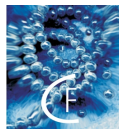


# SAXSPy and Programs for SAXS Data Treatment



Marcelo H. Schwade

Instituto de Física - UFRGS



15 de Setembro de 2021

# Pacote SAXSPy

SAXSPy é um pacote desenvolvido em Python, onde são implementadas classes e funções (métodos) que auxiliam nos processos de tratamento e análise de dados de SAXS.

Para conhecer as opções do pacote e entender seu funcionamento a melhor opção é consultar os códigos que fazem parte do pacote.

Baseados nesse pacote, foram desenvolvidos alguns programas para tratamento de dados SAXS.

## Modo de usar:

O diretório “saxspy” (com os códigos do pacote) deve estar no mesmo diretório do código/programa a ser executado;

No cabeçalho do programa:

```
> import saxspy.saxspy as saxs
```

Observação: o termo “saxs” é apenas um alias para designar o pacote ao longo do programa, pode ser utilizada outra palavra no lugar.

# Pré-requisitos de SAXSPy & Derivados

**Numpy**<sup>1</sup>: para operações de cálculo numérico, estruturas de dados e operações com arrays e matrizes ([<numpy.org/>](http://numpy.org/));

**Matplotlib**<sup>2</sup>: para construção de gráficos com qualidade e versatilidade ([<matplotlib.org/>](http://matplotlib.org/));

**Scipy**<sup>3</sup>: para recursos especiais, como funções especiais e métodos de ajuste e otimização ([www.scipy.org](http://www.scipy.org)).

**Todas essas bibliotecas possuem rica documentação das suas funções na Internet!**

---

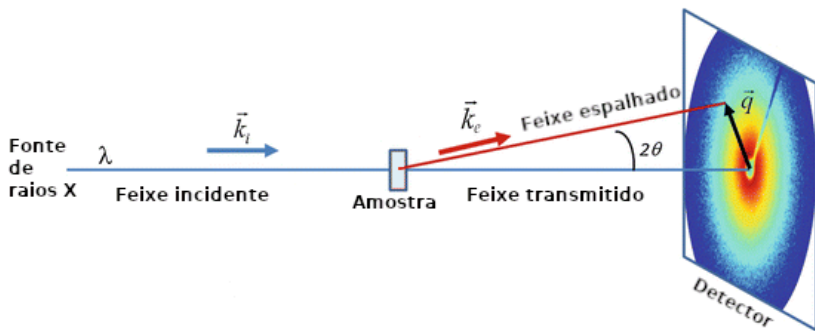
<sup>1</sup> Harris et al. (2020)

<sup>2</sup> Hunter (2007)

<sup>3</sup> Virtanen et al. (2020)

# Tratamento de dados e programas derivados do SAXSPy

# Esquema de uma medida de SAXS



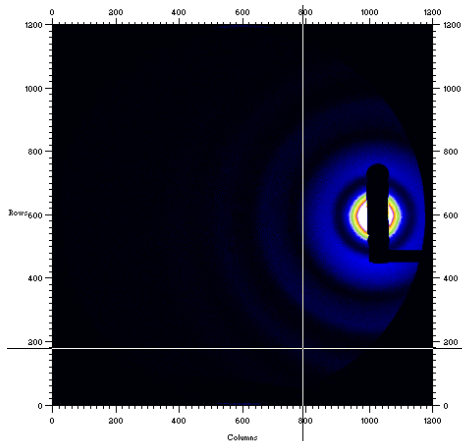
$$q = \frac{4\pi \sin \theta}{\lambda}$$

$2\theta$  é o ângulo de espalhamento e  $\lambda = 1.5419 \text{ \AA}$  ( $\text{CuK}\alpha$ ).

# Medidas de SAXS

As medidas de SAXS obtidas pelo equipamento são um padrão de difração bidimensional, logo, formam uma **imagem**.

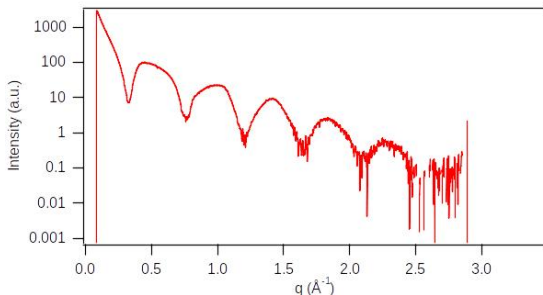
No caso de uma amostra com propriedades isotrópicas (a maioria das nossas amostras de interesse), esse padrão tem **simetria azimutal**: depende apenas do ângulo  $2\theta$  e, portanto, do módulo  $q$ .



# Medidas de SAXS (cont.)

Portanto, a medida de intensidade pode ser integrada azimutalmente para obter uma curva (unidimensional).

Esse processo de integração é, em geral, realizado por programas auxiliares do próprio equipamento, de modo que todo o tratamento discutido daqui por diante é realizado a partir das curvas já integradas (arquivos “.dat”).





# Médias das medidas de SAXS

Muitas vezes as medidas de SAXS que demandam tempos longos para aquisição dos dados são realizadas “quebrando” uma medida longa em várias medidas curtas.

Isso evita que o tempo dedicado a uma medida longa seja inteiramente perdido por conta de incidentes, como quedas de energia elétrica.

Para reunir todos esses conjuntos de medidas curtas em um único conjunto de dados tomamos a média dessas medidas.<sup>4</sup>

---

<sup>4</sup>Adotamos aqui uma média ponderada pelo tempo de exposição de cada medida individual, contudo esse tempo costuma ser o mesmo para as medidas individuais.

## a) MeanMaker

**Faz a média dos arquivos de dados de SAXS experimentais.<sup>5</sup>**

### **Entrada:**

- Arquivo contendo a lista dos arquivos a ser computada a média;
- Diretório onde se encontram os arquivos descritos na lista;
- Nome do arquivo de saída (média).

### **Saída:**

- Arquivo com a média dos dados (q, intensidade, incerteza da intensidade);
- Gráfico na tela com a média calculada.

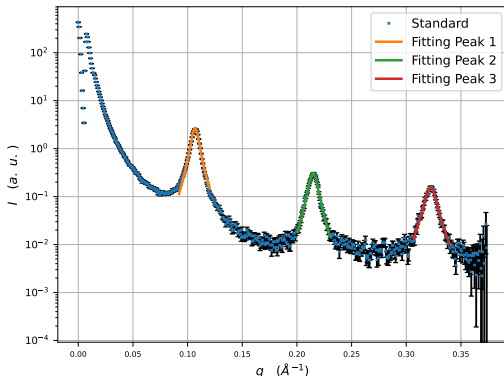
---

<sup>5</sup>Esse programa foi escrito para fazer a média de dados obtidos com o Xenocs Nano-Inxider (CNANO), dados obtidos com outros equipamentos podem demandar ajustes no código.

# Calibração da escala $q$

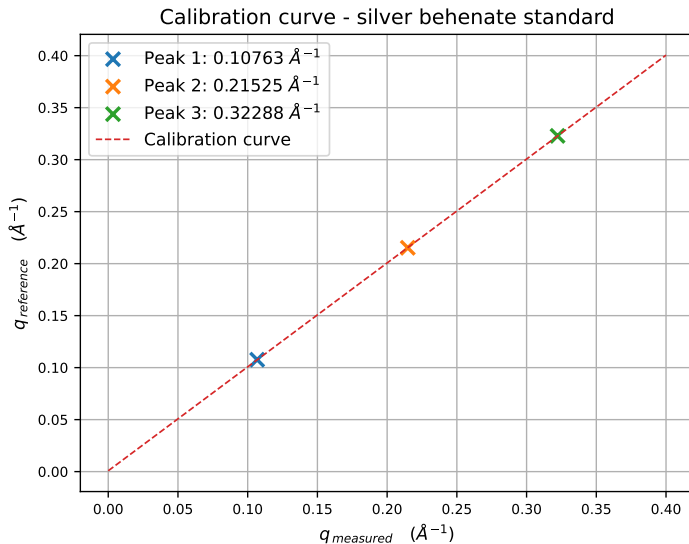
O alinhamento do feixe e a distância amostra/detector podem apresentar variações que alteram a escala de  $q$  (vetor transferência de momentum) pré-calibrada do equipamento.

Para corrigir essas discrepâncias são realizadas medidas de um padrão de behenato de prata<sup>6</sup>.



<sup>6</sup> Huang et al. (1993)

# Calibração da escala q



## b) q-ScaleFit

**Determina os parâmetros para correção/calibração da escala  $q$  de uma medida.**

### **Entrada (principal):**

- Arquivo de dados da medida de behenato de prata (média);
- Nome do arquivo de saída (resultados dos ajustes).

### **Saída (intermediária):**

- Gráfico na tela do espalhamento do behenato de prata.

## b) q-ScaleFit (cont.)

### **Entrada (intermediária):**

- Intervalos de ajuste dos picos;

### **Saída (principal):**

- Gráfico na tela com o resultado do ajuste dos picos.
- Arquivo com os resultados dos ajustes dos picos e da regressão linear de calibração da escala  $q$ ;
- Gráfico na tela com o resultado da regressão de calibração da escala  $q$ .

# Tratamento de dados SAXS

O tratamento da intensidade de uma medida de SAXS é feito conforme a equação (MARIANI; SPINOZZI, 1999, p. 45):

$$I_{abs}(q) = \frac{1}{A} \left\{ \left[ \frac{1}{D_{sample}} \left( \frac{I_{sample}}{t_{sample}} - \frac{I_{cap}}{t_{cap}} \right) \right] - (1 - \phi_V) \left[ \frac{1}{D_{solvent}} \left( \frac{I_{solvent}}{t_{solvent}} - \frac{I_{cap}}{t_{cap}} \right) \right] \right\}$$

# Tratamento de dados SAXS

Identificamos 3 blocos de operações nessa equação:

$$I_{abs}(q) = \frac{1}{A} \left\{ \overbrace{\left[ \underbrace{\frac{1}{D_{sample}} \left( \frac{I_{sample}}{t_{sample}} - \frac{I_{cap}}{t_{cap}} \right)}_{\text{①}} - (1 - \phi_V) \underbrace{\frac{1}{D_{solvent}} \left( \frac{I_{solvent}}{t_{solvent}} - \frac{I_{cap}}{t_{cap}} \right)}_{\text{①}} \right]}_{\text{②}} \right\}^{\text{③}}$$



## Os blocos de operações, na sequência de suas realizações, são:

- ① Correção pela transmissão, pelo capilar e pela espessura;<sup>7</sup>
- ② Correção pelo solvente;
- ③ Correção para escala absoluta.

---

<sup>7</sup>Note que as operações nesse bloco são idênticas para as medidas da amostra e do solvente.

# Programas para tratamento de dados

## Programas principais<sup>8</sup>:

- ① CTTq-Correction
- ② Solvent-Correction
- ③ AbsoluteScale-Correction

## Programas auxiliares<sup>9</sup>:

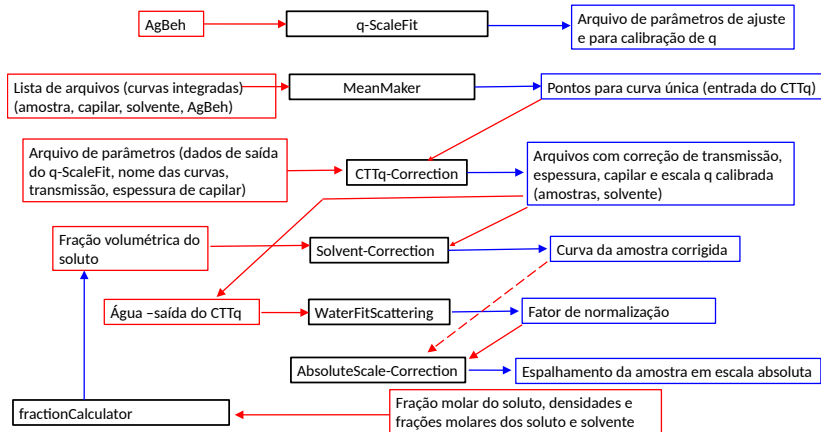
- a) MeanMaker
- b) q-ScaleFit
- c) fractionCalculator
- d) WaterFitScattering

---

<sup>8</sup>Os números correspondem à parte do tratamento de dados que o programa realiza.

<sup>9</sup>As cores indicam em qual etapa do tratamento de dados eles são necessários.

# Fluxograma



# ① CTTq-Correction

**Corrige pelo espalhamento do capilar, pela transmissão, pela espessura da amostra e calibra a escala q.**

## **Entrada:**

- Arquivo de parâmetros contendo:
  - q slope (a);
  - q intercept (b);
  - arquivo da amostra;
  - transmissão da amostra;
  - espessura da amostra (cm);
  - arquivo do capilar da amostra;
  - transmissão do capilar da amostra;
  - nome do arquivo de saída.

# ① CTTq-Correction (cont.)

## Saída:

- Arquivos de dados "pré-corrigidos": é gerado um arquivo de dados "pré-corrigidos" para cada linha de parâmetros;
- Gráfico na tela com todas as curvas "pré-corrigidas".

## c) fractionCalculator

**Calcula a fração volumétrica do soluto<sup>10</sup> a partir da fração molar e dos dados das componentes da amostra (mistura binária).<sup>11</sup>**

### **Entrada:**

- Massa molar do soluto;
- Densidade do soluto;
- Massa molar do solvente;
- Densidade do solvente;
- Fração molar do soluto.

## c) fractionCalculator (cont.)

### Saída:

- Fração volumétrica do soluto.

---

<sup>10</sup>O programa foi feito para amostras de terc-butanol e água, mas alterando os parâmetros pode ser utilizado para qualquer mistura **binária**.

<sup>11</sup>Esse procedimento pode ser realizado manualmente sem dificuldades, o programa apenas simplifica o processo, que em geral exige realizar de manipulações algébricas ou vários cálculos estequiométricos.

## ② Solvent-Correction

**Subtrai o espalhamento do solvente.**

### **Entrada:**

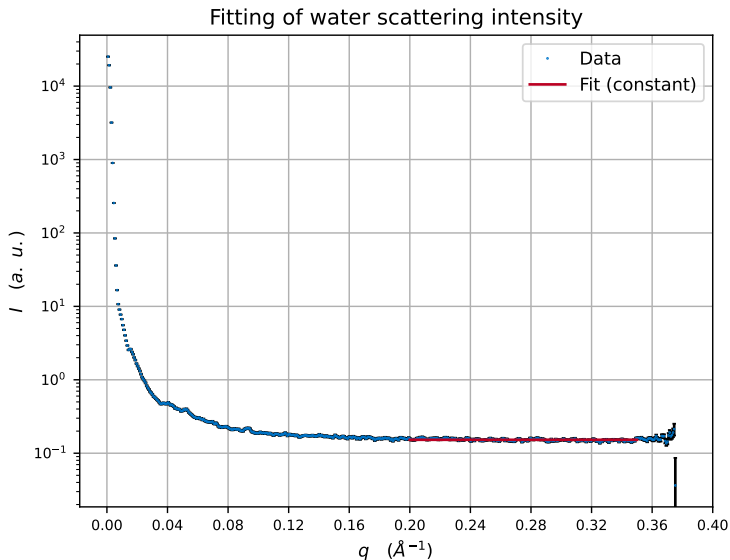
- Arquivo de parâmetros contendo:
  - arquivo de dados (pré-corrigidos) da amostra;
  - arquivo de dados (pré-corrigidos) do solvente;
  - fração volumétrica do soluto;
  - nome do arquivo de saída.

### **Saída:**

- Arquivo de dados corrigido pelo solvente: é gerado um arquivo de dados corrigidos pelo solvente para cada linha de parâmetros;
- Gráfico na tela com todas as curvas corrigidas pelo espalhamento do solvente.



## d) WaterFitScattering



## d) WaterFitScattering

Ajusta uma constante ao **espalhamento da água**<sup>12</sup> e gera uma **constante de normalização para a escala absoluta**.<sup>13</sup>

### Entrada:

- Arquivo de dados pré-corrigidos da água.

### Saída:

- No terminal: resultados do ajuste e Constante de Normalização para a escala absoluta;
- Gráfico na tela com curva de espalhamento da água e a constante ajustada.

---

<sup>12</sup>Seu uso é independente do uso do Solvent-Correction.

<sup>13</sup> Orthaber, Bergmann e Glatter (2000)

### ③ AbsoluteScale-Correction

**Normaliza os dados para a escala absoluta.**

**Entrada:**

- Arquivo de parâmetros contendo:
  - constante de normalização para escala absoluta;
  - nome do arquivo de dados pré-corrigidos e/ou corrigidos pelo solvente;
  - nome do arquivo de saída;

**Arquivo de Saída:**

- Arquivo de dados em escala absoluta: é gerado um arquivo de dados em escala absoluta para cada linha de parâmetros;
- Gráfico na tela com todas as curvas em escala absoluta.

# Orientações de uso dos programas

**Para rodar qualquer um dos programas é necessário:**

1. Possuir os programas (arquivos .py) no diretório onde eles serão chamados;
2. Possuir o diretório com o pacote saxspy no mesmo diretório onde os programas estão;
3. Recomenda-se rodar os programas no mesmo diretório onde se encontram os arquivos de parâmetros e arquivos de entrada.

**Para executar qualquer um dos programas basta seguir a fórmula geral:**

```
$ python3 programa.py
```

# Orientações de uso dos programas (cont.)

Após ser executado cada programa imprime mensagens e solicita dados, parâmetros e outras informações ao usuário.

As informações deveram ser fornecidas ao programa uma a uma, conforme solicitadas.

## Orientações de uso dos programas (cont.)

Caso um programa seja executado várias vezes com parâmetros de entrada iguais ou muito parecidos pode-se utilizar a seguinte estratégia para tornar o procedimento menos exaustivo:

- a) Cria-se um arquivo texto (e. g. `entrada.txt`);
- b) Em cada linha insere-se um dos parâmetros de entrada do programa em questão, e. g.:
  - 1 | `arquivodedados.dat`
  - 2 | `arquivodesaida`
- c) Salva-se o arquivo e executa-se o programa adicionando "`< entrada.txt`", conforme o exemplo.

```
$ python3 programa.py < entrada.txt
```

# Construção dos arquivos de parâmetros

## **Algumas orientações gerais e observações sobre a construção dos arquivos de parâmetros:**

- Os parâmetros devem estar todos na mesma linha e separados por vírgula (com ou sem espaços ou *tabs*);
- A ordem em que os parâmetros são escritos é importante;
- Cada linha de parâmetros no arquivo fornecido é entendida como um conjunto de dados a ser corrigido de forma independente;
- Não há restrições para o número de linhas de parâmetros no arquivo;

# Construção dos arquivos de parâmetros (cont.)

- Linhas em branco não são lidas pelos programas;
- Não use *tabs* ou espaços no início das linhas de parâmetros, pois isso pode fazer com que essas linhas não sejam lidas pelos programas;
- Podem ser escritos comentários no arquivo de parâmetros utilizando o caracter # no **início** da linha;
- O uso do caracter # no meio de uma linha não é entendido como comentário pelos programas.



## ① Arquivo de parâmetros – CTTq

**O arquivo de parâmetros deve ser preenchido com as informações a seguir (nessa ordem) e de acordo com as orientações gerais:**

*q-slope* (a, coeficiente angular), *q-intercept* (b, coeficiente linear)<sup>14</sup>, arquivo da amostra, transmissão da amostra, espessura da amostra (cm), arquivo do capilar da amostra, transmissão do capilar da amostra, nome do arquivo de saída (dados corrigidos).

---

<sup>14</sup>*slope* e *intercept* são dados obtidos através da regressão de calibração da escala *q*, realizada pelo programa q-ScaleFit.

# ① Arquivo de parâmetros – CTTq

## Exemplo de preenchimento de arquivo<sup>15,16</sup>

```
1 | # Amostra
2 | 1.001, 0.0003, amostra_media.dat, 0.2, 0.15,
   | capilar_amostra.dat, 0.8, amostra
3 | # Solvente
4 | 1.001, 0.0003, solvente_media.dat, 0.18, 0.15,
   | capilar_solvente.dat, 0.81, solvente
```

---

<sup>15</sup>As linhas são “quebradas” por conta do limite de tamanho da página, mas cada conjunto de parâmetros ocupa uma única linha no arquivo de parâmetros.

<sup>16</sup>A numeração e o separador “|” apenas ilustram a marcação das linhas, não fazem parte do arquivo de parâmetros.

## ① Arquivo de parâmetros – CTTq

### **Reforçando algumas observações<sup>17</sup>:**

Podem ser adicionadas mais linhas de parâmetros (Amostra 2, Amostra 3, Solvente 2 etc.), sem restrição de número. Cada linha é lida e executada como uma correção independente pelo programa.

Da mesma forma, o arquivo pode conter apenas uma linha, só com os parâmetros da amostra ou só do solvente.

---

<sup>17</sup>Essas observações também são válidas para o Solvent-Correction e para o AbsoluteScale-Correction.

## ② Arquivo de parâmetros (Solvent)

**O arquivo de parâmetros deve ser preenchido com as informações a seguir (nessa ordem):**

Arquivo de dados (pré-corrigidos) da amostra, arquivo de dados (pré-corrigidos) do solvente, fração volumétrica do soluto<sup>18</sup>, nome do arquivo de saída.

---

<sup>18</sup>Pode ser obtido pelo fractionCalculator (misturas binárias), ou manualmente através de cálculos estequiométricos.

## ② Arquivo de parâmetros (Solvent)

**Exemplo de preenchimento de arquivo<sup>19</sup>:**

```
1 | # Amostra 1
2 | amostra.dat, solvente.dat, 0.10, soluto1
3 | # Amostra 2
4 | amostra2.dat, solvente.dat, 0.14, soluto2
```

---

<sup>19</sup>Novamente, o arquivo pode conter apenas uma ou várias linhas de parâmetros.

### ③ Arquivo de parâmetros (AbsoluteScale)

**O arquivo de parâmetros deve ser preenchido com as informações a seguir (nessa ordem):**

Fator de escala absoluta<sup>20</sup>, Nome do arquivo de dados, Nome do arquivo de saída

---

<sup>20</sup>Obtido através do WaterFitScattering.

### ③ Arquivo de parâmetros (AbsoluteScale)

#### Exemplo de preenchimento de arquivo:

```
1 | # Amostra 1
2 | 2.4, soluto1.dat, soluto1_abs
3 | # Amostra 2
4 | 2.4, soluto2.dat, soluto2_abs
5 | # Amostra 3
6 | 2.8, soluto3.dat, soluto3_abs
```

# Considerações finais

O objetivo desses programas é simplificar e padronizar o trabalho com o tratamento de dados de medidas de SAXS.

Para entender o funcionamento deles, o melhor jeito é abrir os códigos e buscar entendê-los lendo e mexendo com eles.

Assim, também pode-se avaliar o que pode ser melhorado, o que certamente inclui muitas coisas.

Não tenham medo de mexer nos códigos e buscar adequá-los às suas necessidades!



# Considerações finais: SAXSPy

O objetivo desse pacote Python é servir de base tanto para o tratamento como para a análise de dados de SAXS.

Para entender seu funcionamento e recursos, a melhor estratégia é abrir os códigos, buscando entendê-los lendo e mexendo com eles.


O SAXSPy é apenas um brotinho e muitos recursos podem ser implementados e adicionados a ele, além de melhorar os já existentes.


Não tenham medo de mexer nos códigos e buscar adequá-los às suas necessidades!


# Considerações finais: SAXSPy (cont.)

Por último, fiquem à vontade se quiserem dar um nome mais criativo para ele :P

# Referências


 HARRIS, C. R. et al. Array programming with NumPy. *Nature*, Springer Science and Business Media LLC, v. 585, n. 7825, p. 357–362, set. 2020. Disponível em: <<https://doi.org/10.1038/s41586-020-2649-2>>.


 HUANG, T. et al. X-ray powder diffraction analysis of silver behenate, a possible low-angle diffraction standard. *Journal of applied crystallography*, International Union of Crystallography, v. 26, n. 2, p. 180–184, 1993.

 HUNTER, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, IEEE COMPUTER SOC, v. 9, n. 3, p. 90–95, 2007.

## Referências (cont.)

 MARIANI, P.; SPINOZZI, F. *Small Angle X-ray Scattering and Particle Shape Reconstruction Methods in Biology*. 1. ed. Ancona: University of Ancona, 1999.

 ORTHABER, D.; BERGMANN, A.; GLATTER, O. Saxes experiments on absolute scale with kratky systems using water as a secondary standard. *Journal of Applied Crystallography*, International Union of Crystallography, v. 33, n. 2, p. 218–225, 2000.

 VIRTANEN, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, v. 17, p. 261–272, 2020.