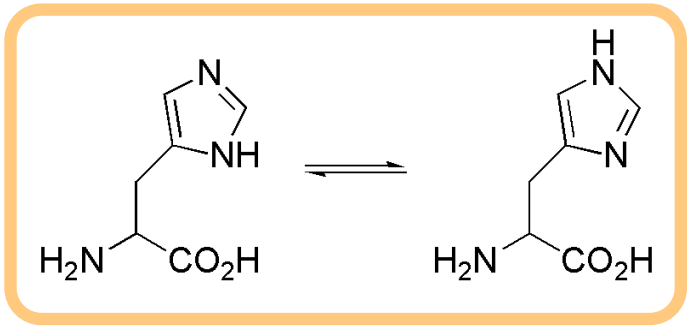
CHEM 436 / 630

Molecular modelling of proteins

Winter 2018 Term

Instructor:
Guillaume Lamoureux
Concordia University, Montréal, Canada

Protonation states of amino acid sidechains



	pK_a
Histidine (H)	6.1
Aspartic acid (D)	3.9
Glutamic acid (E)	4.1
Arginine (R)	12.5
Lysine (K)	10.5
Tyrosine (Y)	10.1

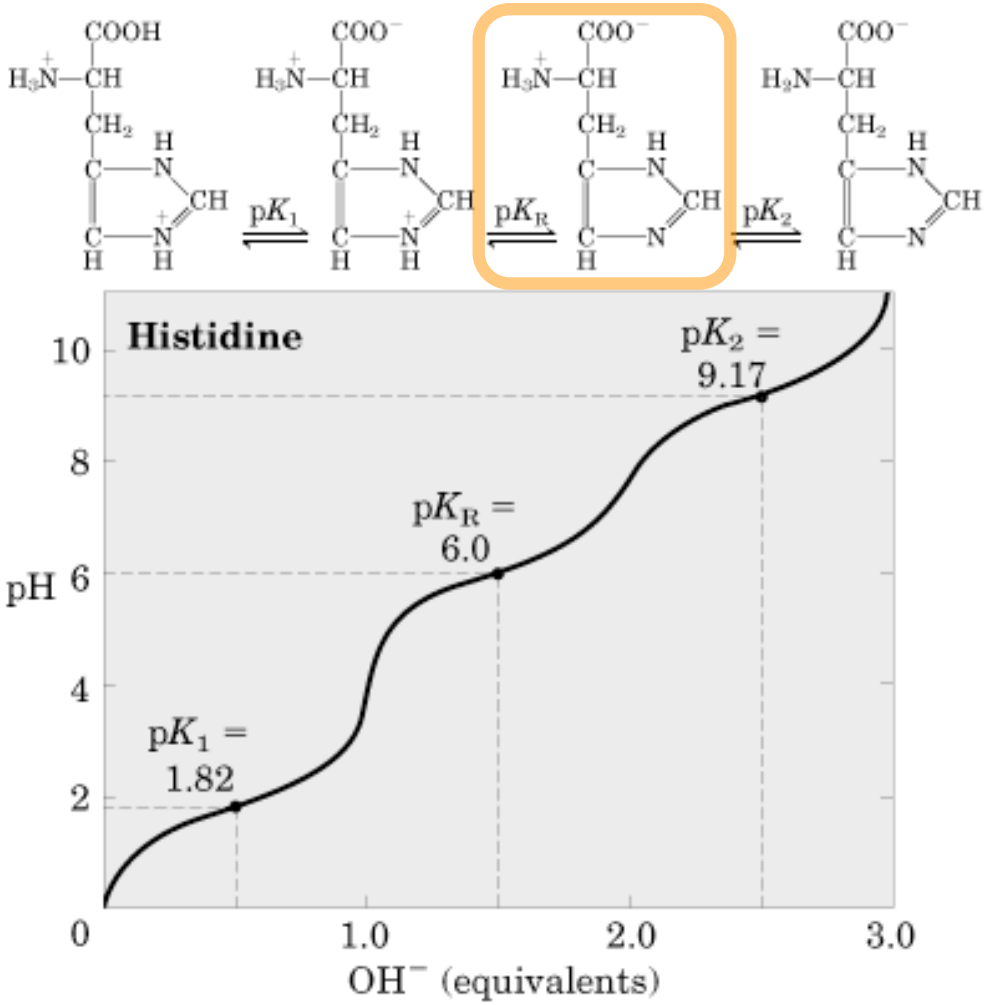


Figure from :
D. L. Nelson & M. M. Cox, Lehninger Principles of Biochemistry, Third Edition.

Definition of pK_a

$$AH \rightarrow A^- + H^+$$

Example:

- Aspartic acid (in protein)
- Aspartate (in protein)
- Proton (in water)

$$pK_a = -\log_{10}\left(\frac{[H_3O^+][A^-]}{[HA]}\right)$$

When the pH is equal to the pKa, $[A^-] = [HA]$

Below the pKa, $[A^-] < [HA]$

Above the pKa, $[A^-] > [HA]$

$$\Delta G^{\text{Protein}}(AH \rightarrow A^- + H^+) = \Delta G^{\text{Water}}(AH \rightarrow A^- + H^+) + \Delta G^{\text{Water}\rightarrow\text{Protein}}_{\text{Solvation}}(A^-) - \Delta G^{\text{Water}\rightarrow\text{Protein}}_{\text{Solvation}}(AH)$$

free energy of reaction **in the protein** = free energy of reaction **in water** + "free energy of transfer" of A^- from water to the protein − "free energy of transfer" of AH from water to the protein

Reference :

M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski & J. H. Jensen. 2011. *J. Chem. Theory Comput.* **7**, 525–537. http://dx.doi.org/10.1021/ct100578z

Definition of ΔpK_a

$$pK_{a,i}^{\text{Protein}} = pK_{a,i}^{\text{Water}} + \Delta pK_{a,i}^{\text{Water}\rightarrow\text{Protein}}$$

pKa of residue ***i*** in the protein

Standard pKa of residue in water

pKa shift for residue ***i*** due to the protein environment

$$\Delta pK_a^{\text{Water}\rightarrow\text{protein}} = \frac{1}{2.30RT} \cdot (\Delta G_{\text{Solvation}}^{\text{Water}\rightarrow\text{Protein}}(A^-) - \Delta G_{\text{Solvation}}^{\text{Water}\rightarrow\text{Protein}}(AH))$$

The direction of the pKa shift depends on whether it is A^- or AH that increases its stability the most by being in the protein.

If the protein stabilizes state A^- more than state AH, ΔpKa is negative and the pKa will decrease.

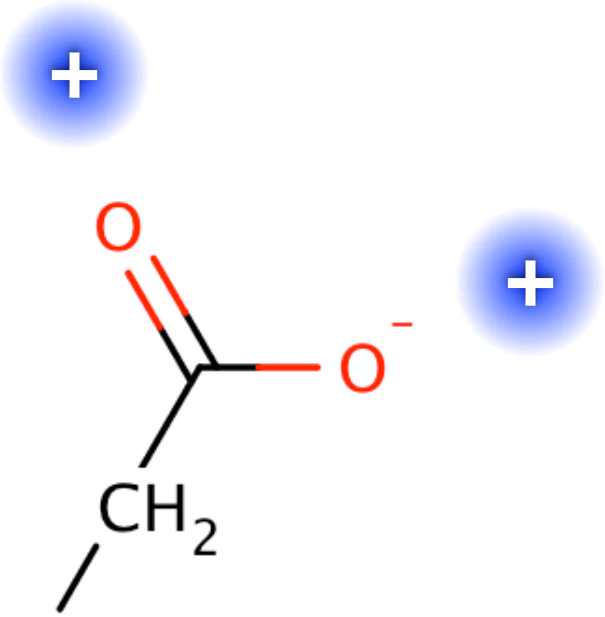
If the protein stabilizes state AH more than state A^-, ΔpKa is positive and the pKa will increase.

Will it increase enough to change the protonation state at physiological pH ?

Reference :

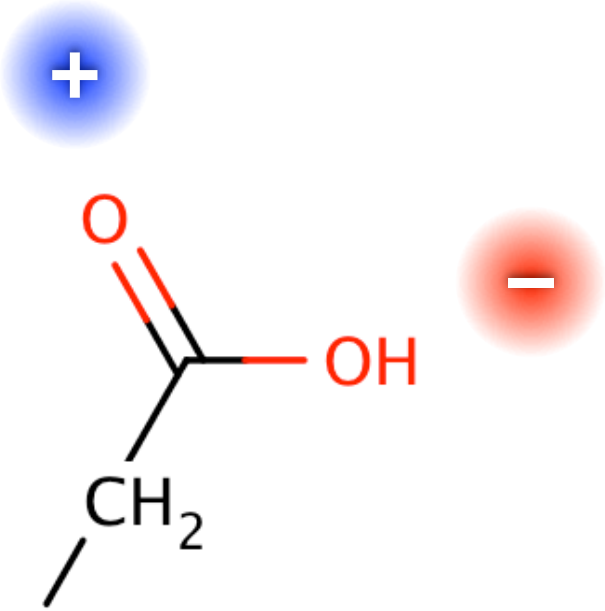
M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski & J. H. Jensen. 2011. *J. Chem. Theory Comput.* **7**, 525–537. http://dx.doi.org/10.1021/ct100578z

Protein environment



$\Delta pK_a < 0$

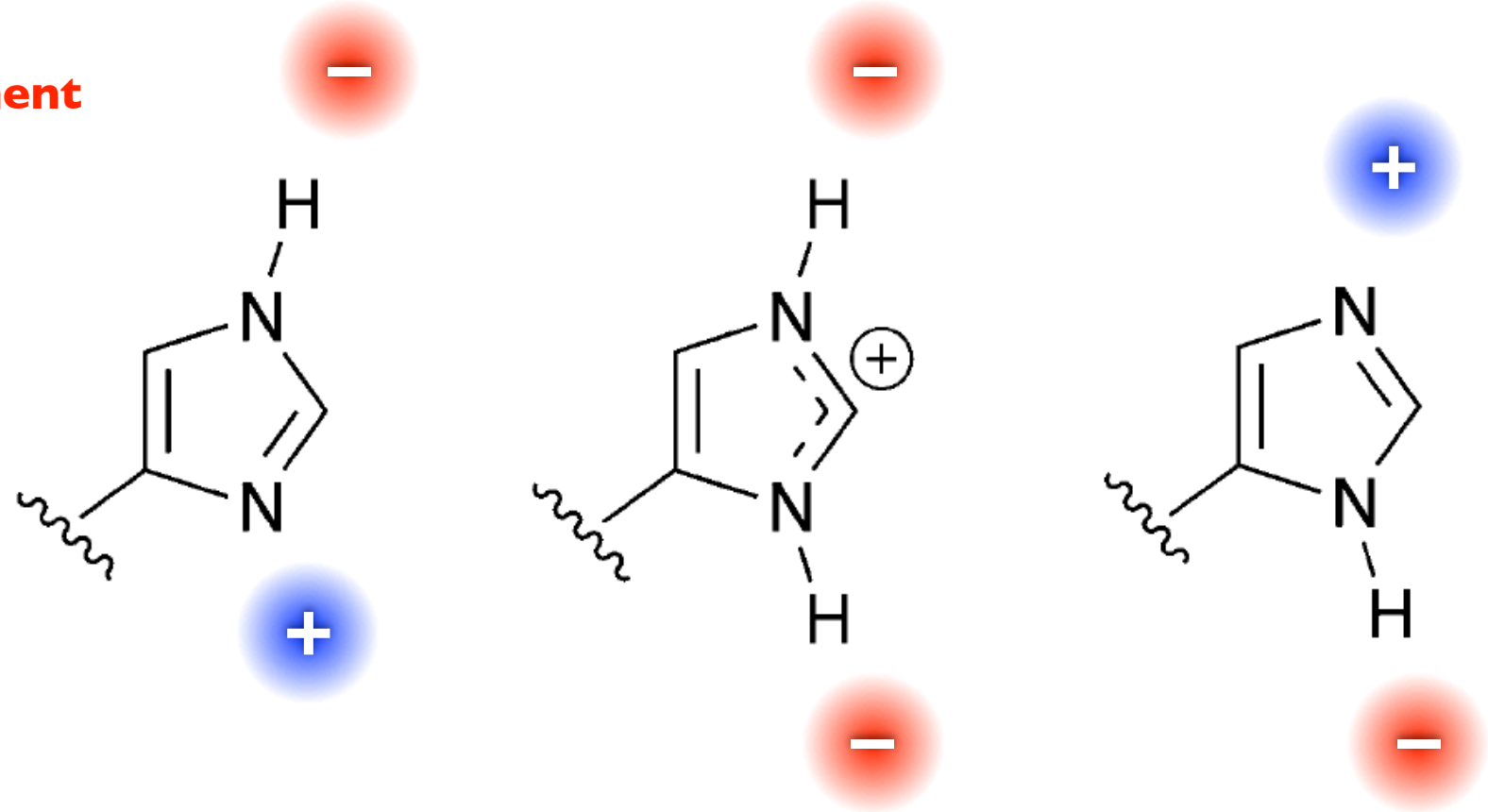
(The pKa decreases.)



$\Delta pK_a > 0$

(The pKa increases.)

Protein environment



$\Delta pK_a < 0$
(The pKa decreases.)

$\Delta pK_a > 0$
(The pKa increases.)

$\Delta pK_a < 0$
(The pKa decreases.)

Based on a figure from :
R. Lonsdale, J. N. Harvey, and A. J. Mulholland. 2012. *Chem. Soc. Rev.* **41**, 3025-3038
http://dx.doi.org/10.1039/C2CS15297E

How is the ΔpK_a estimated in PROPKA?

Charge-charge interactions with other residues

$$\Delta pK_{a,i} = \Delta pK_{a,i}^{Self} + \Delta pK_{a,i}^{QQ}$$

Loss of solvation caused by protein

Intrinsic electrostatics (interactions with surrounding polar groups)

$$\Delta pK_{a,i}^{Self} = \Delta pK_{a,i}^{Desolv} + \Delta pK_{a,i}^{Q\mu}$$

$$\Delta pK_{a,i}^{Desolv} = c \cdot \sum_{j}^{N} \frac{V_j}{r_{ij}^4}$$

$$c = c_{surface} - (c_{surface} - c_{buried}) \cdot w(N)$$

Influence of buried ratio

Hydrogen bonding with neighbouring residues

$$\Delta pK_{a,i}^{Q\mu} \simeq \Delta pK_{a,i}^{HB} + \Delta pK_{a,i}^{RE}$$

Penalty for "reverse" hydrogen bonds

$$\Delta pK_{a,i}^{HB} = \begin{cases} c^{HB} \cdot w(r_{ij}) \cdot \cos\theta & \text{if } \theta \geq 90^\circ \\ 0 & \text{if } \theta < 90^\circ \end{cases}$$

$$w(r_{ij}) = \begin{cases} 1 & \text{if } r_{ij} \leq r_{min} \\ \dfrac{r_{ij} - r_{min}}{r_{max} - r_{min}} & \text{if } r_{min} < r_{ij} < r_{max} \\ 0 & \text{if } r_{ij} \geq r_{max} \end{cases}$$

θ is the donor-hydrogen-acceptor angle. For freely rotating donor-hydrogen bonds, cos θ is set to 1.

Screened Coulomb interaction

$$\Delta pK_{a,i}^{QQ} = \sigma_{ij} \cdot \frac{244}{\varepsilon \cdot r_{ij}} \cdot w(r_{ij})$$

$$\sigma_{ij} = \begin{cases} -1 & \text{if } i \in \text{acids and } j \in \text{bases} \\ & \text{or } i \in \text{bases and } pK_{a,i} < pK_{a,j} \\ +1 & \text{if } i \in \text{bases and } j \in \text{acids} \\ & \text{or } i \in \text{acids and } pK_{a,i} > pK_{a,j} \\ 0 & \text{otherwise} \end{cases}$$

Rule to assign the *sign* of the interaction

$$\varepsilon = \varepsilon_{surface} - (\varepsilon_{surface} - \varepsilon_{buried}) \cdot w_{pair}(N)$$

Effective dielectric constant

Buried ratio

$$w(N) = \begin{cases} 0 & \text{if } N \leq N_{min} \\ \dfrac{N - N_{min}}{N_{max} - N_{min}} & \text{if } N_{min} < N < N_{max} \\ 1 & \text{if } N \geq N_{max} \end{cases}$$

$$w(r_{ij}) = \begin{cases} \dfrac{r_{ij}}{r_{min}} & \text{if } r_{ij} \leq r_{min} \\ \dfrac{r_{ij} - r_{min}}{r_{max} - r_{min}} & \text{if } r_{min} < r_{ij} < r_{max} \\ 1 & \text{if } r_{ij} \geq r_{max} \end{cases}$$

Distance-dependent weight function (makes the screened interaction constant for $r < r_{min}$, and pure Coulomb for $r > r_{max}$)

Reference :

M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski & J. H. Jensen. 2011. *J. Chem. Theory Comput.* **7**, 525–537. http://dx.doi.org/10.1021/ct100578z

How is the ΔpK_a estimated in PROPKA?

Table 1. Nonadjustable Parameters and Descriptors

interaction	type/use	parameter	value
Coulomb	buried ratio	R_C[a]	15.0
	buried ratio	N_{Min}	280
	buried ratio	N_{Max}	560
	$w(r)$	r_{Min}	4.0
	$w(r)$	r_{Max}	10.0
desolvation	VDW volume	V_C	20.58
	VDW volume	V_{C4}	38.79
	VDW volume	V_N	15.60
	VDW volume	V_O	14.71
	VDW volume	V_S	24.43

[a] Contact radius.

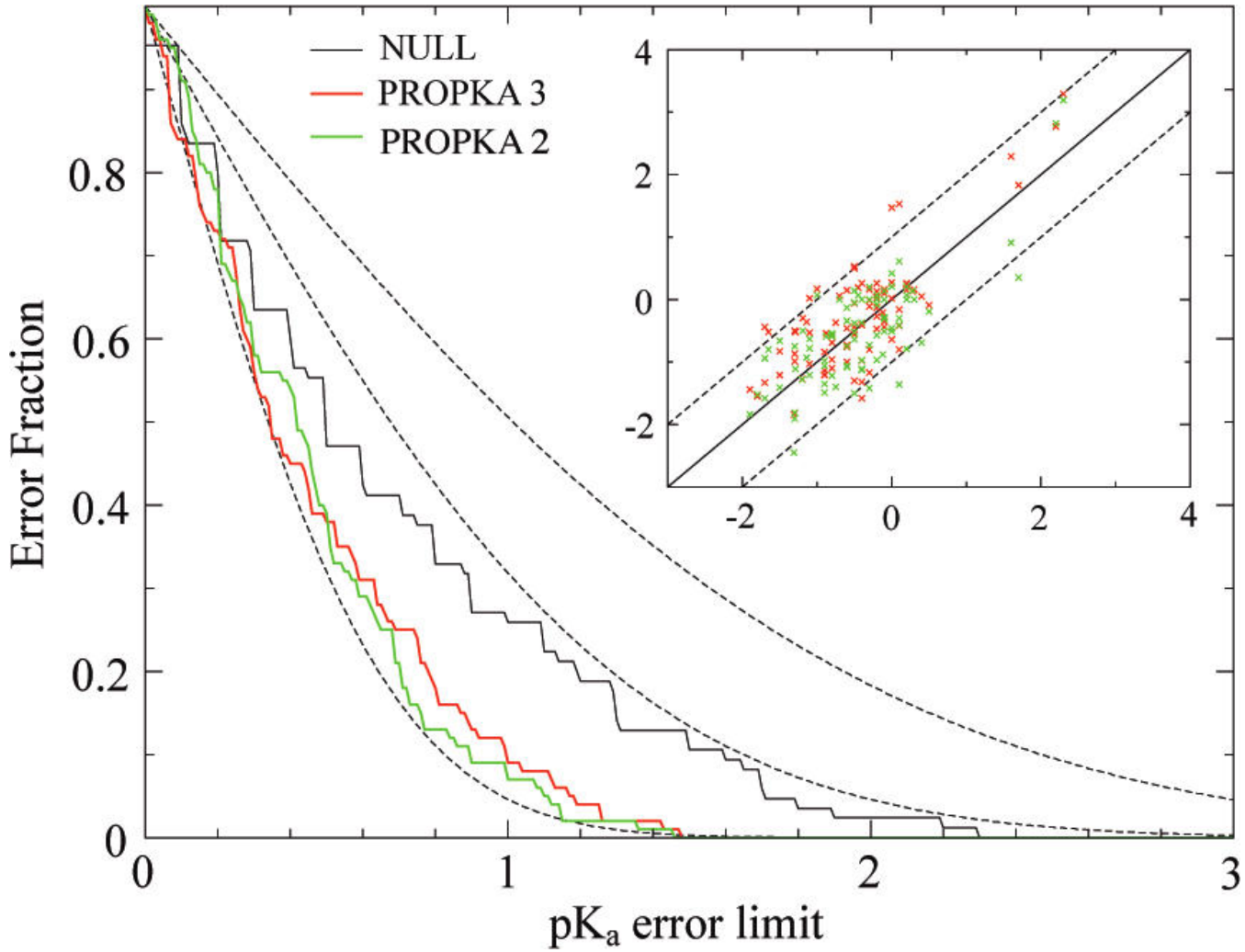
Table 2. Fitted PROPKA3 Parameters

interaction	parameter	value
Coulomb	$\varepsilon_{surface}$	30
	ε_{buried}	160
desolvation	$c_{surface}$	3.375
	c_{buried}	13.5
intrinsic electrostatics	c^{HB}	0.85
	c^{RE}	0.80

Reference :

M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski & J. H. Jensen. 2011. *J. Chem. Theory Comput.* **7**, 525–537. http://dx.doi.org/10.1021/ct100578z

How accurate are the ΔpK_a estimates from PROPKA?



Reference :
M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski & J. H. Jensen. 2011. *J. Chem. Theory Comput.* **7**, 525–537.
http://dx.doi.org/10.1021/ct100578z

How accurate are the ΔpK_a estimates from PROPKA?

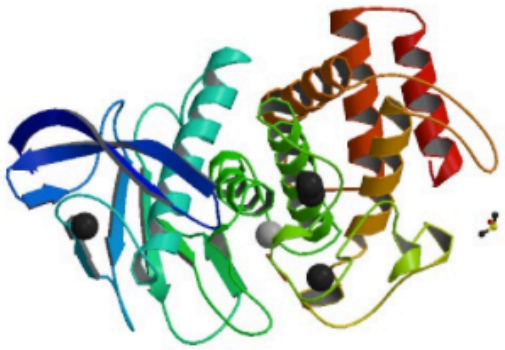
Table 3. rmsd Summarized for Each Residue Type

	COO	ASP	GLU	TYR	LYS	HIS
pK_a values	201	101	100	11	51	30
PROPKA 3	0.79	0.77	0.80	0.75	0.65	1.00
PROPKA 2	0.91	0.94	0.87	0.97	0.72	1.37
Null-model	1.06	1.23	0.86	0.70	1.01	0.93

scary...

Reference :
M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski & J. H. Jensen. 2011. *J. Chem. Theory Comput.* **7**, 525–537.
http://dx.doi.org/10.1021/ct100578z

Biological Assembly 1



3D View: Structure | Electron Density | Ligand Interaction

Standalone Viewers
Protein Workshop | Ligand Explorer

Standalone Viewers
Protein Workshop | Ligand Explorer

Global Symmetry: Asymmetric - C1
Global Stoichiometry: Monomer - A

Biological assembly 1 assigned by authors and generated by PISA (software)

Display Files | Download Files

1LNF

A structural analysis of metal substitutions in thermolysin

DOI: 10.2210/pdb1LNF/pdb

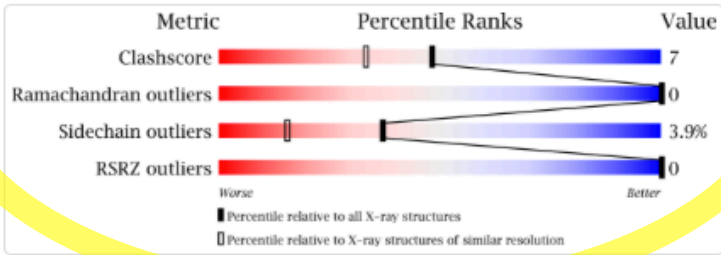
Classification: HYDROLASE (METALLOPROTEASE)
Organism(s): Bacillus thermoproteolyticus

Deposited: 1994-05-13 **Released:** 1995-05-08
Deposition Author(s): Holland, D.R., Matthews, B.W.

Experimental Data Snapshot

Method: X-RAY DIFFRACTION
Resolution: 1.7 Å

wwPDB Validation 3D Report | Full Report



This is version 1.5 of the entry. See complete history.

Literature

Download Primary Citation

Structural analysis of zinc substitutions in the active site of thermolysin.

Holland, D.R., Hausrath, A.C., Juers, D., Matthews, B.W.
(1995) Protein Sci. **4**: 1955-1965

PubMed: 8535232 Search on PubMed | Search on PubMed Central
DOI: 10.1002/pro.5560041001
Primary Citation of Related Structures: 1LNA, 1LNB, 1LNC, 1LND, 1LNE

MolProbity: all-atom structure validation for macromolecular crystallography

Detects:

- atomic clashes
- favorable hydrogen bonds and van der Waals (vdW) contacts
- C_β deviations and bad rotamers
- Ramachandran outliers (φ,ψ)
- Bond-angle and bond-length outliers

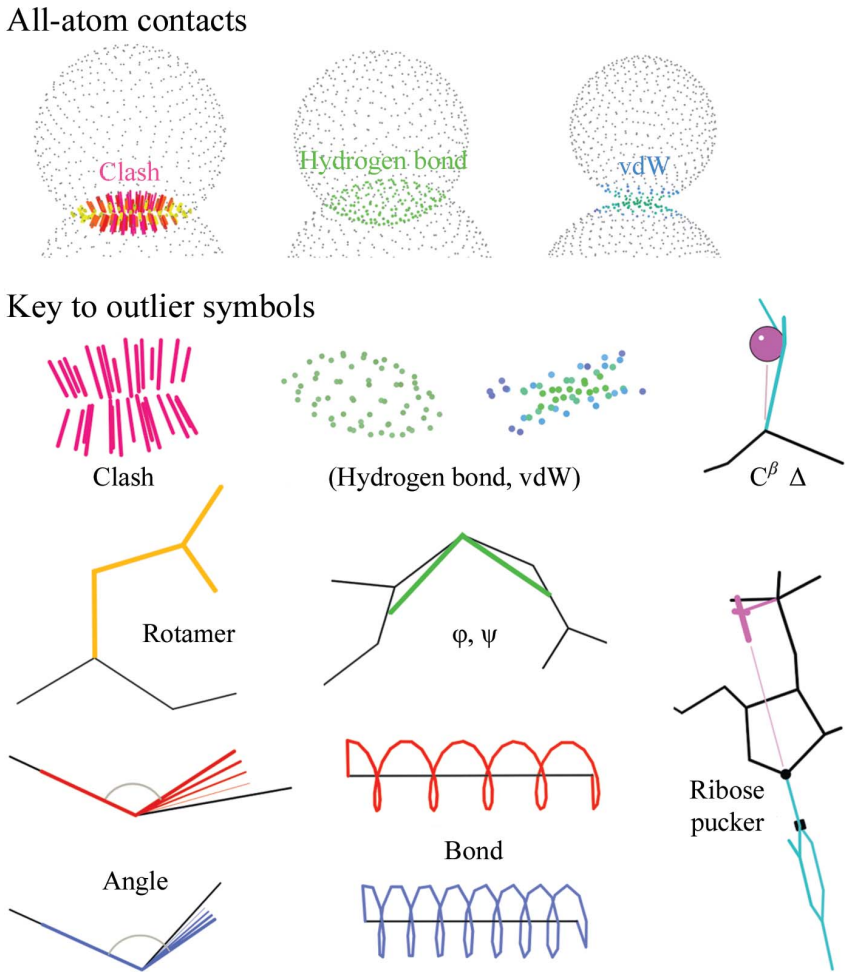


Figure 1
An outlier legend, showing each symbol used in a *MolProbity* multi-criterion kinemage and illustrating the relationship of the three types of all-atom contact to the atomic van der Waals (vdW) surfaces (spheres of small gray dots). The symbols for favorable hydrogen bonds and vdW contacts are included for completeness, as well as the hot-pink spikes of a clash outlier. A C^β deviation of $\geq$0.25 Å is shown as a magenta ball centered on the ideal C^β position and tangent to the modeled position. Bad rotamers are shown as gold side chains and Ramachandran outliers as heavy green lines to the midpoints of the two peptides. Bond-angle outliers are indicated by a fan of lines from the ideal to the modeled bond (red if wide, blue if narrow). Bond-length outliers are indicated as stretched (red) or compressed (blue) springs. A suspicious ribose pucker is diagnosed by the perpendicular distance from the 3′ (following) phosphate to the line of the glycosidic C1′—N1/9 bond and is flagged by a representation of that construction (in magenta if too short, as here, and in purple if too long).

Reference :
V. B. Chen et al. 2010. *Acta Cryst.* D**66**, 12–21. http://dx.doi.org/10.1107/S0907444909042073

The Peptide Bond: Dihedral Angles

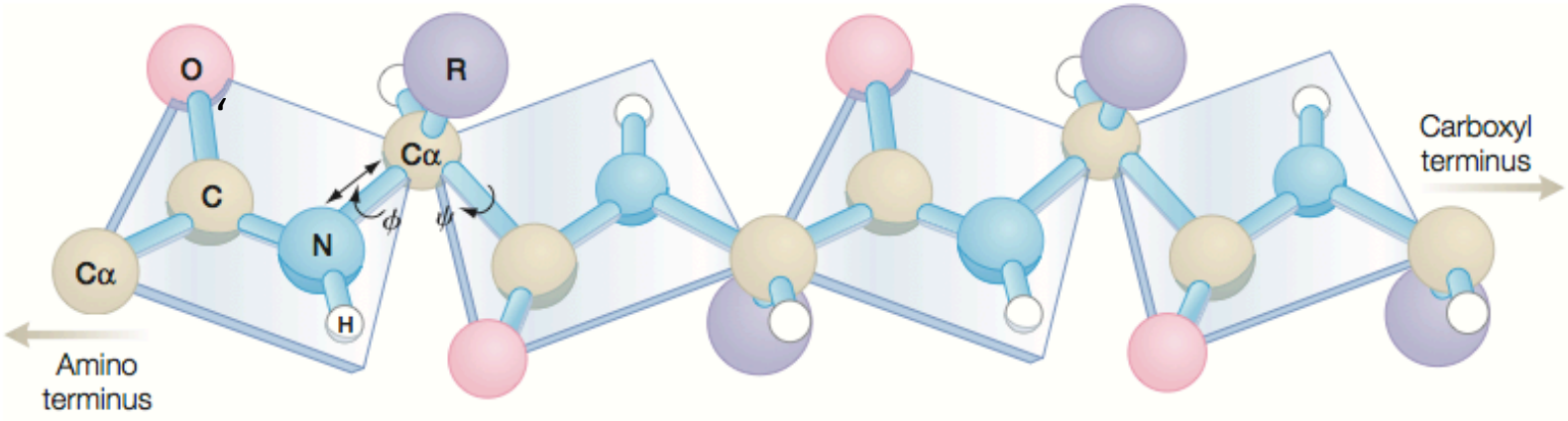


Figure 13.7 The planar peptide groups of a polypeptide chain. Each plane can rotate about the ϕ and ψ angles. Many combinations of angles are forbidden, including $\phi = \psi = 0°$.

Figure from: Allen, J.P. 'Biophysical Chemistry' (2008)

- peptide bonds are planar due to their partial double bond character
- planes are formed from the backbone carbonyl oxygen to the backbone amino hydrogen in each peptide subunit (residue)
- rotations of these planar groups are allowed
- rotations occur around the N-Cα and Cα-C' bonds; these bond angles are called Φ and Ψ, respectively, and are also known as dihedral angles
- due to steric hindrances caused by bulky R groups, the extents of rotations around Φ and Ψ are limited according to the amino acid composition of the polypeptide

Dihedral Angles – Another View of Φ, Ψ, and ω

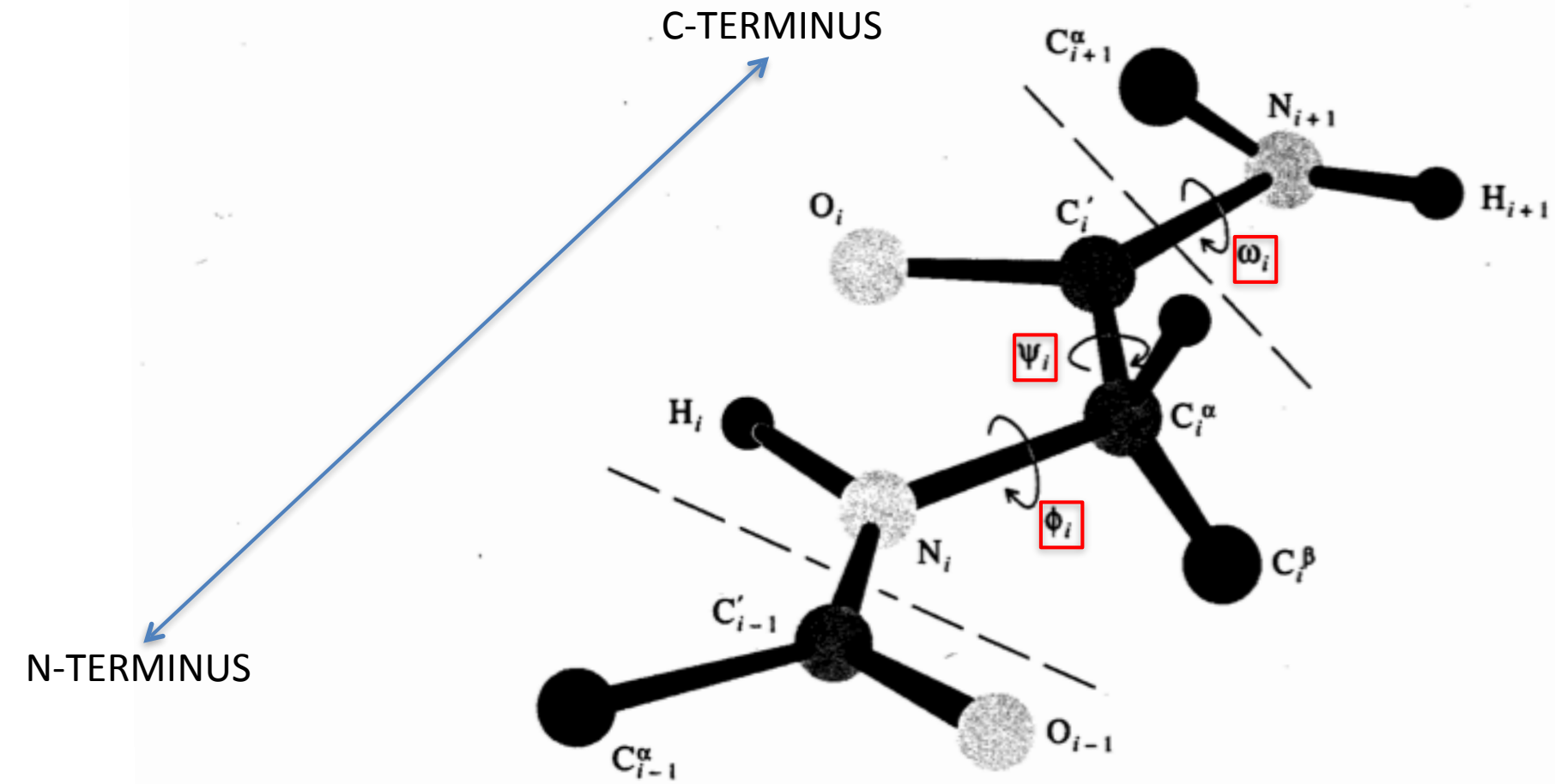


FIGURE 5.1
Perspective drawing of a segment of polypeptide chain comprising two peptide units. Only the C^{β} atom of each side chain is shown. The limits of a single residue (number i of the chain) are indicated by the dashed lines. The recommended notations for atoms and torsion angles are indicated. The polypeptide chain is shown in the fully extended conformation, where $\phi = \psi = \omega = 180°$.

Figure from: 'Proteins 2nd ed.' T.E. Creighton (1993) W.H. Freeman & Co.

- a fully extended polypeptide has Φ = Ψ = ω = 180°
- due to a steric clash between the backbone carbonyl oxygen and the amino hydrogen, Φ = Ψ = 0° is forbidden

Ramachandran Plots

- Ramachandran plots are a convenient way to visualize the distribution of backbone Φ and Ψ angles for all residues in a protein
- the angles on each axis vary from -180° to +180° ; since these angles are equivalent, the plots are continuous
- where a particular residue is allowed depends on its identity: Ala residues, for example, are found mostly in the upper-left quadrant and mid-left regions; Gly residues, due to higher backbone flexibility, are allowed in all four quadrants; Pro residues, due to their constrained nature, are found along Φ = -60°.

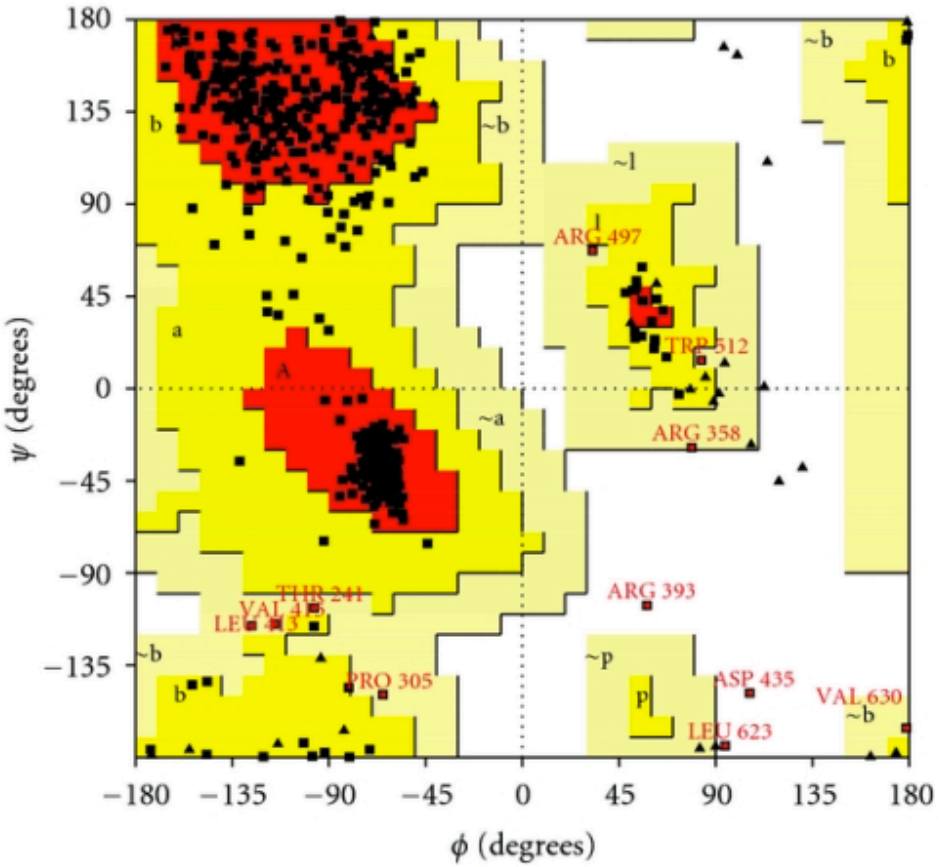
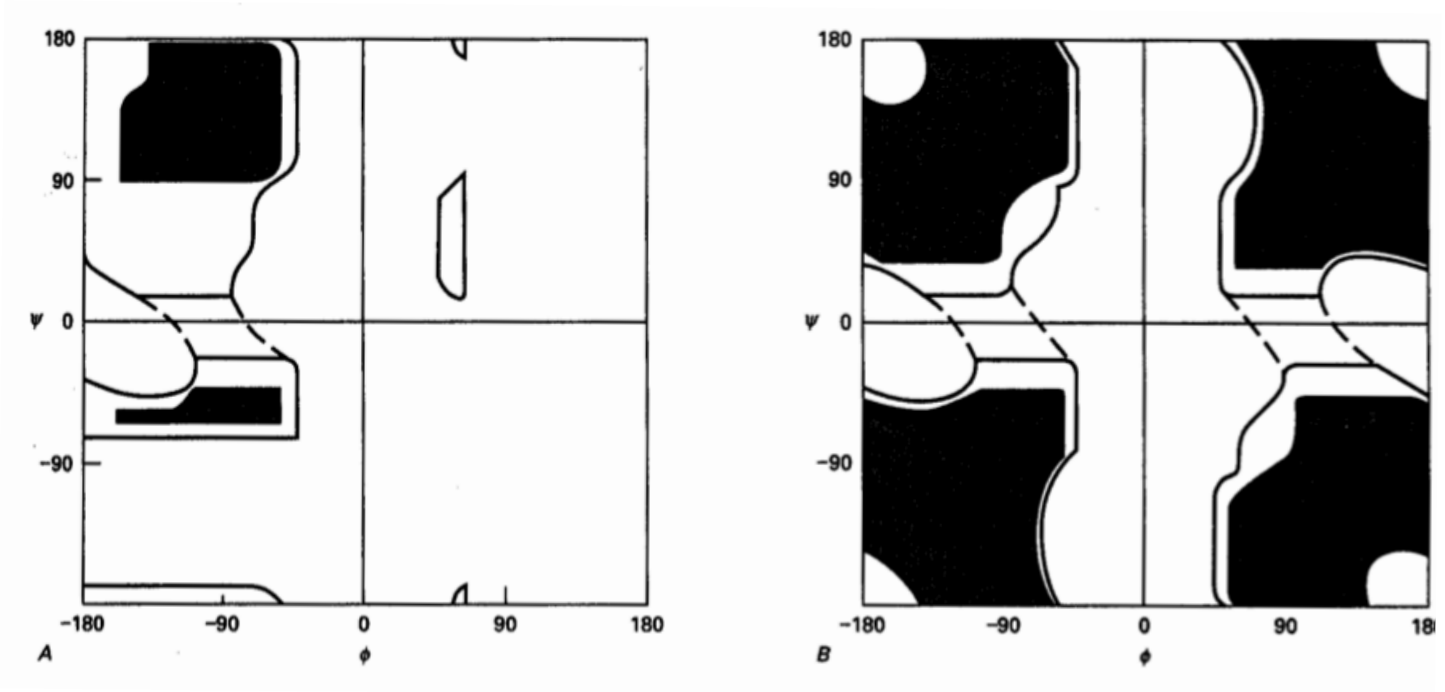


Figure from: http://dx.doi.org/10.4061/2011/278051

Ramachandran Plot of fatty acyl-CoA ligase. The black dots represent individual amino acid residues in the protein. The red regions are know as 'core' or 'favoured' regions. The dark yellow regions are 'allowed'. The light yellow regions are 'generous'. Residues outside these regions are 'disallowed', and are likely in an atypical local geometry, or indicate a problem with the protein structure.

Ramachandran Plots



Ala **Gly**

FIGURE 5.2
Ramachandran plots of the permitted values of ϕ and ψ for different residues. Each two-dimensional plot is continuous at the edges, because a rotation of $-180°$ is the same as one of $+180°$. The original plots that considered only repulsions between hard-sphere atoms are shown in *A* and *B* for Ala and Gly residues, respectively. The fully allowed regions are shaded; the partially allowed regions are enclosed by a solid line. The connecting regions enclosed by the dashed lines are permissible with slight flexibility of bond angles. The much greater flexibility of the Gly residue compared with Ala is apparent, as is the symmetry of the plot for Gly residues resulting from the absence of a chiral side chain
machandran & V. Saisekharan, *Adv. Protein Chem.* 23:283–437, 1968.)

Figure from: 'Proteins 2nd ed.' T.E. Creighton (1993) W.H. Freeman & Co.

MolProbity: all-atom structure validation for macromolecular crystallography

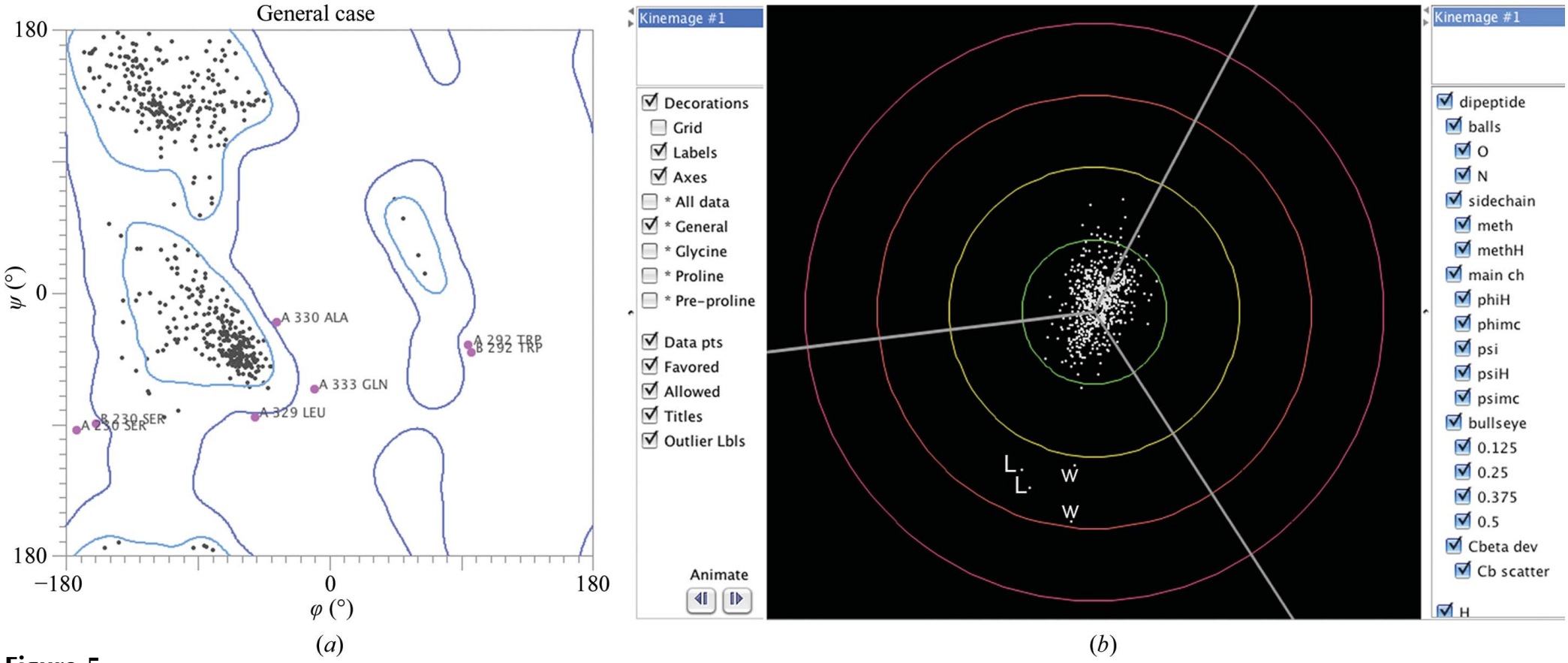


Figure 5

The general case Ramachandran kinemage and the C^{β} deviation kinemage for file 2dq4. In (*a*) the φ, ψ values for each residue are plotted on a background of the smoothed contours from high-quality data (see text). Over 98% lie inside the inner 'favored' 98% contour, but there are seven outliers outside the outer 99.95% contour. Gly, Pro and pre-Pro residues are on separate plots (not shown). In (*b*) the C^{β} deviation kinemage shows each residue's C^{β} position relative to an ideal C^{β} and its three bond vectors (gray lines). Circles mark the deviation distances, with the yellow circle at the 0.25 Å cutoff for outliers. Most of the distribution is good, but an adjacent Leu and Trp in each chain (labeled) are part of an outlier cluster and probably reflect distortions caused by a local fitting problem.

Reference :

V. B. Chen et al. 2010. *Acta Cryst.* D**66**, 12–21. http://dx.doi.org/10.1107/S0907444909042073