

Disulfide by Design Version 1.20
© Wayne State University 2003, All Rights Reserved
09/16/03

Users' Guide



Alan A. Dombkowski, Ph.D., Institute of Environmental Health Sciences,
Wayne State University, Detroit, MI domski@wayne.edu

Contents

1	Description.....	2
2	System Requirements.....	2
3	Installation.....	2
4	Running Disulfide by Design.....	2
4.1	Loading a PDB file.....	3
4.2	Run.....	4
4.3	Output.....	4
4.4	Saving Results.....	4
4.5	Creating a mutant PDB file.....	5
5	Options.....	5
5.1	Inter and Intra-chain.....	5
5.2	Building C _β for glycines.....	5
5.3	Chi3 torsion angle tolerance.....	5
5.4	Ca-Cb-S angle tolerance.....	6
6	Technical Details.....	6
7	Release Notes.....	7
8	References.....	7

1 Description

Disulfide by Design is an application for the rational design of disulfide bonds in proteins. For a given protein structural model, all residue pairs are rapidly assessed for proximity and geometry consistent with disulfide formation, assuming the residues were mutated to cysteines. The output displays residue pairs meeting the appropriate criteria. The input model will typically be a Protein Data Bank (PDB) structure for the protein of interest; however, structures developed through homology modeling may also be used. Engineered disulfides have proven useful for increasing the stability of proteins and to assist the investigation of protein dynamics and interactions. This software was written by Dr. Alan Dombkowski and based on algorithms created for disulfide identification in protein fold recognition methods (Dombkowski & Crippen, 2000). The Disulfide by Design algorithm has been successfully used for disulfide engineering (Anthony et al., 2002; Anthony & Burgess, 2002).

2 System Requirements

Disulfide by Design is currently available for the Windows operating system. It has been tested with Windows 98, NT, 2000 and XP. The application uses about five megabytes of memory and less than one megabyte of disk space. Monitor resolution should be a minimum of 1024 x 768 with 16 or 32 bit color.

3 Installation

Simply double click on “Disulfide by Design Install.exe” and the installer will step you through the installation process. A license agreement will be displayed during the installation process. Proceeding with installation requires agreement with the software license. Upon installation, an icon will be installed under the “Programs” tab of the “Start” menu. To run the program simply double-click the icon. The software can be uninstalled using the Add/Remove Programs tab of the control panel.

4 Running Disulfide by Design

Running Disulfide by Design is as simple as loading a PDB file then pressing “Run.” The default option settings are recommended for general use. However, it may be desirable to change option settings, such as to increase assessment criteria stringency. The options are discussed in greater detail below.

4.1 Loading a PDB file

The input file must be in PDB format with atomic coordinates. File selection is performed with the “Load Structure” button. All residues with a complete main chain backbone and a C_β atom are used in the analysis. In the case of glycine, a C_β will be created if the “Build C_β for Gly” check box is selected. After the structure file has been read a message is displayed showing the number of residues having a complete backbone and C_β. For NMR structures with multiple models only the first model will be used.

1 Load

2 Run

3 Save

Chain	Residue 1	AA	Chain	Residue 2	AA	Chi3	Energy
A	8	PHE	A	12	ILE	+106.11	4.20
A	9	GLN	A	98	LYS	+100.87	3.72
A	15	LEU	A	101	THR	+97.13	1.79
A	20	ILE	A	128	PHE	+115.62	2.80
A	38	ALA	A	79	ASP	-82.54	7.59
A	44	LYS	A	82	ILE	-97.52	2.67
A	58	THR	A	169	ARG	+116.57	3.65
A	59	GLU	A	169	ARG	+101.91	3.56
A	83	GLY	A	140	SER	-78.94	3.32
A	85	THR	A	143	ASP	-78.37	2.27
A	89	LEU	A	148	TRP	+127.80	6.75
A	107	ALA	A	112	ALA	-93.26	2.28
A	118	GLY	A	123	GLY	-88.18	5.02
A	130	VAL	A	181	GLN	-102.05	1.52
A	131	ALA	A	175	LEU	+108.09	3.74
A	133	ARG	A	151	ASN	+125.65	3.49
A	145	GLN	A	161	ASP	-99.15	6.07
A	150	SER	A	156	PHE	+104.32	4.76
A	186	GLU	A	189	ARG	+120.57	2.21
B	2	ALA	B	157	THR	+98.50	2.28
B	8	PHE	B	12	ILE	+112.25	4.91
B	9	GLN	B	98	LYS	+96.48	3.95
B	11	GLU	B	102	LYS	+127.44	5.32
B	15	LEU	B	101	THR	+103.02	2.02
B	16	MET	B	152	ALA	+100.67	2.35
B	20	ILE	B	128	PHE	+106.82	3.49
B	32	ARG	B	197	HIS	-83.06	2.03
B	38	ALA	B	79	ASP	-82.32	6.86
B	44	LYS	B	82	ILE	-98.85	2.33
B	58	THR	B	169	ARG	+121.15	3.87
B	59	GLU	B	169	ARG	+102.53	3.21

Current structure file: F:\MAH3.pdb

4.2 Run

After successful loading of a PDB file, simply press “Run” to start the analysis. Each possible pair of residues will be assessed for potential disulfide formation, assuming the residues were mutated to cysteines. With the default setting, both inter and intra-chain disulfides are analyzed. Deselect the appropriate check box to change the intra/inter-chain setting.

4.3 Output

The output text box displays one line per potential disulfide. Each line shows a pair of residues that have the appropriate geometry to form a disulfide bond if the residues were mutated to cysteines. The analysis is based on the assumption that the mutations would not significantly perturb the protein backbone conformation. Each displayed disulfide meets the minimum geometric criteria, including user-selected parameters. The chain id, residue number, and original amino acid type are shown for each residue in the pair. Chain id and residue number are consistent with the input PDB designations. Also shown are the estimated χ_3 torsion angle and an energy value in kcal/mol. χ_3 angles of known disulfides have a bimodal distribution with peaks observed at +100 and -80 (Petersen *et al.*, 1999), see figure in technical details. These values provide useful selection criteria for disulfide design. The estimated chirality and torsion angle are based on the best possible orientation of putative mutant cysteine S γ atoms, as determined by an energy minimization performed within Disulfide by Design (see technical details below). The energy value is useful for comparison of potential disulfides to select the best possible candidates. Disulfides with a lower energy value are preferential to those with a higher score. The calculated energy is not intended as a comprehensive assessment of conformational energy, but it is provided to allow comparisons of prospective disulfide bonds. Figure 3 is a histogram of energy values calculated for 706 known disulfides using the Disulfide by Design algorithm. This data reveals the distribution of energy values for naturally occurring disulfides and may be helpful when selecting putative bonds for disulfide engineering. The mean energy value is 1.07 kcal/mol and the maximum observed value is 7.91 kcal/mol.

4.4 Saving Results

Results displayed in the output text box can be saved to a user-specified file using the “Save Results” button. Text is saved in tab-delimited format, so the data can easily be imported into Excel or other spreadsheet software.

4.5 Creating a mutant PDB file

A modified form of the original input PDB file can be created by selecting one or more of the potential disulfides then pressing “Create PDB File.” Disulfides are selected by simply pointing the mouse at the desired line then left-clicking. Disulfides may be deselected by clicking a second time. For each disulfide selected, the original amino acids are replaced in the PDB file with cysteines. The respective SSBOND records are inserted near the beginning of the mutant PDB file. The mutant PDB file can be subsequently imported into molecular modeling software such as Rasmol for visualization of the designed disulfides.

5 Options

5.1 Inter and Intra-chain

By default, Disulfide by Design checks all potential inter and intra-chain disulfides. However, it may be desirable to restrict analysis to either inter or intra-chain. This is accomplished by simply deselecting the appropriate check box.

5.2 Building C_{β} for glycines

The disulfide design algorithm requires coordinates for C_{β} atoms to determine the potential for disulfide formation. Since glycine residues do not include a C_{β} atom they cannot be used in the analysis unless a C_{β} is created. The “Build C $_{\beta}$ for Gly” check box enables the construction of C_{β} atoms. The C_{β} location is determined by using the coordinates of the residue backbone atoms. The algorithm was tested by building C_{β} atoms for cysteine residues of proteins with a known structure. The coordinates of the constructed atoms were compared with the actual C_{β} locations, and the average distance was only 0.12 Å indicating accurate modeling of the C_{β} atoms.

5.3 Chi3 torsion angle tolerance

The χ_3 torsion angle is formed by the C_{β} -S γ -S γ -C $_{\beta}$ bonds, with rotation about the S γ -S γ bond (see Figure 1 below). The distribution of χ_3 angles observed in disulfides of known protein structures is bimodal with sharp peaks at +100 and -80 (see Figure 2 below). It may be desirable to restrict candidate disulfides to those having an estimated χ_3 value that falls within the region of observed χ_3 values. The χ_3 angle tolerance can be selected in the options box. The default setting is $+100 \pm 30^\circ$ and $-80 \pm 30^\circ$. If numerous putative disulfides are identified using the default setting it may be useful to decrease the tolerance resulting in a shorter list with preferential characteristics.

5.4 Ca-Cb-S angle tolerance

The distribution of $C_\alpha-C_\beta-S_\gamma$ angles observed in known disulfides has a peak near 115° and covers a range from approximately 105° to 125° (Petersen *et al.*, 1999). The tolerance of this bond angle is selectable, with a default setting of $114.6 \pm 10^\circ$.

6 Technical Details

Figure (1) represents a cysteine pair coupled by a disulfide bond. The approach uses a disulfide model with fixed $C_\beta-S_\gamma$ and $S_\gamma-S_\gamma$ bond lengths, 1.81 and 2.04 Å respectively, and fixed $C_\beta-S_\gamma-S_\gamma$ bond angles of 104.15° . These bond lengths and angles are consistent with values observed in a survey of protein disulfide bonds (Petersen *et al.*, 1999). The χ_3 torsion angle, formed by rotation of the C_β atoms about the $S_\gamma-S_\gamma$ bond, is allowed to vary in the model and can be described as a function of the distance between C_β atoms (Dombkowski & Crippen, 2000). To “fit” the disulfide model between a pair of residues, the algorithm simply rotates the χ_3 angle to obtain a $C_\beta-C_\beta$ distance that matches the $C_\beta-C_\beta$ distance measured between the residues. Numerous S_γ locations are possible, so all possible S_γ orientations are examined and the atomic coordinates providing the lowest energy (E_{ij}) are selected, based on the χ_1 and χ_3 torsion angles and the two $C_\alpha-C_\beta-S_\gamma$ angles. The χ_1 torsion angle is defined by the N- $C_\alpha-C_\beta-S_\gamma$ atoms. E_{ij} is calculated per equations (1-4), where i and j are residue indices, θ is the $C_\alpha-C_\beta-S_\gamma$ angle, and θ_0 is set to 114.6° . Energy units are kcal/mol.

$$E_{ij} = E(\chi_{1,i}) + E(\chi_{1,j}) + E(\chi_3) + E(\theta_i) + E(\theta_j) \quad (1)$$

$$E(\chi_1) = 1.4[1 + \cos(3\chi_1)] \quad (2)$$

$$E(\chi_3) = 4.0[1 - \cos(2\chi_3 + 160)] \quad (3)$$

$$E(\theta) = 55.0[\theta - \theta_0]^2 \quad (4)$$

The energy calculation provides minima at χ_3 values of $+100^\circ$ and -80° and χ_1 values of $\pm 60^\circ$ and $\pm 180^\circ$, consistent with values observed in the latest survey of disulfide bonds (Petersen *et al.*, 1999). Since the disulfide model uses fixed bond lengths, no term is included for bond stretching. The energy calculation provides a means to compare potential disulfides during disulfide design. The distribution of energy values for 706 known disulfide bonds reveals a mean energy value, calculated per equations (1-4), of 1.07 kcal/mol and a maximum of 7.91 kcal/mol.

7 Release Notes

Changes in version 1.20 include:

- Corrected bug that caused text to go beyond the right-hand border of the dialog box with some terminal settings.

Changes in version 1.12 include:

- Energy units are now in kcal/mol.
- For residues with multiple conformers only the first set of coordinates encountered are used.
- For NMR structures with multiple models only the first model found in the PDB file is used.
- Fixed bug in torsion angle calculation that caused crashes for a small number of PDB structures.

8 References

Dombkowski A.A., 2003, Disulfide by Design: A computational method for the rational design of disulfide bonds in proteins, *Bioinformatics*, vol. 19 no. 14, 22 Sep.2003.

Anthony, L.C. and Burgess, R.R., 2002, Conformational Flexibility in sigma 70 Region 2 during Transcription Initiation, *J Biol Chem*. Nov 29;277(48):46433-46441.

Anthony, L.C., Dombkowski, A.A., and Burgess, R.R., 2002, Using disulfide engineering to study conformational changes in the β' 260-309 coiled-coil region of *E. Coli* RNA polymerase during σ 70 binding, *J Bacteriol*. May;184(10):2634-41.

Dombkowski A.A. and Crippen G.M., 2000, Disulfide recognition in an optimized threading potential, *Protein Engineering*, vol. 13 no. 10, 679-689.

Petersen, M., Jonson, P.H., and Petersen, S.B., 1999, Amino acid neighbours and detailed conformational analysis of cysteines in proteins, *Protein Engineering*, vol. 12, no 7, 535-548.

Anatomy of a Disulfide

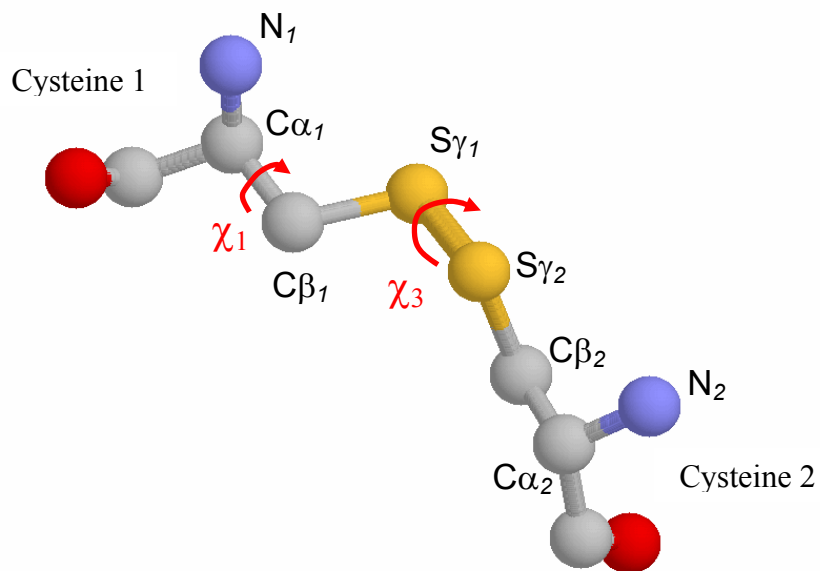


Figure 1

χ_3 distribution in 351 disulfides, Petersen et al., 1999, Protein Engineering

M.T.N.Petersen *et al.*

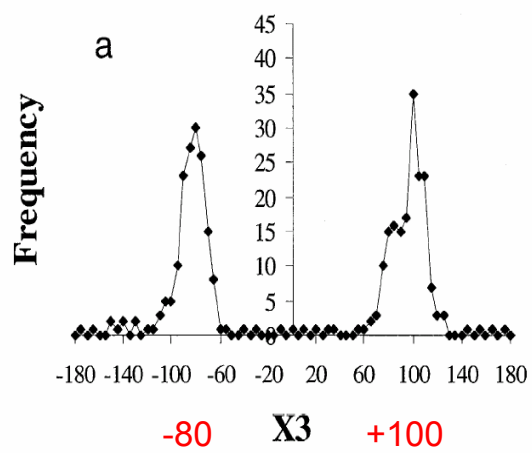


Figure 2

Energy distribution of 706 known disulfide bonds

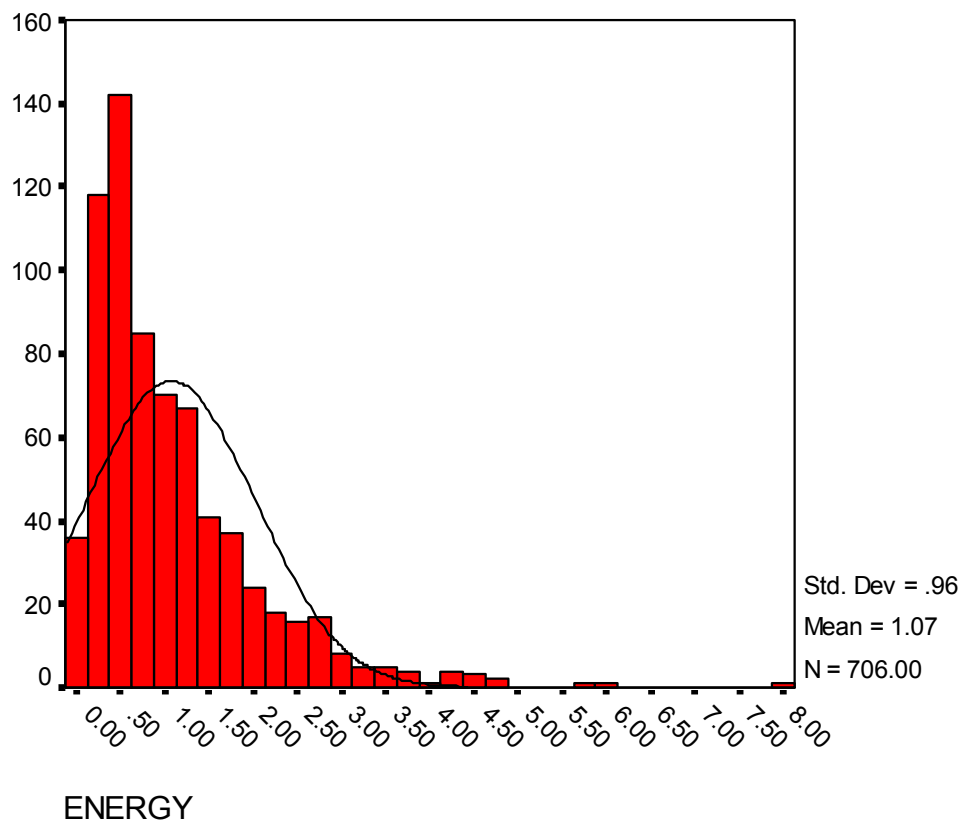


Figure 3