Title

Arginine off-kilter: guanidinium is not as planar as restraints denote

Journal

Acta Crystallographica Section D, Structural Biology, 76(12)

ISSN

2059-7983

Authors

Moriarty, Nigel W

Liebschner, Dorothee

Tronrud, Dale E

et al.

Publication Date

2020-12-01

DOI

10.1107/s2059798320013534

Peer reviewed

Arginine off kilter: guanidinium is not as planar as restraints denote

Authors

Nigel W. Moriarty[a]*, Dorothee Liebschner[a], Dale E. Tronrud[b] and Paul D. Adams[ac]

[a]Molecular Biosciences and Integrated Bioimaging, Lawrence Berkeley National Lab, Berkeley, California, 94720-8235, USA

[b]Department of Biochemistry and Biophysics and the Department of Chemistry, Oregon State University, Corvallis, Oregon, 97331

[c]Department of Bioengineering, University of California, Berkeley, CA, 94720, USA

Correspondence email: nwmoriarty@lbl.gov

Funding information National Institutes of Health (grant No. P01GM063210; grant No. GM083136); US Department of Energy (contract No. DE-AC02-05CH1123).

ynopsis The geometry of arginine shows more complexity than the standard restraints accommodate.

bstract Crystallographic refinement of macromolecular structures relies on stereochemical restraints to mitigate the typically poor data-to-parameter ratio. For proteins, each amino acid has a unique set of geometry restraints, which represent stereochemical information, such as bond lengths, valence angles, torsion angles, dihedrals and planes. It has been shown that the geometry in refined structures can differ significantly from that present in libraries; for example, it was recently reported that the guanidinium moiety in arginine is not symmetric. In this work, we confirm the asymmetry of the Nε-Cζ-Nη1 and Nε-Cζ-Nη2 valence angles in the guanidinium moiety. In addition, we found that the Cδ atom can deviate significantly (more than 20°) from the guanidinium plane. This requires relaxation of the planar restraint for the Cδ atom, as it otherwise causes the other atoms in the group to compensate by distorting the guanidinium core plane. We therefore have formulated a new set of restraints of the arginine side chain, available in the software package *Phenix*, that take into account the asymmetry of the group and the planar deviation of the Cδ atom. This is an example of the need to regularly revisit the geometric restraint libraries used in

macromolecular refinement so that they reflect the best knowledge of the structural chemistry of their components available at the time.

eywords: Arginine, chemical restraints, macromolecular refinement, guanidine, planarity

1. Introduction

With 12 side chain atoms and a molecular weight of about 174 Da, arginine is one of the largest standard amino acids. From a geometric viewpoint, arginine is very interesting because of the guanidinium group (–NH–C–$(NH_2)_2^+$) on its side chain (Fig. 1). At physiological pH, guanidinium is always protonated and positively charged. Arginine is therefore often involved in salt bridges with negatively charged residues, such as aspartic and glutamic acids. As the guanidinium group is hydrophilic, arginine residues are often located at the surface of the protein, so that the side chain can point toward solvent and form hydrogen bonds. Arginine is also very flexible: it has four chi-angles (Cα–Cβ, Cβ–Cγ, Cγ–Cδ, Cδ–Nε) that yield 60 allowed rotameric configurations (Hintze *et al.*, 2016). This inherent flexibility and the fact that arginine is frequently located at the surface where it is not sterically confined by neighbouring residues often causes the density in crystallographic Fourier maps to be partly or completely missing for many atoms of the side chain. As a consequence, arginine side chains can be difficult to model in crystallographic structures. As a result of the typically low observation-to-parameter ratio and the lack of high-resolution data in macromolecular crystallography, stereochemical restraints are required to maintain the correct geometry of arginine residues during crystallographic refinement.

Refinement, which is driven by both restraints and experimental data, should result in a chemically reasonable structure. All refinement programs use geometry restraints, which provide *a priori* stereochemical information about the structural units of macromolecules. At high resolution (better than 1Å), the experimental data typically provide sufficient information to produce accurate atomic coordinates (with the exception of flexible and disordered regions). However, at lower resolution (worse than 2.5-3Å), geometry restraints are especially important because they dominate over the sparse experimental data. Therefore, geometry restraints need to be chemically accurate and their uncertainty, as indicated by the estimated standard deviations (esd), should be

sufficient to allow the experimental data to guide the refinement to a chemically reasonable result.

It is common practice to monitor the root mean squared deviations (*rmsd*) from the geometry restraint targets used in refinement to ensure that the weighting between experimental data and geometric information is reasonable. At low resolution, the *rmsd* values are typically small (approaching zero) as there is insufficient data to determine deviations from ideal geometry. At high resolution, the *rmsd* values can be larger when there is sufficient experimental data to define geometries that truly deviate from the library targets. A related metric for assessing the results of refinement is the *rmsZ* (Z-score), which is an *rmsd* value that is normalized by the standard deviation of the restraint from the library. It is a dimensionless value that ideally should range from near zero for low-resolution models to as large as approaching 1.0 for an ordered model based on high-resolution data.

Some geometry restraints have close chemical equivalents, such as bond lengths and valence angles. In contrast, the planarity restraint is less directly related to chemistry. It is commonly used to enforce planar structures arising from *sp2* hybridisation. The π-orbital electrons of an *sp2*-hybridised atom are repulsed by the bonded atoms resulting in a planar structure. Naïvely, this planarity could be maintained by having the sum of the angle ideal values around the central atom summing to 360°. In practice, this approach fails to enforce planar geometry because each angle is implemented as a statistically independent quantity. This is remedied by adding a harmonic co-planarity restraint defined by reference to the best plane through the atomic positions within the scope of that restraint. The “ideal” position of each atom is in the plane.

Engh & Huber (1991) generated ideal values for bonds and valence angles in standard amino acids to be used in macromolecular refinement with X-plor (Brünger, 1992). The values of the restraints have been updated (Engh & Huber, 2001) but the symmetry of the guanidinium moiety around the Nε–Cζ bond was always enforced, i.e. to nearly identical values for the valence angles between the N*i*–Cζ–N*j* atoms (Nε–Cζ–Nη1: 120.3° (0.5), Nε–Cζ–Nη2: 120.3° (0.5), Nη1–Cζ–Nη2: 119.4° (1.1))[1]. However, the guanidinium group in the arginine side chain is

[1] The choice of having a symmetric guanidinium group possibly originates from the heritage of the library’s use in Molecular Dynamics (MD) based refinement. With the large motions expected in an MD simulation, atom labels would either

not expected to have this symmetric geometry, as the chemical environment of the group is not symmetric: the *cis* configuration leads to repulsion between the Nη1 and Cδ atoms causing a larger bond angle for Nε–Cζ–Nη1. This asymmetry is well known. The Handbook of Biochemistry and Molecular Biology (Vijayan, 1976) reported the Nε–Cζ–Nη1 and Nε–Cζ–Nη2 valence angles as 121.5° and 119.3°, respectively. We also note that the libraries distributed with the refinement packages PROLSQ (Hendrickson & Konnert, 1980) and TNT (Tronrud *et al.*, 1987; Tronrud, 1987) contained asymmetric values that are indistinguishable from Vijayan, 1976. Their valence angles for the *cis* nitrogen atom (Nη1) were therefore slightly larger than the ideal 120.3° proposed by Engh & Huber. However, current widely used macromolecular refinement packages, such as *Phenix* (Liebschner *et al.*, 2019) and *Refmac* (Murshudov *et al.*, 2011) use the Engh & Huber restraints or derivations of them (Vagin *et al.*, 2004). Therefore, they all restrain the guanidinium group with symmetric valence angles.

Recently, Malinska et al. (2016) revisited the geometry of the guanidinium group in arginine. By analyzing high-resolution entries of the Protein Data Bank (PDB), they reported that the moiety is not symmetric. These results from the PDB analysis were corroborated by a search of the Cambridge Structural Database of small molecules.

Guanidine is a planar molecule resulting from resonance structure of the three C–N bonds. This planarity extends to the hydrogen atoms and once the moiety is bonded to the amino acid, the effect of the π-electrons extends to the Cδ atom also. As a consequence, in addition to bonds and valence angles from Engh & Huber, the refinement restraints contain a planarity restraint for the three nitrogen atoms, two carbon atoms and five hydrogen atoms in guanidinium. The group has also a torsion angle restraint involving the Cδ–Nε–Cζ–Nη1 atoms. In particular, the planar restraint includes the Cδ atom that bonds to Nε, thus replacing a hydrogen atom of guanidine. We note that, in contrast to the asymmetry of the valence angles involving Cζ, no refinement package has deviated from this notion of uniform planarity for the guanidinium moiety.

have to be regularly swapped, or a more complex force field would be required that would recognize the dependence of these valence angles on the associated torsion angle. Assigning symmetric valence angles likely avoided this complication. Unfortunately, when the Engh & Huber library was used in other programs, the target values were not updated.

However, instances of non-planar guanidinium groups in arginine can be found in the PDB. A search of the Protein Geometry Database (Berkholz *et al.*, 2010) for arginine residues in models with better than 1.2 Å resolution revealed instances where the Cδ atom deviates more than 20° from the guanidinium plane (Tronrud, D., unpublished results). The non-planarity of the arginine side chains is supported by the electron density.

One example for such a deviation is entry 2xfr (Rejzek *et al.*, 2011), determined at 0.97Å resolution. The Cδ atom in Arg 242 deviates from planarity by approximately 22°, as measured by the Cδ–Nε–Cζ–Nη1 torsion angle. The atomic positions and therefore the distortion of the plane are clearly justified by the $2mF_{obs}$-DF_{model} Fourier map (Fig. 2). Notably, the other atoms in the guanidine group remain visibly planar, indicating that the Cδ atom is more flexible. The residue is otherwise not an outlier, as it is in the favoured region of rotamer conformations and has no clashes.

The examples of planar deviations found in the Protein Geometry database suggested that the planar restraint in guanidine needs modification to account for flexibility. Therefore, we analysed small molecules compounds in the CSD and performed systematic refinements of macromolecular structures in the PDB to quantify the flexibility of the Cδ atom in the guanidine group and to create a revised set of restraints.

1. Methods

To obtain reliable small molecule geometries, the Cambridge Structural Database (CSD, Groom *et al.*, 2016) was searched for the guanidinium moiety (Fig. 1) and the resulting geometries (bond lengths, valence angles, torsion angles) were analysed. The CSD search and the geometry analysis were performed using the programs Conquest and Mercury (Bruno *et al.*, 2002, 2004) from the Cambridge Crystallographic Data Centre software suite. The search can be repeated using the script and settings provided in the supplementary material Fig. S1 and S2. The analysis revealed that in small molecules the Cδ atom can deviate significantly from the plane imposed on the guanidinium moiety by the geometry restraints for proteins (see “Results” section). This led us to formulate a new set of restraints for the guanidine group in arginine and to test the new restraints in refinement.

To test the new arginine restraints, we refined models from the PDB with two different sets of restraints. The first set (“standard”) uses the restraints for arginine from the Monomer Library, which is the standard restraints library for refinement of macromolecules in *Phenix*. Here, the guanidinium group is restrained to be symmetrical and planar (Table 1). The second set of restraints (“flexible”) includes asymmetric valence angles for the guanidinium group and allows the Cδ atom to deviate from the plane by using a larger standard deviation (0.095Å instead of 0.020Å). This corresponds to a Cδ–Nε–Cζ–Nη1 torsion (equivalent to a deviation from the guanidinium plane) of approximately 5°. All refinements were performed using *phenix.refine* (Afonine *et al.*, 2012). Coordinate and experimental data files were obtained from the PDB that met the following criteria: resolution better than 3.05Å; data completeness >90%; data is not twinned; R_{work} <30%; R_{free} <35%; and R_{free} - R_{work} >1.5%. For entries with resolutions better than 1.05Å, the R_{free} - R_{work} criterion was changed to >0.5%. By using these criteria, we excluded suspicious entries and low-resolution data, allowing automatic refinement strategies with default options. Hydrogen atoms were added to the models using *Phenix* ReadySet!. Ligand restraints were generated by *Phenix* eLBOW (Moriarty *et al.*, 2009).

Each model was then subjected to 10 macrocyles of refinement using the default strategy in *phenix.refine* for the refinement of coordinates, atomic displacement parameters (ADP) and occupancies. Non-default refinement options included the optimization of the weight between the experimental data and the geometry restraints. In addition, anisotropic ADP were used for non-hydrogen protein atoms at resolutions better than 1.55Å and for water oxygen atoms at resolutions better than 1.25Å. The quality of the resulting models was assessed numerically using MolProbity (Williams *et al.*, 2018) in *Phenix*. To filter out problematic structures, refined models with a clashscore of greater than 12 were not included in our analysis. Results were grouped into resolution bins of width 0.1Å. Resolution bins with less than 30 refined structures were not taken into account. In total, this led to 26,557 protein structures refined with conventional and modified arginine restraints.

2. Results & Discussion

2.1. Search for guanidinium in the CSD

The search of the CSD for guanidinium (Fig. 1) resulted in 153 entries with 204 instances of the moiety. Some instances had geometric parameters that deviated significantly from the average. Visual inspection of these entries often revealed erroneous results (for example the protonation state in the entry was not consistent with that in the search molecule, e.g. entries CESPAR and COXYET) or an unusual chemical environment (for example a sulphate coordinated to the guanidinium, e.g. entries QAFTUN and SUXYUF).

As the process of manually examining the extrema and removing unreliable entries from the result list is not tractable, a statistically robust outlier rejection method using the interquartile range, Tukey's fences (Beyer, 1981, 2020), was applied to the torsion angles Cδ–Nε–Cζ–Nη1 (T1), Cδ–Nε–Cζ–Nη2 (T2) and T1-T2 (which are the focus of this analysis; Fig. 1). This approach reduced the number of entries to 140 and instances to 180. We note that the automatic process removed the entries already listed above and several more with similar issues. Removal of outliers discarded entries with the most non-planar guanidinium groups. In other words, it led to removing examples that support the flexibility of the moiety. We note that two instances of removed entries – caused by the presence of a sulphate ion causing distortion (QAFTUN, SUXYUF) – are not outside the realm of possibility in proteins. Therefore, the outlier removal process makes the set of entries more planar and, therefore, less extreme.

The values for bonds lengths, valence angles and torsion angles for guanidine moieties in small molecules are summarized in Table 1. Histograms of the internal coordinate values are in the supplemental information Fig. S3-S5. The average bond lengths are essentially identical to the values reported by Malinska *et al.* (2016). The valence angles differ by an insignificant amount, possibly due to models added to the CSD since the study was performed and to different outlier rejection procedures. The guanidinium valence angles are asymmetric, with 121.5° and 119.2° for Nε–Cζ–Nη1 and Nε–Cζ–Nη2, respectively.

To analyse the planarity of the guanidinium group, we examined the torsion angles T1 (Cδ–Nε–Cζ–Nη1) and T2 (Cδ–Nε–Cζ–Nη2). The difference between T1 and T2 measures the planarity of the core moiety (Nε, Cζ, Nη1 and Nη2). The average of T1-T2 in CSD guanidinium structures is 180.0 (1.2)°, meaning that the core moiety is indeed planar, with no entry deviating more than 3.2° from the plane. On the other hand, the torsion angles T1 and T2 have a standard deviation of 6.6° each, with a maximum absolute deviation of 16.2° and 16.7° for

T1 and T2, respectively. This flexibility clearly shows that the Cδ atom has a propensity to deviate from the plane of the core moiety.

2.2. New features added to *Phenix*

New arginine restraints: A feature of the restraints implementation in *Phenix* makes it possible to easily add a flexible guanidinium planar restraint. *Phenix* allows the planar restraint to have a different estimated standard deviation (*esd*) value for each atom in a plane. The *esd* value for all atoms in the guanidinium plane is 0.02Å for the standard restraints. In the case of the flexible restraints, an *esd* of 0.095Å for the Cδ atom allows it to bend out of the plane to approximate the flexibility found in the molecules in the CSD (approximately 5°). The implied *esd* of torsion angle of about 5° ensures that large deviations, as seen in the example described in the introduction (model 2xfr), are allowed if supported by experimental data. Along with the relaxed planarity *esd* for the Cδ atom, the bond and angle values from the CSD analysis are used as a new set of restraints for the arginine guanidine group (Table 1). Note that in both the original arginine restraints and the modified restraints, the T1 torsion angle is restrained to zero degrees with an esd of 10°. This is not a limiting restraint in either case.

Consistent IUPAC atom naming for arginine: One consequence of the asymmetry of the guanidinium group is that the nitrogen atoms Nη1 and Nη2 need to be assigned the appropriate names. In compliance with the IUPAC convention for atom labelling, the Nη1 atom should be always in *cis* configuration in comparison to the Cδ atom. Code was added to *Phenix* that automatically renames the nitrogen atoms (and associated hydrogen atoms) if necessary. The parameter flip_symmetric_amino_acids controls the atom labelling in arginine and defaults to True from *Phenix* version dev-3951.

2.3. Refinement of macromolecules

To test if a planar restraint with more flexibility for the Cδ atom is appropriate for arginine in proteins, we performed test refinements on structures deposited in the PDB. The 26,557 refined structures had experimental data resolutions between 0.85Å and 3.05Å (Fig. 2 inset).

2.3.1. Comparison of standard arginine restraints vs flexible restraints

The interpretation of the refinement results after applying different sets of restraints is a subtle matter. Global quality indicators, such as *R*-factors,

clashscores or global bond and angle *rmsd* values, are only marginally affected if the restraints to a single residue are changed. This is particularly true for arginine residues because they constitute only a small fraction of any given model. Instead, it is more appropriate to analyse bond, valence angle, torsion angle and planar *rmsd* and *rmsZ* values for the modified restraints. While doing this, it is important to remember that *rmsd* and *rmsZ* values should be also interpreted in the context of resolution: At low resolution, the model can be made to agree with any reasonable restraint without violating the fit to the blurry density; *rmsd* values are expected to be low even for restraints that don't completely respect the chemistry. At medium to high resolution (1.5-2.5Å), restraints and data have a similar weight in refinement. The data may contain enough information to drive the model towards a chemically meaningful geometry but if the restraints are chemically unreasonable, the *rmsd*/*rmsZ* values may increase. At ultrahigh resolution (better than 1 Å) the data dominate over the restraints, resulting in a model that is chemically correct, with *rmsd*/*rmsZ* trending towards higher values. However, not all restraints need to deviate from ideality at high resolution. Indeed, if the *rmsd* value of a certain restraint remains low at high resolution, then this ideal value is appropriate for the majority of models. Therefore, when comparing sets of restraints after refinement, it is appropriate to focus on the medium to ultra-high resolution range, as the low-resolution range will generally have low *rmsd*/*rmsZ* values.

If the restraint target values are modified while keeping the esd constant, the *rmsd* and *rmsZ* values are a good indicator that reflects if the new target values lead to a less strained model. This is what can be observed for the bond and angle restraints in the guanidinium group: the bond and angle *rmsd* are lower for flexible restraints (i.e. modified bonds and valence angles) than for standard restraints in the resolution range 3.0 Å and better (Fig. 3).

One must investigate the effect of adding flexibility to a planar restraint in a different way as only the *esd* are changed, not the target value itself. The *esd* of the Cδ atom was increased and as a consequence, the atom can move more freely (i.e. out of plane) during refinement. This means that the *rmsd* in most cases increases as well. Therefore, instead of looking at the *rmsd*/*rmsZ* for a particular restraint (Cδ) only, it is important to analyse the *rmsd*/*rmsZ* for the entire guanidinium group.

Cδ–Nε–Cζ–Nη1 torsion angle: Fig. 4 shows the absolute deviation from zero of the Cδ–Nε–Cζ–Nη1 (T1) torsion angle in resolution bins. We note that even for the standard restraint, where the planar restraint is uniform across the plane, the T1 torsion angle has a mean of approximately 0.25° at resolutions worse than 2Å with a maximum of nearly 2.5° at high resolution. The flexible restraint results in a torsion angle deviation of 1° in the mid resolution range and greater than 4° at high resolution. The torsion angle is therefore systematically larger when the flexible arginine restraints were used. This behaviour is expected. The larger *esd* allows the atoms to deviate from the plane, with the deviation being more pronounced at higher resolution.

Cδ deviations: Fig. 5(a) shows the *rmsd* values of the planar restraint for the Cδ atom that reflect its deviation from the plane. (Fig. S6 includes the Standard Error of the Mean for all results in Fig. 5.) The *rmsd* is relatively close to zero at low resolution for the standard restraints but increases up to 0.025Å at resolutions better than 2Å. The *rmsd* values for the flexible restraints are numerically larger, which is in line with the fact that the Cδ atom can now move more freely. As for the *rmsZ* values of the Cδ planar restraint (Fig. 5(b)), the Cδ atom deviates approximately one sigma value from the mean for both sets of restraints. We note that the standard restraint values have an *rmsZ* of greater than 1.2 at high resolution, which is greater than for the flexible restraints, suggesting that the original restraint is too restrictive.

Nε deviations: The Nε atom *rmsd* and *rmsZ* values show an opposite trend compared to the Cδ atom. The *rmsd* values for the Nε atom (Fig. 5(c)) are similar (close to zero) for both sets of restraints at resolutions worse than 2Å. At resolutions better than 2Å, the Nε atom *rmsd* values are systematically larger for the standard restraints than for the flexible restraints. This indicates that the Nε atom tends to be closer to the plane when the flexible restraints are used. Not surprisingly, there is a similar reduction in *rmsZ* values for the Nε atom (Fig. 5(d)). The *rmsZ* values are systematically smaller for the flexible restraints in all resolution ranges, with a significant reduction at resolution 2Å and better. The drop in Nε atom *rmsd* and *rmsZ* values therefore suggests that the core moiety (Nε–Cζ–Nη1–Nη2) becomes more planar with the flexible restraints

Nη1 and Nη1 deviations: Investigating the Nη1 and Nη2 atoms of the guanidinium moiety provides additional insights (Fig. 6 and S7). The *rmsd* deviations are essentially zero at resolutions worse than 2Å. At resolutions better

than 1.5Å, the standard restraints *rmsd* values slightly increase (approaching 0.003Å) for the Nη1 and Nη2 atoms. Although this is remarkably small, it is still larger than the *rmsd* values using the flexible restraints, which are essentially zero over the entire resolution range. The *rmsZ* values show the same trend: The values for the flexible restraints are systematically smaller. Therefore, as the Nη1 and Nη2 atoms deviate only marginally from the guanidinium plane, this further suggests that the core moiety is flat.

The behaviour of the Cδ, Nε, Nη1 and Nη2 atoms can be summarized as follows: For the standard restraints, the non-Cδ atoms (Nε, Nη1 and Nη2) compensate for the lack of freedom of movement of the Cδ atom by deviating ever so slightly from the guanidinium plane. The new flexible restraints allow the Cδ atom to move away from the plane while the other atoms can relax into the planar core moiety.

High-resolution example: For the Arg 242 residue in 2xfr, refinement with flexible arginine restraints increases the T1 torsion angle from 19° to 24° using the standard and flexible restraints, respectively, while increasing the chemically meaningful planarity of the core group, where the deviation of the Nε atom is reduced from 0.13Å to 0.02Å.

3. Conclusions

Our analysis of small molecules in the CSD reiterates that the guanidinium moiety is asymmetric (Nε–Cζ–Nη1: 121.5°, Nε–Cζ–Nη2: 119.2°). Importantly, this analysis also revealed that the Cδ atom deviates from the plane of the guanidinium group. This plane is typically enforced in crystallographic refinement as a geometry restraint. Based on the bond lengths and valence angles from the CSD, as well as on the propensity of the Cδ atom to deviate from the plane, we formulated a revised set of geometry restraints for the guanidinium group in arginine. To test the impact of these new restraints, we performed refinements of 26,557 PDB entries against X-ray data in the resolution range 0.85Å – 3.55Å. Arginine bond and angle *rmsd* improve with the new sets of restraints. The Cδ atom, which is allowed to deviate more from the plane with an increased esd, indeed has a propensity to move further away from it. However, this increased flexibility of the Cδ atom simultaneously allows the guanidinium core group to become more planar. While the new set of restraints will generally not affect global quality indicators of refinement, it will lead to more chemically meaningful models. We note that the increased flexibility of the arginine side chain can

affect the interpretation of hydrogen-bond networks which are often important for catalytic mechanisms. We therefore suggest that arginine restraints should be updated broadly in refinement and validation programs. The new set of restraints is available in *Phenix* version dev-3951 and later.

Figure 1 Diagram of guanidinium moiety that terminates the side chain of arginine including a schematic representation of the T1 and T2 torsion angles.

Figure 2 The Cδ atom can deviate significantly from the guanidinium plane in arginine. Two views of Arg (A 242) in model 2xfr (0.97 Å resolution). Lightblue: $2mF_{obs}$-DF_{model} map at 1 rms contour. Orange: $2mF_{obs}$-DF_{model} map at 5 rms contour. The location of the Cα and Cβ atoms is shown with lines.

Figure 3 Bond length (Å) and angle (°) *rmsd* values averaged in 0.1Å resolution bins. Refinements with standard arginine restraints are plotted using the green lines and flexible restraints are shown in orange. The *rmsd* values for the whole model are shown with dashed lines, while for the arginine-only *rmsd* values are solid lines. Inset shows the number of refinements in each resolution bin.

Figure 4 Values of the Cδ–Nε–Cζ–Nη1 torsion angle for the standard restraints (green) and flexible restraints (orange) in 0.1Å resolution bins.

Figure 5 Planar *rmsd* in Å (left column, (a,c)) and *rmsZ* (right column, (b,d)) values for Cδ and Nε atoms in the guanidinium moiety averaged in 0.1Å resolution bins. Refinements using the standard restraints have green lines while orange lines denote the flexible restraints. Atom Cδ is the top row (a,b) and atom Nε is bottom row (c,d). The *esd* for Cδ is shown for the standard (0.020Å) and flexible (0.095Å) restraints for reference. The lower row is on the same scale as the upper graphs.

Figure 6 Planar *rmsd* in Å (left column, (a,c)) and *rmsZ* (right column, (b,d)) values for Nη1 and Nη2 atoms in the guanidinium moiety averaged in 0.1Å resolution bins. Refinements using the standard restraints have green lines while orange lines denote the flexible restraints. Atom Nη1 is the top row (a,b) and atom Nη2 is bottom row (c,d). The lower row is on the same scale as the upper graphs.

	Cδ–Nε	Nε–Cζ	Cζ–Nη1	Cζ–Nη2	δ–ε–ζ	ε–ζ–η1	ε–ζ–η2	η1–ζ–η2	δ–ε–ζ–η1 T1	δ–ε–ζ–η2 T2	T1-T2
This work – CSD 2019											
mean	1.458	1.326	1.323	1.330	124.4	121.5	119.2	119.3	0.9	180.7	180.0
rmsd	0.011	0.011	0.010	0.009	1.4	0.9	0.9	0.9	6.7	6.7	1.2
min	1.421	1.292	1.302	1.300	118.9	119.0	116.4	117.2	-16.2	163.3	176.9
max	1.557	1.361	1.370	1.375	127.0	123.7	122.5	122.5	13.3	194.4	183.0
Malinska PDB survey											
mean	1.458	1.327	1.325	1.328	124.9	121.3	119.2	119.6			
rmsd	0.012	0.011	0.013	0.012	1.4	1.0	1.0	1.0[1]			
min	1.390	1.267	1.266	1.294	119.4	118.2	114.2	113.6			
max	1.520	1.384	1.386	1.394	130.2	124.6	123.0	126.1			
Malinska – CSD 2016											
mean	1.456	1.326	1.323	1.329	124.4	121.5	119.2	119.4			
rmsd	0.014	0.011	0.014	0.013	1.4	1.0	0.9	1.3			
GeoStd											
ideal	1.460	1.329	1.326	1.326	124.2	120.0	120.0	119.7			
esd	0.018	0.014	0.018	0.018	1.5	1.9	1.9	1.8			

Table 1 Bond lengths (Å), valence angles (°) and torsion angles (°) with standard uncertainties in the guanidine groups in arginine from various sources. Two sets of geometric values from Malinska et al. (2016).

[1] Appears to be a typographical error in Malinska et al.

cknowledgements This work was supported by the US National Institutes of Health (NIH) (grant P01GM063210) and the Phenix Industrial Consortium. This work was supported in part by the US Department of Energy under Contract No. DE-AC02-05CH1123. DET was supported by NIH grant GM083136 to P.A. Karplus (Oregon State University). Author contributions include – DET performed searches of the Protein Geometry Database; DET, NWM designed the new restraints; DCL, NWM performed searches of the Cambridge Structural Database; DCL, NWM analysed the refinement results; and all authors participated in the manuscript preparation.

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Crystallogr. Sect. -Biol. Crystallogr.* **68**, 352–367.

Berkholz, D. S., Krenesky, P. B., Davidson, J. R. & Karplus, P. A. (2010). *Nucleic Acids Res.* **38**, D320–D325.

Beyer, H. (1981). *Biom. J.* **23**, 413–414.

Brünger, A. T. (1992). X-PLOR: version 3.1: a system for x-ray crystallography and NMR Yale University Press.

Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). *Acta Crystallogr. B*. **58**, 389–397.

Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D. S., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E. & Orpen, A. G. (2004). *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144.

Engh, R. A. & Huber, R. (1991). *Acta Crystallogr. A*. **47**, 392–400.

Engh, R. & Huber, R. (2001). *International Tables for Crystallography*, Vol. *F*, edited by M. Rossmann & E. Arnold, pp. 382–392. Dordrecht: Kluwer Academic Publishers.

Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **72**, 171–179.

Hendrickson, W. & Konnert, J. (1980). *Computing in Crystallography*, Vol. edited by R. Diamond, S. Ramaseshan & K. Venkatesan, p. 13.01-13.26. Bangalore: Indian Academy of Sciences Publishers.

Hintze, B. J., Lewis, S. M., Richardson, J. S. & Richardson, D. C. (2016). *Proteins-Struct. Funct. Bioinforma.* **84**, 1177–1189.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D.

C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Crystallogr. Sect. Struct. Biol.* **75**, 861–877.

Malinska, M., Dauter, M. & Dauter, Z. (2016). *Protein Sci. Publ. Protein Soc.* **25**, 1753–1756.

Moriarty, N. W., Grosse-Kunstleve, R. W. & Adams, P. D. (2009). *Acta Crystallogr. Sect. -Biol. Crystallogr.* **65**, 1074–1080.

Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr. Sect. -Biol. Crystallogr.* **67**, 355–367.

Rejzek, M., Stevenson, C. E., Southard, A. M., Stanley, D., Denyer, K., Smith, A. M., Naldrett, M. J., Lawson, D. M. & Field, R. A. (2011). *Mol. Biosyst.* **7**, 718–730.

Tronrud, D. E. (1987). British Library Document Supply Centre.

Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Crystallogr. A*. **A43**, 489–501.

Vagin, A. A., Steiner, R. A., Lebedev, A. A., Potterton, L., McNicholas, S., Long, F. & Murshudov, G. N. (2004). *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2184–2195.

Vijayan, M. (1976). *CRC Handbook of Biochemistry and Molecular Biology, Proteins*, Vol. *III*, edited by G. Fastman, pp. 742–749. Cleveland: CRC Press.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, D. C. (2018). *Protein Sci.* **27**, 293–315.

(2020). Outlier.

Supporting information

S1. CSD search details

Fig. S1 contains the .con file used to search the CSD. The addition filters are shown in Fig. S2.

S2. CSD results

Results of the CSD structure search for internal coordinates summarised in table 1 are displayed in Fig. S3-S5.

S3. Plane *rmsd* for atoms in guanidinium moiety

Standard error of the mean (SEM) error bars for the plots in Fig. 5 and 6 are shown in Fig. S6 and S7, respectively.

```
T1 *CONN
NFRAG -99
ELDEF CC = AA - D
AT1 C 3  T3       :XY 248 238
AT2 N 2 1        :XY 166 240
AT3 N 1 2        :XY 303 178
AT4 N 1 2        :XY 298 289
AT5 C 2 2 T4       :XY 122 297
AT6 CC 1        :XY 62 297
BO 1 3 99
BO 5 6 1
BO 2 5 1
BO 1 4 99
BO 1 2 99
GEOM
DEFINE V1 4 1 3
DEFINE V2 2 1 3
DEFINE V3 2 1 4
DEFINE V4 5 2 1
DEFINE T1 5 2 1 3
DEFINE T2 5 2 1 4
DEFINE B1 1 3
DEFINE B2 1 4
DEFINE B3 1 2
DEFINE B4 2 5
DEFINE T3 6 5 2 1
SYMCHK ON
ENANT NORMAL
END
```

Figure S1 Conquest .con file used to search the CSD database.

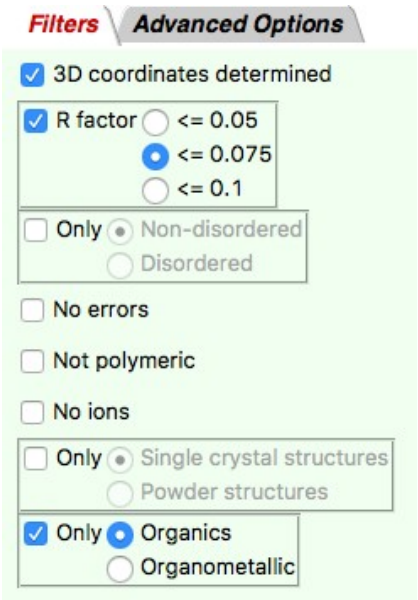

Figure S2 Search filters used for the CSD search.

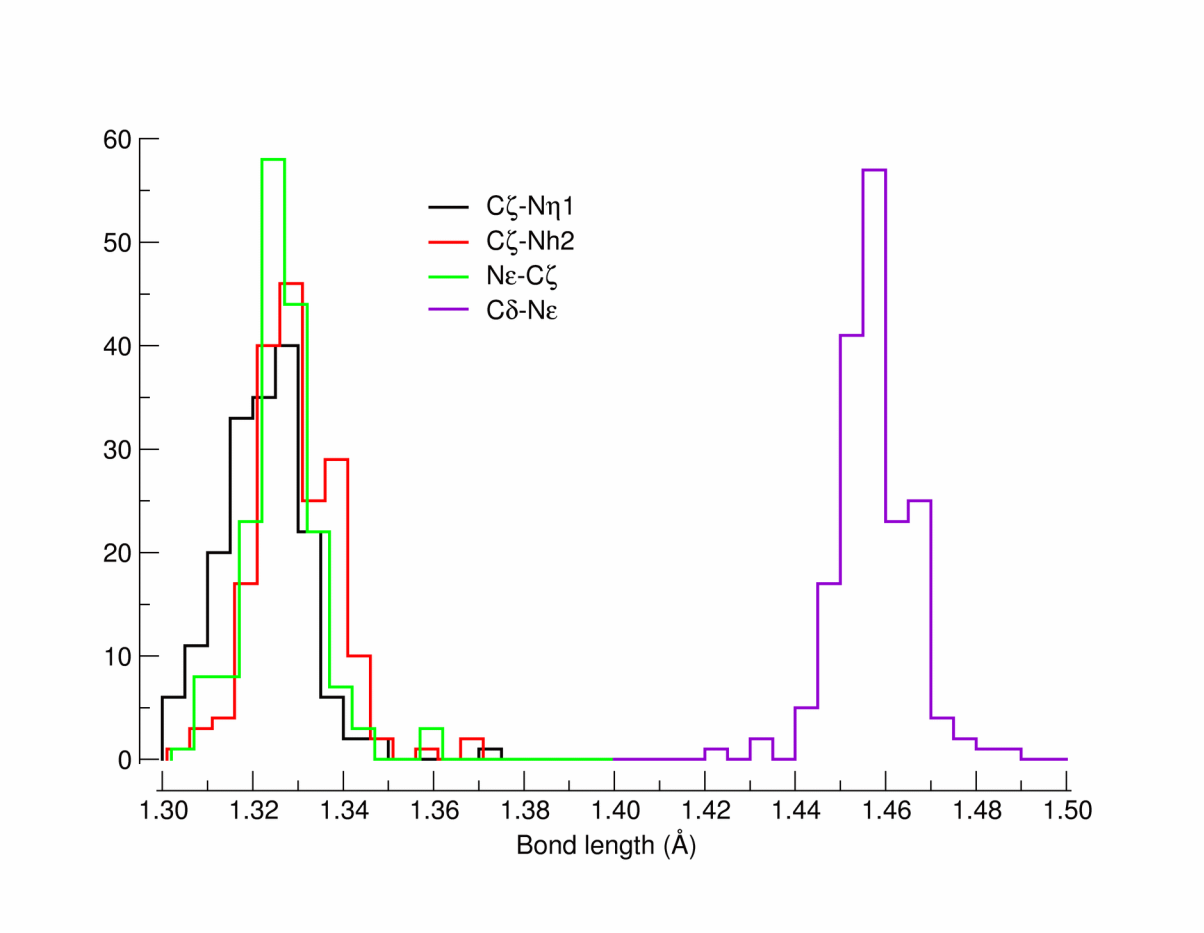

Figure S3 Bond lengths in the guanidinium group from the CSD search.

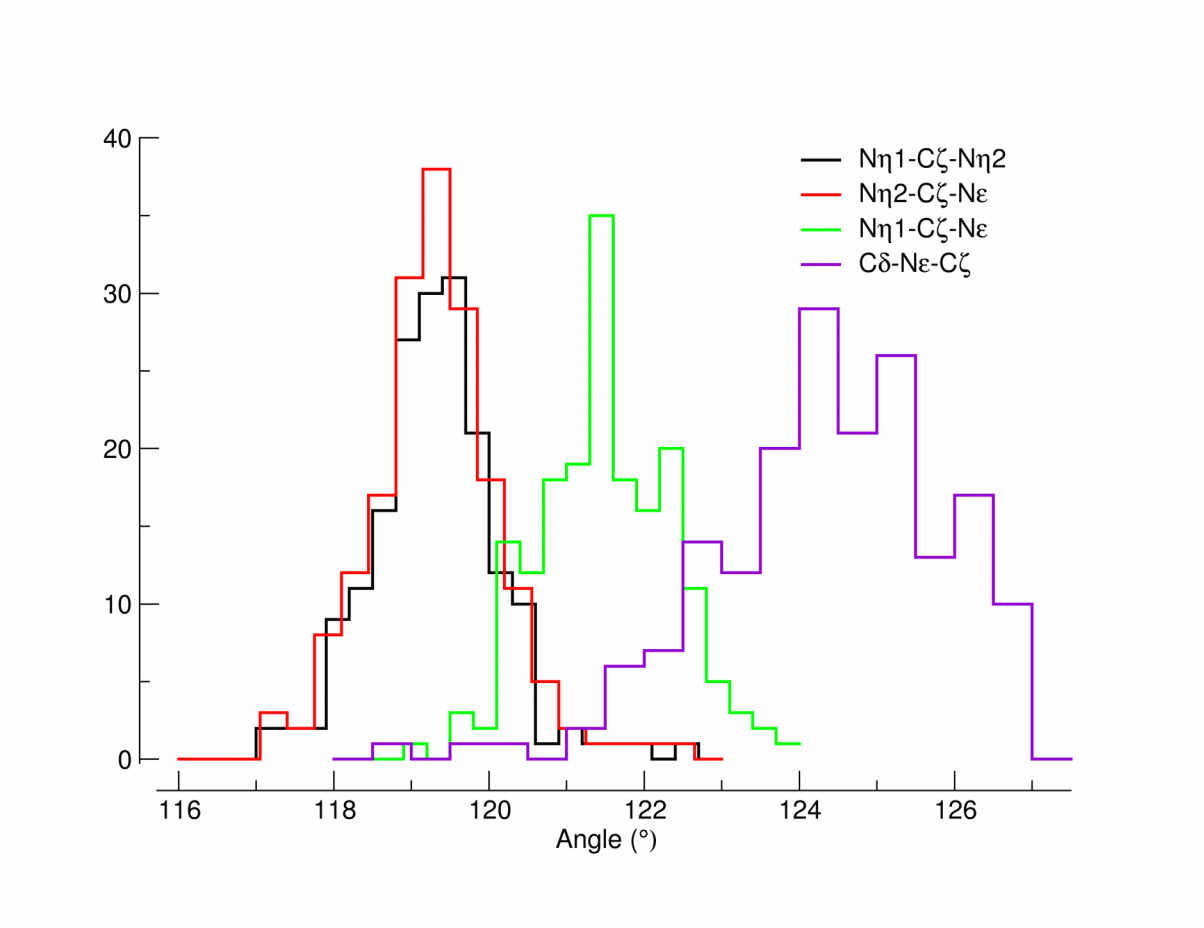

Figure S4 Valence angles in guanidinium from the CSD search.

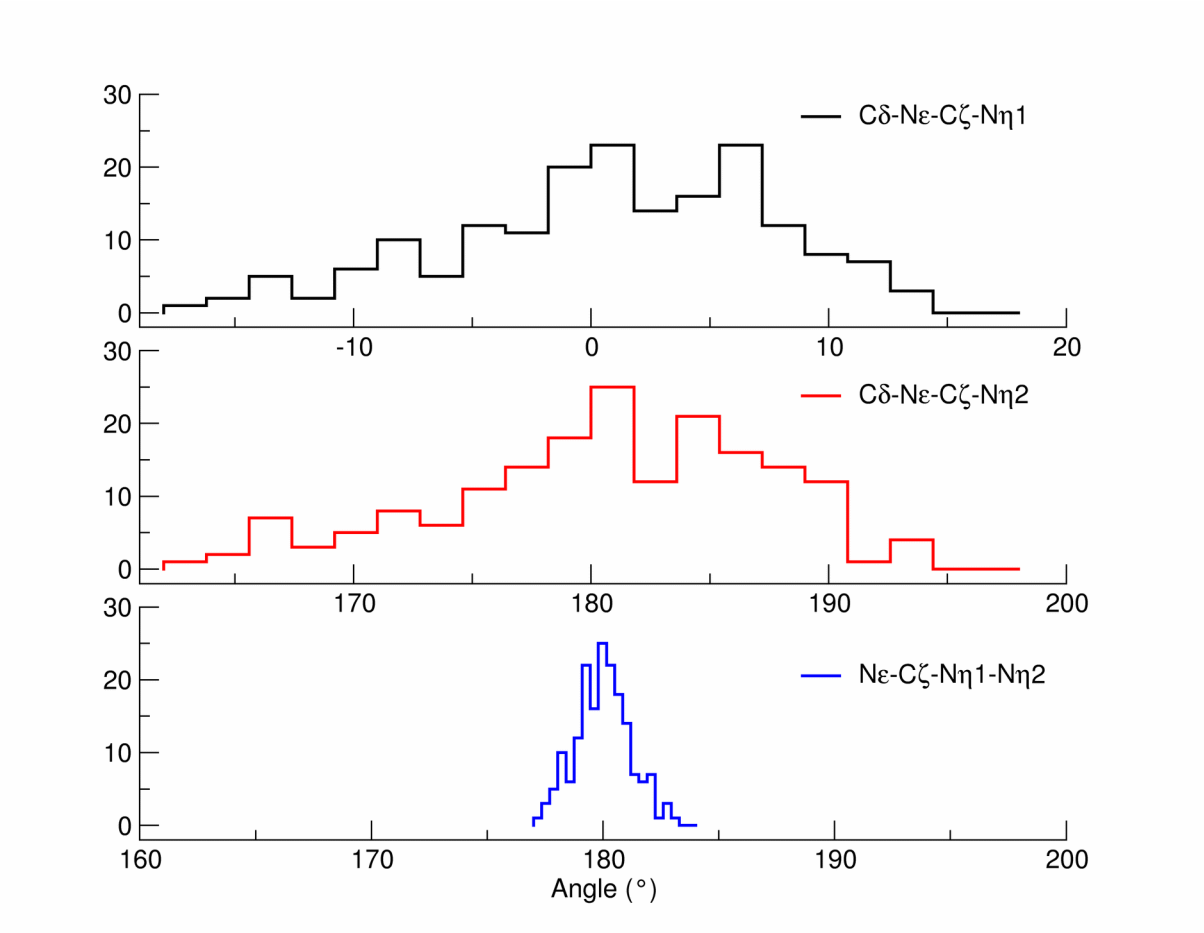

Figure S5 Dihedral angles in the guanidinium group from the CSD search.

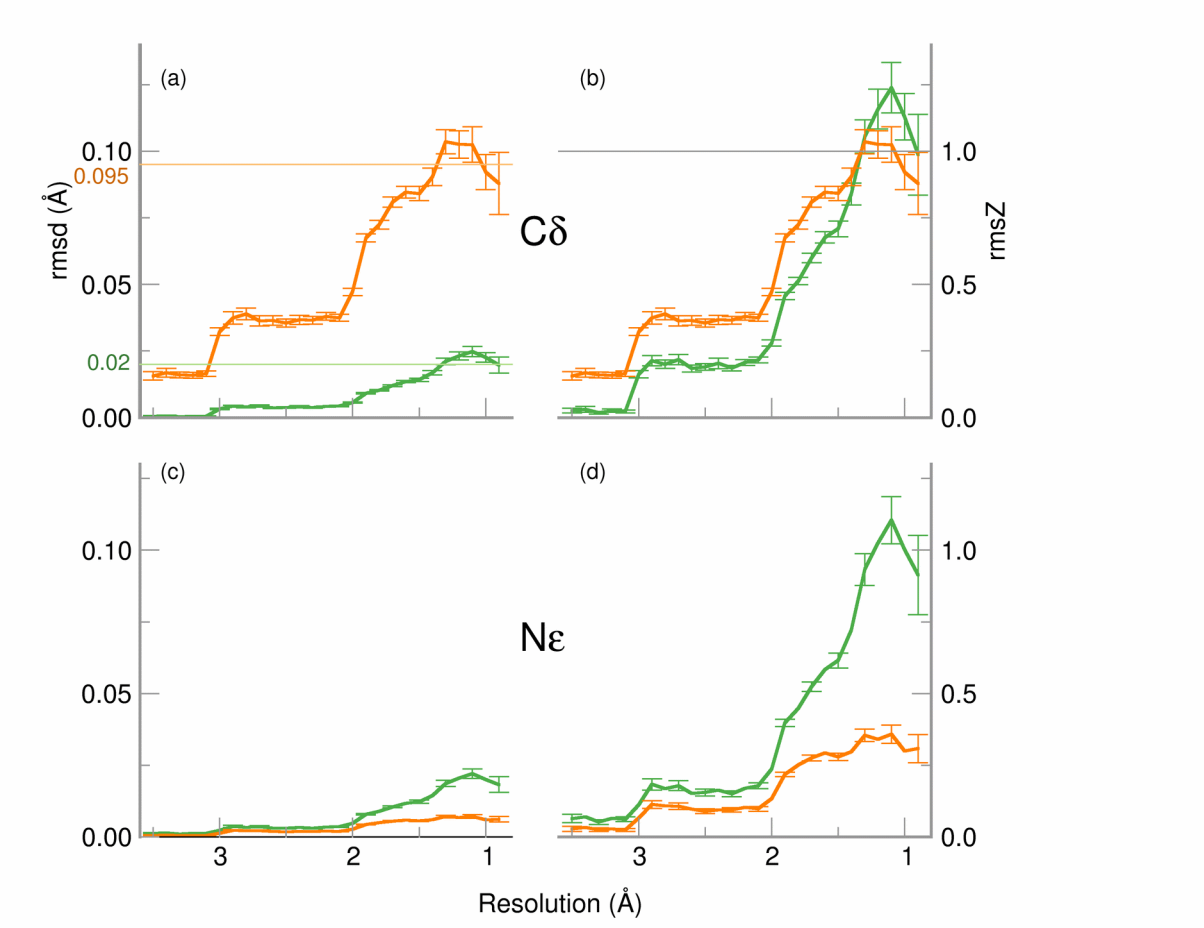

Figure S6 Standard Error of Mean for data points in Fig. 5.

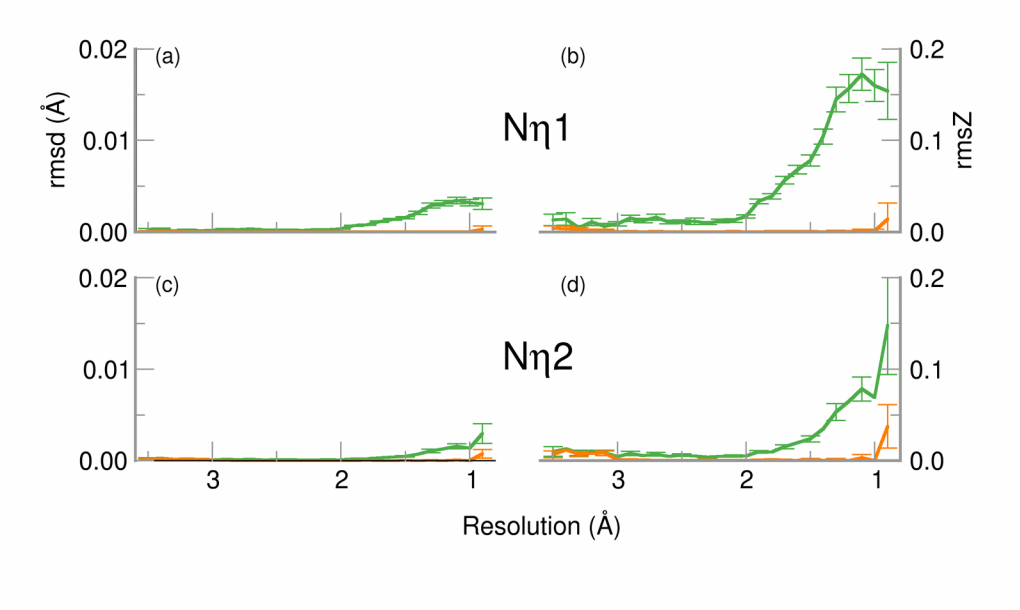

Figure S7 Standard Error of Mean for data points in Fig. 6.