Acad Sci USA 1983; 80:6571–6575.

57. Baysal C, Atilgan AR. Elucidating the structural mechanisms for biological activity of the chemokine family. Proteins 2001; 43:150–160.

58. Crump MP, Rajarathnam K, Kim KS, Clark-Lewis I, Sykes BD. Solution structure of eotaxin, a chemokine that selectively recruits eosinophils in allergic inflammation. J Biol Chem 1998; 273:22471–22479.

59. LiWang AC, Cao JJ, Zheng H, Lu Z, Peiper SC, LiWang PJ. Dynamics study on the anti-human immunodeficiency virus chemokine viral macrophage-inflammatory protein-II (VMIP-II) reveals a fully monomeric protein. Biochemistry 1999; 38:442–453.

60. Kim S, Jao S, Laurence JS, LiWang PJ. Structural comparison of monomeric variants of the chemokine MIP-1beta having differing ability to bind the receptor CCR5. Biochemistry 2001; 40:10782–10791.

61. Rajarathnam K, Kay CM, Dewald B, Wolf M, Baggiolini M, Clark-Lewis I, Sykes BD. Neutrophil-activating peptide-2 and melanoma growth-stimulatory activity are functional as monomers for neutrophil activation. J Biol Chem 1997; 272:1725–1729.

62. Li T, Verma D, Tracka MB, Casas-Finet J, Livesay DR, Jacobs DJ. Thermodynamic stability and flexibility characteristics of antibody fragment complexes. Protein Pept Lett 2013;21:752–765.

63. Carlson J, Baxter SA, Dreau D, Nesmelova IV. The heterodimerization of platelet-derived chemokines. Biochim Biophys Acta 2013; 1834:158–168.

64. Mayo KH, Barker S, Kuranda MJ, Hunt AJ, Myers JA, Maione TE. Molten globule monomer to condensed dimer: role of disulfide bonds in platelet factor-4 folding and subunit association. Biochemistry 1992; 31:12255–12265.

65. Fernandez EJ, Lolis E. Structure, function, and inhibition of chemokines. Annu Rev Pharmacol Toxicol 2002; 42:469–499.

66. Yan Z, Zhang J, Holt JC, Stewart GJ, Niewiarowski S, Poncz M. Structural requirements of platelet chemokines for neutrophil activation. Blood 1994; 84:2329–2339.

67. Yang LW, Bahar I. Coupling between catalytic site and collective dynamics: a requirement for mechanochemical activity of enzymes. Structure 2005; 13:893–904.

68. Crump MP, Spyracopoulos L, Lavigne P, Kim KS, Clark-lewis I, Sykes BD. Backbone dynamics of the human CC chemokine

eotaxin: fast motions, slow motions, and implications for receptor binding. Protein Sci 1999; 8:2041–2054.

69. Eigenbrot C, Lowman HB, Chee L, Artis DR. Structural change and receptor binding in a chemokine mutant with a rearranged disulfide: X-ray structure of E38C/C50AIL-8 at 2 A resolution. Proteins 1997; 27:556–566.

70. Rajarathnam K, Sykes BD, Dewald B, Baggiolini M, Clark-Lewis I. Disulfide bridges in interleukin-8 probed using non-natural disulfide analogues: dissociation of roles in structure from function. Biochemistry 1999; 38:7653–7658.

71. Prado GN, Suetomi K, Shumate D, Maxwell C, Ravindran A, Rajarathnam K, Navarro J. Chemokine signaling specificity: essential role for the N-terminal domain of chemokine receptors. Biochemistry 2007; 46:8961–8968.

72. Rajagopalan L, Rajarathnam K. Ligand selectivity and affinity of chemokine receptor CXCR1. Role of N-terminal domain. J Biol Chem 2004; 279:30000–30008.

73. Rajagopalan L, Chin CC, Rajarathnam K. Role of intramolecular disulfides in stability and structure of a noncovalent homodimer. Biophys J 2007; 93:2129–2134.