1 SUPPLEMENTARY MATERIAL

1.1 Mathematical properties of the Cauchy-Binet kernel

The BC scores are maximum for identical structures.

$$\text{RMSD}\,(X,Y) = 0 \implies \text{BC}\,(X,Y) = 1 \tag{1}$$

However, the reciprocal statement is not true. It is possible that $\text{BC}\,(X,Y) = 1$ for two different fragments with $RMSD(X,Y) > 0$.

To have a better understanding of what is measured by BC score, we perform a scaling on protein structures. The structures are scaled along their axis of inertia so that each one contributes equally to the radius of gyration.This normalisation makes the structure spherical in shape. The matrix coordinates of the scaled structures are

$$\hat{X} = X(X^\top X)^{-1/2}$$

and

$$\hat{Y} = Y(Y^\top Y)^{-1/2}$$

From this, we have:

$$\text{BC}\,(X,Y) = \det((X^\top X)^{-1/2} X^\top Y (Y^\top Y)^{-1/2})$$

$$\iff \text{BC}\,(X,Y) = \frac{\det(X^\top Y)}{\det((X^\top X)^{1/2}) \det((Y^\top Y)^{1/2})}$$

and

$$\text{BC}\,(X,Y) = \det(\hat{X}^\top \hat{Y}) \tag{2}$$

From RMSD formula, we have

$$RMSD^2(\hat{X}, \hat{Y}) = \frac{1}{N}(\|\hat{X}\|^2 + \|\hat{Y}\|^2 - 2(s\sigma_1 + \sigma_2 + \sigma_3)) \tag{3}$$

with $\hat{\sigma}_i$ denoting the singular values of the correlation matrix $\hat{X}^\top \hat{Y}$.

The scale matrices $\hat{X}$ and $\hat{Y}$ are orthormals $\hat{X}^\top \hat{X} = I$, $\hat{Y}^\top \hat{Y} = I$ The singular values of $\hat{X}$ and $\hat{Y}$ are equal to 1, so that $\|\hat{X}\|^2 = \sum_i \hat{\sigma}_i{}^2(X) = 3$, $\|\hat{X}\|^2 = 3$.

Using formula (3) we have:

$$\text{RMSD}^2(\hat{X}, \hat{Y}) = \frac{1}{N}(6 - 2(s\hat{\sigma}_1 + \hat{\sigma}_2 + \hat{\sigma}_3))$$

s is equal to the sign of $\det(\hat{X}^\top \hat{Y})$ and $\det(X^\top Y)$. Let us consider positive value of BC or equivalently positive s.

$$\text{RMSD}^2(\hat{X}, \hat{Y}) = \frac{1}{N}(6 - 2(\hat{\sigma}_1 + \hat{\sigma}_2 + \hat{\sigma}_3) \tag{4}$$

Using the concavity of the log function, we have for any positive x

$$\log(x) \leq x - 1.$$

This inequality is valid for any positive x and is tight for x close to 1. From 4

$$\text{RMSD}^2(\hat{X}, \hat{Y}) \leq \frac{1}{N}(6 - 2\sum_i (1 + \log(\hat{\sigma}_i))$$

$\iff$

$$\text{RMSD}^2(\hat{X}, \hat{Y}) \leq -\frac{2}{N}(\sum_i \log(\hat{\sigma}_i))$$

$\iff$

$$\textbf{RMS}D^2(\hat{X}, \hat{Y}) \leq -\frac{2}{N}(\log \prod_i \hat{\sigma}_i)$$

$$\text{RMSD}^2(\hat{X}, \hat{Y}) \leq -\frac{2}{N}(\log |\det(\hat{X}^\top \hat{Y})|)$$

$\iff$

$$\text{RMSD}^2(\hat{X}, \hat{Y}) \leq -\frac{2}{N} \log(\text{BC}\,(X,Y))$$

Finally,

$$\text{RMSD}^2(\hat{X}, \hat{Y}) \leq -\frac{2}{N} \log(\text{BC}\,(X,Y))$$

1.1.1 *BC score is a flexible score*   The above inequality implies that

$$\mathrm{BC}\,(X,Y) = 1 \Rightarrow \mathrm{RMSD}\,(\hat{X}, \hat{Y}) = 0$$

The reverse statement is obviously true.

$$\mathrm{BC}\,(X,Y) = 1 \Leftrightarrow \mathrm{RMSD}\,(\hat{X}, \hat{Y}) = 0 \tag{5}$$

This implies that it exists a 3×3 rotation matrix R such that

$$\hat{X} = \hat{Y} R$$

By definition, the Grassman vector representing R is equal to $\det(R) = 1$.

$$\Phi(\hat{X}) = \Phi(\hat{Y} R) = \Phi(\hat{Y}) \Phi(R)$$

Finally

$$\mathrm{BC}\,(X,Y) = 1 \Leftrightarrow \phi(X) = \phi(Y) \tag{6}$$

We also have

$$\hat{X} = \hat{Y} R$$
$$\Leftrightarrow X(X^\top X)^{-1/2} = Y(Y^\top Y)^{-1/2} R$$
$$\Leftrightarrow X = Y(Y^\top Y)^{-1/2} R (X^\top X)^{1/2}$$

Therefore, it exists a real 3×3 matrix A

$$X = YA$$

This implies the structure X is a linear transformation of the structure Y. As $\det(A) > 0$, this transformation is a combination of different scaling along given axes and a rotation. This excludes symmetries and scaling by a negative number. Otherwise stated, the BC score is invariant to positive linear transformation of the fragments.

$$\mathrm{BC}\,(X,Y) = 1 \Leftrightarrow \text{it exists an invertible matrix such that} X = YA \tag{7}$$

More generally, we have

$$\mathrm{BC}\,(XA, YB) = \mathrm{BC}\,(X,Y) \tag{8}$$

Suppose $\det(A) > 0$ and $\det(B) > 0$ then we have:

$$\Longrightarrow \mathrm{BC}\,(XA, YB) = \frac{\det(A^\top X^\top Y B)}{\sqrt{\det(A^\top X^\top X A)\det(B^\top Y^\top Y B)}}$$

$$\Longrightarrow \mathrm{BC}\,(XA, YB) = \frac{\det(X^\top X)\det(A)\det(B)}{\sqrt{\det(A)^2 \det(X^\top X)\det(B)^2 \det(Y^\top Y)}}$$

$$\Longrightarrow \mathrm{BC}\,(XA, YB) = \mathrm{BC}\,(X,Y)$$

In particular, we have $BC(X, XA) = 1$.

1.2 Distribution of simulated fragments for different lengths

Table 1. Mean and standard deviation of BC score for generated and real fragments

Length	gen. fr. mean	gen. frag. sd	real fr. mean	real fr.sd
10				
20	1.98e-03	0.158	1.43e-02	0.213
30	2.37e-03	0.144	4.19e-03	0.234
40	4.06e-03	0.139	1.74e-03	0.273
50	4.60e-03	0.140	1.46e-03	0.257
60	3.94e-03	0.141	6.57e-03	0.226

Table 2. Estimated parameters of the exceedance distribution relative to different fragment lengthes

Length	scale	shape	F(u)
10			
20	$0.1069 \pm 3.15e\text{-}3$	$-0.2136 \pm 6.29e\text{-}3$	$0.9794 \pm 4.36e\text{-}4$
30	$0.0788 \pm 2.39e\text{-}3$	$-0.1576 \pm 4.90e\text{-}3$	$0.9820 \pm 4.15e\text{-}4$
40	$0.0774 \pm 1.74e\text{-}3$	$-0.1549 \pm 3.45e\text{-}3$	$0.9719 \pm 4.95e\text{-}4$
50	$0.0739 \pm 1.88e\text{-}3$	$-0.1479 \pm 3.75e\text{-}3$	$0.9790 \pm 4.52e\text{-}4$
60	$0.0689 \pm 2.48e\text{-}3$	$-0.1376 \pm 5.07e\text{-}3$	$0.9877 \pm 3.40e\text{-}4$

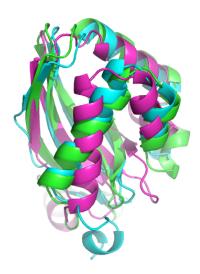## 1.3   BC scoring of CASP10 models

Fig. 1. Green: CASP10 T0743d1 experimental conformation Cyan: Server model. BC score 0.69. TM-score 0.74 Magenta: Server model. BC score 0.70. TM-score 0.46 The mismatch in the helix orientations (magenta model) and the mismatch in the bottom helix (cyan model) result in similar BC scores over the complete structures.